THÈSE

préparée à

L'INRIA Sophia-Antipolis

et présentée à

L'ÉCOLE POLYTECHNIQUE

pour obtenir le grade de

DOCTEUR de L'ÉCOLE POLYTECHNIQUE

Spécialité

Informatique

par

Xavier PENNEC

Sujet de la thèse :

L'INCERTITUDE DANS LES PROBLÈMES DE RECONNAISSANCE ET DE RECALAGE

APPLICATION EN IMAGERIE MÉDICALE ET BIOLOGIE MOLÉCULAIRE

Soutenue le 5 décembre 1996 devant un jury composé de :

MM.	**Stéphane**	**MALLAT**	Président
	Michael	**BRADY**	Rapporteurs
	Jean-Marie	**MORVAN**	
	Nicholas	**AYACHE**	Directeur
	Philippe	**CINQUIN**	Examinateurs
	Olivier	**FAUGERAS**	

« The average Ph.D. thesis is nothing but a transference of bones from one graveyard to another. »

J. Frank Dobie, “A Texan in England”, 1945

Remerciements

Je tiens tout d'abord à adresser ma plus vive gratitude aux membres de mon jury :

- à Nicholas Ayache, mon directeur de thèse, qui m'a accueilli dans son laboratoire dans des conditions plus que favorables à une recherche fructueuse et qui a su, pendant ces trois ans, conseiller et orienter habilement mes travaux vers leur aboutissement,
- à Jean-Marie Morvan et Mike Brady, qui ont eu la gentillesse d'accepter la lourde tâche de rapporteurs auprès du jury; je les remercie tout particulièrement pour l'attention et le temps qu'ils y ont consacrés, ainsi que pour leurs encouragements chaleureux et la lucidité de leurs conseils,
- à Olivier Faugeras, qui m'a initié il y a quelques années aux mystères de la vision par ordinateur et de la géométrie; je le remercie également pour les discussions fructueuses que nous avons pu avoir et pour ses conseils judicieux,
- à Stéphane Mallat, qui a grâcieusement accepter de présider ce jury, et à Philippe Cinquin qui, en me faisant l'honneur de siéger à ce jury, ont fait preuve de leur intérêt pour mon travail.

Je ne peux pas passer sous silence l'immense reconnaissance que j'ai envers mes deux relecteurs anonymes, qui ont certainement plus influencé l'écriture de ce manuscrit que ce qu'ils pensent :

- Vicente Cervera a su, malgré l'obstacle de la langue, porter le regard acéré d'un mathématicien sur toute la partie théorique. Je le remercie également pour les nombreuses discussions que nous avons eues et la patience avec laquelle il a supporté mes élucubrations mathématiques.
- Jérôme Declerck, véritable détecteur de zones d'ombres, a eu la patience infinie de corriger les fautes d'orthographe, de français, de style et de m'écouter parler. Il a aussi impitoyablement souligné les passages difficiles à comprendre et m'a grandement aidé à les clarifier.

J'exprime également toute ma sympathie et ma gratitude à mes collègues et néanmoins amis qui ont du supporter cette thèse, et en particulier à mes chefs de bureau successifs : Hervé Delingette et Jean-Philippe Thirion, dont le bon sens et les conseils ont été proverbiaux. Je voudrais joindre à eux Grégoire Malandain et Gérard Subsol pour la qualité du support matériel et intellectuel qu'ils m'ont prodigué pendant ces années. Je remercie également Olivier Dourthe pour ses idées et son expérience de médecin, ainsi que tous ceux qui animent régulièrement la vie du projet Epidaure.

Je voudrais remercier tous mes amis qui ont supporté mes divagations et su m'encourager et me soutenir dans mes égarements métaphysiques : Sylvain Lazard, Mireille Bossy, Bruno Vasselle, Lucien Nocera, Guy Perrin, Jean Marc et Sandrine Phelippeau et Johan Montagnat. Que ceux que j'ai oublié me pardonnent.

Enfin, je voudrais remercier ma famille et en particulier mes parents qui m'ont donné l'éducation sans laquelle je n'en serais pas là aujourd'hui.

Table des matières

Résumé

Il est souvent utile en traitement d'image de ne conserver qu'un ensemble restreint de caractéristiques localisées que nous appelons primitives géométriques, censées caractériser la plus grande partie de l'information. Ces primitives sont habituellement des contours, des points de coin ou leur équivalent en imagerie 3D : des surfaces, des lignes de crête ou des points extrémaux, etc.

Cependant, ces primitives sont souvent plus complexes que de simples points : on peut ainsi associer un vecteur normal à un point d'une surface, ce qui en fait un point orienté, ou considérer un trièdre composé de la normale et des deux directions principales en plus d'un point extrémal, ce qui forme un repère. Ces primitives géométriques forment une variété qui n'est généralement pas un espace vectoriel, sur lequel agit un groupe de transformation qui modélise les différentes « prises de vue possibles » de l'image. De plus, il est impératif de gérer l'incertitude sur ces primitives le plus finement possible pour obtenir des résultats robustes en contrôler la précision. Le problème que l'on se pose ici est donc de pouvoir travailler avec ces primitives comme on travaille d'habitude sur les points, en particulier pour pouvoir faire de la reconnaissance, du recalage et des statistiques permettant d'inférer la précision de nos résultats.

La première partie de la thèse est consacrée au développement d'outils mathématiques génériques sur les primitives qui permettent de résoudre ces problèmes. Nous présentons tout d'abord quelques paradoxes qui montrent que l'on ne peut pas considérer impunément ces primitives comme de simples vecteurs, puis, sur des bases de géométrie riemannienne et en se restreignant aux variétés homogènes ayant une métrique invariante pour le groupe considéré, nous développons une notion de moyenne cohérente, puis de matrice de covariance. D'autres opérations statistiques peuvent ensuite être généralisées aux variétés, telles que la distance de Mahalanobis et le test du χ^2. Nous montrons ensuite comment cette théorie peut être appliquée et implémentée en machine dans une structure orientée objet ne dépendant pas du type de primitive considéré.

Dans la seconde partie de la thèse, nous développons les calculs relatifs à ce formalisme mathématique pour les rotations et les transformations rigides agissant sur des repères orientés, semi-orientés et non-orientés. Après analyse des algorithmes classiques sur les points, nous développons des algorithmes de recalage génériques basés sur les primitives, et nous proposons en particulier des méthodes pour estimer l'incertitude sur le recalage obtenu. Nous exposons parallèlement une méthode de validation statistique pour confirmer la précision de cette analyse, qui montre qu'un recalage d'une précision bien inférieure à la taille du voxel peut être obtenue dans le cas des images médicales 3D. Un deuxième volet de l'analyse statistique concerne la robustesse des algorithmes de reconnaissance.

Le champ applicatif que nous considérons va du recalage d'images médicales tridimensionnelles à la reconnaissance de sous-structures (3D) dans les protéines et souligne la validité de l'approche générique « orientée primitive » que nous avons choisie pour le traitement haut niveau de données géométriques.

Chapitre 1

Introduction

« La science naît du jour où des erreurs, des échecs, des surprises désagréables, nous poussent à regarder le réel de plus près. »
René Thom, Modèles mathématiques de la morphogénèse

1.1 Cadre général

La vision humaine est certainement celui des cinq sens sur lequel l'esprit humain se repose le plus pour analyser, comprendre et interpréter l'espace qui nous entoure. C'est par ailleurs le seul sens qui puisse nous fournir une information globale sur la géométrie de notre environnement, ce qui nous permet de prévoir et donc d'interagir avec le monde extérieur. La vision artificielle est une branche de la science informatique qui s'est donné pour but de simuler, si ce n'est de comprendre, le phénomène de la vision, pour pouvoir en doter les ordinateurs. La robotique, quant à elle, utilise la modélisation fournie par la vision, combinée avec des informations provenant d'autres capteurs, pour se déplacer et interagir avec le monde réel. Au centre de ces disciplines, il y a la modélisation géométrique du monde, forcément simplificatrice mais qui permet d'extraire rapidement l'information pertinente.

Depuis plusieurs années déjà, de nouvelles techniques d'acquisitions de données sont apparues, produisant non plus des images en deux dimensions comme la caméra ou l'appareil photo, mais directement en trois dimensions, comme le scanner X ou l'imagerie par résonance magnétique (IRM). Ces images ne mesurent plus la réflectance de la surface des objets, mais l'atténuation d'un rayonnement en un point intérieur d'un objet (c'est le cas du scanner X), ou la densité de matière en un point (IRM densité de proton). Permettant de voir « au travers » des objets quasiment sans interagir avec eux, ces nouvelles techniques ont trouvé une application naturelle dans le domaine médical et sont devenues des extensions du système de perception des médecins et des chirurgiens pour le diagnostic et la thérapeutique. Cependant, les images tridimensionnelles ainsi obtenues sont porteuses d'une quantité d'information énorme qu'il est très difficile d'appréhender dans sa totalité. Il faut donc traiter ces images, cerner et souligner l'information pertinente pour le médecin (la surface d'un organe, le positionnement d'une tumeur), ou bien quantifier les changements (l'évolution d'une tumeur, le mouvement relatif de deux os) pour comparer avec la « normalité ». Toutes ces opérations sont rarement possibles directement sur l'image et c'est la modélisation géométrique qui

va nous permettre de concentrer l'information intéressante, noyée dans la masse de voxels de l'image, en un petit nombre de caractéristiques significatives.

Le même type de techniques, résonance magnétique nucléaire ou rayons X, peut être utilisée à une toute autre échelle, celle de l'atome, pour étudier et tenter de visualiser les structures moléculaires d'un cristal, permettant, entre autres, de déterminer la position dans l'espace des atomes d'une protéine : c'est le domaine de la crystallographie et de la biologie moléculaire structurale. Pour interpréter et comprendre la structure d'une protéine, il faut non seulement regarder le type des atomes (s'agit-il d'un oxygène, d'un carbone ou d'un azote ?) mais aussi leurs positions relatives dans l'espace : un groupe d'atomes sera ainsi capable d'interagir avec une autre structure dans une configuration donnée, mais il en serait bien incapable si l'un ou l'autre de ces atomes était déplacé de quelques angströms : une fois encore, la géométrie est au cœur du problème, même si les interactions sont d'une autre nature.

Ces processus d'imagerie s'organisent en deux grandes étapes successives dans l'analyse hiérarchique ascendante proposée par (Marr, 1982) (voir figure 1.1) : on effectue tout d'abord des traitements de bas niveau sur les données (les images) pour obtenir une information sémantiquement riche mais non structurée, formée « d'atomes » que nous appellerons **primitives géométriques** . Ceci correspond, dans le processus de vision humaine, à la construction de *percepts*. L'ensemble des primitives géométriques produites par les traitements bas niveau à partir d'un jeu de données forme une **scène** .

L'important, pour pouvoir ensuite organiser ces primitives, est de noter qu'un **objet** n'est par caractérisé par un ensemble de primitives *dans des positions fixées*, mais plutôt par un ensemble de primitives dans une *configuration* fixée [1]. La position de nos primitives dépend en effet de la position de l'objet original et de la position du système d'acquisition dans l'espace ambiant et, pour pouvoir comparer ou reconnaître des objets dans des positions ou lors d'acquisitions différentes, il nous faut modéliser les différentes « prises de vues » possibles, ce qui se traduit dans le cadre géométrique par le choix d'un **groupe de transformation**.

- En vision artificielle, les primitives géométriques sont ainsi des contours, des points de jonction, etc., symbolisant la projection dans l'image 2D des arêtes et des coins des objets 3D. Si l'on s'intéresse aux objets rigides plans vus par un système oculaire assimilable à un modèle sténopé, on pourra considérer ces primitives comme des points du plan projectif soumis à l'action des homographies. On pourra se reporter à (Faugeras, 1993) pour d'autres exemples de primitives et d'autres types de transformations.

- En imagerie médicale, les primitives peuvent être des points sur une surface, des lignes de crête, des points extrémaux, généralisant en un certain sens les points de coin, etc. Les systèmes d'acquisition étant calibrés, on pourra considérer que les transformations sont rigides entre deux images du même patient, même si les images proviennent de deux modalités différentes. Par contre, il faut passer à des groupes de transformation plus généraux, tels que les similitudes, les transformations affines ou des déformations encore plus générales, si l'on veut comparer le crâne, le cerveau ou le foie de deux patients.

- En biologie moléculaire, les primitives géométriques sont déjà d'un niveau plus élevé puisqu'il s'agit non seulement des coordonnées des atomes formant une protéine (ou une autre structure moléculaire telle l'ARN), mais aussi des relations entres ces atomes (les liaisons). On pourra

1. Cette notion d'objet (ou de modèle) comme configuration d'un ensemble de primitives est sans doute la plus simple. On peut bien sûr ajouter des relations entre primitives, des probabilités d'observation corrélées, etc.

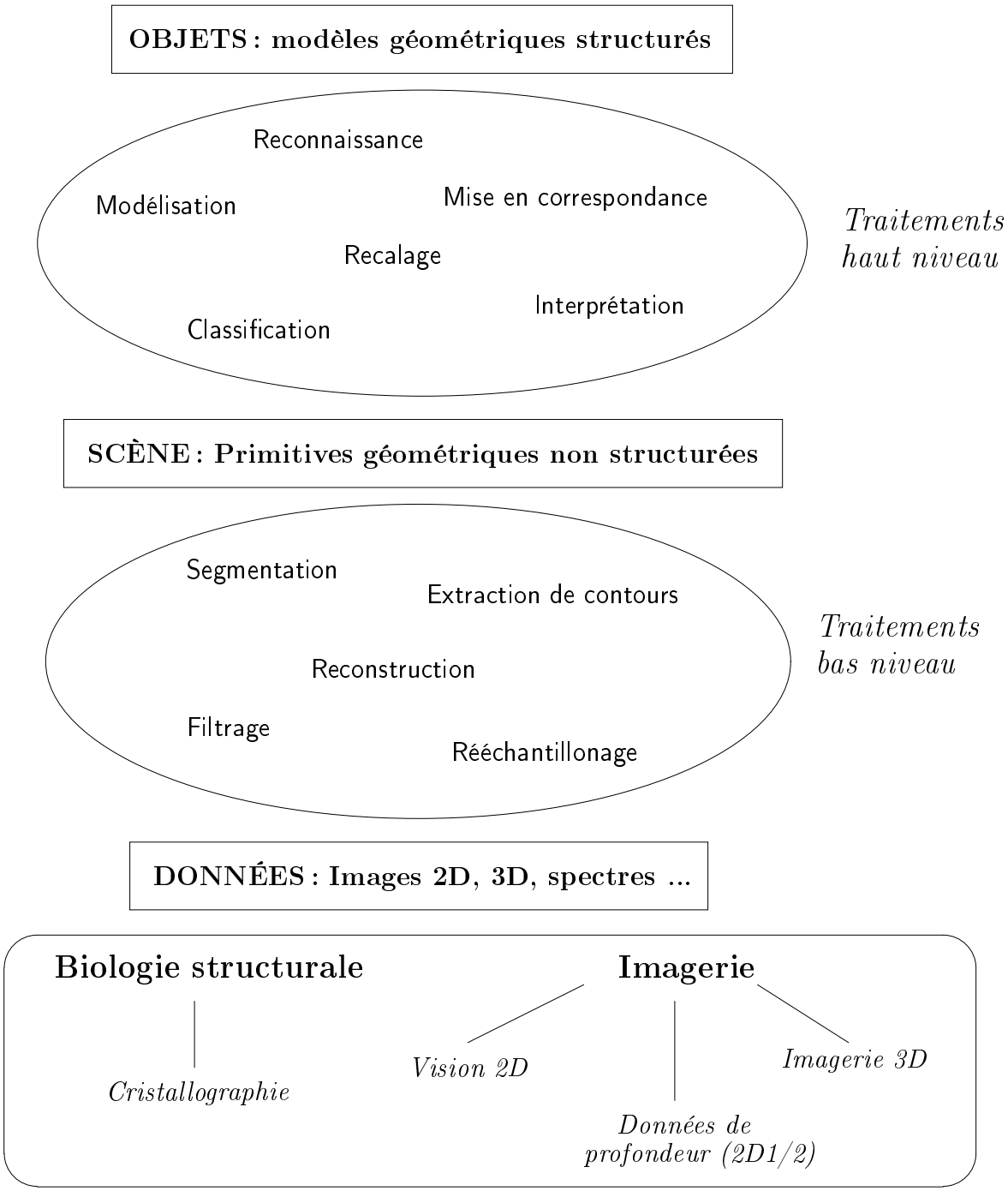


FIG. 1.1 – *Modèle de structuration de l'information en traitement d'images (approche hiérarchique ascendante ou « bottom-up ») : on extrait de la masse de données brutes un ensemble de primitives géométriques censées caractériser la majeure partie de l'information présente. On traite ensuite cet ensemble de primitives par des algorithmes dits de haut niveau pour l'organiser, le comparer avec d'autres acquisitions, créer des modèles structurés du monde réel, reconnaître un modèle structuré déjà identifié dans ces données.*

considérer en première approche que les protéines sont animées de mouvements rigides par bloc.

L'étape qui nous intéresse et dans le cadre de laquelle se place le travail présenté dans ce manuscrit est le traitement haut niveau qui va permettre de structurer l'information contenue dans l'ensemble des primitives géométriques extraites des données (une scène). Nous utilisons ici le qualificatif *géométrique* pour souligner que les traitements bas niveau nous fournissent des types particuliers de primitive (points orientés, droites, repères, points projectifs...) implicitement soumis à un groupe de transformation. La détermination de ce groupe est sans doute l'une des étapes les plus importantes dans la modélisation d'un problème géométrique.

On peut distinguer plusieurs grandes classes de traitements de haut niveau : la **reconnaissance** d'objets (la dénomination anglaise « feature-based recognition » est plus précise), se propose de reconnaître, à partir d'une bibliothèque d'objets, ceux qui sont présents dans une scène. Ce processus se décompose souvent en deux étapes couplées : la **mise en correspondance** (« matching ») vise à apparier les primitives d'un modèle avec celles de la scène (en supposant éventuellement une transformation), et le **recalage** (« registration ») calcule la meilleure transformation entre le modèle et la scène à partir de ces appariements.

On peut bien sûr comparer deux scènes sans avoir de modèle a priori, ou comparer un nombre quelconque de scènes pour en extraire les parties communes : on parle alors de **recalage multiple** et de **mise en correspondance multiple**. C'est une étape essentielle dans le processus de **modélisation** des objets, puisqu'il nous permet de ne conserver que les primitives qui sont représentatives d'un objet, mais aussi de caractériser et de réduire l'incertitude sur la mesure de ces primitives grâce à la **fusion** de plusieurs mesures (après recalage).

Notons que les dénominations *recalage* et *mise en correspondance* sont souvent considérées comme synonymes, comprenant toutes les deux des phases d'appariement des primitives et de calcul de la transformation. Toutefois, *recalage* insiste plutôt sur le calcul de la transformation, tandis que la *mise en correspondance* met en avant la recherche des appariements. Dans ce manuscrit, nous utiliserons en général ces termes dans leur sens restreint.

1.2 Motivations

Un grand nombre d'algorithmes ont été développés pour comparer deux scènes ou reconnaître des objets, le plus souvent rigides ou affines, dont on possède un modèle géométrique *a priori* basé sur des des primitives géométriques simples : les points. On peut citer par exemple les algorithmes de hachage géométrique (Lamdan et Wolfson, 1988; Rigoutsos et Hummel, 1993), du plus proche voisin itératif (Besl et McKay, 1992; Zhang, 1994) ou de prédiction-vérification (Ayache et Faugeras, 1986; Huttenlocher et Ullman, 1987).

On s'aperçoit cependant que si ces algorithmes fonctionnent vite et bien dans le cas de données bien individualisées et peu bruitées, il n'en va pas de même quand il s'agit de retrouver de petits objets dans une scène complexe très bruitée. La complexité maximale des algorithmes est alors atteinte et le problème de la gestion de l'incertitude se pose de manière cruciale.

Par ailleurs, les modèles géométriques du monde réel amènent souvent à considérer des primitives plus complexes, par exemple des droites (Grimson, 1990), des plans (Faugeras, 1993), des points ou des plans orientés (Feldmar et al., 1997), des repères (Pennec et Ayache, 1998; Pennec et Thirion, 1997), etc., et la généralisation hâtive des techniques ponctuelles à ces primitives plus complexes peut conduire à des paradoxes (voir (Pennec et Ayache, 1996a) et section (2.3)).

Nous pensons que ces problèmes en vision par ordinateur et plus généralement en traitement d'image ou de donnée géométrique, peuvent trouver une solution acceptable en prenant en compte de manière intrinsèque à la fois la nature de **variété différentielle** de l'ensemble des primitives d'un type donné et le fait que les mesures de ces primitives soient **incertaines**. Ce type d'approche permettrait de plus de concevoir des **algorithmes génériques**, évitant ainsi la profusion de méthodes ad hoc spécifiques à chaque type de primitive.

1.2.1 Au delà des points : pourquoi utiliser des primitives géométriques ?

La modélisation bas niveau nous amène souvent à considérer des primitives plus complexes que les points : nous verrons ainsi à la section (2.3.1) qu'un acide aminé dans une protéine se modélise en première approche par un repère tridimensionnel (un point et un trièdre orthonormé direct), et que les **points extrémaux** et plus généralement les points sur une surface sont munis de deux directions principales et d'une normale pour former des repères semi-orientés (un vecteur normal et deux *directions* orthogonales).

L'utilisation des primitives les plus informatives possibles nous permet de réduire considérablement la **complexité** des algorithmes de reconnaissance et de recalage : on passe par exemple d'un algorithme d'une complexité de $O(n^3)$ à $O(n^2)$ dans le chapitre (11) pour la reconnaissance de sous-structures dans les protéines en utilisant des repères au lieu de points. Dans (Feldmar et al., 1997), l'introduction des normales en plus des points autorise la recherche d'appariements initiaux avec les bitangentes, ce qui réduit considérablement la complexité de l'étape d'appariement.

Le deuxième effet est une augmentation de la **robustesse** des algorithmes grâce à la sélectivité accrue des primitives : on peut ainsi éliminer 80% des appariements aberrants en utilisant des trièdres très incertains en plus des points (chapitre 10). Un autre effet qui nous a motivé, mais qui n'est finalement pas le plus convaincant, est l'augmentation de la **précision** de la transformation dans un recalage : on obtient par exemple un gain de 10 à 20% dans le cas des images médicales au chapitre 9.

Enfin, les transformations mises en jeu dans ces problèmes géométriques ont une nature intrinsèque de **groupe de Lie** et sont donc des variétés différentielles. Si l'on veut pouvoir estimer l'incertitude sur le résultat d'un recalage, il faut donc en particulier pouvoir travailler avec ces primitives géométriques.

1.2.2 Gestion fine de l'incertitude

La modélisation de bas niveau en terme de primitives est intrinsèquement simplificatrice et est sujette à de multiples erreurs : incertitude sur la nature et la position des primitives, défaut d'observation (occlusion) ou extraction de primitives parasites sont les principales sources d'erreur.

La présence de ces incertitudes a des effet très importants sur des algorithmes de haut niveau comme la mise en correspondance : considérer que des primitives (ou des invariants) ne sont identiques que si elles sont égales (à la discrétisation près) peut amener à rejeter beaucoup d'appariements corrects, rendant par là même hasardeuse voire improbable la détection d'un objet pourtant présent dans l'image. On appelle cela un **faux négatif** .

D'un autre côté, dès que l'on autorise une certaine erreur sur les mesures (ne serait-ce que par la discrétisation des images ou celle des nombres réels), des primitives qui n'ont rien à voir avec l'objet recherché peuvent être appariées et ainsi « conspirer » pour amener à la reconnaissance d'un objet qui n'existe pas : c'est un **faux positif** . Ces reconnaissances fantômes sont en général identifiables grâce à une vérification plus approfondie utilisant des informations supplémentaires pouvant avoir été éliminées lors de la modélisation en primitives géométriques. Ces vérifications sont toutefois

coûteuses en temps de calcul et peuvent induire une explosion de la complexité de l'algorithme (Lamdan et Wolfson, 1989; Grimson et Huttenlocher, 1990a; Lamdan et Wolfson, 1991).

Il est donc important de gérer l'erreur pour éviter les faux négatifs, mais de le faire au plus juste pour limiter la probabilité ou le nombre de faux positifs. De même, dans un problème de recalage, il est faux (et éventuellement dangereux dans le cas de l'imagerie médical) de dire que la transformation estimée est exacte, mais il est en général visuellement évident que l'incertitude est inférieure (en terme d'écart-type) à la taille de l'image.

1.2.3 Des méthodes génériques

Il existe souvent une solution qui semble simple pour traiter un problème particulier avec un type particulier de primitive. Malheureusement, cette solution est rarement applicable pour un autre type de primitive ou un autre problème. Nous voudrions obtenir un ensemble de modules génériques permettant de traiter la plupart des problèmes (recalage, mise en correspondance, fusion, etc.) et reposant sur un minimum de modules particuliers à chaque type de primitive.

Pour cela, nous avons observé que la plupart des algorithmes haut niveau sur les points pouvaient s'exprimer à partir des quelques opérations suivantes :

- composition et inversion des transformations,
- action d'une transformation sur une primitive,
- comparaison de primitives : distance,
- estimation d'une primitive à partir de plusieurs observation : moyenne,
- estimation d'une transformation à partir d'appariements de primitives : recalage,
- test de compatibilité (peut-on apparier deux primitives en supposant une certaine transformation) : zone de compatibilité,
- calcul et comparaison des invariants binaires, ternaires, etc. (invariants obtenus pour deux, trois primitives, etc.).

Pour pouvoir gérer l'incertitude, il faut de plus pouvoir propager l'information d'incertitude associée aux mesures dans les opérations ci-dessus, et éventuellement rajouter les notions de :

- modèle de bruits (en remplacement du bruit additif) : définition et estimation,
- comparaison statistique des primitives : équivalent de la distance de Mahalanobis,
- test de vraisemblance : équivalent du test du χ^2.

L'idée est donc d'exprimer au maximum ces opérations comme des modules génériques basés sur le nombre minimum d'opérations spécifiques. La conception d'un nouvel algorithme de haut niveau pour un nouveau problème ne nécessiterait alors qu'une étape de modélisation où l'on définirait le type de primitives en ayant un nombre minimal d'opérations à implémenter, et une organisation facile des briques algorithmiques de moyen niveau définies ci-dessus.

1.3 Organisation du manuscrit

« Pour expliquer un brin de paille, il faut démonter tout un univers. »
Rémy de Gourmont

Le manuscrit est divisé en deux grandes parties : la première partie s'attache au côté théorique de l'incertitude sur les variétés différentielles et les groupes de Lie et la seconde est axée sur les applications de cette théorie dans les problèmes de recalage et de reconnaissance.

Le thème majeur de la partie théorique est : comment faire des probabilités et des statistiques sur une variété différentielle comme on le fait dans $\mathbb{R}^n$? L'objectif est double : il s'agit non seulement d'étudier cette question d'un point de vue mathématique, mais aussi de développer les résultats nécessaires à l'application de cette théorie pour construire des algorithmes de haut niveau dans la seconde partie. Selon le point de vue, théorique ou applicatif, on pourra ainsi considérer le chapitre 6 comme une synthèse de la partie théorique, ou au contraire comme le plan de travail guidant les développements de cette partie.

Ce chapitre est tout à fait central dans le manuscrit puisqu'il constitue la charnière entre la théorie mathématique et les applications informatiques. Il répond également à la motivation « méthodes génériques » de la section précédente car nous y présentons les résultats dans une structure orientée objet générique, ne dépendant, pour chaque type de primitive, que d'une phase mathématique et de l'implémentation de quelques opérations atomiques.

Le chapitre 7, commençant la seconde partie, s'attache d'ailleurs à dériver et implémenter ces opérations atomiques pour les primitives tridimensionnelles que nous utilisons dans les applications. Les chapitres concernant les algorithmes haut niveau (reconnaissance et recalage) sont toutefois conçus de manière générique et simplement appliqués à l'imagerie médicale et à la biologie moléculaire. Les questions de précision étant particulièrement importantes (voire vitales) en imagerie médicale, nous avons plus insisté sur le recalage que sur la mise en correspondance, en développant, entre autres, des méthodes de validation de nos estimations d'incertitude (chap. 9). Cette partie applicative montre que la théorie que nous avons développée est non seulement implémentable avec des temps de calculs acceptables, mais produit aussi des résultats prenant en compte les erreurs au plus juste, répondant ainsi à l'objectif de « gestion fine de l'incertitude » de la section précédente.

1.3.1 Théorie

Nous cherchons donc dans la première partie à développer les outils mathématiques pour pouvoir gérer l'incertitude sur les primitives géométriques. Pour cela, Nous analysons rapidement au **chapitre 2** l'origine des erreurs de mesures et les façons classiques de les quantifier dans $\mathbb{R}^n$: il apparaît que le meilleur compromis est de considérer la mesure d'un vecteur (ou d'un point) comme la réalisation d'un vecteur aléatoire dont on ne conserve que la moyenne et la matrice de covariance. Nous rappelons donc les bases de la théorie des probabilités et en particulier comment on peut généraliser la notion de **variable aléatoire** à celle de **vecteur aléatoire**, ainsi que la propagation l'incertitude sur ces vecteurs aléatoires. Le but de cet exposé est bien sûr de pouvoir généraliser toutes ces opérations à des **primitives aléatoires**. Malheureusement, nous montrons que ce n'est pas si simple et que nombre de paradoxes peuvent apparaître : le paradoxe de Bertrand est sans doute le plus connu et montre l'ancienneté du problème puisqu'il date de 1907. Nous en avons développé d'autres pour mettre en évidence les problèmes posés par le calcul de la moyenne et la modélisation du bruit sur les primitives.

En fait, si l'on peut utiliser la théorique classique des probabilités pour gérer l'incertitude sur les

points, c'est parce que l'on peut associer aux points une structure d'espace vectoriel, chose que l'on ne peut pas faire de manière évidente pour des primitives géométriques plus générales. Nous nous focalisons donc au **chapitre 3** sur la structure intrinsèque de l'ensemble des primitives : celle de **variété différentielle**. L'ensemble des transformations implicitement associées aux primitives rentre également dans ce cadre pour constituer un **groupe de Lie**. Nous nous concentrons plus particulièrement dans la suite sur la partie des primitives affectée par l'action du groupe : c'est la notion de variété homogène. En suivant l'exemple du paradoxe de Bertrand, nous nous cherchons alors à déterminer la mesure invariante qui constitue la base de la **théorie des probabilités géométriques**. Cette théorie est toutefois insuffisante puisqu'elle ne concerne que des distributions uniformes. Pour pouvoir aller plus loin, nous avons besoin d'une distance invariante sur notre variété, ce qui est l'objet de la section (3.4). Toutefois, les résultats ainsi obtenus sont toujours insuffisants et nous devons faire intervenir des notions de **géométrie différentielle** et de **géométrie riemannienne** pour pouvoir caractériser les conditions d'existence d'une distance invariante. Cela nous permet également de construire la représentation exponentielle de la variété, qui constitue la représentation « la plus linéaire possible » de la variété et qui s'avère posséder des propriétés remarquables.

Nous pouvons alors développer au **chapitre 4** une théorie des probabilités sur les primitives géométriques, en définissant tout d'abord la **densité de probabilité** d'une **primitive aléatoire**, puis avec Fréchet et Karcher la manière de définir la moyenne d'une telle primitive. Cette moyenne étant définie au moyen d'une minimisation, il nous faut développer un algorithme pour pouvoir la calculer effectivement. A partir de la **primitive moyenne**, il n'est pas très dur de définir une **matrice de covariance** grâce au développement de la variété dans le plan tangent (la carte exponentielle). Avec ces « moments du premier et second ordre », nous pouvons traiter l'essentiel de nos problèmes de probabilité sur les primitives géométriques.

Dans le **chapitre 5**, les problèmes statistiques sont plus ouverts : si nous pouvons définir proprement les **modèles de bruit isotropes et homogènes**, il est clair que la définition de la **distribution gaussienne** sur une variété que nous proposons n'est pas la seule possible, mais elle a l'avantage de pouvoir être approchée par une gaussienne classique pour une covariance faible. De même, le choix de la définition de la **distance de Mahalanobis** est plus guidé par des considérations applicatives que théoriques.

Nous résumons dans le **chapitre 6** les principaux résultats obtenus dans cette partie, mais vus cette fois-ci du côté applicatif : la **phase mathématique** de la section permet de savoir s'il existe une distance invariante pour un type de primitives soumis à l'action d'un groupe de transformation donné, puis de déterminer les géodésiques et donc la carte principale. Il ne reste alors qu'à implémenter les **opérations atomiques** pour que ce type de primitives forme un objet de la classe « **primitives probabilistes** ». Les opérations de base sont alors génériques et permettent déjà de construire quelques briques d'algorithmes haut niveau.

1.3.2 Applications

Pour pouvoir appliquer la théorie de la première partie, nous commençons par étudier au **chapitre 7** les **rotation 3D** sous la forme matricielle, puis en utilisant les **quaternions**. Cette étude dépasse la cadre de la méthodologie présentée dans la synthèse de la partie théorique, mais autorise ainsi une approche plus appliquée des techniques générales sur la variétés différentielles présentées au chapitre 3. A partir du **vecteur rotation** et de ses opérations atomiques, nous pourrons développer relativement simplement les calculs nécessaires aux transformations rigides, donc

aux **repères**. Nous envisagerons alors le cas des **repères semi et non-orientés**.

Dans le **chapitre 8**, nous nous intéressons à deux problèmes d'estimation : le recalage et la fusion de primitives. Nous analysons tout d'abord les techniques de **recalage** classiques aux **moindres carrés** à partir d'appariements de points, puis nous développons trois techniques de recalage à partir d'appariements de primitives quelconques, techniques qui s'adaptent également très bien à la **fusion** de primitives. La difficulté dans toutes ces méthodes est d'évaluer l'incertitude sur la transformation ou la primitive estimée. Enfin, nous abordons le problème de l'estimation du **modèle de bruit** sur les primitives, ce qui permet de construire un algorithme de **recalage robuste** estimant tous les paramètres.

Les questions de **précision** sont particulièrement importantes, voire vitales, en imagerie médicale : nous nous posons dans la première partie du **chapitre 9** la question de la **validation** de notre estimation de l'incertitude sur le recalage. Nous développons pour cela une méthode statistique qui fonctionne sur des données synthétiques, mais aussi sur des données réelles. Les expériences sur les donnés synthétiques montrent une validation parfaite dans le cas d'un bruit connu sur les primitives et permettent d'affiner les domaines de validité dans le cas de l'estimation de ce bruit lors du recalage. Nous présentons une expérience de recalage mono-patient sur des données IRM qui valide pleinement notre estimation de l'erreur sur le recalage et prédit un écart-type de 0.1 mm, soit un dixième de voxel. L'incertitude sur le recalage peut aussi servir à décider statistiquement s'il y a un mouvement relatif entre deux structures : nous présentons une telle expérience sur le bassin.

Nous abordons dans le **chapitre 10** les principaux algorithmes de **reconnaissance** de **mise en correspondance** habituellement utilisés sur les points, et nous les formalisons en terme de primitives géométriques. Nous nous attachons ensuite aux modifications nécessaires dans ces algorithmes pour que l'erreur soit prise en compte et que l'on soit sur de reconnaître un objet s'il est présent. La contrepartie de cette **correction** des algorithmes est l'apparition de **faux positifs**, c'est-à-dire de reconnaissances fantômes, et nous en analysons la probabilité d'apparition. Cette analyse nous permet en particulier de mesurer la **sélectivité** des primitives utilisées et de valider la mise en correspondance des images médicales précédemment utilisées. Enfin, pour finir, nous identifions quatre point clés pour l'utilisation pratique des algorithmes précités avec des primitives génériques : il s'agit du calcul des invariants, du clustering et de trouver des solutions efficaces et correctes pour l'indexation et le plus proche voisin.

Le **chapitre 11** met en oeuvre l'une de ces techniques en biologie moléculaire pour comparer des structures tridimensionnelles de protéines et en extraire la partie commune. Étant basé sur les repères au lieu des points, cet algorithme réduit considérablement la complexité par rapport à l'état de l'art : de $O(n^3)$ à $O(n^2)$. Les résultats expérimentaux confirment la validité de l'approche et le gain en robustesse grâce à l'utilisation des repères.

Enfin, nous nous intéressons dans le **chapitre 12** aux problèmes de **modélisation** et en particulier au **recalage multiple**, qui vise à mettre dans le même repère plusieurs observations d'un même objet. Nous analysons plusieurs algorithmes sur les points et nous présentons des exemples d'application en modélisation à la fois pour la biologie moléculaire et pour l'imagerie médicale.

1.4 Contributions

Mises à part les synthèses sur les probabilités dans $\mathbb{R}^n$ (chap. 2) et sur la géométrie riemannienne (section 3.5), la première partie est dans l'ensemble originale. Dans la seconde partie, nous analysons en général les techniques connues utilisant les points et nous les généralisons aux primitives

quelconques. Dans le détail, les contributions majeures sont les suivantes :

- Chapitre 3 : la théorie mathématique des mesures invariantes est bien établie dans (Santalo, 1976). Nous en présentons à la section (3.3) une formalisation plus accessible en terme de représentation. Si la métrique invariante sur un groupe de Lie est une notion connue en géométrie riemannienne, la formalisation de la distance invariante par une « norme » (section 3.4) est nouvelle, et nous n'avons trouvé aucun travail sur les conditions d'existence d'une métrique invariante sur une variété (section 3.5). L'utilisation du lieu de coupure pour créer la carte principale semble également originale. Une partie des travaux de ce chapitre a fait l'objet de publications dans (Pennec et Ayache, 1996a; Pennec et Ayache, 1996b).
- Chapitre 4 : la densité de probabilité d'une primitive aléatoire est évoquée dans plusieurs travaux mathématiques, mais nous avons entièrement développé les propriétés de propagation présentées dans la section (4.1). L'espérance au sens de Fréchet est très peu connue, nous n'avons trouvé sa caractérisation comme un barycentre exponentiel que dans (Emery et Mokobodzki, 1991). L'algorithme que nous présentons pour son obtention à la section (4.3) est original, ainsi que la définition de la matrice de covariance et les résultats associés de la section (4.4).
- Chapitre 5 : les aspect statistiques sont des contributions inédites qui posent d'ailleurs plus de questions qu'elle n'en résolvent, en particulier pour les définitions d'une distribution gaussienne et de la distance de Mahalanobis.
- Chapitre 6 : l'idée d'une structure « orientée objet » générique pour gérer les primitives géométriques n'est pas neuve, mais elle n'était pas vraiment réalisable jusqu'à présent.
- Chapitre 7 : si les deux premières section, concernant les matrices de rotation et les quaternions, ne réalisent qu'une synthèse, la section (7.3) sur le vecteur rotation comporte des résultats nouveaux, et en particulier la dérivation de la composition. Nous nous sommes de plus attachés à résoudre efficacement toutes les instabilités numériques. Les résultats présentés à la section (7.5) sur les repères semi et non-orientés sont également entièrement nouveaux.
- Chapitre 8 : les algorithmes de recalage à partir d'appariements de primitives (section 8.2) sont originaux, ainsi que les algorithmes de fusion de primitives (8.3). La méthode de recalage robuste estimant de plus l'incertitude sur les données et la transformation est également une contribution.
- Chapitre 9 : la méthode statistique pour valider l'estimation de l'incertitude sur le recalage est, à notre connaissance, nouvelle et unique. Ce chapitre, avec quelques éléments des précédents a été publié dans (Pennec et Thirion, 1997; Pennec et Thirion, 1995). La section concernant les mouvements relatifs des du bassin dans les images IRM est également novatrice.
- Chapitre 10 : les résultats de statistiques précédemment développés nous permettent de formaliser correctement la gestion de l'erreur dans les algorithmes de reconnaissance et la notion de sélectivité des primitives. L'analyse du nombre moyen de faux positifs par intégration sur les transformations admissibles est nouvelle.
- Chapitre 11 : l'utilisation d'algorithmes de vision pour la biologie moléculaire est déjà présente dans (Fischer et al., 1992a; Fischer et al., 1992b), mais l'utilisation des repères au lieu des points permet de réduire la complexité de $O(n^3)$ à $O(n^2)$, ce qui est une contribution significative par rapport à l'état de l'art. Publication dans (Pennec et Ayache, 1998; Pennec et Ayache, 1994).

- Chapitre 12 : nous proposons une nouvelle technique pour recaler simultanément plusieurs objets formés de points en dimension quelconque et calculer l'objet moyen sans favoriser l'un des objets et en présence de multiples occlusions. L'inclusion de la probabilité d'observation d'un point dans le modèle ainsi obtenu est également originale. Publication dans (Pennec, 1996).

Première partie

Primitives géométriques et incertitude

Chapitre 2

Le problème de l'incertitude

« Il n'est pas certain que tout soit certain. »
B. Pascal, Pensées, 1670

Le but de ce chapitre est de faire le point sur les les techniques usuelles utilisées pour gérer les erreurs de mesures (i.e. l'incertitude) sur les points et les vecteurs, et de donner les bases nécessaires en probabilité pour pouvoir généraliser ces méthodes à d'autres types d'objets géométriques haut niveau, comme des droites, des points avec un vecteur unitaire, etc., objets que nous appellerons primitives géométriques.

Un rapide état de l'art fera apparaître l'importance et la prédominance de la gestion probabiliste de l'erreur, et nous examinerons donc ce qu'est un espace probabilisé, ce qu'est une variables aléatoire, et comment on peut la manipuler. La généralisation de la notion de *variable aléatoire* (dans $\mathbb{R}$) à celle de *vecteur aléatoire* (dans $\mathbb{R}^n$) est intéressante et nous donne un exemple de la marche à suivre pour étendre ces notions et les méthodes associées à nos primitives géométriques. Cependant, nous montrerons dans la dernière section que la généralisation hâtive de ces techniques fournit nombre de paradoxes, et qu'il faut être particulièrement précautionneux dans cette tâche : nous nous y emploierons dans les chapitres suivants.

2.1 Erreurs de mesure

« L'imprévisible est dans la nature des choses »
Mencius, Philosophe Confucéen, IV^{e}-III^{e} siècle av. J.C.

2.1.1 Provenance et influence

Les mesures que l'on peut effectuer en vision par ordinateur sur une image ne sont jamais parfaites et, de l'objet réel aux primitives que l'on mesure, on peut distinguer plusieurs sources d'erreur dans la chaîne de traitement :

- Les déformations introduites par le système de vision (que ce soit une caméra comme en vision classique, ou un scanner, ou encore une IRM), qui ne sont en général pas linéaires.

- La discrétisation de l'image, à la fois en positionnement (pixels) et en niveaux de gris.
- Les pré-traitements effectués sur l'image avant la reconnaissance : filtrages, extraction de contours... et l'extraction des primitives proprement dite. En particulier, il existe souvent une « échelle » privilégiée pour observer et mesurer les primitives, et le traitement de la même image à deux échelles différentes fournira un positionnement différent des primitives observables à ces deux échelles, mais aussi des primitives observables à l'une des deux échelles seulement.
- La variabilité du modèle et l'approximation de la transformation : c'est par exemple le cas en reconnaissance 2D d'objets 3D si l'on utilise une similitude pour approcher une projection orthogonale, ou lorsqu'on considère des transformations euclidiennes pour des objets très légèrement déformables (imagerie médicale).

Ces erreurs vont induire des changements importants dans les algorithmes de vision haut niveau. En effet, différentes mesures d'une même primitive positionnée exactement au même endroit dans plusieurs images seront différentes. Dans un problème de reconnaissance, considérer que deux primitives ou deux invariants sont identiques (à la discrétisation près) peut donc amener à discréditer beaucoup d'appariements, rendant par là même hasardeuse voire improbable la détection d'un objet pourtant présent dans l'image. On appelle cela un **faux négatif**.

D'un autre côté, dès que l'on autorise une certaine erreur sur les mesures (ne serait-ce que la discrétisation des tables de hachage ou celle des nombres réels), des primitives qui n'ont rien à voir avec l'objet recherché peuvent être appariées et ainsi « conspirer » pour amener à la reconnaissance d'un objet qui n'existe pas : c'est un **faux positif**. On notera que ceux-ci sont identifiables grâce à une vérification plus approfondie utilisant des informations supplémentaires (non utilisées dans la modélisation de l'image par primitives). Ces vérifications sont toutefois coûteuses en temps de calcul et peuvent induire une explosion de la complexité de l'algorithme.

Il est donc important de gérer l'erreur pour éviter les faux négatifs, mais de la faire au plus juste pour limiter la probabilité ou le nombre de faux positifs.

2.1.2 Erreur bornée

Une approche simple pour modéliser l'incertitude consiste à supposer que l'on connaît la position exacte[1] x d'un point (ou qu'on a pu l'approximer) et une borne ε sur la norme de l'erreur : nos mesures suivent alors le modèle

$$\hat{x} = x + \delta x \qquad \text{avec} \qquad \|\delta x\| \leq \varepsilon$$

On notera qu'en général il n'est pas supposé que la distribution de l'erreur soit uniforme sur la boule $\mathcal{B}(0, \varepsilon)$, mais c'est la distribution qui minimise l'information quand on ne connaît que la borne ε (voir sections (2.2.2.7) et (2.2.3.6)). Cette approche est utilisée intensivement dans (Grimson, 1990) pour la reconnaissance avec la technique des arbres d'interprétation et les calculs de probabilité de faux positif basés sur ce modèle pour l'algorithme de « geometric hashing » sont développés pour différents environnements dans (Grimson et al., 1994) et (Pennec, 1993a).

Toutefois, d'importants problèmes sont rencontrés dès que l'on veut manipuler nos primitives pour en extraire des informations plus synthétiques ou les recaler : comment propager cette borne sur l'incertitude dans les opérations utilisées ?

Afin d'éviter complètement les faux négatifs, il est tentant de calculer une borne conservative, c'est-à-dire stricte : si $y = f(x)$, alors $\|\hat{x} - x\| \leq \varepsilon_x$ implique que $\|\hat{y} - y\| \leq \varepsilon_y$. Les quelques études

1. Nous laissons de côté le problème philosophique de savoir s'il existe réellement une position exacte, dont nous ne pouvons de toute façon pas obtenir la valeur à partir d'images discrètes d'objets continus.

susnommées ont cependant montré que l'obtention d'une telle borne est un problème complexe et mène souvent à une surestimation importante de l'erreur, augmentant dangereusement la probabilité de faux-positifs. Par ailleurs, l'utilisation d'une borne sur la *norme* de l'erreur suppose une certaine isotropie de la zone d'erreur, ce qui est souvent abusif. Pour relâcher cette hypothèse, on peut affecter à chaque composante d'un vecteur une marge d'erreur indépendante : $|\hat{x}_i - x_i| \leq \varepsilon_{x_i}$ et on travaille alors avec des zones d'erreur qui sont des pavés de $\mathbb{R}^n$. C'est l'idée de base de l'arithmétique des intervalles (Moore, 1966). Les calculs restent cependant lourds et complexes dès que l'on utilise des fonctions non linéaires. De plus, si les erreurs de mesure sur les composantes sont indépendantes dans un repère particulier et induisent une zone d'erreur particulièrement agréable (allongée suivant une axe et plate suivant un autre, par exemple), ce n'est généralement pas le cas dans un autre repère ou après une transformation. Une étude expérimentale sur l'estimation des rotations de (Orr et al., 1991) montre d'ailleurs que la méthode probabiliste donne des résultats meilleurs, plus rapidement et avec une meilleure estimation de l'incertitude que la méthode basée sur les intervalles.

2.1.3 Erreur probabiliste

En supposant que l'on puisse acquérir autant d'images que l'on veut d'un objet, il serait (théoriquement) possible de caractériser toutes les valeurs possibles de la position d'une primitive, ou plus spécialement, comme on est dans le cadre continu, la distribution des mesures. Si la primitive est un point ou un vecteur x, un outil parfaitement adapté pour cela est de considérer que la mesure $\hat{x}$ de x est la réalisation d'un **vecteur aléatoire $\boldsymbol{x}$** et de caractériser la distribution de l'erreur par la densité de probabilité de ce vecteur aléatoire (abrégé par la suite en densité).

Cependant, d'un point de vue calculatoire ou informatique, il est peu aisé de manipuler des fonctions sur $\mathbb{R}^n$: on est donc amené à approximer cette densité par ses premiers moments (valeur moyenne et matrice de covariance par exemple) considérés habituellement comme suffisamment représentatifs. On peut alors dériver un certain nombre de règles de calcul relativement simples pour calculer la propagation de ces moments de manière exacte dans certaines opérations et de manière approchée dans le cas général. D'un point de vue statistique, on peut également exprimer les principales hypothèses sur les modèles de bruits en terme de moyenne et de covariance. On peut donc obtenir un ensemble cohérent d'outils probabilistes et statistiques pour travailler avec l'approximation au second ordre des vecteurs aléatoires.

2.2 Probabilités dans $\mathbb{R}^n$

« Le hasard n'est que la mesure de notre ignorance. Les phénomènes fortuits sont, par définition, ceux dont nous ignorons les lois. »
H. Poincaré, Calcul des probabilités, 1912

Cette section rappelle dans un premier temps un certain nombre de notions de base de probabilité ainsi que quelques problèmes de statistiques. Concernant les probabilités, on pourra consulter par exemple (Neveu, 1990; Papoulis, 1991; Pelat, 1992) pour de plus amples développements. Pour la propagation des moments, on pourra se reporter à (Zhang et Faugeras, 1992; Faugeras, 1993; Csurka et al., 1995).

2.2.1 Espaces probabilisés

Un espace probabilisé est un espace Ω d'événements élémentaires auquel on adjoint une tribu $\mathcal{B}$ qui contient les parties de Ω et une mesure de probabilité Pr sur les éléments de cette tribu.

2.2.1.1 Évènements élémentaires

Les événements élémentaires $\omega \in \Omega$ sont les résultats possibles d'une expérience aléatoire. L'espace Ω regroupe l'ensemble des résultats possibles de l'expérience. Dans le cas d'un jet de dé, le résultat sera par exemple un chiffre entre 1 et 6 : $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si l'expérience consiste à jeter plusieurs dés (différentiables), le résultat sera un n-uplet de tels chiffres.

Dans notre cas, l'expérience aléatoire sera plutôt la mesure d'une primitive géométrique, et l'ensemble des résultats possibles sera donc l'ensemble de ces primitives géométriques. Ainsi, si l'on mesure un point de $\mathbb{R}^n$, on a $\Omega = \mathbb{R}^n$.

2.2.1.2 Évènements et parties

Plutôt que de travailler sur chacun des résultats possibles, on caractérise un **événement** par un ensemble de résultats possibles (un sous-ensemble ou une partie de Ω) : la somme de deux jets de dés est supérieure à 10, le premier jet de dé est supérieur au second, ou dans notre cas : la mesure de x est égale à y à une distance ε près... sont des événements caractérises par des sous ensembles $A, B \ldots \subset \Omega$. Dans cette optique, les événements élémentaires sont évidemment les singletons $\{\omega\}$.

Les opérations logiques qui nous intéressent sur les événements sont très proches des opérateurs ensemblistes : à tout événement A, on peut associer son contraire $\bar{A} = \Omega - A$, la réalisation de A **ou** celle de B est l'événement $A \cup B$ tandis que la réalisation simultanée de A **et** de B est l'événement $A \cap B$. L'événement impossible est l'ensemble vide $\varnothing$ et un événement certain est représenté par l'ensemble Ω au complet.

Deux relations importantes entre les événements sont encore à noter : $A \subset B$ exprime que A implique B (si le résultat de l'expérience est dans A, alors il est aussi dans B), et $A \cap B = \varnothing$ signifie que les événements A et B sont incompatibles et s'excluent mutuellement.

Un système exhaustif d'événements est une partition de Ω, c'est-à-dire une suite dénombrable A_i d'événements deux à deux incompatibles ($A_i \cap A_j = \varnothing$ si $i \neq j$) couvrant toutes les possibilités : $\cup_i A_i = \Omega$.

2.2.1.3 Tribu d'événements

A priori, il pourrait sembler naturel de considérer que toute partie de Ω représente un événement, mais cela n'est possible que si Ω est dénombrable. Pour des espaces plus grands, il est impossible de définir des probabilité intéressantes sur toutes les parties de Ω. On est donc réduit à se cantonner à une famille de parties, et l'on imposera de plus que cette famille soit stable par les opérations de base : on appelle **tribu** ou σ-algèbre $\mathcal{A}$ sur Ω une famille de parties de Ω contenant l'événement impossible $\varnothing$, stable par complémentation, stable par réunion et intersection dénombrable.

La construction explicite des tribus sur un espace quelconque n'est pas une chose facile, mais considérant une famille $\mathcal{C}$ de parties, on peut monter qu'il existe une plus petite tribu qui contienne cette famille, que l'on l'appelle tribu engendrée par la famille $\mathcal{C}$. Ce procédé de construction peu explicite permet toutefois de définir sur tout espace topologique, et en particulier métrique, la **tribu borélienne** de Ω comme la tribu engendrée par la famille des ouverts de l'espace.

2.2.1.4 Mesure de probabilité

Pour compléter la description de notre expérience aléatoire, il nous reste à quantifier la probabilité d'occurrence ou d'observation d'une événement. On appelle **mesure** sur notre espace Ω muni de la tribu $\mathcal{A}$ une fonction Pr σ-additive de $\mathcal{A}$ dans $\mathbb{R}^+$: si A_i est une suite dénombrable d'événements deux à deux disjoint, on doit avoir

$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i)$$

Pour que cette mesure soit appelée **probabilité**, il faut de plus qu'elle soit normalisée : $\Pr(\Omega) = 1$. Le triplet $(\Omega, \mathcal{A}, \Pr)$ est alors un espace probabilisé.

On dit qu'un ensemble (ou un événement) A est de **mesure nulle** si sa probabilité est nulle $\Pr(A) = 0$. A l'opposé, cet événement est presque certain si $\Pr(\Omega - A) = 0$. Une propriété est vraie presque partout si l'ensemble des points où elle n'est pas réalisée est de mesure nulle.

Une relation importante lie les probabilités de deux événements quelconques :

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Cette relation est à l'origine de la sous additivité des probabilités et elle met en valeur le couplage des événements A et B, qui peut être mesuré par l'événement $A \cap B$ (réalisation simultanée de A *et* de B). Si les événements sont disjoints ($\Pr(A \cap B) = 0$), alors on retrouve l'additivité.

2.2.1.5 Probabilités conditionnelles et règle de Bayes

Dans l'examen des expériences aléatoires, il est souvent intéressant de prendre en compte une connaissance partielle sur le résultat. On cherche par exemple à étudier la probabilité d'un événement A *sachant que* l'événement B est réalisé (par hypothèse ou par observation). On note $\Pr(A|B)$ cette **probabilité conditionnelle**, et en supposant que l'événement conditionnant B ne soit pas de mesure nulle, on a la relation :

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

qui exprime en quelque sorte la normalisation due à la réduction de l'espace des résultats de Ω à B. La probabilité de l'intersection s'écrit alors :

$$\Pr(A \cap B) = \Pr(A|B).\Pr(B) = \Pr(B|A).\Pr(A)$$

d'où l'on obtient la règle de Bayes :

$$\Pr(A|B) = \frac{\Pr(B|A).\Pr(A)}{\Pr(B)}$$

On dit que deux événements A et B sont **indépendants** si la connaissance de l'un n'apporte aucune information sur l'autre :

$$\Pr(A|B) = \Pr(A) \quad \text{et} \quad \Pr(B|A) = \Pr(B) \quad \text{d'où} \quad \Pr(A \cap B) = \Pr(A).\Pr(B)$$

C'est une sorte de notion d'orthogonalité sur les événements : par exemple, savoir qu'un point aléatoire dans le plan a pour abscisse x ne nous apporte pas d'information sur son ordonnée y.

2.2.2 Variable aléatoire (observable)

Soit $\omega \in \Omega$ un événement élémentaire et $\boldsymbol{x} = x(\omega)$ une variable réelle dépendant de l'événement ou x est une application de Ω dans $\mathbb{R}$. La variable $\boldsymbol{x}$ est en général un paramètre mesurable de notre expérience, que l'on appelle en physique une **observable**. Pour que l'on puisse, a proprement parler, appeler $\boldsymbol{x}$ une **variable aléatoire réelle** (ou v.a.r. en abrégé), if faut que, quelque soient les réels a et b, on puisse parler de la probabilité $\Pr(a < x < b)$. Il faut donc que $x^{(-1)}(]a, b[)$ soit un événement de la tribu $\mathcal{A}$ considérée sur Ω. On appelle plus généralement fonction borélienne une fonction qui respecte ainsi la structure de tribu borélienne, et toutes les variables aléatoires ou changements de variables que nous considéreront dans la suite seront de telles fonction. Afin d'alléger l'écriture, on note souvent de façon identique l'application et le réel qui en est le résultat. On écrira donc indifféremment une v.a.r. : $\boldsymbol{x}$ ou $x(\omega)$.

Le formalisme des variables aléatoires permet de s'abstraire (quasiment) totalement de l'espace probabilisé d'origine Ω, puisque l'on travaille maintenant dans $\mathbb{R}$, et d'associer à chaque variable $\boldsymbol{x}$ ou $\boldsymbol{y}$ une mesure de probabilité différente (mais théoriquement liées par le fait qu'elles proviennent d'un même espace probabilisé). Ainsi, on oublie souvent de spécifier l'espace Ω pour ne travailler que sur des variables aléatoires (réelles) de probabilité données.

2.2.2.1 Densité de probabilité

On dit qu'une variable aléatoire $\boldsymbol{x}$ possède ou suit une **densité de probabilité** $p_{\boldsymbol{x}}$ sur $\mathbb{R}$ si, pour tout intervalle $]a, b[$, on a :

$$\Pr(\boldsymbol{x} \in]a, b[) = \int_a^b p_{\boldsymbol{x}}(y).dy$$

On parle également de distribution de probabilité, en sachant que $p_{\boldsymbol{x}}$ peut parfois être une distribution et non une fonction, en incluant par exemple des Diracs. Nous nous cantonnerons ici à des densités qui sont des fonctions, le plus souvent continues, le cas des distributions étant plus difficilement généralisable sur les variétés différentielles.

Notons que dans la définition ci-dessus, la contrainte originelle $\Pr(\Omega) = 1$ et le fait que la probabilité soit positive se traduisent en

$$p_{\boldsymbol{x}}(y) > 0 \qquad \text{et} \qquad \int_{-\infty}^{\infty} p_{\boldsymbol{x}}(y).dy = 1$$

2.2.2.2 Espérance

Nous avons maintenant une mesure Pr sur l'espace Ω, et des fonction à valeurs réelles sur cet espace : les variables aléatoires réelles. En nous restreignant aux fonction mesurable, nous pouvons donc intégrer ces fonctions : c'est la définition de l'espérance mathématique.

$$\mathbf{E}\,[\,\boldsymbol{x}\,] = \int_\Omega x(\omega).\Pr(d\omega) = \int \boldsymbol{x}.\mathrm{dPr} \tag{2.1}$$

Si la v.a.r. admet une densité de probabilité $p_{\boldsymbol{x}}$, cette définition se ramène à une intégration classique dans $\mathbb{R}$, et l'on peut encore une fois s'affranchir de l'espace d'origine Ω :

$$\mathbf{E}\,[\,\boldsymbol{x}\,] = \int_{\mathbb{R}} y.p_{\boldsymbol{x}}(y).dy$$

Notons cependant que l'on peut, grâce à cet opérateur, retrouver la mesure de probabilité induite par la variable aléatoire réelle sur $\mathbb{R}$. Si $\mathcal{U}$ est un ouvert de $\mathbb{R}$:

$$\mathbf{E}\,[\,\mathbf{1}_{\mathcal{U}}(\boldsymbol{x})\,] = \Pr(\mathcal{U})$$

L'espérance est un opérateur linéaire sur les variables aléatoires : si λ_1 et λ_2 sont deux réels et $\boldsymbol{x}$, $\boldsymbol{y}$ deux v.a.r., on a :

$$\mathbf{E}\,[\,\lambda_1.\boldsymbol{x} + \lambda_2.\boldsymbol{y}\,] = \lambda_1.\mathbf{E}\,[\,\boldsymbol{x}\,] + \lambda_2.\mathbf{E}\,[\,\boldsymbol{y}\,]$$

Notons encore que si φ est une fonction (borélienne mesurable) de $\mathbb{R}$ dans $\mathbb{R}$, l'espérance de la variable composée $\varphi(\boldsymbol{x})$ est donnée par :

$$\mathbf{E}\,[\,\varphi(\boldsymbol{x})\,] = \int_{\mathbb{R}} \varphi(y).p_{\boldsymbol{x}}(y).dy$$

2.2.2.3 Information et entropie

Le but final d'une observation dans un système aléatoire est de d'extraire les données les plus pertinentes, c'est-à-dire le plus d'**information** possible ou encore réduire au maximum l'incertitude de l'observation. A partir de la probabilité $p_{\boldsymbol{x}}$ d'une v.a.r. $\boldsymbol{x}$ on peut avoir une idée intuitive de l'incertitude de notre observation : plus la distribution est étalée, plus le résultat de l'expérience est incertain. Shannon a formalisé cette notion intuitive en 1948 (voir aussi (Jaynes, 1996, ch. 11)) et a montré que la seule mesure de l'incertitude adaptée (à un facteur d'échelle près) est l'entropie :

$$\mathbf{H}\,[\,\boldsymbol{x}\,] = -\mathbf{E}\,[\,\log(p_{\boldsymbol{x}}(\boldsymbol{x}))\,] = -\int \log(p_{\boldsymbol{x}}(x)).p_{\boldsymbol{x}}(x).dx$$

L'apparition du logarithme dans cette formule provient de la condition d'additivité imposée sur la notion d'incertitude : si deux variables aléatoires $\boldsymbol{x}$ et $\boldsymbol{y}$ sont indépendantes, leur densité conjointe est $p_{(\boldsymbol{x},\boldsymbol{y})}(x,y) = p_{\boldsymbol{x}}(x).p_{\boldsymbol{y}}(y)$ (voir section 2.2.3.7), et l'entropie conjointe est simplement l'addition des entropies :

$$\mathbf{H}\,[\,(\boldsymbol{x},\boldsymbol{y})\,] = -\int \left(\log\left(p_{\boldsymbol{x}}(x)\right) + \log\left(p_{\boldsymbol{y}}(y)\right)\right).p_{\boldsymbol{x}}(x).p_{\boldsymbol{y}}(y).dx.dy = \mathbf{H}\,[\,\boldsymbol{x}\,] + \mathbf{H}\,[\,\boldsymbol{y}\,]$$

Dans ce manuscrit, nous préférons utiliser la notion opposée d'information :

$$\mathbf{I}\,[\,\boldsymbol{x}\,] = -\mathbf{H}\,[\,\boldsymbol{x}\,] = \int \log(p_{\boldsymbol{x}}(x)).p_{\boldsymbol{x}}(x).dx$$

et conserver le terme incertitude pour les caractéristiques numériques. Il est difficile de donner une interprétation à une valeur donnée de l'information, excepté que plus elle est petite, plus on se rapproche de la distribution uniforme. On regarde plutôt en général la différence d'information entre une distribution *a priori* et une distribution *a posteriori*, ce qui quantifie l'information que l'on a gagné lors de notre traitement.

2.2.2.4 Moments d'une variable aléatoire

Une v.a.r. $\boldsymbol{x}$ est entièrement décrite par sa densité de probabilité (si elle est possède une), mais cette information est souvent trop riche pour être appréhendée ou bien traitée de manière

informatique. On souhaite donc caractériser cette v.a.r. par un ensemble de paramètres restreints, et l'on utilise le plus souvent les moments, définis à partir de l'espérance par :

$$m_{\boldsymbol{x}}^k = \mathbf{E}\left[\, \boldsymbol{x}^k \,\right]$$

le moment d'ordre 0 est toujours égal à 1 (contrainte de normalisation) mais les moments d'ordre supérieur n'existent pas forcément.

Moyenne Le moment d'ordre 1 est une valeur centrale de la distribution particulièrement important est porte le nom de **moyenne** ou d'espérance de la v.a.r. $\boldsymbol{x}$:

$$\overline{x} = \mathbf{E}\left[\, \boldsymbol{x} \,\right]$$

C'est en quelque sorte la « meilleure » approximation de la distribution par une valeur déterministe (ou constante).

Variance Afin de pouvoir comparer les distributions centrées autour de valeurs différentes, on centre en général les moments d'ordre supérieur autour de la moyenne :

$$\bar{m}_{\boldsymbol{x}}^k = \mathbf{E}\left[\, (\boldsymbol{x} - \mathbf{E}\left[\, \boldsymbol{x} \,\right])^k \,\right]$$

Parmi ces caractéristiques numériques, on distingue plus particulièrement le moment centré d'ordre 2 que l'on appelle **variance** :

$$\sigma_{\boldsymbol{x}}^2 = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x})^2 \,\right]$$

et qui indique la dispersion de la distribution autour de la valeur centrale.

2.2.2.5 Médiane et quantiles

D'autre caractéristiques numériques sont encore utilisées : la **médiane** se définit comme l'élément médian ou milieu d'une distribution. On a donc :

$$\Pr(\boldsymbol{x} \leq x_{med}) = \Pr(\boldsymbol{x} > x_{med}) = 0.5$$

Notons que la médiane est comme la moyenne une valeur centrale de la distribution.

Il n'est pas difficile de montrer que la médiane réalise le minimum du critère suivant, ce qui permettra de généraliser cette notion aux vecteurs et aux éléments aléatoires :

$$x_{med} = \arg\min_y \mathbf{E}\left[\, |\boldsymbol{x} - y| \,\right]$$

La médiane sépare le support de la distribution en deux parties de probabilités égales. On appelle **quantile** x_α la valeur qui sépare le support en deux parties inégales de probabilités $1-\alpha$ du côté inférieur et α de l'autre :

$$\Pr(\boldsymbol{x} < x_\alpha) = 1 - \alpha \qquad \text{et} \qquad \Pr(\boldsymbol{x} > x_\alpha) = \alpha$$

Cette notion de quantile fait cependant appel à une relation d'ordre totale qui ne sera pas facilement généralisable dans des espaces de dimension supérieure.

2.2.2.6 Loi gaussienne

Supposons maintenant que, par hypothèse ou par mesure, nous ne connaissions que la moyenne $\overline{x}$ et l'écart-type σ d'une v.a.r. $\boldsymbol{x}$. Pour pouvoir utiliser un certain nombre d'outils probabilistes et statistiques, nous avons besoin d'assigner à $\boldsymbol{x}$ une densité de probabilité. En l'absence d'information supplémentaire, quelle fonction $p(x)$ choisir ? Celle-ci doit vérifier les conditions suivantes :

- C'est une densité de probabilité : $p(x) \geq 0$ et $\int p(x).dx = 1$,
- de moyenne μ : $\int x.p(x).dx = \mu$,
- et de variance σ^2 : $\int (x-\mu)^2.p(x).dx = \sigma^2$.

Il est raisonnable de choisir parmi les fonctions qui vérifient ces propriétés celle qui est la moins informative, c'est-à-dire qui minimise l'information $\mathbf{I}\left[\, \boldsymbol{x} \mid \overline{x} = \mu,\ \sigma^2_{\mathbf{x}} = \sigma^2 \,\right]$. Grâce au calcul des variations, c'est un exercice que l'on peut résoudre en écrivant le lagrangien :

$$\Lambda(p) = \int \left(p.\log(p) + \alpha.p + \beta.x.p + \frac{\gamma}{2}.(x-\mu)^2.p \right).dx$$

où α, β et γ sont les multiplicateurs de Lagrange permettant de prendre en compte les contraintes. L'équation d'Euler caractérisant l'optimum est obtenue en dérivant par rapport à p sous le signe somme et en égalant à zéro : $\log(p) + 1 + \alpha + \beta.x + \frac{\gamma}{2}.(x-\mu)^2 = 0$, ce qui se réécrit :

$$p(x) = \exp(-1-\alpha)\,.\exp\left(-\beta.x - \frac{\gamma}{2}.(x-\mu)^2\right)$$

En reportant dans les contraintes, et après quelques intégrations (que Maple fait très bien), on trouve les valeurs des multiplicateurs :

$$\alpha = \frac{\log(2\,\pi)}{2} + \log(\sigma) - 1 \qquad \beta = 0 \qquad \gamma = \frac{1}{\sigma^2}$$

ce qui donne la densité de probabilité de la loi gaussienne $N(\mu, \sigma^2)$:

$$p(x) = \frac{1}{\sigma\sqrt{2\,\pi}}.\exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right) \tag{2.2}$$

Information de la gaussienne Il est intéressant de calculer l'information de cette distribution. En notant k la constante de normalisation, on a :

$$\mathbf{I}\left[\, N(\mu,\sigma^2) \,\right] = -\log(k) - \frac{1}{2.\sigma^2}.\int (x-\mu)^2.p(x).dx$$

Cette dernière intégrale étant justement la variance, on trouve

$$\mathbf{I}\left[\, N(\mu,\sigma^2) \,\right] = \frac{-1-\log(2.\pi)}{2} - \log(\sigma) \qquad \text{soit} \qquad \mathbf{I}\left[\, N(\mu,\sigma^2) \,\right] \propto -\log(\sigma) \tag{2.3}$$

L'information est donc inversement proportionnelle au log de la variance.

2.2.2.7 Loi uniforme

Si la loi normale correspond classiquement à une erreur de mesure, la loi uniforme correspond à l'idée de hasard au sens commun, c'est-à-dire d'étalement maximal. Afin d'avoir une densité de probabilité, on se limite habituellement à un ensemble mesurable, ou plus simplement dans le cas d'une v.a.r. à un intervalle $\mathcal{U} =]a, b[$ (comme la taille d'une image, par exemple). On définit alors la loi uniforme sur un tel ensemble $\mathcal{U}$ par la densité :

$$p(x) = \frac{\mathbf{1}_{\mathcal{U}}(x)}{\int_{\mathcal{U}} dx} \qquad \text{où} \qquad \mathbf{1}_{\mathcal{U}}(x) = \begin{cases} 1 & \text{si} \quad x \in \mathcal{U} \\ 0 & \text{sinon} \end{cases} \tag{2.4}$$

Notons la loi uniforme sur $\mathcal{U}$ minimise l'information conditionnelle : $\mathbf{I}\,[\,\mathbf{x} \,|\, \mathbf{x} \in \mathcal{U}\,]$. Les caractéristiques numériques sur un intervalle $\mathcal{U} =]a, b[$ sont simples:

$$\overline{x} = \frac{a+b}{2} \qquad \text{et} \qquad \sigma^2 = \frac{(a-b)^2}{12}$$

2.2.3 Vecteur aléatoire

Supposons maintenant que l'on ait un ensemble de mesures simultanées sur notre expérience aléatoire. Si l'on range ces n variables aléatoires réelles $\boldsymbol{x}_i$ dans un vecteur $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots \boldsymbol{x}_n)^{\mathrm{T}}$, on obtient naturellement un **vecteur aléatoire** (abréviation v.a.), qui est donc une application $\boldsymbol{x} = x(\omega)$ de Ω dans $\mathbb{R}^n$ dont les composantes sont des v.a.r.. Observons que cette dernière condition équivaut à imposer que, pour tout ouvert $\mathcal{U}$ de $\mathbb{R}^n$, $\{\boldsymbol{x} \in \mathcal{U}\}$ est un événement de la tribu $\mathcal{A}$ sur l'espace d'origine Ω.

2.2.3.1 Densité de probabilité

C'est la généralisation directe de la densité des v.a.r. On dit qu'un v.a. $\boldsymbol{x}$ possède ou suit une densité de probabilité $p_{\boldsymbol{x}}$ sur $\mathbb{R}^n$ si, pour tout ouvert $\mathcal{U} \in \mathbb{R}^n$, on a :

$$\Pr(\boldsymbol{x} \in \mathcal{U}) = \int_{\mathcal{U}} p_{\boldsymbol{x}}(y).dy$$

Les contraintes de normalisation se traduisent trivialement par :

$$p_{\boldsymbol{x}}(y) > 0 \qquad \text{et} \qquad \int_{\mathbb{R}^n} p_{\boldsymbol{x}}(y).dy = 1$$

2.2.3.2 Espérance d'une observable

Soit φ une fonction borélienne (intégrable) de $\mathbb{R}^n$ dans $\mathbb{R}$, et $\boldsymbol{x}$ un vecteur aléatoire de densité $p_{\boldsymbol{x}}$. Alors $\varphi(\boldsymbol{x})$ est une variable aléatoire réelle dont on peut calculer l'espérance :

$$\mathbf{E}\,[\,\varphi(\boldsymbol{x})\,] = \int_{\mathbb{R}^n} \varphi(y).p_{\boldsymbol{x}}(y).dy$$

En particulier, on peut retrouver la mesure de probabilité induite par le v.a. $\boldsymbol{x}$ sur $\mathbb{R}^n$ grâce à l'espérance de la fonction indicatrice. Si $\mathcal{U}$ est un ouvert de $\mathbb{R}^n$:

$$\mathbf{E}\,[\,\mathbf{1}_{\mathcal{U}}(\boldsymbol{x})\,] = \Pr(\mathcal{U})$$

2.2.3.3 Information et entropie

La définition est une transposition directe du cas réel :

$$\mathbf{I}[\,\boldsymbol{x}\,] = -\mathbf{H}[\,\boldsymbol{x}\,] = \int_{\mathbb{R}^n} \log(p_{\boldsymbol{x}}(x)).p_{\boldsymbol{x}}(x).dx$$

2.2.3.4 Moments de la densité

En ensemble particulier de fonctions réelles est formé par les puissances des composantes, et l'espérance de ces fonctions défini les moments de la distributions. Soit $k = (k_1, \ldots k_n)$ un vecteur d'entiers. Les moments sont alors :

$$m^k_{\boldsymbol{x}} = \mathbf{E}\left[\, \boldsymbol{x}_1^{k_1}.\boldsymbol{x}_2^{k_2} \ldots \boldsymbol{x}_n^{k_n} \,\right]$$

L'ordre du moment est maintenant la somme des puissances. Il n'y a donc qu'un seul moment d'ordre 0, et il est toujours égal à 1 (contrainte de normalisation). Les moments d'ordre supérieur n'existent pas forcément.

Moyenne ou espérance d'un vecteur aléatoire Il y a n moments d'ordre 1, correspondants à $\overline{x}_1, \ldots \overline{x}_n$. Ordonnés dans un vecteurs, il forment un vecteur $\overline{x}$ que l'on appelle **moyenne** ou d'espérance du v.a. $\boldsymbol{x}$:

$$\overline{x} = \mathbf{E}[\,\boldsymbol{x}\,] = (\mathbf{E}[\,\boldsymbol{x}_1\,], \ldots, \mathbf{E}[\,\boldsymbol{x}_n\,])^{\mathrm{T}} = (\overline{x}_1, \ldots, \overline{x}_n)$$

Notons que l'espérance d'un vecteur aléatoire (ici définie) et l'espérance d'une fonction réelle de ce vecteur sont deux notions distinctes : le premier opérateur est à valeur dans $\mathbb{R}^n$ et le second dans $\mathbb{R}$. Les propriétés de linéarité de l'espace vectoriel $\mathbb{R}^n$ nous permettent cependant de généraliser la notion d'intégrale à valeur dans $\mathbb{R}$ à une intégrale à valeur dans $\mathbb{R}^n$: l'intégrale d'un vecteur étant le vecteur de l'intégrale des composantes. Ceci permet d'écrire l'espérance comme pour le cas réel :

$$\mathbf{E}[\,\boldsymbol{x}\,] = \int_{\mathbb{R}^n} y.p_{\boldsymbol{x}}(y).dy = \int_{\Omega} x(\omega).\Pr(d\omega)$$

Cette façon de définir l'espérance d'un vecteur aléatoire n'est utilisable que dans les espaces de Banach (espaces vectoriel normés complets) et l'intégrale correspondante est **l'intégrale de Pettis** (voir par exemple (Grenander, 1981, chap. 1)).

Matrice de covariance Les moments centrés sont alors définis par

$$m^k_{\boldsymbol{x}} = \mathbf{E}\left[\, (\boldsymbol{x}_1 - \overline{x}_1)^{k_1}.(\boldsymbol{x}_2 - \overline{x}_2)^{k_2} \ldots (\boldsymbol{x}_n - \overline{x}_n)^{k_n} \,\right]$$

Si l'on regarde les moments (centrés) d'ordre 2, on peut en dénombrer n^2 ne faisant intervenir à chaque fois qu'une ou deux composantes : on les nomme respectivement **variances** et **covariances**. Il peuvent tous s'exprimer sous la forme suivante :

$$\sigma^2_{ij} = \mathbf{E}[\,(\boldsymbol{x}_i - \overline{x}_i).(\boldsymbol{x}_j - \overline{x}_j)\,]$$

On les rassemble en général dans une matrice symétrique $\Sigma_{\boldsymbol{xx}}$ appelée matrice de variance-covariance ou covariance en abrégé. Une fois encore, les propriétés d'espace vectoriel nous permettent de généraliser les notions d'intégrale et d'espérance à valeur dans l'espace des matrices et d'écrire :

$$\Sigma_{\boldsymbol{xx}} = \mathbf{E}[\,(\boldsymbol{x} - \overline{x}).(\boldsymbol{x} - \overline{x})^{\mathrm{T}}\,] = \int_{\mathbb{R}^{n^2}} (y - \overline{x}).(y - \overline{x})^{\mathrm{T}}.p_{\boldsymbol{x}}(y).dy$$

2.2.3.5 Loi gaussienne

Les moments d'ordres supérieur peuvent être exprimés de manière similaire mais nécessitent l'emploi de tenseurs pour les regrouper. Pour des applications numériques, on se contente en général de la moyenne et de la covariance. Cependant, pour certaines estimations (de type maximum de vraisemblance, par exemple), nous avons besoin de supposer une densité de probabilité ayant ces caractéristiques numériques. Pour trouver la densité qui minimise l'information conditionnelle $\mathbf{I}\left[\, \boldsymbol{x} \mid \overline{x} = \mu,\ \Sigma_{\boldsymbol{xx}} = \Sigma \,\right]$, on écrit le lagrangien :

$$\Lambda(p) = \int \left(p.\log(p) + \alpha.p + \langle \beta \mid x \rangle .p + \frac{1}{2}.(x-\mu)^{\mathrm{T}}.\Gamma.(x-\mu).p \right) .dx$$

où les multiplicateurs de Lagrange sont cette fois un réel α pour la contrainte de normalisation, un vecteur β pour la contrainte sur la moyenne, et une matrice Γ qui se révélera symétrique pour la contrainte sur la covariance.

En écrivant l'équation d'Euler pour l'optimum du lagrangien et en utilisant les contraintes (pour un calcul détaillé, voir la section 5.2.3), on trouve comme valeur des multiplicateurs :

$$\alpha = \frac{n}{2}\log(2\,\pi) + \frac{1}{2}\log(\det(\Sigma)) - 1 \qquad \beta = 0 \qquad \Gamma = \Sigma^{(-1)}$$

ce qui donne la densité de probabilité de la loi gaussienne $N(\mu, \Sigma)$:

$$p(x) = \frac{1}{\sqrt{(2\,\pi)^n.\det(\Sigma)}}.\exp\left(-\frac{(x-\mu)^{\mathrm{T}}.\Sigma^{(-1)}.(x-\mu)}{2} \right) \tag{2.5}$$

Information de la gaussienne L'information de cette distribution est intéressante pour nous indiquer sa pertinence. On note k la constante de normalisation :

$$\mathbf{I}\left[\, N(\mu,\sigma^2) \,\right] = -\log(k) - \frac{1}{2}.\int (x-\mu)^{\mathrm{T}}.\Sigma^{(-1)}.(x-\mu).p(x).dx$$

En notant que $x^{\mathrm{T}}.V.y = \mathrm{Tr}(V.y.x^{\mathrm{T}})$, on peut sortir la covariance de l'intégrale :

$$\mathbf{I}\left[\, N(\mu,\sigma^2) \,\right] = -\log(k) - \mathrm{Tr}\left(\frac{\Sigma^{(-1)}}{2}.\int (x-\mu).(x-\mu)^{\mathrm{T}}.p(x).dx \right)$$

L'intégrale donne alors la covariance Σ et comme $\mathrm{Tr}(\Sigma^{(-1)}.\Sigma) = \mathrm{Tr}(I_n) = n$, on obtient

$$\mathbf{I}\left[\, N(\mu,\Sigma) \,\right] = \frac{n}{2}\left(-1 - \log(2.\pi)\right) - \frac{1}{2}\log\left(\det(\Sigma)\right) \quad \propto \quad \log\left(\det(\Sigma^{(-1)})\right) \tag{2.6}$$

L'incertitude croit donc linéairement avec le logarithme du volume de la région de confiance délimitée par l'ellipsoïde d'équation $(x-\mu)^{\mathrm{T}}.\Sigma^{(-1)}.(x-\mu) = \chi^2$.

2.2.3.6 Loi uniforme

La loi uniforme sur un ouvert $\mathcal{U}$ de $\mathbb{R}^n$ est donnée par la densité :

$$p(x) = \mathbf{1}_{\mathcal{U}}(x) \ \Big/ \int_{\mathcal{U}} dx \tag{2.7}$$

Notons cette loi minimise l'information conditionnelle : $\mathbf{I}\left[\, \mathbf{x} \mid \mathbf{x} \in \mathcal{U} \,\right]$.

2.2.3.7 Probabilités marginales

En statistiques, on supposera souvent plusieurs mesures simultanées de v.a. pouvant être corrélées ou indépendantes. Nous examinons ici le cas de deux v.a. $\boldsymbol{x}$ et $\boldsymbol{y}$ de dimension m et n, la généralisation à un nombre plus grand de mesures étant similaire. On peut ranger ces deux variables dans un seul vecteur aléatoire $\boldsymbol{z}^{\mathrm{T}} = (\boldsymbol{x}^{\mathrm{T}}, \boldsymbol{y}^{\mathrm{T}})$, dont la densité $p_{\boldsymbol{z}}(z) = p_{(\boldsymbol{x},\boldsymbol{y})}(x, y)$ est appelée **densité conjointe**. Les **densités marginales** de $\boldsymbol{x}$ et $\boldsymbol{y}$ sont obtenue par intégration partielle :

$$p_{\boldsymbol{x}}(x) = \int_{\mathbb{R}^n} p_{(\boldsymbol{x},\boldsymbol{y})}(x, y).dy \qquad \text{et} \qquad p_{\boldsymbol{y}}(y) = \int_{\mathbb{R}^m} p_{(\boldsymbol{x},\boldsymbol{y})}(x, y).dx$$

Les moments sont reliés par les relations

$$\overline{z} = \mathbf{E}\left[\, \boldsymbol{z} \,\right] = \left| \begin{array}{c} \overline{x} \\ \overline{y} \end{array} \right. \qquad \text{et} \qquad \Sigma_{\boldsymbol{z}\boldsymbol{z}} = \left[\begin{array}{cc} \Sigma_{\boldsymbol{x}\boldsymbol{x}} & \Sigma_{\boldsymbol{x}\boldsymbol{y}} \\ \Sigma_{\boldsymbol{y}\boldsymbol{x}} & \Sigma_{\boldsymbol{y}\boldsymbol{y}} \end{array} \right]$$

où $\Sigma_{\boldsymbol{x}\boldsymbol{y}} = \Sigma_{\boldsymbol{y}\boldsymbol{x}}^{\mathrm{T}}$ est la matrice de covariance croisée

$$\Sigma_{\boldsymbol{x}\boldsymbol{y}} = \mathbf{E}\left[\, (\boldsymbol{x} - \overline{x}).(\boldsymbol{y} - \overline{y})^{\mathrm{T}} \,\right]$$

Si l'on note (de manière un peu abusive) $\Pi_{\boldsymbol{x}} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} = [I_m\,;\, 0]$ la matrice de projection qui fait passer du vecteur aléatoire $\boldsymbol{z}$ au vecteur $\boldsymbol{x}$, on peut exprimer directement les premiers moments de ce dernier par :

$$\overline{x} = \Pi_{\boldsymbol{x}}.\overline{z} \qquad \text{et} \qquad \Sigma_{\boldsymbol{x}\boldsymbol{x}} = \Pi_{\boldsymbol{x}}.\Sigma_{\boldsymbol{z}\boldsymbol{z}}.\Pi_{\boldsymbol{x}}^{\mathrm{T}}$$

La dérivation est similaire pour les moments de $\boldsymbol{y}$.

Mesures indépendantes Notons que la densité conjointe $p_{\boldsymbol{z}}(z)$ se factorise en

$$p_{(\boldsymbol{x},\boldsymbol{y})}(x, y) = p_{\boldsymbol{x}}(x).p_{\boldsymbol{y}}(y)$$

si et seulement si les v.a. $\boldsymbol{x}$ et $\boldsymbol{y}$ sont indépendants, ce qui sera souvent supposé en statistiques. La matrice de covariance se simplifie alors puisque l'on a $\Sigma_{\boldsymbol{y}\boldsymbol{x}} = 0$.

2.2.3.8 Densités conditionnelles

Nous utiliserons relativement peu de probabilités conditionnelles dans la suite, mais par souci de complétude nous incluons ici quelques formules importantes. Soient $\boldsymbol{x}$ et $\boldsymbol{y}$ deux v.a.. On note $p_{(\boldsymbol{x}|\boldsymbol{y})}(x|y)$ la densité de probabilité de $\boldsymbol{x}$ sachant que $\boldsymbol{y} = y$. Cette densité est donnée à partir de la densité conjointe et de la densité de $\boldsymbol{y}$ par

$$p_{(\boldsymbol{x}|\boldsymbol{y})}(x|y) = \frac{p_{(\boldsymbol{x},\boldsymbol{y})}(x, y)}{p_{\boldsymbol{y}}(y)} = \frac{p_{(\boldsymbol{x},\boldsymbol{y})}(x, y)}{\int p_{(\boldsymbol{x},\boldsymbol{y})}(x, y).dx}.$$

Comme on a les formules symétriques pour la densité de $\boldsymbol{y}$ sachant $\boldsymbol{x}$, on éliminer la densité conjointe de ces formules pour obtenir la règle de Bayes :

$$p_{(\boldsymbol{x}|\boldsymbol{y})}(x|y) = \frac{p_{(\boldsymbol{y}|\boldsymbol{x})}(y|x).p_{\boldsymbol{x}}(x)}{p_{\boldsymbol{y}}(y)} = \frac{p_{(\boldsymbol{y}|\boldsymbol{x})}(y|x).p_{\boldsymbol{x}}(x)}{\int p_{(\boldsymbol{y}|\boldsymbol{x})}(y|x).p_{\boldsymbol{x}}(x).dx}$$

Supposons que $\boldsymbol{x}$ soit une grandeur physique non directement observable mais que l'on puisse mesurer $\boldsymbol{y}$ dont le lien est connu avec $\boldsymbol{x}$ (par l'intermédiaire de la loi conditionnelle $p_{(\boldsymbol{y}|\boldsymbol{x})}(y|x)$). La densité *a priori* $p_{\boldsymbol{x}}(x)$ est une mesure de la connaissance que l'on a sur la grandeur non observable $\boldsymbol{x}$ *avant* l'expérience, et la loi *a posteriori* $p_{(\boldsymbol{x}|\boldsymbol{y})}(x|y)$ rend compte de la connaissance gagnée sur cette grandeur *après* l'observation $\boldsymbol{y} = y$.

2.2.4 Propagation de l'incertitude

2.2.4.1 Changement de vecteur aléatoire bijectif

Soit $\boldsymbol{x}$ un vecteur aléatoire de dimension n et h une fonction bijective C^1 de $\mathbb{R}^n$ dans $\mathbb{R}^n$. Le vecteur $\boldsymbol{y} = h(\boldsymbol{x})$ est encore une vecteur aléatoire dont on veut caractériser la densité de probabilité.

La probabilité pour que $\boldsymbol{y}$ soit dans un ouvert $\mathcal{Y}$ est donnée par

$$\Pr(\boldsymbol{y} \in \mathcal{Y}) = \Pr\left(h(\boldsymbol{x}) \in \mathcal{Y}\right) = \Pr\left(\boldsymbol{x} \in h^{(-1)}(\mathcal{Y})\right) = \int_{h^{(-1)}(\mathcal{Y})} p_{\boldsymbol{x}}(x).dx$$

Pour faire le changement de variable $y = h(x)$ dans cette intégrale, on a besoin du jacobien J_h de la fonction vectorielle h (voir à ce sujet l'appendice (A) pour une explication plus détaillée). Notons que la matrice jacobienne J_h dépend en général du point x où elle est calculée. Nous noterons $J_h(x)$ ou $J_h|_x$ cette valeur s'il y a lieu de le préciser.

Le changement de volume infinitésimal de la forme différentielle est alors donné par $dy = |J_h|.dx$ et notre probabilité se réécrit :

$$\Pr(\boldsymbol{y} \in \mathcal{Y}) = \int_{\mathcal{Y}} \frac{p_{\boldsymbol{x}}(h^{(-1)}(y))}{|J_h(h^{(-1)}(y))|}.dy$$

La densité de probabilité de $\boldsymbol{y} = h(\boldsymbol{x})$ est donc par définition :

$$p_{\boldsymbol{y}}(y) = \frac{p_{\boldsymbol{x}}(x)}{|J_h(x)|} \qquad \text{avec} \qquad x = h^{(-1)}(y) \tag{2.8}$$

2.2.4.2 Propagation des moments

En fait, d'un point de vue informatique, nous approximerons un vecteur aléatoire $\boldsymbol{x}$ par sa moyenne et sa covariance : on notera $\boldsymbol{x} \sim (\overline{x}, \Sigma_{\boldsymbol{xx}})$. Nous sommes donc plus intéressés par la propagation des moments que par celle des densités.

Le cas affine Considérons une transformation affine représentée par une matrice inversible et une translation : $\boldsymbol{y} = A.\boldsymbol{x} + t$. La densité de $\boldsymbol{y}$ est donc simplement $p_{\boldsymbol{y}}(y) = |A|^{(-1)}.p_{\boldsymbol{x}}(A^{(-1)}.(y-t))$ et les moments sont obtenus par les intégrales :

$$\overline{y} = \mathbf{E}\left[\,\boldsymbol{y}\,\right] = \int y.p_{\boldsymbol{y}}(y).dy = \int y.p_{\boldsymbol{x}}(A^{(-1)}.(y-t)).\frac{dy}{|A|}$$

$$\Sigma_{\boldsymbol{yy}} = \mathbf{E}\left[\,(\boldsymbol{y}-\overline{y}).(\boldsymbol{y}-\overline{y})^{\mathrm{T}}\,\right] = \int (y-\overline{y}).(y-\overline{y})^{\mathrm{T}}.p_{\boldsymbol{x}}(A^{(-1)}.(y-t)).\frac{dy}{|A|}$$

En faisant le changement de variable $y = A.x + t$ dans ces intégrales, on trouve de manière **exacte** les moments de $\boldsymbol{y} = A.\boldsymbol{x} + t$:

$$\overline{y} = A.\overline{x} + t \qquad \text{et} \qquad \Sigma_{\boldsymbol{yy}} = A.\Sigma_{\boldsymbol{xx}}.A^{\mathrm{T}} \tag{2.9}$$

Notons que si $\boldsymbol{x} \sim N(\mu, \Sigma)$ est gaussien, alors le vecteur $\boldsymbol{y} = A.\boldsymbol{x} + t$ est également gaussien avec les paramètres trouvés : $\boldsymbol{y} \sim N(A.\mu + t, A.\Sigma.A^{\mathrm{T}})$.

Le cas non-linéaire En général, il n'existe pas de formule exacte pour la propagation des moments. On peut cependant en calculer une valeur approchée. Considérons le développement en série de Taylor de la fonction h à l'ordre 1 autour de la moyenne :

$$h(x) = h(\overline{x}) + J_h(\overline{x}).(x - \overline{x}) + O\left(\|x - \overline{x}\|^2\right)$$

Comme l'espérance est un opérateur linéaire, on a :

$$\overline{y} = \mathbf{E}\left[\, h(\boldsymbol{x}) \,\right] = h(\overline{x}) + J_h(\overline{x}).\mathbf{E}\left[\, x - \overline{x} \,\right] + O\left(\mathbf{E}\left[\, \|x - \overline{x}\|^2 \,\right]\right)$$

soit $\overline{y} \simeq h(\overline{x})$. En reportant dans l'équation de calcul de la covariance :

$$\begin{array}{rcl} \Sigma_{\boldsymbol{yy}} & = & \mathbf{E}\left[\, (h(\boldsymbol{x}) - \overline{x}).(h(\boldsymbol{x}) - \overline{x})^{\mathrm{T}} \,\right] \\ & = & J_h(\overline{x}).\mathbf{E}\left[\, (x - \overline{x}).(x - \overline{x})^{\mathrm{T}} \,\right].J_h(\overline{x})^{\mathrm{T}} + O\left(\mathbf{E}\left[\, \|x - \overline{x}\|^3 \,\right]\right) \end{array}$$

Au final, on obtient donc :

$$\overline{y} \simeq h(\overline{x}) \qquad \text{et} \qquad \Sigma_{\boldsymbol{yy}} \simeq J_h(\overline{x}).\Sigma_{\boldsymbol{xx}}.J_h(\overline{x})^{\mathrm{T}}$$

On doit cependant bien faire attention que ces équations ne sont que des approximations au 1er ordre et que négliger les termes d'ordre supérieur dans le développement limité peut amener un biais tout à fait non négligeable, en particulier au voisinage de points de forte courbure ou de singularité de la fonction considérée.

2.2.4.3 Propagation pour une fonction C^1 non bijective

On considère ici une fonction h de classe C^1 de $\mathbb{R}^m$ dans $\mathbb{R}^p$, qui n'est donc plus bijective pour $p \neq m$. Pour calculer la densité de $\boldsymbol{y} = h(\boldsymbol{x})$, il nous faut combiner les différents outils déjà développés.

Si la dimension p de $\boldsymbol{y}$ (l'espace but) est inférieure à celle de l'espace d'origine, cela veut dire qu'il y a réduction d'information. On construit un vecteur $\boldsymbol{z}^{\mathrm{T}} = (\boldsymbol{y}^{\mathrm{T}}, \pi(\boldsymbol{x})^{\mathrm{T}})$ où $\pi(\boldsymbol{x})$ est une projection de $\boldsymbol{x}$ qui compense les coordonnées de $\boldsymbol{y}$ « manquantes » de manière à ce que $\boldsymbol{z} = h'(\boldsymbol{x})$ soit une bijection. On peut alors calculer la densité de $\boldsymbol{z}$ en utilisant la formule (2.8) et celle de $\boldsymbol{y}$ est obtenue comme une probabilité marginale.

Si la dimension de $\boldsymbol{y}$ est supérieure à celle de $\boldsymbol{x}$, alors la distribution de $\boldsymbol{y} = h(\boldsymbol{x})$ est contrainte à n'occuper qu'un sous-espace de dimension m dans $\mathbb{R}^p$, et on ne pourra pas à proprement parler définir sa densité comme une fonction mais plutôt comme une distribution faisant intervenir des « Diracs ». Les moments sont cependant toujours définis mais la matrice de covariance n'est plus de rang plein (elle présente des valeurs propres nulles), ce qui pose un certain nombre de problèmes pour d'autres applications.

Cependant, les formules développées ci-dessus pour la propagation des moments se généralisent aisément à ces deux cas et on obtient le théorème suivant.

Théorème 2.1 *Soit $\boldsymbol{x} \sim (\overline{x}, \Sigma_{\boldsymbol{xx}})$ un vecteur aléatoire de $\mathbb{R}^m$ et h une fonction C^1 de $\mathbb{R}^m$ dans $\mathbb{R}^p$. Alors le v.a. $\boldsymbol{y} = h(\boldsymbol{x})$ a pour moments :*

$$\overline{y} = h(\overline{x}) \qquad et \qquad \Sigma_{\boldsymbol{yy}} = J_h(\overline{x}).\Sigma_{\boldsymbol{xx}}.J_h(\overline{x})^{T} \tag{2.10}$$

Cette formule est exacte si h est une fonction affine mais n'est qu'une approximation au premier ordre dans le cas général.

Cette formule est d'un intérêt pratique énorme, puisqu'elle nous permet d'obtenir un ensemble complet d'opérations pour travailler avec les vecteurs aléatoires approximés par leur moyenne et leur covariance. Cependant, nous devons faire très attention que cet ensemble d'opérations devient de moins en moins cohérent avec la non linéarité des fonctions utilisées. Cela prendra une importance particulière lorsque le vecteur aléatoire $\boldsymbol{x}$ sera la représentation d'une primitive aléatoire $\mathbf{x}$ et la fonction h une transformation géométrique : voir par exemple la figure (4.1).

Fonction de deux variables indépendantes C'est un cas particulier du théorème précédent. En utilisant les résultats sur les covariances obtenues avec les probabilités marginales ($\Sigma_{\boldsymbol{xy}} = 0$) et en scindant le jacobien de h en deux parties

$$J_{h_x} = \frac{\partial h(x,y)}{\partial x} \qquad \text{et} \qquad J_{h_y} = \frac{\partial h(x,y)}{\partial y}$$

on peut écrire les moments de $\boldsymbol{z} = h(\boldsymbol{x},\boldsymbol{y})$ comme :

$$\overline{\boldsymbol{z}} = h(\overline{\boldsymbol{x}},\overline{\boldsymbol{y}}) \qquad \text{et} \qquad \Sigma_{\boldsymbol{zz}} = J_{h_x}.\Sigma_{\boldsymbol{xx}}.J_{h_x}^{\mathrm{T}} + J_{h_y}.\Sigma_{\boldsymbol{yy}}.J_{h_y}^{\mathrm{T}}$$

De même que précédemment, ces formules ne sont exactes que si h est linéaire (ou affine).

Addition de deux vecteurs indépendants Cette opération apparaît comme une l'application simple du paragraphe précédent. Elle donne cependant avec l'opposé d'un vecteur une structure de groupe à l'espace des vecteurs (que l'on peut voir comme l'espace des translations) et il est donc intéressant de noter ces résultats avant de généraliser à des groupes de transformation plus complexes.g

La densité du vecteur aléatoire $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ est donné par le produit de convolution suivant :

$$p_{(\boldsymbol{x}+\boldsymbol{y})}(z) = p_{\boldsymbol{x}} \circledast p_{\boldsymbol{y}}(z) = \int_{\mathbb{R}^n} p_{\boldsymbol{x}}(t).p_{\boldsymbol{y}}(z-t).dt = \int_{\mathbb{R}^n} p_{\boldsymbol{y}}(t).p_{\boldsymbol{x}}(z-t).dt \tag{2.11}$$

Les moments de cette variable sont donnés par

$$\overline{\boldsymbol{z}} = \overline{(\boldsymbol{x}+\boldsymbol{y})} = \overline{\boldsymbol{x}} + \overline{\boldsymbol{y}} \qquad \text{et} \qquad \Sigma_{\boldsymbol{zz}} = \Sigma_{\boldsymbol{xx}} + \Sigma_{\boldsymbol{yy}}$$

Opposé d'un vecteur La densité du vecteur opposé $-\boldsymbol{x}$ est quant à elle simplement : $p_{-\boldsymbol{x}}(x) = p_{\boldsymbol{x}}(-x)$, sa moyenne est $\overline{(-\boldsymbol{x})} = -\overline{\boldsymbol{x}}$ et la covariance est inchangée.

2.2.4.4 Vecteur aléatoire défini par une fonction implicite

Soit $\boldsymbol{x} \sim (\overline{x}, \Sigma_{\boldsymbol{xx}})$ une vecteur aléatoire de dimension m et $\Phi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^p$ une fonction C^1. On s'intéresse ici au v.a. $\boldsymbol{y}$ de dimension p défini implicitement par l'équation

$$\Phi(\boldsymbol{x},\boldsymbol{y}) = 0$$

Soit (x_0, y_0) un point solution (vérifiant $\Phi(x_0, y_0) = 0$). D'après le théorème des fonction implicites, il existe autour de ce point une fonction explicite $y = \varphi(x)$ *si et seulement si* le jacobien $\dfrac{\partial \Phi}{\partial y}$ est inversible, et on a dans ce cas :

$$\frac{\partial y}{\partial x} = \frac{\partial \varphi}{\partial x} = -\left(\frac{\partial \Phi}{\partial y}\right)^{(-1)} \frac{\partial \Phi}{\partial x}$$

En utilisant un développement limité au premier ordre, on détermine que la moyenne $\overline{\boldsymbol{y}}$ est définie implicitement par

$$\Phi(\overline{x}, \overline{y}) = 0$$

et la covariance est donnée par

$$\Sigma_{\boldsymbol{yy}} = \left(\frac{\partial \Phi}{\partial y}\right)^{(-1)} \frac{\partial \Phi}{\partial x} \, \Sigma_{\boldsymbol{xx}} \, \frac{\partial \Phi}{\partial x}^{\mathrm{T}} \left(\frac{\partial \Phi}{\partial y}\right)^{(-\mathrm{T})} \tag{2.12}$$

Les jacobiens de Φ sont estimés ici en $(x, y) = (\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$.

2.2.4.5 Minimisation d'un critère

Soit C un critère, c'est-à-dire une fonction de classe C^2 de $\mathbb{R}^m \times \mathbb{R}^p$ dans $\mathbb{R}^+$. Regardons tout d'abord le cas déterministe : on défini le vecteur p-dimensionnel y comme l'argument pour lequel le critère est minimum pour un vecteur donné x :

$$y = \mathrm{ArgMin}_z \left(C(x, z)\right)$$

Une condition nécessaire et suffisante pour obtenir un minimum local est

$$\Phi(x, y)^{\mathrm{T}} = \frac{\partial C}{\partial z}(x, y) = 0 \qquad \text{et} \qquad H = \frac{\partial^2 C}{\partial z^2}(x, y) \quad \text{définie positive} \tag{2.13}$$

La matrice H des dérivées secondes est symétrique et est appelée matrice hessienne du critère. Si l'on remplace maintenant x par un vecteur aléatoire $\boldsymbol{x}$, l'argument minimisant notre critère est à son tour un vecteur aléatoire :

$$\boldsymbol{y} = \mathrm{ArgMin}_z \left(C(\boldsymbol{x}, z)\right)$$

défini de manière implicite par les équations (2.13). On peut donc utiliser les résultats de la section précédente pour calculer la propagation des moments (toujours au premier ordre) :

$$\overline{y} = \mathrm{ArgMin}_z \left(C(\overline{x}, z)\right) \qquad \text{et} \qquad \Sigma_{\boldsymbol{yy}} = H^{(-1)}.J_{\Phi_x}.\Sigma_{\boldsymbol{xx}}.J_{\Phi_x}^{\mathrm{T}}.H^{(-1)} \tag{2.14}$$

où H est calculée en $(x, y) = (\overline{x}, \overline{y})$ et J_{Φ_x} est le jacobien :

$$J_{\Phi_x} = \frac{\partial \Phi(x, y)}{\partial x}(\overline{x}, \overline{y}) = \frac{\partial^2 C(x, z)}{\partial x \, \partial z}(\overline{x}, \overline{y})$$

Cette formule sera utilisée par exemple dans la section (8.1.3.1) pour calculer l'incertitude sur la transformation rigide minimisant les moindres carrés entre points appariés.

2.2.5 Modèle de bruit

Nous avons vu au début du chapitre que toute mesure était inévitablement bruitée. Si x est le vecteur exact que l'on cherche à mesurer, on caractérise sa mesure par un vecteur aléatoire $\boldsymbol{x}$ dont nous ne retiendrons que la moyenne et la covariance $(\overline{\boldsymbol{x}}, \Sigma_{\boldsymbol{xx}})$ pour des raisons d'efficacité informatique et parce que c'est le plus souvent suffisant pour caractériser la majeur partie de l'information exploitable.

Nous avons développé jusqu'ici un certain nombre d'outils pour propager l'incertitude dans nos calculs, mais il nous faut maintenant regarder de plus les hypothèses raisonnables que nous pouvons faire sur les mesures. En effet, nous avons avec notre approximation n paramètres pour la moyenne (n étant la dimension de notre espace), et $n.(n+1)/2$ paramètres pour la matrice de covariance (car elle est symétrique). Quelle sont les hypothèses raisonnables que l'on peut faire sur ces $n.(n+3)/2$ paramètres par mesure ?

2.2.5.1 Bruit additif

Le v.a. $\boldsymbol{x}$ caractérise la mesure du vecteur exact x. On appelle bruit l'écart entre la mesure et la valeur exacte. Dans le cas de vecteurs, on caractérise cet écart par la différence des valeurs :

$$\boldsymbol{e_x} = \boldsymbol{x} - x$$

Observons que cette formulation permet de dissocier la valeur exacte du processus aléatoire qui la perturbe. On peut alors envisager de comparer les bruits $\boldsymbol{e_x}$ et $\boldsymbol{e_y}$ de mesure en deux points différents x et y : on dit que les bruits de mesure de x et y sont identiques si les distributions de $\boldsymbol{e_x}$ et $\boldsymbol{e_y}$ sont identiques.

Plus généralement, on appelle **bruit additif** un processus de mesure dont le bruit est identique en tout point de l'espace $\mathbb{R}^n$. La mesure de n'importe quel vecteur x est donc caractérisée en utilisant nos notations par un vecteur aléatoire

$$\boldsymbol{x} = x + \boldsymbol{e}_x \qquad \text{avec} \qquad \boldsymbol{e}_x \sim (\overline{e}, \Sigma_{\boldsymbol{ee}})$$

On conserve ici l'indexation sur x du bruit additif pour bien montrer que les v.a. sont différents en chaque point, même si leur distribution est identique. Notons que si les moyennes du bruit et de la mesure sont différentes, leurs matrices de covariance sont identiques.

2.2.5.2 Distributions identiques et indépendantes

Si l'on répète une expérience et que l'on mesure plusieurs vecteurs $\boldsymbol{x}_i$ (un **échantillon** suivant la terminologie statistique), on doit normalement ranger ces mesures dans un seul grand vecteur et étudier la loi conjointe. Si les conditions de mesure sont les mêmes, il est souvent justifié de supposer que les mesures suivent la même loi et qu'elles ne sont pas corrélées (n'ont pas d'influence l'une sur l'autre). Ceci se caractérise par

$$\boldsymbol{x}_i \sim (\mu, \Sigma) \qquad \text{et} \qquad \mathbf{E}\,[\,(\boldsymbol{x}_i - \mu).(\boldsymbol{x}_j - \mu)^{\mathrm{T}}\,] = 0 \quad \text{si } i \neq j$$

Lorsque l'on mesure des objets géométriques, on répète souvent une expérience (par exemple la mesure d'un point), mais en des emplacements différents. On inclus donc dans la notion de distribution identique indépendante une hypothèse de bruit additif qui a pour effet de laisser libre la moyenne de chaque mesure (la localisation), mais qui contraint les covariances :

$$\mathbf{E}\,[\,\boldsymbol{x}_i\,] = x_i + \overline{e} \qquad \text{et} \qquad \mathbf{E}\,[\,(\boldsymbol{x}_i - x_i).(\boldsymbol{x}_j - x_j)^{\mathrm{T}}\,] = \delta_{ij}.\Sigma_{\boldsymbol{ee}}$$

2.2.5.3 Bruit centré

On distingue deux types d'erreur de mesure : les erreur systématiques et les erreurs dites aléatoires. L'erreur systématique se modélise par une déviation de la moyenne $\overline{\boldsymbol{x}}$ de notre mesure par rapport à la valeur exacte x. Puisque c'est en fait la valeur exacte que l'on veut mesurer, on considère généralement que ce type d'erreur doit être corrigé, au besoin par une calibration des instruments, de manière à n'observer que des mesures centrées autour de la valeur exacte. On dit donc que le bruit est centré si

$$\overline{\boldsymbol{x}} = \mathbf{E}\,[\,\boldsymbol{x}\,] = x$$

Dans le cas d'un bruit additif, cela revient à supposer une moyenne nulle sur le bruit :

$$\overline{e} = 0 \qquad \text{soit} \qquad \boldsymbol{e}_x \sim (0, \Sigma_{\boldsymbol{ee}})$$

2.2.5.4 Bruit gaussien

Comme nous l'avons déjà remarqué, il peut être parfois utile de connaître la densité de probabilité et non plus simplement ses moments, et l'hypothèse qui minimise l'information $\mathbf{I}[\,\boldsymbol{x}\,|\,\overline{x}=\mu,\ \Sigma_{\boldsymbol{xx}}=\Sigma\,]$ est la distribution gaussienne $N(\mu,\Sigma)$:

$$p(x)=\frac{1}{\sqrt{(2\,\pi)^n.\det(\Sigma)}}.\exp\left(-\frac{(x-\mu)^{\mathrm{T}}.\Sigma^{(-1)}.(x-\mu)}{2}\right) \tag{2.15}$$

Par ailleurs, la loi normale apparaît comme très proche de la distribution empiriques observées sur de nombreuses mesures (en particulier d'erreur) et elle est de plus la limite de nombreuses lois dans certaines conditions. On a en particulier le **théorème de la limite centrale** qui assure que, sous des conditions très générales, la somme de n v.a.r. indépendantes tend vers la loi normale lorsque n tend vers l'infini. On explique ainsi que la somme de petites erreurs indépendantes sur nos mesures produise une erreur finale presque gaussienne.

2.2.5.5 Mesures aberrantes : gaussienne contaminée

En fait, il peut arriver que le système de mesure utilisé produise de temps en temps des mesures aberrantes, que l'on appelle d'habitude par anglicisme **outliers** par opposition à **inliers**. Dans le cas d'une chaîne de de mesure en traitement d'image, divers inexactitudes peuvent se combiner pour induire un module en erreur, produisant ainsi une erreur grossière sur une estimation au module suivant. Il est ainsi courant (quoique l'on essaie d'en diminuer au maximum l'occurrence) qu'une erreur d'appariement (un faux positif) fournisse une mesure aberrante pour l'algorithme de recalage.

On modélise habituellement ces erreurs en « contaminant » le bruit de mesure par une densité de très forte variance : si p est la densité du vecteur de mesure normalement bruité et p_{out} est une densité de forte variance modélisant les erreurs grossières, on écrira notre densité : $p_{\boldsymbol{x}}=(1-\varepsilon).p+\varepsilon.p_{out}$ où ε règle la probabilité que notre mesure soit aberrante. Dans le cadre gaussien habituel, on a $p=N(\overline{x},\Sigma)$ et il semble naturel de modéliser également la contamination par une gaussienne : $p=N(\mu,\Sigma_{out})$. On obtient ainsi le modèle standard de la gaussienne contaminée.

En fait, les mesures aberrante n'ont aucune raison d'être gaussiennes et il est plus raisonnable de les modéliser avec une distribution uniforme sur l'image (ou la zone de mesure). On obtient alors la densité : $p_{\boldsymbol{x}}=(1-\varepsilon).N(\overline{x},\Sigma)+\frac{\varepsilon}{\mathcal{V}_{Im}}.\mathbf{1}_{Im}$, où $\mathcal{V}_{Im}$ est le volume de l'image. Si l'on a une estimation de la distribution théorique du bruit de mesure $N(\overline{x},\Sigma)$, on peut alors éliminer la plupart des mesures aberrantes grâce à un test du χ^2 à $\alpha\%$. Un bon compromis pour classer les mesures comme correctes ou aberrantes est $\alpha=\frac{\varepsilon}{\mathcal{V}_{Im}}$.

Cette modélisation des mesures aberrantes sera utilisée à la section (8.5.1) pour éliminer dans le recalage les erreurs provenant de la mise en correspondance et au chapitre (10) pour calculer la probabilité de faux positif et de faux négatif.

2.3 Quelques problèmes pour généraliser

« La géométrie a beau être, selon Pascal, la science la plus parfaite pour nous autres hommes, elle n'en prend pas moins, du point de vue même de démonstration, des aspects forts variés, jusque dans les travaux mathématiques de cet auteur : on ne démontre pas, même lorsqu'on est entré dans le style géométrique, un problème de probabilité comme un problème de centre de gravité. »
J.P. Cléro, Épistémologie des mathématiques.

Nous avons développé jusqu'ici quasiment tous les outils pour gérer l'incertitude sur des mesures de *vecteurs*. Ces outils seront utilisés dans la seconde partie de ce manuscrit pour introduire les statistiques dans les algorithmes de reconnaissance et de recalage.

Le problème que nous nous posons maintenant est leur généralisation à la mesure de *primitives géométriques* telles que des droites, des repères, des rotations... Nous présentons dans cette section quelques paradoxes potentiels qui illustrent les difficultés que nous avons rencontrées et montrent que l'on ne peut pas considérer impunément ce type de primitives comme de simples vecteurs.

2.3.1 Utilité des primitives dans les algorithmes géométriques haut niveau

S'il est évident que les transformations ont leur rôle dans ce type d'algorithme, l'emploi de primitives géométriques plus complexes que les points est sans doute d'une nécessité moins reconnue, certainement à cause des difficultés que posent leur gestion. Nous pouvons évoquer ici trois raisons principales pour l'utilisation de telles primitives.

Tout d'abord, ce type de primitive émerge naturellement dans la modélisation de certains problèmes. En reconnaissance de sous-structures dans les protéines, la configuration dans l'espace d'un acide aminé est caractérisée (en première approche) par la position de 3 atomes du squelette de la protéine. Ces trois atomes ont toujours la même géométrie (voir figure (2.1) à droite), et les contraintes sur la position de ces atomes ne sont donc pas discriminantes. Le seul descripteur géométrique d'un acide aminé est donc sa pose, c'est-à-dire sa position et son orientation dans l'espace, soit encore : un repère.

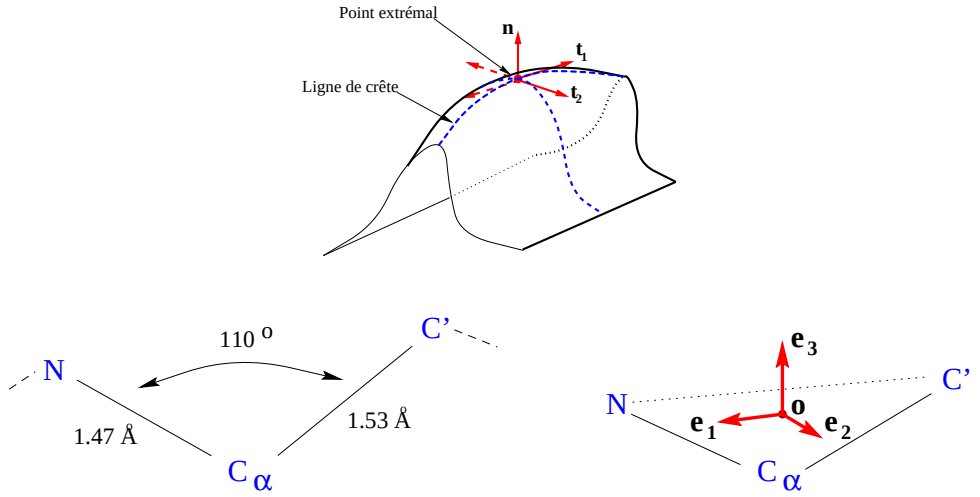


FIG. 2.1 – *Les repères sont des primitives géométriques qui émergent naturellement dans la modélisation de problèmes géométriques. En haut, les points extrémaux sur une iso-surface. En bas : description géométrique de la configuration d'un acide aminé.*

En imagerie médicale volumique, les **points extrémaux** définis dans (Thirion et Gourdon, 1995) sont des points sur une iso-surface qui optimisent un critère de géométrie différentielle basé sur la courbure. Certains de ces points peuvent ainsi être vu comme la généralisation des points de coin. Étant définis sur une surface, ces points sont munis de deux directions principales de courbure t_1, t_2 et de la normale à la surface (figure (2.1) à gauche), ce qui forme encore une fois un repère car ces trois vecteurs sont orthonormés. En fait, ce n'est pas tout à fait un repère, mais ce que l'on appellera un repère semi orienté car les directions principales t_1 et t_2 ne sont que des *directions*, c'est-à-dire que l'on mesure indifféremment t_i ou $-t_i$.

La seconde raison est qu'en utilisant le maximum d'information sur les primitives, qui sont les objets de base dans notre modélisation haut niveau du « monde » dans lequel on travaille, on obtiendra forcément une **sélectivité** et une **précision** plus importante pour la reconnaissance et le recalage. Ainsi, l'introduction des normales en plus des points dans (Feldmar et al., 1997) autorise la recherche d'appariements initiaux avec les bitangentes, ce qui réduit considérablement la complexité de l'étape d'appariement. De la même façon, l'utilisation des repères au lieu des points pour modéliser les acides aminés permet dans (Pennec et Ayache, 1998) de réduire la complexité de la reconnaissance des sous-structures communes de $O(n^3)$ à $O(n^2)$ (n étant le nombre d'acides aminés).

Enfin, même des transformations aussi simples que les transformations rigides ne sont pas des vecteurs, et si l'on veut pouvoir estimer l'incertitude sur une transformation, c'est-à-dire sa précision, il nous faut généraliser les méthodes statistiques usuelles.

2.3.2 Le paradoxe de Bertrand

Le problème soulevé par J. Bertrand en 1907 est de calculer la probabilité qu'une « corde aléatoire » d'un cercle ait une longueur supérieure au côté d'un triangle équilatéral inscrit dans ce cercle. En supposant un cercle de rayon 1, le côté de ce triangle est $\sqrt{3}$. Trois méthodes sont envisagées pour résoudre ce problème, illustrées figure (2.2).

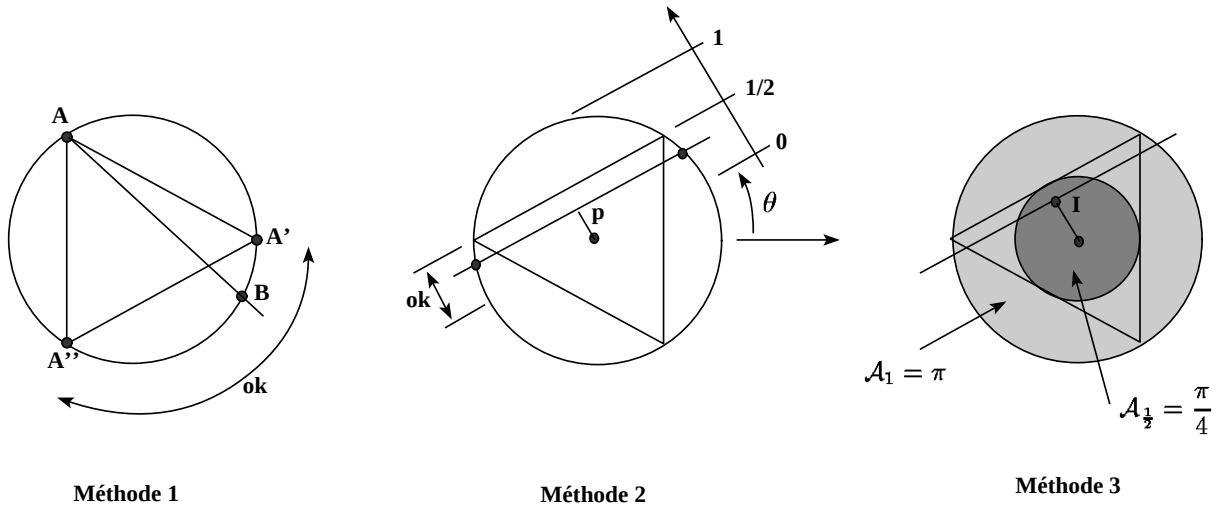


FIG. 2.2 – *Trois méthodes pour calculer la probabilité qu'une « corde aléatoire » ait une longueur supérieure au côté d'un triangle équilatéral inscrit. De gauche à droite, une probabilité de* $\frac{1}{3}$, $\frac{1}{2}$ *et* $\frac{1}{4}$.

Méthode 1: Une corde intersecte par définition le cercle en deux points que l'on peut supposer indépendants et distribués uniformément. En supposant que l'un des points soit A sur la figure (2.2), le second point doit se trouver entre A' et A'' pour que la longueur de la corde soit supérieure au côté du triangle. Comme ceci représente un tiers de la circonférence du cercle, la probabilité cherchée est $\frac{1}{3}$.

Méthode 2: Une corde est caractérisée par sa distance p au centre du cercle (entre 0 et 1) et son orientation θ (entre 0 et 2π) par rapport à une droite fixe. Si l'on trace un triangle équilatéral dont

un côté est parallèle à la corde, on voit que la distance p doit être inférieure à un demi pour que la corde ait une longueur suffisante. En supposant une orientation et une distance à l'origine uniforme, on trouve une probabilité normalisée de $\frac{1}{2}$.

Méthode 3: Une corde est définie de manière unique par le pied I de la perpendiculaire joignant le centre du cercle. Ce point doit être dans un disque de rayon $\frac{1}{2}$ pour assurer une longueur suffisante à la corde. En supposant une distribution uniforme de I dans l'intérieur du cercle original, la probabilité normalisée est de $\frac{1}{4}$.

Les trois solutions présentées sont correctes, mais ne se réfèrent pas à la même notion d'uniformité dans la façon de choisir une corde. En utilisant la représentation (p, θ) décrite dans la seconde méthode, on peut calculer (Kendall et Moran, 1963) que les mesures de probabilité associées sont respectivement :

$$d\sigma_1 = \frac{dp.d\theta}{2\pi\sqrt{1-p^2}} \qquad d\sigma_2 = \frac{dp.d\theta}{2\pi} \qquad d\sigma_3 = p.\frac{dp.d\theta}{\pi}$$

Seule la seconde est invariante par l'action des rotations, translations et réflexions.

Ce paradoxe illustre le problème de la **mesure** utilisée, qui représente également la notion d'uniformité ou de hasard pur. Comme nous travaillons avec des primitives géométriques, nous avons toujours plus ou moins un groupe de transformation qui agit sur ces primitives et nous pensons que, *en l'absence d'information supplémentaire*, il est raisonnable de supposer que la mesure uniforme est invariante par l'action de ce groupe, comme il paraît raisonnable de supposer que la probabilité demandée ci-dessus ne dépend pas du positionnement du cercle dans le plan. Comme toute la théorie des variables et vecteurs aléatoires que nous avons vue repose sur cette mesure d'uniformité (la mesure de Lebesgue dans ce cas), il est important de voir cela en détail pour définir des primitives géométriques aléatoires ayant un sens. Nous développerons le calcul des mesures invariantes dans la section (3.3).

2.3.3 Espérance d'une droite 2D

La méthode classique pour travailler sur des primitives géométriques (ou plus généralement une variété différentielle) utilise une **représentation** de ces primitives, c'est-à-dire une carte locale dans laquelle une primitive est représentée par un vecteur. Cette équivalence n'est possible que sur une domaine bien précis (le domaine de définition de la carte) et ne couvre pas forcément toutes les primitives, ou alors inclus des singularités. Une introduction plus détaillée de ces notion fera l'objet de la section (3.2).

Pour gérer l'incertitude, une solution simple consiste à modéliser une primitive aléatoire $\mathbf{x}$ comme un *vecteur* aléatoire $\boldsymbol{x}$ dans la représentation (voir par exemple (Durrant-Whyte, 1988; Ayache, 1991; Zhang et Faugeras, 1992)). La moyenne ou l'espérance d'une primitive aléatoire est alors définie comme la primitive correspondant à la moyenne du vecteur aléatoire $\boldsymbol{x}$ dans la représentation. Si $\mathcal{D}$ est le domaine de définition de notre représentation, on écrira donc :

$$\overline{x} = \mathbf{E}\,[\,\boldsymbol{x}\,] = \int_{\mathcal{D}} y.p_{\boldsymbol{x}}(y).dy$$

Nous pensons que cet opérateur est mal défini dans le cadre des variétés. La première raison est que le résultat de l'intégrale n'est pas obligatoirement dans le domaine de définition : une somme de matrices de rotations n'a pas de raison d'être orthogonale dans le cas général. La seconde raison

est que cet opérateur d'espérance ne commute pas (en général) avec un changement de repère ou l'application d'une transformation. Nous développons ce dernier point dans l'exemple suivant.

Considérons par exemple une droite vectorielle 2D orientée : elle peut être représentée par un vecteur unitaire ou plus simplement par son angle $\theta \in \mathcal{D} =]-\pi\ ;\ \pi[$ par rapport à un axe donné. On définit donc une droite aléatoire $\boldsymbol{\theta}$ par une densité $p(\theta)$, et sa moyenne par

$$\overline{\theta} = \mathbf{E}\,[\,\boldsymbol{\theta}\,] = \int_{\mathcal{D}} \theta.p(\theta).\frac{d\theta}{2\pi}$$

Notons que $d\theta$ est la mesure uniforme pour les droites soumises à l'action des rotations. L'action d'une rotation d'angle λ est simplement l'addition des angles (modulo 2π) et peut également être vue comme un changement de repère.

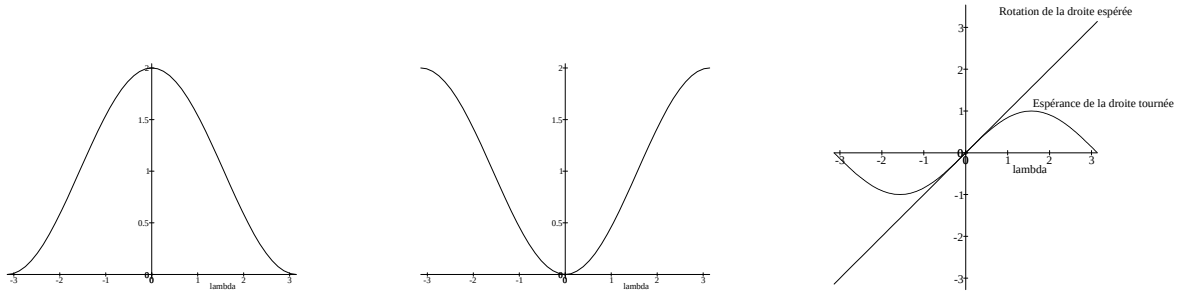


FIG. 2.3 – *Gauche: Densité originale* $p_0(\theta) = 2\cos(\theta/2)^2$. *L'espérance est* $\overline{\theta}(0) = 0$. *Milieu: densité après rotation d'un angle* $\lambda = \pi$. *L'espérance* $\overline{\theta}(\pi)$ *est encore* 0, *alors que la rotation de l'espérance est* $\underline{\theta}_\pi = \pi$. *Droite : comparaison de l'espérance de la droite tournée et de la rotation de la droite espérée.*

Nous avons tracé en figure (2.3) une densité de moyenne $\overline{\theta}_0 = 0$. Si nous changeons de référentiel, c'est-à-dire que nous appliquons une rotation d'angle λ, on peut voir que l'espérance de la droite tournée devient $\overline{\theta}(\lambda) = \sin(\lambda)$, alors que la rotation de l'espérance est $\underline{\theta}(\lambda) = \lambda$. En particulier, pour une rotation de π, on trouve que $\underline{\theta}(\pi) = \pi$ et $\overline{\theta}(\pi) = 0$!

En fait, en utilisant l'espérance dans une carte, nous sommes en train d'essayer de définir une intégrale *à valeur dans la variété*, ce qui n'a pas de fondement mathématique. Pour remédier à ce problème, nous définirons dans la section (4.2) l'espérance au sens de Fréchet (proposée en 1948), qui généralise l'espérance classique dans les espaces vectoriels aux espaces métriques quelconques. Le lecteur pourra se référer à la figure (4.1) pour trouver une comparaison visuelle de la moyenne classique et de la moyenne de Fréchet sur des vecteurs rotation.

2.3.4 Le paradoxe de la droite la plus proche

On utilise beaucoup la distance pour classifier, quantifier et minimiser sur les points. La distance est même au cœur de certains algorithmes comme le point le plus proche itératif (**ICP** pour Iterative Closest Point), ou les algorithmes d'extraction du squelette (Medial Axis Transform). Pour pouvoir comparer nos primitives (et définir l'espérance de Fréchet), nous aurons également besoin d'une fonction de distance entre les primitives. L'exemple de cette section montre que la définition d'une telle distance doit être faite avec précaution. En particulier, on peut rarement définir la distance entre deux primitives comme la distance entre les deux vecteurs correspondants dans une carte.

Nous avons vu avec le paradoxe de Bertrand plusieurs représentations d'une droite 2D. Nous utilisons ici une autre représentation minimale, introduite dans (Ayache, 1991), fondée sur l'équation

de la droite : $a.x + b.y + c = 0$. Pour obtenir une représentation minimale, on doit éliminer un paramètre :

- Carte 1: Les droites non parallèles à l'axe des x sont représentées par $d = (a, p) \in \mathbb{R}^2$ et ont pour équation : $a.x + y + p = 0$.
- Carte 2: Les droites non parallèles à l'axe des y sont représentées par $d' = (a', p') \in \mathbb{R}^2$ et ont pour équation : $x + a'.y + p' = 0$.

Dans la première carte, la droite $d = (a, p)$ coupe l'axe des y au point $(0, -p)$ et a pour vecteur directeur $(1, -a)$. La représentation est symétrique dans la seconde carte : la droite $d' = (a', p')$ coupe l'axe des x au point $(-p', 0)$ et a pour vecteur directeur $(-a', 1)$.

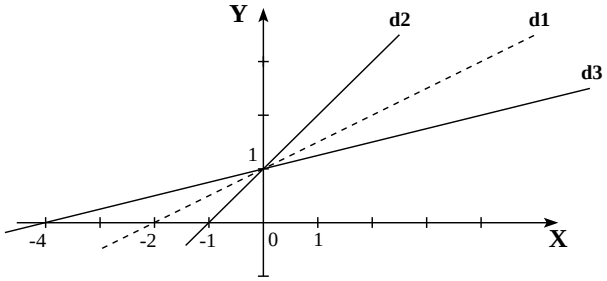


FIG. 2.4 – *Trois droites planaires.*

Nous avons tracé trois droites dans la figure (2.4). Le problème est de trouvé laquelle, de d_2 ou de d_3, est la plus proche de d_1. Si l'on définit la distance entre deux droites comme la distance euclidienne entre leurs représentations, on a $\text{dist}(d_1, d_2) = \sqrt{(a_1 - a_2)^2 + (p_1 - p_2)^2}$.

Les coordonnées des trois droites dans la première carte sont $d_1 = (-1/2\,;\, -1)$, $d_2 = (-1\,;\, -1)$, $d_3 = (-1/4\,;\, -1)$ et on obtient donc :

$$\text{dist}(d_1, d_2) = 1/2 \quad > \quad \text{dist}(d_1, d_3) = 1/4$$

Si nous considérons les droites dans la seconde carte, leur coordonnées sont $d_1 = (-2\,;\, 2)$, $d_2 = (-1\,;\, 1)$, $d_3 = (-4\,;\, 4)$ et les distances sont maintenant :

$$\text{dist}(d_1, d_2) = \sqrt{2} \quad < \quad \text{dist}(d_1, d_3) = 2\sqrt{2}$$

Nous avons donc le « paradoxe » suivant : d_3 est la droite la plus proche de d_1 dans une carte, alors que c'est d_2 dans la seconde. En fait, visuellement, la première carte a raison : il y a un angle de 12.5^o entre d_1 et d_3 contre un angle de 18.5^o entre d_1 et d_2 (ces angles sont évidement invariants par transformation rigide).

Si cet exemple semble trivial, il montre une fois de plus qu'on ne peut généralement pas définir des opérations sur la variété en utilisant directement leur équivalent vectoriel dans une carte. De plus, il semble souhaitable de choisir une distance invariante par le groupe de transformation que l'on considère afin de pouvoir comparer nos objets géométriques indifféremment avant ou après transformation. Une méthode générique pour obtenir une distance Riemannienne invariante sera proposée dans la section (3.5).

2.3.5 Le paradoxe du bruit additif

On se pose ici la question de savoir comment définir un bruit identique et indépendant sur plusieurs primitives. La méthode « classique » consisterait à dire que les représentations $\boldsymbol{x}_i$ des

primitives aléatoires $\mathbf{x}_i$ ont un bruit additif identique et indépendant :

$$\boldsymbol{x}_i = x_i + \boldsymbol{e}_i \qquad \text{avec} \qquad \boldsymbol{e}_i \sim (\mu, \Sigma)$$

Pour simplifier encore le problème, on suppose que le bruit est centré. Les mesures sont donc distribuées comme $\boldsymbol{x}_i \sim (x_i\ ,\ \Sigma)$.

Les problèmes surgissent lorsque l'on veut changer de « point de vue », c'est-à-dire appliquer une transformation globale f aux primitives exactes et mesurées. Les nouvelles mesures sont maintenant représentées par $\boldsymbol{y}_i = f \star \boldsymbol{x}_i$ (on note $f \star x$ l'action d'une transformation f sur la représentation x d'une primitive). En général, il n'y a aucune raison pour que l'action d'une transformation soit linéaire dans la carte considérée et le jacobien $J_f(x) = \partial(f \star x)/\partial x$ est une fonction non constante de x. Les nouvelles mesures sont donc caractérisées par les distributions :

$$\boldsymbol{y}_i \sim (f \star x_i\ ,\ J_f(x_i).\Sigma.J_f(x_i)^{\mathrm{T}})$$

et elles ne sont plus du tout identiques puisqu'elles ne partagent pas la même matrice de covariance.

Un exemple avec les vecteurs de rotation Le vecteur rotation sera introduit au chapitre (7). Il nous suffit pour le moment de savoir qu'un vecteur rotation $r = \theta.n$ représente une rotation tri-dimensionelle d'angle θ autour du vecteur unitaire n.

Soient $\boldsymbol{r_1}$ et $\boldsymbol{r_2}$ des mesures de $r_1 = (\frac{\pi}{2}, 0, 0)^{\mathrm{T}}$ (rotation de 90 degrés autour de l'axe x) et $r_2 = (0, 0, 0)^{\mathrm{T}}$ (rotation identité). Supposons un bruit additif centré indépendant et identique sur ces deux mesures, de covariance

$$\Sigma_1 = \Sigma_2 = \Sigma = \sigma^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Nous avons choisi une covariance singulière pour faciliter la représentation visuelle (figure (2.5) à gauche).

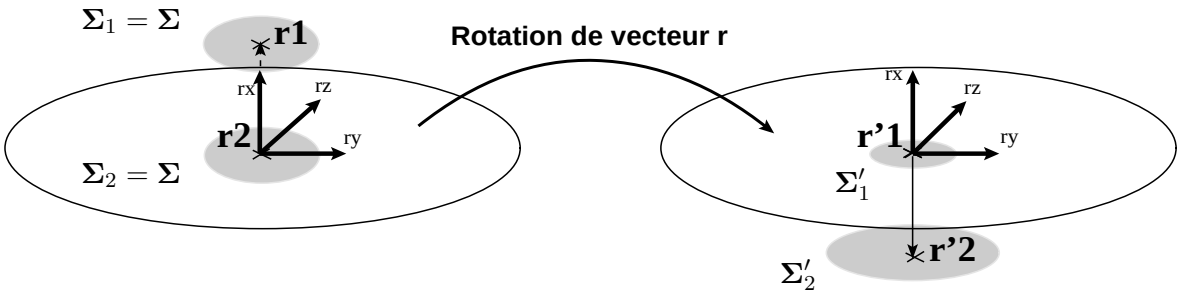


FIG. 2.5 – *La même matrice de covariance sur deux vecteurs de rotation dans un système de coordonnées conduit généralement à des matrices de covariance différentes dans un autre système de coordonnées.*

Si l'on applique une rotation globale de $\pi/2$ autour de l'axe des x (dans le sens des aiguilles d'une montre), représentée par le vecteur rotation $r = (-\frac{\pi}{2}, 0, 0)^{\mathrm{T}}$, on obtient les vecteurs de rotation exacts $r'_1 = r \star r_1 = (0, 0, 0)^{\mathrm{T}}$ et $r'_2 = r \star r_2 = (-\frac{\pi}{2}, 0, 0)^{\mathrm{T}}$.

Le calcul des jacobiens et la propagation des covariances montre que la mesure de ces vecteurs est $\boldsymbol{r_1} \sim (r'_1\ ,\ \Sigma'_1)$ et $\boldsymbol{r_2} \sim (r'_2\ ,\ \Sigma'_2)$ avec

$$\Sigma'_1 = \frac{8}{\pi^2}\Sigma \qquad \text{et} \qquad \Sigma'_2 = \frac{\pi^2}{8}\Sigma$$

Il y a donc un facteur 1.5 entre les deux covariances, ce qui ne peut plus être considéré comme négligeable (figure (2.5) à droite).

Nous verrons à la section (5.1) plusieurs manières de comparer les distribution en différents points, ce qui conduira à la notion de bruit isotrope ou homogène.

2.4 Résumé

Nous avons résumé dans ce chapitre un certain nombre de résultats de probabilités qui permettent de gérer efficacement l'incertitude sur des mesures de vecteurs. On peut en particulier modéliser une telle mesure comme un vecteur aléatoire $\boldsymbol{x}$ dont on ne conserve, pour des raisons d'efficacité informatiques, que la moyenne et la covariance :

$$\boldsymbol{x} \sim (\overline{x}\ ,\ \Sigma_{\boldsymbol{xx}})$$

Ces moments sont calculés à partir de la probabilité de densité $p_{\boldsymbol{x}}$ du vecteur aléatoire grâce à une généralisation de l'opérateur d'espérance à valeur dans l'espace des vecteurs et des matrices :

$$\overline{x} = \mathbf{E}\,[\,\boldsymbol{x}\,] = \int_{\mathbb{R}^n} y.p_{\boldsymbol{x}}(y).dy$$

$$\Sigma_{\boldsymbol{xx}} = \mathbf{E}\,[\,(\boldsymbol{x}-\overline{x}).(\boldsymbol{x}-\overline{x})^{\mathrm{T}}\,] = \int_{\mathbb{R}^{n^2}} (y-\overline{x}).(y-\overline{x})^{\mathrm{T}}.p_{\boldsymbol{x}}(y).dy$$

La propagation de l'incertitude se déduit alors (pour une fonction implicite) en utilisant le jacobien de la transformation :

$$h(\boldsymbol{x}) \sim (h(\overline{x})\ ,\ J_h(\overline{x}).\Sigma_{\boldsymbol{xx}}.J_h(\overline{x})^{\mathrm{T}})$$

Cette formule n'est toutefois exacte que si h est une fonction *affine*. C'est une approximation au premier ordre dans le cas général, comme les formules que l'on a pu déduire pour des vecteurs aléatoires définis par une fonction implicite ou le minimum d'un critère.

Nous avons passé en revue divers hypothèses sur les bruits de mesure, et en particulier celle des distribution identiques et indépendantes (IID), liée pour les vecteurs à celle de bruit additif. L'hypothèse de centrage étant également souvent admise, on obtient alors le modèle de mesure suivant :

$$\boldsymbol{x}_i = x_i + \boldsymbol{e}_i \qquad \text{avec} \qquad \boldsymbol{e}_i \sim (0, \Sigma) \qquad \text{et} \qquad \Sigma_{\boldsymbol{e_i e_j}} = \mathbf{E}\left[\ \boldsymbol{e}_i.\boldsymbol{e}_j^{\mathrm{T}}\ \right] = 0$$

Enfin, nous avons montré par quelques exemples que la généralisation des notions de mesure uniforme, d'espérance, de distance et de bruits identiques n'était pas triviale pour des primitives géométriques non vectorielles et pouvait donner des résultats contradictoires.

Chapitre 3

Outils de base sur les primitives géométriques

« Le travail des physiciens modernes consiste
à découvrir les symétries du monde. »
Heinz Pagels, L'Univers Quantique.

3.1 Introduction

3.1.1 Un peu d'histoire

La géométrie est communément définie comme la science des figures de l'espace. Bien que déjà étudiée en Egypte ancienne, ce sont les *éléments* d'Euclide (fin du IV^{e} siècle av. J.C.) qui marquent les fondements de la géométrie. Il faut attendre Descartes (1596-1650) pour que la géométrie analytique prenne son essor, géométrie que l'on peut caractériser comme l'expression d'une réalité géométrique par une relation entre quantités variables, l'usage des coordonnées, et le principe de la représentation graphique.

C'est un jeune géomètre français, Evariste Galois (1811-1832) qui a introduit en mathématiques, à propos de la résolution des équations algébriques, la notion de groupe d'opérations. Cette idée a conduit Sophus Lie (1842-1899) et Félix Klein (1849-1925) à formuler dans le « programme d'Erlang » le concept très général de géométrie ainsi :

- Soit une variété $\mathcal{M}$ d'éléments que nous conviendrons d'appeler des points.
- Soit un ensemble d'opérations effectuées sur ces points, formant un groupe $\mathcal{G}$.
- La géométrie de la variété $\mathcal{M}$ ayant pour groupe principal le groupe $\mathcal{G}$ est l'ensemble des propriétés de la variété invariantes pour les opérations de ce groupe.

La synthèse de Klein ne porte cependant que sur les espaces homogènes, et Riemann (1826-1866) introduit en 1854 la notion de métrique sur les éléments différentiels d'une variété, définissant ainsi un type d'espace très général (voir section 3.5). Les deux conceptions ne furent raccordées que par les travaux de Elie Cartan (1869-1951) qui généralisa la notion d'espace de Riemann par l'introduction d'une connexion définie par le groupe.

3.1.2 Organisation du chapitre

Nous présentons tout d'abord les raisons qui nous conduisent à modéliser le monde géométrique que l'on veut traiter sous forme de variétés différentielles sur lesquelles agissent un groupe de transformation, et nous développons le cas important des variétés homogènes.

L'étape suivante est ensuite de développer des outils permettant de gérer l'incertitude dans ce formalisme. Or, nous avons vu au chapitre précédent que cela pouvait poser problème (section 2.3). En particulier, le paradoxe de Bertrand pose le problème de la mesure uniforme à choisir sur la variété de nos primitives ou notre groupe de transformation, mesure qui est à la base de toute théorie des probabilités. Ce point essentiel a été étudié tout au long de ce siècle (le paradoxe date de 1907), et a conduit au développement de la **théorie des probabilités géométriques** (voir en particulier (Kendall et Moran, 1963; Santalo, 1976)).

Nous nous focalisons à la section (3.3) sur la détermination de la mesure invariante, les *probabilités géométriques* en question étant en fait plus reliées à la stéréologie qu'à la gestion de l'erreur sur les mesures de primitives géométriques. Une application directe de ces *probabilités géométriques* sera cependant évoquée dans la seconde partie avec l'étude de la robustesse des algorithmes d'appariement et le calcul du nombre moyen de faux positifs.

De manière similaire, nous chercherons à la section (3.4) la caractérisation d'une distance invariante sur notre variété. Cependant, pour avoir des résultats suffisants, nous devrons nous intéresser aux aspect métriques et différentiels des ensembles d'objets géométriques.

Nous rappelons ainsi les bases de la **géométrie différentielle** et de la **géométrie riemannienne**, à la section (3.5) mais nous renvoyons le lecteur à des ouvrages plus spécialisés comme (Spivak, 1979; Klingenberg, 1982; Carmo, 1992) pour la théorie. Ces résultats nous permettent de caractériser les conditions d'existence d'une distance invariante sur nos ensembles de primitives, ainsi que de construire la représentation exponentielle qui s'avérera posséder des propriétés très intéressantes nous permettant, entre autres, de généraliser de nombreux résultats de probabilités dans le chapitre suivant.

3.2 Variétés différentielles et groupes de Lie

« Traiter la nature par la sphère, le cylindre et le cône... »
P. Cézanne (1839-1906), Lettre à Emile Bernard

Lorsque l'on considère des objets géométriques, la première remarque est que l'on veut pouvoir les comparer dans différentes configurations ou positions. On considère donc implicitement un ensemble de transformations de l'espace qui nous permettent d'identifier nos objets, et le choix du type de transformations détermine les propriétés des objets : voir par exemple les triangles de la figure (3.1).

La seconde remarque est que les objets géométriques avec lesquels on va travailler sont catégorisés et constituent des ensembles bien distincts : on sent bien qu'un carré ou qu'une droite n'a rien à faire avec des triangles. Il nous faut donc déterminer et modéliser l'ensemble des objets qui nous intéressent, ce qui sera fait à la section (3.2.1), puis déterminer le groupe de transformation qui agit sur ces objets (section 3.2.2) et enfin les propriétés induites par l'action du groupe sur les objets (section 3.2.3).

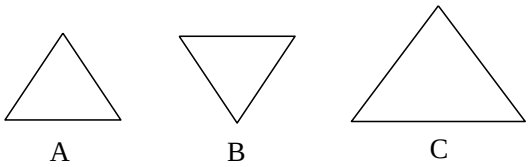


FIG. 3.1 – *Comparaison d'objets géométriques: on peut dire que les trois triangles sont similaires (relativement au groupe des similitudes), ou bien que* **A** *et* **B** *seulement sont congruents (pour les transformations rigides), ou encore qu'ils sont tous différents (en translation).*

3.2.1 Primitives géométriques : variétés différentielles

3.2.1.1 Une première approche ensembliste

Quand on parle de primitives géométriques, on pense tout d'abord à des points, bien sûr, mais aussi à des droites, des segments, des plans... La plupart de ces primitives constituent un ensemble de points dans l'**espace euclidien** ambiant ayant au besoin des caractéristiques supplémentaires. On pourra ainsi considérer des points attribués ou numérotés, ou encore affublés d'une échelle si l'on s'intéresse à l'espace multi-échelle, mais cela revient à rajouter une dimension à l'espace ambiant.

Ces primitives peuvent être décrites par un vecteur de paramètres p et une fonction β qui associe à ces paramètres la primitive géométrique x constituée des points vérifiant l'équation $\beta(p, x) = 0$. Une première représentation ensembliste s'écrit donc :

$$\mathrm{x} = \{x \in \mathbb{R}^n \ / \ \beta(p, x) = 0\}$$

La fonction β associe une **représentation** spécifique p à un type particulier de primitive (droites, plans, courbes, triangles...). Les plans orientés sont par exemple représentés par les paramètres $p = (n, d)$ où n est un vecteur unitaire normal au plan et pointant vers la partie de l'espace ne contenant pas l'origine et $d \geq 0$ est la distance du plan à l'origine. L'équation du « plan p » est donc :

$$\beta(p, x) = \langle n \mid x \rangle - d = 0$$

3.2.1.2 Variété et représentations

Les ensembles usuels de primitives sont réguliers et constituent des **variétés différentielles**. Cela signifie que cet ensemble est localement difféomorphe à un espace vectoriel $\mathbb{R}^m$ de dimension constante, c'est-à-dire qu'il existe en tout point x de la variété $\mathcal{M}$ une bijection différentiable [1] entre un voisinage de x et un ouvert de $\mathbb{R}^m$. La dimension m est justement la dimension de la variété. Dans l'exemple ci-dessus, nous pouvons voir que p est de dimension 4 avec une contrainte quadratique de dimension 1. La variété des plans est donc de dimension 3. Cette formulation mathématique un peu complexe signifie simplement qu'une variété n'est pas un espace vectoriel, mais on peut la traiter **localement** comme tel, en particulier pour dériver. Pour simplifier l'écriture, le qualificatif différentiel sera souvent omis mais sous-entendu lorsque l'on parlera d'une variété.

Il existe de nombreuses façons de représenter une variété, avec différentes propriétés. Dans l'exemple ci-dessus, la représentation est **ambiguë** puisque, pour les plans passant par l'origine,

1. On devrait en fait définir un système de cartes locales et imposer que le changement de carte soit localement difféomorphe. La bijection différentiable de $\mathcal{M}$ dans la carte fait intervenir la notion d'espace tangent à la variété et de différentielle d'une fonction de la variété dans une autre, qui ne peuvent se définir que postérieurement. Par souci de simplicité nous ne développerons que succinctement ces notions à la section (3.5.1), et nous reportons le lecteur à (Bourguignon, 1990) pour en avoir une présentation rigoureuse et plus détaillée.

$p = (n, 0)$ et $p' = (-n, 0)$ représentent le même plan. Par contre, elle est **complète** car chaque plan est représenté par (au moins) un jeu de paramètres. Bien qu'elle ne soit pas **minimale** puisqu'elle nécessite 4 paramètres liés par une contrainte, elle est dérivable, sauf en $d = 0$.

3.2.1.3 Les variétés sont des surfaces lisses

Il est souvent utile de représenter une variété comme une « surface lisse » de dimension n dans $\mathbb{R}^k$ avec une correspondance bijective et des contraintes dérivables non dégénérées (de dimension constante)[2]. Cela permet en particulier de montrer que l'ensemble de primitives qui nous intéresse possède bien une structure de variété (voir par exemple (Bourguignon, 1990, chap. VII)). A l'inverse, le théorème de Whitney assure que n'importe quelle variété de dimension n peut être plongée de cette manière dans un espace $\mathbb{R}^k$ où la dimension k est bornée par $k \leq 2n + 1$ (voir (Spivak, 1979, chap. 9)). Cette façon de voir les choses nous sera en particulier utile pour définir simplement une métrique riemannienne à la section (3.5).

Plans : Reprenons l'exemple des plans, mais en rajoutant une orientation : le plan divise maintenant l'espace en un « intérieur » et un « extérieur ». Si nous choisissons toujours la normale extérieure dans la représentation $p = (n, d)$, la distance d à l'origine devient algébrique et peut être négative. La seule contrainte restante est donc $\|n\| = 1$. La représentation est maintenant **non ambiguë** et nous avons obtenu une bijection entre la variété des plans orientés et $\mathcal{S}_2 \times \mathbb{R}$, où $\mathcal{S}_2$ est la sphère unité dans $\mathbb{R}^3$, ce qui est une surface on ne peut plus lisse (infiniment différentiable) de $\mathbb{R}^4$.

Matrices orthogonales : De la même manière, les matrices orthogonales de $\mathbb{R}^n$ vérifient la contrainte $R.R^{\mathrm{T}} = R^{\mathrm{T}}.R = I_n$ et constituent donc une surface lisse de $\mathbb{R}^{n^2}$ que l'on nomme $\mathcal{O}_n$ pour (groupe) Orthogonal de $\mathbb{R}^n$. En fait, cette surface comprend deux feuilles ou composantes non connexes. En effet, la contrainte ci-dessus impose que $\det(R)^2 = 1$, soit $\det R = \pm 1$, et comme le déterminant est une fonction continue de $\mathbb{R}^{n^2}$ dans $\mathbb{R}$, on ne peut pas passer continûment de $\det R = -1$ à $\det R = +1$. La feuille de déterminant positif rassemble les rotations de $\mathbb{R}^n$, et cette variété est notée $\mathcal{SO}_n$ pour (groupe) Spécial Orthogonal. Nous verrons au chapitre 7 que la variété des rotations de $\mathbb{R}^3$, qui nous intéressent plus particulièrement, est aussi isomorphe à $\mathcal{P}_3$, l'espace projectif de $\mathbb{R}^4$, grâce à l'équivalence entre une rotation R d'angle θ autour de l'axe (unitaire) n et les deux quaternions unitaires $q = \pm(\cos(\theta/2)\ ,\ \sin(\theta/2).n)$.

Espace projectif : Rappelons que l'espace projectif $\mathcal{P}_n$ est l'ensemble des **directions**, c'est-à-dire l'ensemble des droites vectorielles, de $\mathbb{R}^{n+1}$. On peut obtenir une « visualisation » de cet espace en normalisant n'importe quel point x de la droite (sauf l'origine) pour obtenir un point $n = x/\|x\|$ sur la sphère $\mathcal{S}_n$, auquel on doit associer son point antipodal $-n$ pour obtenir une représentation non ambiguë (le point $-x$ appartient aussi à la droite).

Points : Pour revenir à des primitives plus simples, notons que les points de $\mathbb{R}^n$ constituent évidemment une variété puisqu'ils forment déjà un espace vectoriel. Un exemple plus intéressant est donné avec les points orientés, c'est-à-dire munis d'un vecteur unitaire. On peut modéliser ainsi des points sur une surface orientable : si on peut extraire efficacement l'information nécessaire, autant associer à chaque point la normale extérieure à la surface. La représentation triviale $u = (x, n)$, où x

2. Le terme de surface sera dorénavant pris dans ce sens général, l'analogie avec les surfaces lisses ou les courbes de $\mathbb{R}^3$ étant une bonne façon de voir les choses dans la plupart des cas.

est le point et n le vecteur unitaire, identifie cette variété à $\mathbb{R}^n \times \mathcal{S}_{n-1}$. Si la surface (ou la courbe en 3 dimensions) n'est pas orientable, on ne peut pas décider d'un intérieur et d'un extérieur cohérent, et on ne peut associer qu'une direction. On utilise alors la représentation $u = (x, \pm n)$ qui identifie cette variété à $\mathbb{R}^n \times \mathcal{P}_{n-1}$.

Nous supposons simplement pour l'instant que nous disposons d'une représentation complète et non ambiguë de la variété sous forme de surface lisse dans $\mathbb{R}^k$, que nous identifions avec la variété.

3.2.2 Transformations et groupes de Lie

Un certain nombre de groupes de transformations sont familiers : translations, rotations, similitudes ou transformations affines... De façon plus générale, une transformation d'un ensemble $\mathcal{X}$ est une bijection g de $\mathcal{X}$ dans $\mathcal{X}$. Comme l'ensemble $\mathcal{X}$ sera pour nous une variété, nous supposerons de plus que la transformation $\mathrm{g}\star\mathrm{x} = \mathrm{g(x)}$ est continue et différentiable. Par ailleurs, g est une bijection, et il existe donc une transformation **inverse** que l'on note $\mathrm{g}^{(-1)}$. Notons encore que l'on peut **composer** deux transformation g_1 et g_2 pour former une nouvelle fonction $\mathrm{g(x)} = \mathrm{g}_2(\mathrm{g}_1(\mathrm{x}))$ qui est aussi une transformations. On notera $\mathrm{g} = \mathrm{g}_2 \circ \mathrm{g}_1$. L'ensemble de toutes les transformations, muni de ces deux opérations d'inversion et de composition constitue donc une groupe appelé **groupe de transformation général** de l'ensemble $\mathcal{X}$. L'élément neutre est la transformation identité Id qui laisse chaque élément de $\mathcal{X}$ inchangé.

Ce groupe étant de dimension infinie dans les cas intéressants, on se restreint en général à un groupe de transformation particulier $\mathcal{G}$ qui peut être vu comme un sous-groupe du groupe général, ou bien comme un ensemble de transformations stable par la composition et l'inversion.

3.2.2.1 Groupes de Lie

On obtient la classe importante des **groupes de Lie** si $\mathcal{G}$ a une structure topologique séparée[3] et si les opérations de composition et d'inversion sont continues et différentiables. L'espace $\mathcal{G}$ est alors une variété différentielle. En fait, la plupart des groupes de transformation usuels sont des groupes de Lie dès lors qu'ils ont des opérations raisonnables.

Dans ce manuscrit, notre principal groupe d'intérêt sera le groupe des transformations rigides. Un élément de ce groupe est constitué d'une matrice de rotation et d'une translation et peut donc être représenté par $\mathrm{f} = (R\,,\, t)$. En utilisant la section précédente, on peut donc dire que la variété des transformations rigides de $\mathbb{R}^n$ est $\mathcal{SO}_n \times \mathbb{R}^n$. Pour obtenir la structure de groupe de Lie, il nous faut ajouter les opérations de composition et d'inversion :

$$\mathrm{f}^{(-1)} = (R^{\mathrm{T}},\, -R^{\mathrm{T}}.t) \qquad \text{et} \qquad \mathrm{f} = \mathrm{f}_2 \circ \mathrm{f}_1 = (R_2.R_1,\, R_2.t_1 + t_2) \tag{3.1}$$

Il est intéressant de remarquer que les opérations que nous avons proposé dans l'équation (3.1) *ne correspondent pas* au produit direct des opérations des groupe $\mathcal{SO}_n$ et $\mathbb{R}^n$ qui auraient été $(R^{\mathrm{T}}\,,\, -t)$ pour l'inversion et $(R_2.R_1\,,\, t_1 + t_2)$ pour la composition. Nous avons ici un **produit semi-direct** $\mathcal{SO}_n \ltimes \mathbb{R}^n$ dans le sens où les rotations sont indépendantes des translations, mais l'inverse est faux. Une même structure de variété peut donc être munie de structures de groupes différentes.

3. Un espace est dit de Hausdorff ou possédant une structure topologique séparée si pour tout point x et y, il existe deux voisinage $\mathcal{U}$ et $\mathcal{V}$ de x et y qui soient disjoints : ces voisinages *séparent* les deux points. Les espaces non séparés sont en fait de peu d'intérêt pour nous et il est assez difficile d'imaginer une signification physique à un tel espace.

3.2.2.2 Action du groupe sur la variété

La structure de groupe de Lie définie ci-dessus est indépendante de la notion de transformation d'un ensemble. Pour revenir à un groupe de transformation, il nous reste à préciser comment un élément g du groupe $\mathcal{G}$ agit sur un point x de la variété $\mathcal{M}$ considérée. En d'autres termes, il faut donner un sens à l'expression : $\mathrm{g} \star \mathrm{x} = \mathrm{g(x)}$.

Notons que les opérations internes et l'action du groupe sur la variété ne sont pas indépendants puisque l'on doit avoir la propriété de distributivité

$$(\mathrm{g}_1 \circ \mathrm{g}_2) \star \mathrm{x} = \mathrm{g}_1 \star (\mathrm{g}_2 \star \mathrm{x})$$

sans oublier l'action neutre de l'identité :

$$\forall \mathrm{x} \in \mathcal{M} \qquad \mathrm{e} \star \mathrm{x} = \mathrm{x}$$

Ainsi, les opération proposés à l'équation (3.1) pour le groupe des transformations rigides sont compatibles avec l'action suivante sur $\mathbb{R}^n$:

$$\mathrm{f} \star x = R.x + t \tag{3.2}$$

3.2.2.3 De l'action dans l'espace euclidien à l'action sur les primitives

Dans le cas des primitives géométriques, la transformation s'applique en général à l'espace euclidien ambiant dans lequel sont définis ces objets géométriques, mais nous voudrions travailler directement sur les primitives en faisant abstraction de cet espace d'origine. On doit cependant faire bien attention à ce que la nature des primitives définies soit conservée par les transformations considérées : deux vecteurs orthonormés ne sont plus ni orthogonaux ni normalisés après une transformation affine générale. La première contrainte est donc que la variété $\mathcal{M}$ des primitives soit globalement invariante par le groupe de transformation $\mathcal{G}$.

Reprenons la définition « ensembliste » d'une primitive $\mathrm{x} = \{x \in \mathbb{R}^n \,/\, \beta(p, x) = 0\}$ et appliquons une transformation g à cet ensemble :

$$\mathrm{g} \star \mathrm{x} = \{y \in \mathbb{R}^n \;/\; \beta(p, \mathrm{g}^{(-1)} \star y) = 0\}$$

Si nos paramètres sont bien choisis (i.e. la représentation est non ambiguë, complète, continue et dérivable), alors il ne doit exister qu'un seul paramètre q tel que $\beta(q, x) = 0$ représente $\mathrm{g} \star \mathrm{x}$. On notera alors $q = \mathrm{g} \star p$.

Le groupe $\mathcal{G}$ est alors un groupe de transformation sur la variété $\mathcal{M}$ des primitives. L'écriture explicite de l'action du groupe sur certaines primitives peut être très compliquée dans certaines représentations et conduire à des équations non-linéaires (voir par exemple l'action d'un vecteur rotation et surtout la composition de vecteurs rotations au chapitre 7). Cependant, les cas usuels sont relativement simples. Reprenons l'exemple des plans orientés identifiés avec $\mathcal{S}_2 \times \mathbb{R}$ par $\mathrm{p} = (n, d)$ (n unitaire). Faisons agir la transformations rigide f sur le plan $\beta(\mathrm{p}, x) = 0$:

$$\beta(\mathrm{p}, \mathrm{f}^{(-1)} \star x) = \langle n \mid R^{\mathrm{T}}.(x - t) \rangle - d = \langle x \mid R.n \rangle - d - \langle t \mid R.n \rangle$$

On a donc un seul et unique q vérifiant $\beta(\mathrm{q}, x) = \beta(\mathrm{p}, \mathrm{f} \star x)$, et l'action est donnée par :

$$\mathrm{f} \star \mathrm{p} = \mathrm{q} = (R.n,\ -d - \langle t \mid R.n \rangle)$$

De même, on détermine que l'action d'une transformation rigide f sur le point orienté $\mathrm{u} = (x, n)$ (où n est unitaire) est donnée par

$$\mathrm{f} \star \mathrm{u} = (R.x + t,\ R.n)$$

Nous nous intéresserons dans ce manuscrit à un autre type de primitives : les repères, constitués d'un point et d'un trièdre orthonormé direct. Nous avons déjà observé que nous ne pouvions pas utiliser le groupe des transformations affines puisque l'orthonormalité du trièdre ne serait pas conservée, mais les transformations rigides sont appropriées. Une particularité des repère est d'ailleurs leur équivalence avec ces transformations.

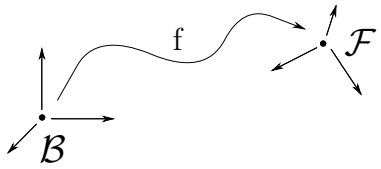


FIG. 3.2 – *Repères et transformations rigides.*

Soit $\mathcal{B} = \{o, e_1, e_2, e_3\}$ le repère canonique de l'espace euclidien $\mathbb{R}^3$ (on a donc en terme de coordonnées : $o = (0,0,0)^{\mathrm{T}}$ et $[e_1, e_2, e_3] = I_3$), et $\mathcal{F} = \{x, e'_1, e'_2, e'_3\}$ un autre repère dont les coordonnées sont exprimées dans le repère canonique $\mathcal{B}$. Le mouvement rigide qui amène $\mathcal{B}$ sur $\mathcal{F}$ est unique et donné par

$$\mathrm{f} = (U, x) \qquad \text{où} \qquad U = \left[e'_1, e'_2, e'_3\right]$$

L'orthonormalité du repère $\mathcal{F}$ assure en effet que U soit une rotation. La variété des repères (3D) est donc $\mathcal{S}O_3 \times \mathbb{R}^3$, et l'action d'une transformation rigide $\mathrm{g} = (R,\ t)$ sur un repère $\mathrm{f} = (U,\ x)$ est « équivalente » à la composition :

$$\mathrm{g} \star \mathrm{f} = (R.U,\ R.x + t) \equiv \mathrm{g} \circ \mathrm{f}$$

3.2.3 Variétés homogènes et invariants unaires

Nous aurons besoin par la suite d'une relation particulière entre le groupe de transformation et la variété des primitives : soit $\mathrm{o} \in \mathcal{M}$ un point particulier que l'on appelle **l'origine**. La variété $\mathcal{M}$ est dite **transitive** ou **homogène** pour le groupe $\mathcal{G}$ si n'importe quel point de la variété peut être atteint à partir de l'origine par une transformation du groupe, autrement dit si :

$$\mathcal{G} \star \mathrm{o} = \{\mathrm{x} = \mathrm{g} \star \mathrm{o} \ /\ \mathrm{g} \in \mathcal{G}\} = \mathcal{M}$$

Cela veut dire en fait que les primitives n'ont pas d'**invariants** par le groupe. Si l'on utilise par exemple comme primitive les repères d'orientation directe *et* indirecte, aucune transformation rigide ne pourra nous amener d'une orientation à l'autre : l'orientation est donc invariante, et la variété n'est pas homogène. Si par contre on ne prend que les repères d'orientation directe (ce que l'on suppose dans ce manuscrit par l'appellation *repère*) ou si l'on ajoute les symétries au groupe de transformation, alors la variété est homogène.

En fait, nous supposerons que nos primitives $\mathrm{x}' \in \mathcal{M}'$ peuvent être factorisées en une partie homogène $\mathrm{x} \in \mathcal{M}$, qui varie sous l'action du groupe $\mathcal{G}$, et une partie invariante $\mathrm{i} \in \mathcal{I}$. Une primitive s'écrit donc

$$\mathrm{x}' = (\mathrm{x},\ \mathrm{i}) \in \mathcal{M}' = \mathcal{M} \times \mathcal{I}$$

Ces invariants sont particuliers à chaque primitive et ne dépendent pas d'un doublet ou d'un triplet de primitives comme pour l'indexation géométrique (on va même jusqu'à 4 points pour trouver un invariant projectif en vision par ordinateur). Nous les appelons donc **invariants unaires**. Notons que ces invariants unaires ne sont pas forcément différentiels. Les **invariants différentiels** seraient obtenus en considérant des quantités différentielles invariantes le long d'une courbe, d'une surface ou plus généralement d'une sous-variété de la variété de nos primitives.

Par exemple, les points extrémaux sont constitués de quantités variables sous l'action du groupe des transformations rigides : la position et l'orientation du trièdre, et de deux invariants unaires : les courbures principale (qui sont ici des invariants différentiels). Dans le cas des protéines, la primitive originale est un acide aminé que nous caractérisons par la position des trois atomes C_α, C' et N du squelette (nous rappelons cette modélisation dans la figure 3.3). Pour ces 9 variables, n'y a que 6 degrés de liberté sous l'action des transformations rigides : nous factorisons donc la primitive originale en un repère (partie homogène) et 3 invariants unaires qui apparaissent naturellement : un angle et deux distances. Malheureusement, ces trois invariants unaires nous sont inutiles pour la reconnaissance puisqu'ils sont (théoriquement) identiques dans tous les acides aminés.

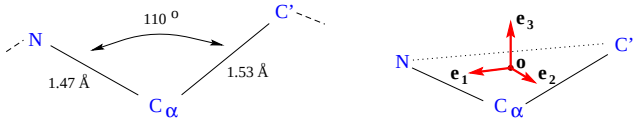


FIG. 3.3 – *Description géométrique de la configuration d'un acide aminé.*

Comme le groupe n'agit pas sur les invariants (par définition), la connaissance du groupe et de son action ne nous apporte aucune information pour gérer les invariants, en particulier sur le choix de la métrique (voir section 3.5). Nous supposons donc dans la suite que nous avons factorisé les invariants unaires et que la variété $\mathcal{M}$ est homogène. On peut dans ce cas identifier les éléments $x \in \mathcal{M}$ de la variété avec des sous-ensembles du groupe $\mathcal{G}$ de la manière suivante. Soit $\mathcal{H}$ le sous-ensemble des transformations qui laisse l'origine o invariante :

$$\mathcal{H} = \{h \in \mathcal{G} \ / \ h \star o = o\} \tag{3.3}$$

$\mathcal{H}$ porte le nom de **sous-groupe d'isotropie** ou de stabilité du groupe $\mathcal{G}$ au point o. Dans un groupe, la **translation à gauche** est la composition à gauche par un élément fixe g. Les ensembles de transformations obtenus par la translation à gauche de $\mathcal{H}$ (les éléments de l'espace quotient $\mathcal{G}/\mathcal{H}$) peuvent être identifiés avec les éléments de la variété $\mathcal{M}$. En effet, si g est *une* transformation qui amène l'origine sur $x = g \star o$, alors $g \star \mathcal{H}$ est *l'ensemble* des transformations qui amènent l'origine sur x. On appelle **coset** [4] de x cet ensemble et nous le notons $\mathcal{F}_x$:

$$\mathcal{F}_x = \{g \in \mathcal{G} \ / \ g \star o = x\} \tag{3.4}$$

Pour en finir avec les notations, nous aurons souvent besoin d'un représentant du coset $\mathcal{F}_x$ que nous noterons en général f_x. D'une manière ple générale, nous appelons **fonction de placement** une fonction qui a tout point x de la variété $\mathcal{M}$ associe un représentant f_x du coset du point.

3.2.3.1 Le cas des repères

Pour les repères, l'origine est choisie au repère canonique : $o = (Id, 0)$, et comme la variété est en bijection avec le groupe rigide $\mathcal{G}$, le groupe d'isotropie et tous les cosets se réduisent à un seul

4. Nous prenons ici la dénomination anglaise, plus concise. Nous devrions en fait dire : les classes à gauche de $\mathcal{G}$ modulo $\mathcal{H}$.

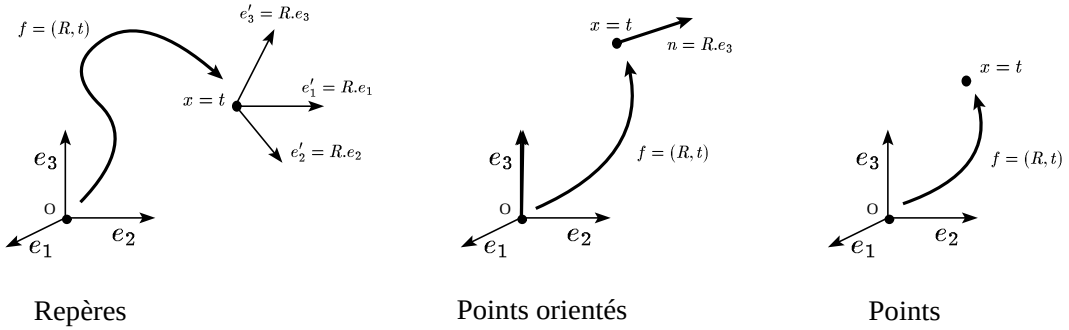


FIG. 3.4 – *Les origines choisies pour les repères, les points orientés et les points et leur déplacement par une transformation rigides. Ces trois types de primitives n'ont pas d'invariants unaires.*

élément :

$$\mathcal{H} = \{\, Id\} \qquad \text{et} \qquad \mathcal{F}_{\mathrm{f}} = \{\mathrm{f}\}$$

Ce cas particulier où la variété est « équivalente » au groupe conduit à des simplifications importantes dans la théorie qui suit.

3.2.3.2 L'exemple des points

Il est naturel de prendre comme origine $\mathrm{o} = (0,0,0)^{\mathrm{T}}$. Le groupe d'isotropie est obtenu en résolvant l'équation $R.0 + t = 0$, ce qui impose une translation nulle :

$$\mathcal{H} = \{(R,0) \;/\; R \in \mathcal{S}O_3\}$$

Une transformation qui amène l'origine au point x est la translation $\mathrm{f}_x = (I_3, x)$, et les cosets sont donc les ensembles

$$\mathcal{F}_{\mathrm{x}} = \{(R,x) \;/\; R \in SO_3\}$$

Notons que les points sont « équivalents » aux translations et que c'est cela qui permet d'identifier points et vecteurs et de parler de différences ou d'addition de points.

3.2.4 Cartes et atlas

Pour pouvoir travailler localement dans une variété comme dans l'espace vectoriel $\mathbb{R}^n$, et en particulier pour pouvoir dériver, nous aurons besoin d'un système de coordonnées locales, autrement dit une représentation **minimale**, dont la dimension est celle de la variété. En général, nous ne pouvons pas couvrir toute la variété de manière différentiable avec une seule carte (il y a dans le cas contraire un difféomorphisme **global** avec un espace vectoriel). On demande donc un ensemble de **cartes** couvrant la variété et formant un **atlas**, exactement comme un atlas géographique représente la surface de la terre de manière continue partout, en changeant au besoin de carte de temps en temps. Une carte est définie par une bijection $\mathrm{x} = \varphi_i(x)$ d'un ouvert $\mathcal{D}_i$ de la représentation dans un ouvert de la variété $\mathcal{M}$. L'ensemble des cartes doit se recouvrir pour qu'il soit possible de passer de l'une à l'autre. On trouvera par exemple une étude de différentes représentations des droites 2D et 3D, des plans et des rotations dans (Ayache, 1991). Nous ne présentons ici que quelques exemples de cartes qui nous serviront par la suite.

3.2.4.1 Représentation des rotations

Nous verrons à la section (7.1) qu'une rotation 3D peut être caractérisée par un angle θ et un axe n (un vecteur unitaire). Cette représentation n'est cependant pas minimale puisque nous avons une contrainte unitaire sur l'axe, axe qui n'est de plus pas défini pour la rotation identité. On obtient une représentation minimale mais ambiguë en considérant le vecteur rotation $r = \theta.n$. Pour définir un atlas, nous avons besoin au minimum de 4 cartes :

- **Carte 1 :** Les rotations qui ne sont pas des réflexions sont représentées par les vecteurs rotation de la boule ouverte $\mathcal{B}_3(0, \pi)$.
- **Carte 2, 3 et 4 :** Les rotations différentes de l'identité dont l'axe n'est pas orthogonal à l'axe des x (respectivement des y et des z) sont représentées par les vecteurs rotation de la demi-boule ouverte $\mathcal{B}_3^{x^+} = \{r \in \mathcal{B}_3(0, 2\pi) \ / \ r_x > 0\}$ (respectivement $\mathcal{B}_3^{y^+}$, $\mathcal{B}_3^{z^+}$).

En théorie, nous devrions utiliser les 4 cartes et indexer le vecteur rotation sur la carte utilisée. En pratique, nous pourrons nous contenter d'utiliser la **carte principale** (la première) en se rappelant que, sur la frontière du domaine, les vecteurs $r = \pi.n$ et $r = -\pi.n$ sont identifiés. Si l'on note $R(r)$ et $r(R)$ les fonctions qui relient la carte principale à la variété, on peut écrire directement les opérations de composition et d'inversion dans la carte :

$$r^{(-1)} = r(R(r)^{\mathrm{T}}) \qquad \text{et} \qquad r_2 \circ r_1 = r(R(r_2).R(r_1))$$

Le détail de ces fonctions et de leur dérivation sera présenté à la section (7.3).

3.2.4.2 Transformations rigides, repères et points

A partir de la représentation des rotations et des vecteurs unitaires, et de leurs opérations de base, on peut construire facilement des représentations pour les transformations rigides, les repères et les point orientés. Nous résumons simplement ici les représentations utilisées et les opérations de base.

Transformations rigides

- Représentation : on utilise la carte principale qui représente une transformations rigides $\mathrm{f} = (R, t)$ par le vecteur rotation r correspondant à la matrice de rotation R et la translation t : $f = (r, t)$.
- Composition : $f_2 \circ f_1 = (r_2 \circ r_1 \ , \ r_2 \star t_1 + t_2)$
- Inversion: $f^{(-1)} = (r^{(-1)} \ , \ r^{(-1)} \star (-t))$

Repères

- Représentation : elle est calquée sur celle des transformations rigides. On représente le repère $\mathrm{f} = (R, x)$ par le vecteur $f = (r, x)$.
- Action des transformations rigides : $f \star f_1 \equiv f \circ f_1$
- Groupe d'isotropie : $\mathcal{H} = \{o\}$
- Représentant des repères : $f_f = f \in \mathcal{F}_f$.

Points

- Représentation : un point x est évidemment représenté par son vecteur de coordonnées.
- Action de la transformation rigide $f = (r, t)$: $f \star x = r \star x + t$.
- Groupe d'isotropie : $\mathcal{H} = \{f = (r, 0) \ / \ R(r) \in \mathcal{SO}_3\}$
- Représentant des repères : $f_x = (0, x) \in \mathcal{F}_x$.

3.3 Mesure invariante ou uniforme

« Mesurer ce qui est mesurable, et rendre mesurable ce qui ne l'est pas. »
Galilei, Galileo (1564 - 1642), cité par H. Weyl

« The author discusses valueless measures in pointless spaces. »
P.R. Halmos, I want to be a Mathematician, 1985

3.3.1 Probabilités géométriques classiques

L'objet des probabilités géométriques classiques est de déterminer la probabilité d'observer un événement quand un certain nombre d'éléments géométriques sont répartis au hasard. Nous avons vu avec le paradoxe de Bertrand que le problème de base est la définition de la **mesure uniforme** sur la variété, et qu'une hypothèse raisonnable est de supposer l'invariance de cette mesure par l'action du groupe de transformation considéré.

Nous ne nous intéressons ici qu'à la détermination de la mesure invariante, qui sera utilisée au chapitre suivant pour définir des densités de probabilité sur la variété ayant de bonnes propriétés, les *probabilités géométriques* en question étant en fait plus reliées à la stéréologie qu'à la gestion de l'erreur sur les mesures de primitives géométriques. Une application directe de ces *probabilités géométriques* sera cependant évoquée dans la seconde partie avec l'étude de la robustesse des algorithmes d'appariement et le calcul du nombre moyen de faux positifs.

Pour en revenir à la mesure invariante, observons qu'un réel aléatoire $\boldsymbol{x}$ a une probabilité uniforme sur $\mathbb{R}$ si la probabilité qu'il soit dans un intervalle $]x_0 \ ; \ x_0 + d[$ est la même pour tout x (dans un certain ouvert avec d suffisamment faible). Ceci se traduit au niveau de la mesure par $d(x_0 + x) = dx$ pour x_0 constant, ce que l'on peut interpréter comme l'invariance en translation de la mesure de Lebesgue. De la même façon, on définit la **mesure uniforme ou invariante** sur la variété $\mathcal{M}$ sous l'action du groupe $\mathcal{G}$ comme la mesure (ou l'élément de volume infinitésimal) qui est invariant par l'action de n'importe quel élément fixe f du groupe. Soit $d\mathcal{M}$ une telle mesure, cela implique que $d\mathcal{M}(\mathrm{f} \star \mathrm{x}) = d\mathcal{M}(\mathrm{x})$ quelque soit le point x de $\mathcal{M}$.

On peut rechercher l'expression de cette mesure invariante directement dans une représentation donnée (voir (Kendall et Moran, 1963)), mais un formalisme général est développé dans (Santalo, 1976) pour l'extraire à partir de la mesure $d_L\mathcal{G}$ invariante à gauche sur le groupe $\mathcal{G}$. Nous en donnons ici une traduction simplifiée en utilisant une représentation minimale.

3.3.2 Mesure invariante sur un groupe de Lie (mesure de Haar)

Dans un groupe de Lie, la composition peut être vue vue comme l'action à droite ou à gauche du groupe sur lui-même. Nous pouvons donc nous intéresser cette fois-ci à la mesure invariante à

gauche ($d_L\mathcal{G}(\mathrm{g} \circ \mathrm{f}) = d_L\mathcal{G}(\mathrm{f})$ pour toute transformation $\mathrm{g} \in \mathcal{G}$ fixée) ou à la mesure invariante à droite ($d_R\mathcal{G}(\mathrm{f} \circ \mathrm{g}) = d_R\mathcal{G}(\mathrm{f})$). Comme le groupe agit à gauche sur les variétés de primitives dans notre contexte, nous sommes principalement intéressés par la mesure invariante à gauche, que nous appellerons simplement mesure invariante.

Pour être mathématiquement correct, nous devrions définir l'espace $C_o(\mathcal{G})$ des fonctions réelles continues sur $\mathcal{G}$ à support compact[5] et étudier les fonctionnelles I sur cet espace, c'est-à-dire les fonctions de $C_o(\mathcal{G})$ dans $\mathbb{R}$, qui sont homogènes ($I(k.\alpha) = k.I(\alpha)$ pour $k \in \mathbb{R}$) et additives ($I(\alpha+\beta) = I(\alpha) + I(\beta)$). Cela justifie la notation intégrale :

$$I(\alpha) = \int_{\mathrm{f}\in\mathcal{G}} \alpha(\mathrm{f}).d\mathcal{G}(\mathrm{f})$$

et la condition d'invariance à gauche s'écrit alors :

$$\forall \mathrm{g} \in \mathcal{G} \qquad \int_{\mathrm{f}\in\mathcal{G}} \alpha(\mathrm{g} \circ \mathrm{f}).d_L\mathcal{G}(\mathrm{g} \circ \mathrm{f}) = \int_{\mathrm{f}\in\mathcal{G}} \alpha(\mathrm{f}).d_L\mathcal{G}(\mathrm{f}) \tag{3.5}$$

(Hochschild, 1965) montre que, si le groupe est localement compact, il existe une et une seule fonctionnelle invariante à gauche (à un facteur d'échelle près) vérifiant les propriétés ci-dessus. Cette intégrale est appelée l'intégrale de Haar (à gauche). De manière symétrique, il existe une unique intégrale de Haar à droite.

Il est intéressant de noter que ces mesures invariantes à droite et à gauche sont généralement différentes : cette différence est quantifiée par le **module** $\Delta\mathcal{G}$ défini par la relation : $d_L\mathcal{G}(\mathrm{f}) = \Delta\mathcal{G}(\mathrm{f}).d_R\mathcal{G}(\mathrm{f})$. Le groupe est dit uni-modulaire si $\Delta\mathcal{G}$ est constant en tout point du groupe (i.e. égal à 1 à une constante de normalisation près), c'est-à-dire quand les mesures invariantes à droite et à gauche sont identiques. Un groupe compact est toujours uni-modulaire, mais les groupes seulement localement compacts peuvent avoir des mesures de Haar différentes à droite et à gauche. Les mesures de Haar à droite et à gauche sont ainsi identiques sur $\mathcal{SO}_3$ puisque le groupe des rotations 3D est compact (c'est un sous groupe de l'ensemble des matrices vérifiant $\det(A) = 1$, et l'image réciproque d'un compact par une application continue est compacte). Les transformation rigides 3D sont également uni-modulaires (voir section 7.4.1.2) bien que le groupe ne soit plus que localement compact à cause de l'introduction des translations.

Les mesures invariantes peuvent être calculées dans le cas général par les équations de Maurer-Cartan (Santalo, 1976), mais un théorème très intéressant nous permet de les exprimer simplement dans une représentation *minimale* : supposons que le domaine de définition de la carte couvre presque toute la variété[6] (pour n'avoir qu'une seule carte à utiliser) et que le jacobien de la translation à gauche $J_L(f)$ existe et soit continu presque partout. Alors la mesure invariante s'exprime dans cette carte par :

$$d_L\mathcal{G}(f) = \frac{df}{|J_L(f)|} \qquad \text{avec} \qquad J_L(f) = \left.\frac{\partial(f \circ e)}{\partial e}\right|_{e=\,Id} \tag{3.6}$$

La mesure invariante à droite s'exprime de la même façon en utilisant le jacobien de la translation à droite J_R . Le module du groupe est donc :

$$\Delta\mathcal{G}(f) = \frac{|J_R(f)|}{|J_L(f)|}$$

5. Un espace topologique E ou un sous-ensemble E d'un espace topologique est dit **compact** si l'on peut trouver dans toute union d'ouverts contenant E un nombre fini de ces ouverts dont l'union contient E. Intuitivement, cela signifie que E ne s'étend pas jusqu'à l'infini.

6. Comme nous intégrons des fonctions et non pas des distributions, nous pouvons « oublier » un sous-ensemble du groupe de mesure nulle. On pourrait également partitionner la variété en plusieurs domaines complémentaires dans différentes cartes avec les propriétés requises.

Une propriété intéressante de ce module est que $\Delta\mathcal{G}(f^{(-1)}) = \Delta\mathcal{G}(f)^{(-1)}$.

Preuve :
Soit $d_L\mathcal{G}(f)$ la mesure proposée dans l'équation (3.6), où l'on note $|J| = |\det(J)|$. On a :

$$d_L\mathcal{G}(g \circ f) = \frac{d(g \circ f)}{\left|\frac{\partial((g\circ f)\circ e)}{\partial e}\Big|_{e=Id}\right|} \qquad \text{et} \qquad d(g \circ f) = \left|\det\left(\frac{\partial(g \circ f)}{\partial f}\right)\right| .df$$

On obtient par la dérivation en chaîne :

$$\frac{\partial((g \circ f) \circ e)}{\partial e}\Big|_{e=Id} = \frac{\partial(g \circ f)}{\partial f} . \frac{\partial(f \circ e)}{\partial e}\Big|_{e=Id}$$

Comme $\det(A.B) = \det(A).\det(B)$ pour les matrices carrées, on peut conclure que $d_L\mathcal{G}(g\circ f) = d_L\mathcal{G}(f)$.
La preuve est identique pour l'invariance à droite de la mesure $d_R\mathcal{G}$ proposée. ■

3.3.3 Exemple sur les rotations et les transformation rigides

On utilise le jacobien de la translation de l'identité dérivé à la section (7.3.5) pour déterminer la mesure uniforme sur le vecteur rotation :

$$d_L\mathcal{G}(r) = d_R\mathcal{G}(r) = \frac{4\sin^2(\theta/2)}{\theta^2}\, dr \tag{3.7}$$

où $\theta = \|r\|$. Pour les transformation rigides, la section (7.4.1.2) montre que la mesure invariante (à droite comme à gauche) dans la représentation $f = (r, t)$ est

$$d_L\mathcal{G}\,(f) = d_R\mathcal{G}\,(f) = \frac{4\sin^2(\theta/2)}{\theta^2}\, dr.dt \tag{3.8}$$

Les rotations et les transformation rigides sont donc uni-modulaires.

3.3.4 Mesure invariante sur une variété homogène

Nous avons vu à la section (3.2.3) comment identifier une variété homogène $\mathcal{M}$ avec l'espace quotient $\mathcal{G}/\mathcal{H}$. On peut trouver grâce à la section précédente les mesures invariantes (à gauche) $d_L\mathcal{G}$ et $d_L\mathcal{H}$ sur les groupes $\mathcal{G}$ et $\mathcal{H}$, et écrire que $d_L\mathcal{G} = d\mathcal{M}.d_L\mathcal{H}$ où $d\mathcal{M}$ est une mesure sur la variété $\mathcal{M}$ (ou l'espace quotient $\mathcal{G}/\mathcal{H}$). (Santalo, 1976) donne plusieurs formes d'une condition nécessaire et suffisante pour que $d\mathcal{M}$ soit une mesure invariante (i.e. $d\mathcal{M}(\mathrm{g} \star \mathrm{x}) = d\mathcal{M}(\mathrm{x})$). L'une d'elle peut se formuler ainsi (e est ici un élément de la variété $\mathcal{M}$) : supposons que nous utilisons une représentation minimale dont le domaine de définition couvre presque toute la variété et que le jacobien de la translation de l'origine $J(f)$ existe et soit continu presque partout. Alors la mesure est invariante si et seulement si :

$$\forall h \in \mathcal{H}, \qquad |J(h)| = \left|\frac{\partial(h \star e)}{\partial e}\Big|_{e=\mathrm{o}}\right| = 1 \tag{3.9}$$

Cela signifie que la mesure d'un élément de volume infinitésimal à l'origine ne change par sous l'action du groupe d'isotropie, c'est-à-dire avec les transformations qui laissent l'origine inchangée. Si cette condition n'est pas vérifiée, il n'existe pas de mesure invariante sur le groupe, et dans le cas contraire on peut la calculer comme nous l'avons fait pour le groupe :

$$d\mathcal{M}(x) = \frac{dx}{|J(f_x)|} \qquad \text{avec} \qquad J(f_x) = \frac{\partial(f_x \star e)}{\partial e}\Big|_{e=\mathrm{o}} \qquad \text{et} \qquad f_x \in \mathcal{F}_x \tag{3.10}$$

Preuve : Soit $d\mathcal{M}(x)$ la mesure proposée ci-dessus. La condition d'existence (3.9) est en effet requise pour que $d\mathcal{M}$ soit invariante vis à vis du choix de $f_x \in \mathcal{F}_x$: soit f_x et $f'_x = f_x \circ h$ (avec $h \in \mathcal{H}$) deux transformations de $\mathcal{F}_x$. Par la règle de dérivation en chaîne, on peut écrire :

$$J(f'_x) = \left.\frac{\partial(f_x \star (h \star e))}{\partial e}\right|_{e=\mathrm{o}} = \left.\frac{\partial(f_x \star e')}{\partial e'}\right|_{e'=\mathrm{o}} . \left.\frac{\partial(h \star e)}{\partial e}\right|_{e=\mathrm{o}} = J(f_x). \left.\frac{\partial(h \star e)}{\partial e}\right|_{e=\mathrm{o}}$$

et on a $|J(f_x)| = |J(f'_x)|$ si et seulement si $\left|\left.\frac{\partial(h\star e)}{\partial e}\right|_{e=\mathrm{o}}\right| = 1$. La preuve de l'invariance de la mesure $d\mathcal{M}(x)$ est alors très similaire à celle de la mesure du groupe et peut être obtenue en remplaçant $(f \circ e)$ par $(f_x \star e)$. ■

3.3.4.1 Exemple 1 : les repères

Puisque l'action des transformations rigides sur les repères est identifiable à la composition à gauche dans le groupe, la mesure invariante sur les repères (pour les transformations rigides) est identique. Si l'on note $f = (r, x)$, on a :

$$d\mathcal{M}(f) = \frac{4\sin^2(\theta/2)}{\theta^2}\, dr.dx$$

Notons que la condition d'existence est remplie puisque le groupe d'isotropie $\mathcal{H}$ est réduit à l'identité :

$$\left|\left.\frac{\partial(\ Id \star e)}{\partial e}\right|_{e=\mathrm{o}}\right| = |\det(I_6)| = 1$$

3.3.4.2 Exemple 2 : le cas des points

Le groupe d'isotropie est caractérisé par une translation nulle et on a $h \star x = R.x$. La condition d'existence s'écrit donc $|R| = 1$, ce qui est toujours vérifié (pour des rotations). En utilisant la translation $f_x = (0, x)$ comme fonction de placement (représentant du coset $\mathcal{F}_x$), on a $f_x \star e = x$ et le jacobien se réduit donc à $J(f_x) = I_3$. La mesure invariante est donc la mesure de Lebesgue :

$$d\mathcal{M}(x) = dx$$

Notons que si l'on considère le groupe de transformation affine, le groupe d'isotropie est toujours caractérisé par une translation nulle et $h \star x = A.x$, où A est maintenant une matrice quelconque de déterminant non nul. Il n'existe donc pas de mesure invariante puisque $|A| \neq 1$ en général.

3.4 Distance invariante

> *« We come now to the question: what is a priori certain or necessary, respectively in geometry (doctrine of space) or its foundations? Formerly we thought everything; nowadays we think nothing. Already the distance-concept is logically arbitrary; there need be no things that correspond to it, even approximately. »*
> A. Einstein, "Space-Time." Encyclopaedia Britannica, 14th ed

Dans la section précédente, nous avons pu déterminer les conditions d'existence de la mesure invariante en utilisant les jacobiens des opérations de base dans une représentation donnée. Nous nous intéressons ici aux questions de métrique sur le groupe et la variété.

La distance est très souvent utilisée dans les algorithmes géométriques pour classifier et quantifier des différences ou comme critère de minimisation, par exemple pour le recalage. Elle est même au cœur de certains algorithmes comme le plus proche voisin itératif (Iterative Closest Point: (Besl et McKay, 1992; Zhang, 1994)). Pour pouvoir comparer des primitives géométriques et généraliser ainsi un grand nombre de méthodes utilisées sur les points, nous avons besoin d'une fonction de distance sur nos primitives.

Cependant, nous avons vu à la section (2.3.4) qu'une distance sur la représentation d'une variété n'est généralement pas une distance sur la variété. Rappelons tout d'abord qu'une distance sur la variété $\mathcal{M}$ est une fonction de $\mathcal{M}^2$ dans $\mathbb{R}$ vérifiant les quatre propriétés suivantes.

- **Positivité** : $\mathrm{dist}(\mathrm{x},\mathrm{y}) \geq 0$.
- **Symétrie** : $\mathrm{dist}(\mathrm{x},\mathrm{y}) = \mathrm{dist}(\mathrm{y},\mathrm{x})$.
- **Séparation**, aussi appelé caractère **défini** de la distance : $(\mathrm{dist}(\mathrm{x},\mathrm{y}) = 0) \Leftrightarrow (\mathrm{x} = \mathrm{y})$. Sans cette propriété, la fonction dist prend le nom d'**écart**.
- **Inégalité triangulaire** : $\mathrm{dist}(\mathrm{x},\mathrm{y}) + \mathrm{dist}(\mathrm{y},\mathrm{z}) \geq \mathrm{dist}(\mathrm{x},\mathrm{z})$

3.4.1 Utilité d'une distance invariante

Comme nous avons vu que l'hypothèse d'invariance sur la mesure permet de lever les paradoxes potentiels sur les probabilités, il est très souhaitable que le résultat de nos algorithmes basés sur la distance ne dépende ni de la représentation choisie, ni du système de coordonnée que l'on a choisis pour la variété. Si un objet (ou une primitive) est le plus proche d'un autre depuis un point de vue, il est souhaitable qu'il le soit toujours depuis un autre point de vue, i.e. après une transformation globale. Dans cette optique, l'utilisation d'une distance invariante garantie la stabilité du résultats de nos algorithmes.

Un autre exemple est la méthode classique aux moindres carrés pour calculer la transformation entre un ensemble de primitives x_i dans une image avec un ensemble de primitives y_i dans une autre image : le critère à minimiser est la somme des distances au carré $C(\mathrm{f}) = \sum_i \mathrm{dist}^2(\mathrm{f} \star \mathrm{x}_i, \mathrm{y}_i)$. Soit $\bar{\mathrm{f}}$ la transformation qui minimise ce critère. Par simplicité, on supposera qu'elle est unique. Si maintenant les primitives x_i sont déplacées globalement par une transformation g ($\mathrm{x}'_i = \mathrm{g} \star \mathrm{x}_i$), le critère devient

$$C'(\mathrm{f}) = \sum_i \mathrm{dist}^2(\mathrm{f} \star \mathrm{x}'_i, \mathrm{y}_i) = \sum_i \mathrm{dist}^2((\mathrm{f} \circ \mathrm{g}) \star \mathrm{x}_i, \mathrm{y}_i) = C(\mathrm{f} \circ \mathrm{g})$$

Avec ou sans la contrainte d'invariance sur la distance, le nouveau résultat est $\bar{\mathrm{f}}' = \bar{\mathrm{f}} \circ \mathrm{g}$. Si par contre ce sont les primitives de la seconde image qui sont globalement déplacées ($\mathrm{y}'_i = \mathrm{g} \star \mathrm{y}_i$), le critère est :

$$C''(\mathrm{f}) = \sum_i \mathrm{dist}^2(\mathrm{f} \star \mathrm{x}_i, \mathrm{y}'_i) = \sum_i \mathrm{dist}^2(\mathrm{f} \star \mathrm{x}_i, \mathrm{g} \star \mathrm{y}_i)$$

L'invariance de la distance est une condition suffisante pour conclure que

$$C''(\mathrm{f}) = \sum_i \mathrm{dist}^2\left((\mathrm{g}^{(-1)} \circ \mathrm{f}) \star \mathrm{x}_i, \mathrm{y}_i\right) = C(\mathrm{g}^{(-1)} \circ \mathrm{f})$$

Le nouveau minimum est donc $\bar{\mathrm{f}}'' = \mathrm{g}^{(-1)} \circ \bar{\mathrm{f}}$, ce qui donne le résultat intuitivement escompté : $\bar{\mathrm{f}} = \mathrm{g} \circ \bar{\mathrm{f}}''$. La même expérience peut être faite si les deux images sont déplacées par la *même transformation* (ce qui correspond à un changement de repère global), et l'invariance de la distance

est requise pour montrer que la transformation trouvée est $\bar{f}''' = g^{(-1)} \circ \bar{f} \circ g$, c'est-à-dire l'expression de l'ancienne transformation après changement de repère.

Dans le reste de cette section, nous tentons de caractériser les propriétés d'une distance invariante sur le groupe et sur la variété.

3.4.2 Distance invariante sur une variété

soient $x, y \in \mathcal{M}$ deux primitives et $g \in \mathcal{G}$ une transformation. L'invariance de la distance s'exprime par $\text{dist}(x, y) = \text{dist}(g \star x, g \star y)$, ce qui signifie en particulier que cette distance est complètement définie par la distance $N(z)$ d'une primitive $x = {f_y}^{(-1)} \star x$ à l'origine. En utilisant comme transformation ${f_y}^{(-1)}$ ou ${f_x}^{(-1)}$ dans la formule ci-dessus, on obtient en effet :

$$\text{dist}(x, y) = \text{dist}({f_y}^{(-1)} \star x, o) = N({f_y}^{(-1)} \star x) = N({f_x}^{(-1)} \star y) \tag{3.11}$$

Les axiomes de la distance se traduisent alors, avec l'hypothèse d'invariance, en les propriétés suivantes :

- $N(x) \geq 0$ et $(N(x) = 0) \Leftrightarrow (x = o)$.
- $N({f_x}^{(-1)} \star o) = N(x)$ for $f_x \in \mathcal{F}_x$ et donc $N(h \star x) = N(x)$ pour tout $h \in \mathcal{H}$.
- $N(x) + N(y) \geq N({f_y}^{(-1)} \star x) = N({f_x}^{(-1)} \star y)$ pour tout $f_x \in \mathcal{F}_x$ et $f_y \in \mathcal{F}_y$

Ces propriétés sont très proches de celles requises pour définir une norme sur un espace vectoriel, excepté cependant l'homogénéité pour la multiplication par un scalaire positif. Pour distinguer la fonction N de la distance invariante associée, nous appelons N la « norme » de la variété. Notons que nous avons ici travaillé directement dans la variété et pas dans une carte particulière.

3.4.3 Distance invariante sur un groupe de Lie

Si nous travaillons maintenant sur le groupe de transformation $\mathcal{G}$, on peut demander à la distance d'être invariante à droite où à gauche. Comme dans le cas de la variété ci-dessus, la distance invariante à gauche est déterminée par une « norme » N_L sur le groupe satisfaisant les propriété suivantes :

- $N_L(f^{(-1)}) = N_L(f)$.
- $N_L(f) \geq 0$ et $(N_L(f) = 0) \Leftrightarrow (f = \text{Id})$.
- $N_L(f) + N_L(g) \geq N_L(g^{(-1)} \circ f) = N_L(f^{(-1)} \circ g)$.

L'inégalité triangulaire devenant $N_R(f) + N_R(g) \geq N_R(g \circ f^{(-1)}) = N_R(f \circ g^{(-1)})$ pour une distance invariante à droite. Les distances invariantes à droite et à gauche correspondantes sont

$$\text{dist}_L(f, g) = N_L(g^{(-1)} \circ f) = N_L(f^{(-1)} \circ g) \qquad \text{et} \qquad \text{dist}_R(f, g) = N_R(g \circ f^{(-1)}) = N_R(f \circ g^{(-1)})$$

Nous ne nous intéressons ici qu'à la distance invariante à gauche car elle induit une distance invariante sur une variété homogène.

3.4.4 Distance induite sur la variété par celle du groupe

Soit N_L une « norme » sur le groupe $\mathcal{G}$. On défini la « semi-norme » induite sur la variété homogène $\mathcal{M}$ par :

$$N(x) = \inf_{(h \in \mathcal{H},\, f_x \in \mathcal{F}_x)} (N_L(h \circ f_x)) = \inf_{(h_1, h_2) \in \mathcal{H}^2} (N_L(h_1 \circ f_x \circ h_2)) \tag{3.12}$$

Si l'infimum de $N_L(\mathrm{h}_1 \circ \mathrm{f} \circ \mathrm{h}_2)$ est atteint pour chaque transformation f par un couple $(\mathrm{h}_1, \mathrm{h}_2) \in \mathcal{H}^2$, alors la « semi-norme » est séparable et est donc une « norme » (voir preuve ci-dessous). Cette propriété est toujours vraie si le groupe d'isotropie $\mathcal{H}$ est compact mais n'est pas automatiquement vérifié sinon. Par exemple, il n'y a pas de distance invariante induite sur les points par les similitudes ou les transformation affines. En particulier, si le groupe d'isotropie H est composé d'un nombre fini d'éléments discrets, alors il y une distance invariante sur la variété. C'est le cas des repères semi et non orientés décrits à la section (7.5) où le groupe d'isotropie est composé de 2 et 4 transformations.

En supposant qu'il existe une norme possédant les propriétés requises, la distance associée est automatiquement invariante et satisfait :

$$\mathrm{dist}(\mathrm{x},\mathrm{y}) = \inf_{\{\mathrm{f_x}\in\mathcal{F}_\mathrm{x}\,;\,\mathrm{f_y}\in\mathcal{F}_\mathrm{y}\}} (\,\mathrm{dist}_L(\mathrm{f_x},\mathrm{f_y})) = \inf_{\{(\mathrm{h}_1,\mathrm{h}_2)\in\mathcal{H}^2\}} \left(N_L(\mathrm{h}_1 \circ {\mathrm{f_x}}^{(-1)} \circ \mathrm{f_y} \circ \mathrm{h}_2)\right) \tag{3.13}$$

Preuve : Soit N_L une norme sur le groupe $\mathcal{G}$ et N la semi norme de l'équation (3.12) induite sur la variété $\mathcal{M}$. La positivité de N provient directement de la positivité de N_L, et la symétrie découle de celle de N_L et de la symétrie de la définition :

$$N({\mathrm{f_x}}^{(-1)} \star \mathrm{o}) = \inf_{(\mathrm{h}_1,\mathrm{h}_2)\in\mathcal{H}^2} \left(N_L(\mathrm{h}_1 \circ {\mathrm{f_x}}^{(-1)} \circ \mathrm{h}_2)\right) = \inf_{(\mathrm{h}'_1,\mathrm{h}'_2)\in\mathcal{H}^2} \left(N_L(\mathrm{h}'_1 \circ {\mathrm{f_x}}^{(-1)} \circ \mathrm{h}'_2)\right) = N(\mathrm{x})$$

L'inégalité triangulaire est préservée par l'infimum :

$$\begin{aligned} \inf_{(\mathrm{h}_1,\mathrm{h}_2)\in\mathcal{H}^2} \left(N_L(\mathrm{h}_1^{(-1)} \circ {\mathrm{f_x}}^{(-1)} \circ \mathrm{f_y} \circ \mathrm{h}_2)\right) &\leq \inf_{\mathrm{h}_1\in\mathcal{H}} \{N_L(\mathrm{f_x} \circ \mathrm{h}_1)\} + \inf_{\mathrm{h}_1\in\mathcal{H}} \{N_L(\mathrm{f_y} \circ \mathrm{h}_2)\} \\ &\leq \inf_{\mathrm{h}_1,\mathrm{h}_2} (N_L(\mathrm{h}_1 \circ \mathrm{f_x} \circ \mathrm{h}_2)) + \inf_{\mathrm{h}_1,\mathrm{h}_2} (N_L(\mathrm{h}_1 \circ \mathrm{f_y} \circ \mathrm{h}_2)) \end{aligned}$$

et on obtient au final :

$$N(\mathrm{x}) + N(\mathrm{y}) \geq N({\mathrm{f_x}}^{(-1)} \star y)$$

C'est le caractère défini de la distance qui pose problème, sauf si l'infimum de $N_L(\mathrm{h}_1 \circ \mathrm{f} \circ \mathrm{h}_2))$ est atteint pour toute transformation f par un couple $(\mathrm{h}_1, \mathrm{h}_2) \in \mathcal{H}^2$:

$$\begin{aligned} N(\mathrm{x}) = 0 &\Leftrightarrow \exists(\mathrm{h}_1,\mathrm{h}_2) \in \mathcal{H}^2 \,/\, N_L(\mathrm{h}_1 \circ \mathrm{f_x} \circ \mathrm{h}_2) = 0 \\ &\Leftrightarrow \exists(\mathrm{h}_1,\mathrm{h}_2) \in \mathcal{H}^2 \,/\, \mathrm{f_x} = \mathrm{h}_1^{(-1)} \circ \mathrm{h}_2 \\ &\Leftrightarrow \mathrm{f_x} \in \mathcal{H} \Leftrightarrow \mathcal{F}_\mathrm{x} = \mathcal{H} \\ &\Leftrightarrow \mathrm{x} = \mathrm{o} \end{aligned}$$

■

3.4.5 Distance invariante sur les transformation rigides 3D

La distance Euclidienne sur $\mathbb{R}^3$ est induite par la norme L_2 : $d_t(x,y) = \|x - y\|$. Par ailleurs, on montre dans (Altmann, 1986) ou à la section (7.1) que l'angle θ de la rotation 3D est une « norme » qui induit une distance invariante à gauche et à droite sur $\mathcal{SO}_3$. En utilisant le vecteur rotation comme représentation, on a $d_\theta(\,Id, r) = \|r\| = \theta$ et donc $d_\theta(r_1, r_2) = \|r_2^{(-1)} \circ r_1\| = \|r_1 \circ r_2^{(-1)}\|$ (le dernier terme provenant de l'égalité avec l'invariance à droite).

On définit donc la « norme » sur le groupe des transformations rigides (voir section 7.4.1.1) comme :

$$N_\lambda(\mathrm{f}) = N_\lambda((r,t)) = \|f\| = \sqrt{\lambda^2\,\|r\|^2 + \|t\|^2}$$

où λ est un paramètre fixé qui permet de régler l'importance du trièdre (partie rotation) par rapport à la position (partie translation). En effet, l'angle de la rotation est en radians (ou en degrés...) et la translation en millimètres, kilomètres ou inches... On pondère habituellement les deux termes par l'inverse de leur domaine de variation (π pou θ et le diamètre l_0 de l'image ou de l'objet d'intérêt pour la translation : $\lambda = l_0/\pi$). Quand on a une information sur le niveau de bruit des mesures (i.e. des écart-types σ_θ et σ_t), on peut aussi utiliser $\lambda = \sigma_t/\sigma_\theta$.

La distance invariante à gauche est donc :

$$\mathrm{dist}_L(\mathrm{f}_1,\mathrm{f}_2)^2 = \|f_2^{(-1)} \circ f_1\|^2 = \lambda^2 \|r_2^{(-1)} \circ r_1\|^2 + \|t_1 - t_2\|^2$$

tandis que la distance invariante à droite est :

$$\mathrm{dist}_R(\mathrm{f}_1,\mathrm{f}_2)^2 = \|f_1 \circ f_2^{(-1)}\|^2 = \lambda^2 \|r_1 \circ r_2^{(-1)}\|^2 + \|t_1 - (r_1 \circ r_2^{(-1)}) \star t_2\|^2$$

Bien que le groupe des transformations rigides soit uni-modulaire (les mesures invariantes à droite et à gauche sont donc identiques), les distance invariantes à droite et à gauche présentées sont visiblement différentes. Cela reflète le fait que $N_\lambda(\mathrm{f}_1 \circ \mathrm{f}_2) \neq N_\lambda(\mathrm{f}_2 \circ \mathrm{f}_1)$.

3.4.6 Discussion sur les distances invariantes

A partir d'une distance invariante (à gauche) sur le groupe de transformation, on peut donc déterminer une distance invariante induite sur la variété. Toutefois, si nous avons une condition suffisante pour l'existence (l'infimum de $N_L(\mathrm{h}_1 \circ \mathrm{f} \circ \mathrm{h}_2)$ doit être atteint quelque soit f), cette condition ne semble pas nécessaire. De plus elle est quelque peu difficile à manipuler.

Le second problème de cette approche est qu'elle nécessite de connaître une distance invariante sur le groupe. Or, si l'on peut caractériser une telle distance, rien dans cette section ne nous aide à la construire.

3.5 Métrique riemannienne

Pour pallier les défauts de l'approche précédente, nous devons prendre en compte la structure de l'espace sur lequel on travaille : groupe de Lie et variété. Les questions métriques sur de tels espaces sont étudiées depuis longtemps dans le cadre de la géométrie riemannienne. Nous résumons dans la première section quelques résultats importants de cette théorie. Pour plus de détail, on pourra consulter les ouvrages qui ont inspiré cette compilation : (Spivak, 1979, chap. 9), (Klingenberg, 1982) et (Carmo, 1992).

Nous appliquerons ensuite ces résultats au cas des groupes de Lie connexes puis des variétés homogènes, ce qui nous donnera les conditions d'existence d'une distance invariante, mais aussi son expression et la représentation exponentielle pour cette métrique. Cette représentation prendra une importance capitale au chapitre suivant pour caractériser l'espérance d'une primitive aléatoire et permettra de définir une matrice de covariance associée.

3.5.1 Espace vectoriel tangent

Nous avons déjà vu que les variétés n'étaient pas en général des espaces vectoriels. Pour pouvoir mesurer la vitesse le long d'une courbe sur une variété, et donc la longueur de cette courbe, il est nécessaire d'introduire l'espace vectoriel tangent à la variété.

Soit γ une courbe différentiable sur $\mathcal{M}$ passant par un point p en $t = 0$. Dans un système de coordonnées locales $(x,\mathcal{U})$, γ est représentée par la courbe $\gamma_x(t) = x(\gamma(t)) \in \mathcal{U}$ et a pour dérivée $\dot{\gamma}_x(0) = v_x \in \mathbb{R}^n$. Si l'on considère maintenant une fonction f sur $\mathcal{M}$ différentiable autour de p et à valeur dans $\mathbb{R}$ (une observable), on peut écrire cette fonction dans la carte x : $f_x(x) = f(\mathrm{x})$. On peut ainsi considérer la dérivée de la fonction composée $f \circ \gamma$ dans la carte x :

$$\partial_\gamma f = \frac{\partial (f \circ \gamma)}{\partial t} = \frac{\partial f_x(x)}{\partial x} . \frac{d\gamma_x(t)}{dt} = \frac{\partial f_x}{\partial x} . v_x$$

La valeur de $\partial_\gamma f$ ne dépend pas de la carte choisie car $f \circ \gamma$ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Elle est appelée **dérivée directionnelle** de f selon γ. On appelle **vecteur tangent** à la courbe γ (au point p) l'application ∂_γ qui fait correspondre à toute observable f sa dérivée directionnelle $\partial_\gamma f$.

L'ensemble des vecteurs tangents en un point $\mathrm{p} \in \mathcal{M}$ constitue un espace vectoriel que l'on appelle **espace vectoriel tangent** et que l'on note $T_{\mathrm{p}}\mathcal{M}$.

Remarquons encore que la dérivée directionnelle $\partial_\gamma f$ ne dépend pas de la courbe γ choisie mais uniquement du vecteur v_x (si l'on exprime tout dans la carte x) : celle-ci est égale à $\partial_\delta f$ du moment que δ est une courbe différentiable telle que $\delta(0) = \mathrm{p}$ et $\dot{\delta}_x(0) = v_x$. Il est donc raisonnable de représenter le vecteur tangent $\partial_v = \partial_\gamma = \partial_\delta \in T_{\mathrm{p}}\mathcal{M}$ par son expression $v_x = \dot{\gamma}_x = \dot{\delta}_x$ dans la carte locale. Plus précisément, la carte x induit en tout point $\mathrm{p} \in \mathcal{M}$ une base de l'espace vectoriel tangent $T_{\mathrm{p}}\mathcal{M}$

$$\left.\frac{\partial}{\partial x}\right|_{x(\mathrm{p})} = \left[\left.\frac{\partial}{\partial x_1}\right|_{x(\mathrm{p})}, \quad \cdots \quad \left.\frac{\partial}{\partial x_n}\right|_{x(\mathrm{p})}\right]$$

dans laquelle on peut exprimer le vecteur tangent ∂_v :

$$\partial_v = \frac{\partial}{\partial x}.v_x = \sum_{i=1}^{n} v_{x_i} \frac{\partial}{\partial x_i}$$

Si l'on change maintenant de carte et que l'on utilise le système de coordonnées locales $y = \varphi(x)$, la courbe γ s'exprime par $\gamma_y(t) = \varphi(\gamma_x(t))$ et les dérivées sont reliées par le jacobien du changement de coordonnées : $v_y = \dot{\gamma}_y = \frac{\partial \varphi(x)}{\partial x}.\dot{\gamma}_x$. Les notations sont bien cohérentes puisque l'on a

$$\partial_v = \frac{\partial}{\partial x}.v_x = \frac{\partial}{\partial y}.\frac{\partial y}{\partial x}.v_y$$

3.5.2 Métrique riemannienne

Nous avons maintenant un espace vectoriel tangent $T_{\mathrm{x}}\mathcal{M}$ en tout point de la variété dans lequel nous pouvons mesurer la vitesse instantanée le long d'une courbe (la norme du vecteur vitesse), pourvu que l'on se dote d'un produit scalaire $\langle . \mid . \rangle$. Il faut cependant réaliser que les espaces vectoriels tangents en deux point différents de la variété ne sont pas les mêmes et qu'il est difficile de les comparer. On peut par exemples penser aux différents plans tangents de la sphère $\mathcal{S}_2$ plongée dans $\mathbb{R}^3$: à l'exception des points antipodaux, ces plans sont tous différents.

Une **métrique riemannienne** est une application qui donne à chaque espace vectoriel tangent $T_{\mathrm{x}}\mathcal{M}$ un produit scalaire $\langle . \mid . \rangle_{\mathrm{x}}$, avec cependant une contrainte de continuité : si $\partial_{v_1(\mathrm{x})}$ et $\partial_{v_2(\mathrm{x})}$ sont deux champs de vecteurs différentiables sur la variété, alors $\left\langle \partial_{v_1(\mathrm{x})} \mid \partial_{v_2(\mathrm{x})} \right\rangle_{\mathrm{x}}$ doit être une fonction différentiable de x. On peut exprimer ce produit scalaire dans une carte locale x en remarquant que la i$^{\mathrm{e}}$ coordonnée x_i est une observable dont on peut calculer le vecteur tangent ∂_{x_i}. Son produit scalaire avec la dérivée de la j$^{\mathrm{e}}$ coordonnée $\left\langle \partial_{x_i} \mid \partial_{x_j} \right\rangle_{\mathrm{x}} = q_{ij}(x)$ est par définition une application différentiable sur le domaine de définition de la carte. En rassemblant ces fonctions dans une matrice $Q(x) = [q_{ij}(x)]$ et en notant que l'expression d'un vecteur tangent dans cette carte est $\partial_v = \frac{\partial}{\partial x}.v_x = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}.v_{x_i}$, on obtient :

$$\langle \partial_v \mid \partial_w \rangle_{\mathrm{x}} = \sum_{i,j=1}^{n} v_{x_i}.q_{ij}(x).w_{x_j} = v_x^{\mathrm{T}}.Q(x).w_x \tag{3.14}$$

La matrice symétrique définie positive $Q(x)$ est appelée **représentation locale de la métrique riemannienne** dans la carte x. Si l'on change de carte pour le système de coordonnées locales y, on peut relier les représentations locales de la métrique (aux points de chevauchement) par

$$Q(y) = \frac{\partial y}{\partial x}^{\mathrm{T}}.Q(x).\frac{\partial y}{\partial x}$$

Pour finir, notons qu'il existe toujours une métrique riemannienne sur une variété puisque l'on peut plonger cette variété dans un espace $\mathbb{R}^k$ (avec $k \leq 2n+1$) et ainsi doter les espaces tangents des produits scalaires induits par le produit scalaire canonique de l'espace ambiant $\mathbb{R}^k$.

3.5.2.1 Forme volume ou mesure riemannienne

Dans une espace vectoriel de base $\mathcal{A} = (a_1, \dots a_n)$, le volume du parallélépipède formé par les vecteurs de base est $\mathcal{V} = |A| = |\det(A)|$ si l'on note $A = [a_1, \dots a_n]$ la matrice de passage de la base $\mathcal{A}$ à la base orthonormée canonique. Pour calculer le volume d'un objet décrit dans la base $\mathcal{A}$, nous devons donc utiliser la mesure $|A|.dy$ pour obtenir le même volume que dans la base canonique. On peut relier « volume unitaire infinitésimal » à la métrique de notre espace vectoriel de la façon suivante : le produit scalaire de deux vecteurs y_1 et y_2 dans la base $\mathcal{A}$ est donné par

$$\langle y_1 \mid y_2 \rangle_{\mathcal{A}} = y_1^{\mathrm{T}}.(A^{\mathrm{T}}.A).y_2$$

et a donc pour matrice symétrique définie positive associée : $Q = A^{\mathrm{T}}.A$ (cette matrice est indépendante de la position puisque l'on est dans un espace vectoriel). La forme volume (que l'on appelle également mesure) peut donc s'écrire : $\sqrt{|Q|}.dy$.

Si l'on se place maintenant dans une variété riemannienne, cet élément de volume infinitésimal peut être défini dans chaque espace vectoriel tangent et est de plus une fonction différentiable de la position sur la variété. On obtient donc une mesure sur $\mathcal{M}$ qui s'exprime dans une carte locale grâce à la représentation locale de la métrique :

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|Q(x)|}.dx \tag{3.15}$$

3.5.2.2 Longueur d'une courbe

Pour simplifier les notations, on omettra l'indice x indiquant la carte utilisée pour l'expression d'une courbe γ ou d'un vecteur tangent v s'il n'y a pas de doute possible.

Soit $\gamma : [a,b] \in \mathbb{R} \mapsto \mathcal{M}$ une courbe différentiable par morceau. On note $\partial_\gamma(t) \in T_{\gamma(t)}\mathcal{M}$ son vecteur tangent ou vecteur vitesse et $\dot{\gamma}$ son expression dans la carte locale. Avec la métrique riemannienne, on peut mesurer la vitesse instantanée le long de cette courbe par la norme du vecteur tangent et calculer la longueur de la courbe en intégrant cette vitesse :

$$\mathcal{L}_a^b(\gamma) = \int_a^b \|\partial_\gamma\| .dt = \int_a^b \left(\langle \partial_{\gamma(t)} \mid \partial_{\gamma(t)} \rangle_{\gamma(t)} \right)^{\frac{1}{2}} .dt \tag{3.16}$$

Dans la carte locale, cela s'exprime par :

$$\mathcal{L}_a^b(\gamma) = \int_a^b (\dot{\gamma}(t)^{\mathrm{T}}.Q(\gamma(t)).\dot{\gamma}(t))^{\frac{1}{2}} .dt$$

L'abscisse curviligne se définit comme $\sigma_\gamma(t) = \mathcal{L}_a^t(\gamma)$, et l'on dit qu'une courbe est paramétrée par l'abscisse curviligne si $\sigma_\gamma(t) = t$.

3.5.2.3 Distance riemannienne sur une variété connexe

Par définition, la connexité de la variété implique que l'on puisse joindre deux points quelconques de celle-ci par une courbe[7]. On définit alors la distance entre deux points x et y de $\mathcal{M}$ comme la

7. La variété étant localement euclidienne, elle est localement connexe par arcs. Si de plus elle est connexe, alors on peut joindre un nombre fini d'arcs définis localement pour relier deux points : la variété est donc connexe par arc.

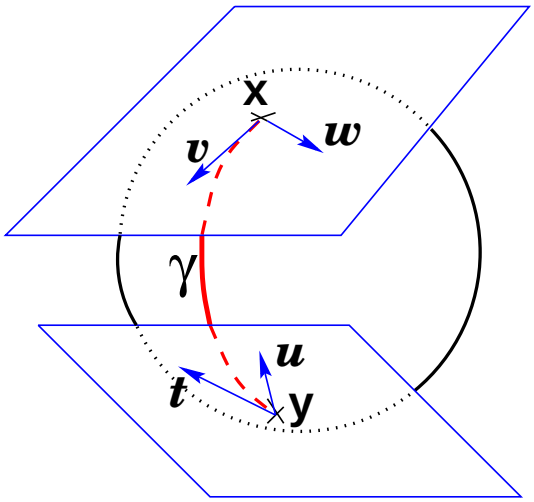


FIG. 3.5 – *Les plans tangents aux points* x *et* y *sont différents : les vecteurs* v *et* w *de* $T_x\mathcal{M}$ *ne sont pas comparables aux vecteurs* t *et* u *de* $T_y\mathcal{M}$ *et il est donc naturel que le produit scalaire soit défini sur chaque plan tangent. Avec cette collection de produits scalaires, on peut calculer la norme du vecteur tangent à une courbe* γ *(le vecteur vitesse instantané) et l'intégration de ce vecteur le long de la courbe donne la longueur de celle-ci.*

longueur du plus court chemin entre ces deux points :

$$\text{dist}(x,y) = \inf\{\mathcal{L}(\gamma) \,/\, \gamma \text{ est une courbe lisse par morceau de x à y}\} \tag{3.17}$$

3.5.2.4 Géodésiques

Trouver une expression explicite pour la distance est donc un problème lié à la détermination des courbes qui minimisent la distance entre deux points. Le calcul des variations s'intéresse de manière plus générale à l'optimisation, le long d'une courbe γ, d'un **lagrangien** $F(\gamma, \partial_\gamma)$. On cherche donc la courbe qui optimise la fonctionnelle $\mathcal{J}(\gamma) = \int_a^b F(\gamma, \partial_\gamma).dt$.

Dans notre cas, nous voulons minimiser la **fonctionnelle longueur** $\mathcal{L}(\gamma) = \int_a^b \|\partial_\gamma\| .dt$, mais la **fonctionnelle énergie** est en fait plus facile à utiliser :

$$\mathcal{E}(\gamma) = \frac{1}{2}\int_a^b \|\partial_\gamma\|^2 .dt$$

Ce sont les points critiques[8] de cette fonctionnelle que l'on appelle **géodésiques**. En fait, on peut montrer qu'une géodésique γ est aussi un point critique de la fonctionnelle longueur $\mathcal{L}(\gamma)$ et est de plus paramètrée par l'abscisse curviligne. Réciproquement, si γ est un point critique de $\mathcal{L}$ (dont le vecteur tangent ne s'annule pas), alors le paramétrage en abscisse curviligne $\gamma(\sigma(t))$ est une géodésique.

On peut donc se focaliser sur les géodésiques. En utilisant une carte locale, on peut écrire le lagrangien d'énergie :

$$F(x, \partial_v) = \frac{1}{2}\langle \partial_v \mid \partial_v \rangle_x = \frac{1}{2} v^T.Q(x).v = \frac{1}{2}\sum_{i,j} q_{ij}(x).v_i.v_j$$

8. Les points critiques d'une fonction sont les points pour lesquelles la différentielle s'annule. C'est la formalisation mathématique de la phrase : « les optimums sont caractérisés par une dérivée nulle ».

où x est un point de $\mathcal{M}$, ∂_v un vecteur de $T_{\mathrm{x}}\mathcal{M}$ et x, v leur expression dans la carte locale. On introduit alors les symboles de Christoffel (on note $q^{ij} = [Q^{(-1)}]_{ij}$ les éléments de l'inverse de la représentation locale de la métrique Q) :

$$\Gamma^i_{j,k} = \frac{1}{2}\sum_{m=1}^{n} q^{im}\left(\frac{\partial q_{mj}}{\partial x_k} + \frac{\partial q_{mk}}{\partial x_j} - \frac{\partial q_{jk}}{\partial x_m}\right) \tag{3.18}$$

Le calcul des variations montre que les géodésiques (les points critiques de l'énergie $\mathcal{E}$) sont les courbes qui satisfont, *dans la carte* $x = (x_1, \dots x_n)$, le système suivant d'équations différentielles du second ordre :

$$\frac{d^2\gamma_i}{dt^2} + \sum_{j,k=1}^{n} \Gamma^i_{j,k}.\frac{d\gamma_j}{dt}.\frac{d\gamma_k}{dt} = 0 \tag{3.19}$$

Supposons maintenant que nous ayons déterminé les géodésiques dans un ensemble de cartes couvrant la variété (un atlas). On a donc un ensemble de courbes $\gamma : [a, b] \mapsto \mathcal{M}$. La variété est dite **géodésiquement complète** si toute géodésique peut être étendue à un domaine de définition couvrant $\mathbb{R}$ tout entier : $\gamma : \mathbb{R} \mapsto \mathcal{M}$. Une conséquence importante de la complétude géodésique est donnée par le théorème de Hopf-Rinow-De Rham qui assure alors que la variété $\mathcal{M}$ est complète pour la distance induite, définie à l'équation (3.17), et qu'il existe toujours au moins une géodésique de longueur minimale entre deux points de la variété (dont la longueur est alors la distance entre ces points).

D'un point de vue pratique, les géodésiques sont déterminées dans une carte x et on obtient les courbes $\gamma_{(x,v)}$. Si on a une expression explicite de la variété $\mathcal{M}$ et donc de $\mathrm{x}(x)$, comme dans la plupart des cas pour nous, on peut alors écrire l'équation des géodésiques directement sur la variété : $\gamma(\mathrm{x}, \partial_v) = \mathrm{x}(\gamma_{(x,v)})$. On peut alors vérifier simplement s'il y a ou non des singularités sur ces géodésiques qui nous empêchent d'étendre leur domaine de définition à $\mathbb{R}$. Notons que cette vérification doit se faire avec l'équation de la géodésique sur la variété et non pas dans la carte. Dans le cas où l'on a une définition implicite de la variété, la fonction $\mathrm{x}(x)$, est elle aussi implicite, et on doit vérifier l'extension des géodésiques dans les cartes, en cherchant à raccorder les morceaux dans différentes cartes. Pour simplifier la suite, nous supposerons dorénavant que la variété est géodésiquement complète.

3.5.2.5 Application exponentielle

Pour finir notre résumé de géométrie différentielle, nous introduisons une fonction importante qui réalise localement un difféomorphisme entre l'espace tangent en un point de la variété et un voisinage de ce point : la fonction exponentielle. Dans le cas d'une variété géodésiquement complète, cette fonction permet de couvrir presque toute la variété et nous fournit une carte très adaptée au traitement informatique.

La théorie des équation différentielles du second ordre nous apprend qu'en tout point $\mathrm{x} \in \mathcal{M}$, il existe une et une seule géodésique $\gamma_{(\mathrm{x},\partial_v)}$ ayant le vecteur tangent $\partial_v \in T_{\mathrm{x}}\mathcal{M}$. Celle-ci est théoriquement définie dans un intervalle suffisamment petit, mais puisque l'on travaille avec une variété géodésiquement complète, elle peut être étendue à $\mathbb{R}$.

La fonction exponentielle au point x associe à tout vecteur ∂_v de l'espace tangent $T_{\mathrm{x}}\mathcal{M}$ le point de la variété atteint au bout d'un temps 1 par l'unique géodésique démarrant de x avec ce vecteur tangent.

$$\begin{array}{rccc} \exp_{\mathrm{x}} & : & T_{\mathrm{x}}\mathcal{M} & \longrightarrow & \mathcal{M} \\ & & \partial_v & \longmapsto & \exp_{\mathrm{x}}(\partial_v) = \gamma_{(\mathrm{x},\partial_v)}(1) \end{array} \tag{3.20}$$

On obtient donc une description très simple de la géodésique $\gamma_{(\mathrm{x},\partial_v)}$ par l'intermédiaire de cette fonction : $\gamma_{(\mathrm{x},\partial_v)}(t) = \exp_{\mathrm{x}}(t.\partial_v)$. La fonction exponentielle peut donc être vue comme le développement de la variété dans l'espace tangent *le long des géodésiques.*

Une propriété très intéressante de la fonction exponentielle est qu'elle réalise localement un difféomorphisme d'un voisinage (suffisamment petit) de $0 \in T_{\mathrm{x}}\mathcal{M}$ dans un voisinage $\mathcal{U}_{\mathrm{x}}$ du point $\mathrm{x} \in \mathcal{M}$. On note $\log_{\mathrm{x}}$ la fonction inverse. Nous avons donc obtenue une carte de la variété centrée en x que l'on appellera carte exponentielle en x et où les géodésiques passant par x sont les droites passant par l'origine :

$$\log_{\mathrm{x}}(\gamma_{(\mathrm{x},\partial_v)}) = t.\partial_v$$

Dans un système de coordonnées locales x, le vecteur tangent ∂_v est représenté par v dans la base $\frac{\partial}{\partial x}$. On peut donc utiliser simplement l'espace $\mathbb{R}^n$ muni de la base induite par la carte locale pour représenter l'espace vectoriel tangent et noter la carte exponentielle dans cette base :

$$\exp_x \quad : \quad \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathcal{M} \\ v & \longmapsto & \exp_x(v) = \exp_{\mathrm{x}}\left(\frac{\partial}{\partial x}.v\right) = \gamma_{(\mathrm{x},\partial_v)}(1) \end{array}$$

Notons que l'exponentielle est à valeur dans la variété et pas dans la carte locale. Dans un voisinage de l'origine $0 \in \mathbb{R}^n$, c'est-à-dire pour des vecteurs vitesse initiaux suffisamment faibles, cette application est un difféomorphisme et l'on notera $\overrightarrow{\mathrm{xy}} = \log_x(\mathrm{y})$ l'inverse.

Pour une gestion informatique simple de la variété, il nous reste maintenant à déterminer un domaine de définition ayant une extension maximale.

3.5.2.6 Lieu de coupure et domaine maximal de la carte exponentielle

Puisque l'on considère une variété géodésiquement complète, il existe toujours au moins une géodésique minimisante joignant deux points x et y de la variété. Cette géodésique part de x avec un vecteur ∂_v dont la norme est la distance entre les points :

$$\forall(\mathrm{x},\mathrm{y}) \in \mathcal{M} \quad : \quad \exists \partial_v \in T_{\mathrm{x}}\mathcal{M} \quad / \quad \exp_{\mathrm{x}}(\partial_v) = \mathrm{y} \qquad \text{et} \qquad \mathrm{dist}(\mathrm{x},\mathrm{y}) = \|\partial_v\| = (\langle \partial_v \mid \partial_v \rangle_{\mathrm{x}})^{1/2}$$

Si l'on normalise cette géodésique, (on considère maintenant que $\|\partial_v\| = 1$), il existe un temps $t_0 \in \mathbb{R} \cup \{+\infty\}$ tel que la géodésique $\gamma_{(\mathrm{x},\partial_v)}(t) = \exp_{\mathrm{x}}(t.\partial_v)$ soit minimisante pour $t \in [0, t_0]$ et ne le soit plus après pour $t \in (t_0, +\infty)$. Si ce temps est fini, on appelle **point de coupure** de la géodésique le point $\gamma_{(\mathrm{x},\partial_v)}(t_0)$, et le vecteur correspondant $t_0.\partial_v$ est le **point de coupure tangentiel**.

L'ensemble des points de coupures de toutes les géodésiques partant de x est le **lieu de coupure** (ou cut locus) $C(\mathrm{x})$ de ce point et l'ensemble des vecteurs correspondants est le **lieu de coupure tangentiel** $\mathcal{C}(\mathrm{x})$. On a donc : $C(\mathrm{x}) = \exp_{\mathrm{x}}(\mathcal{C}(\mathrm{x}))$, et le domaine maximal sur lequel la carte exponentielle $\exp_{\mathrm{x}}$ est bijective est donc l'ouvert qui est à « l'intérieur » du lieu de coupure tangentiel :

$$\mathcal{D}(\mathrm{x}) = \{t.\partial_v \quad \text{avec} \quad t \in [0, t_o) \quad \text{où} \quad t_0.\partial_v \in \mathcal{C}(\mathrm{x})\}$$

Avec cette définition, il est aisé de voir que ce domaine $\mathcal{D}(\mathrm{x})$ est connexe et étoilé par rapport à l'origine de l'espace tangent $T_{\mathrm{x}}\mathcal{M}$. Cependant, la propriété sans doute la plus importante est que l'image de ce domaine par l'application exponentielle couvre toute la variété $\mathcal{M}$ à l'exception du lieu de coupure: $\exp_{\mathrm{x}}(\mathcal{D}(\mathrm{x})) = \mathcal{M} - C(\mathrm{x})$, et le segment de droite de 0 à $\partial_v \in \mathcal{D}(\mathrm{x})$ est transformé en l'unique géodésique joignant x à $\mathrm{y} = \exp_{\mathrm{x}}(\partial_v)$. De plus, la frontière de $\mathcal{C}(\mathrm{x}) = \partial\mathcal{D}(\mathrm{x})$ est continue et est constituée d'un ensemble dense de points où plusieurs géodésiques minimisantes se rencontrent.

En utilisant un système de coordonnées locales pour donner une base à l'espace vectoriel tangent, on a donc obtenu une carte $(\exp_x, \mathcal{D}(\mathrm{x}))$ centrée en x (c'est-à-dire telle que $\exp_x(0) = \mathrm{x}$), dont le

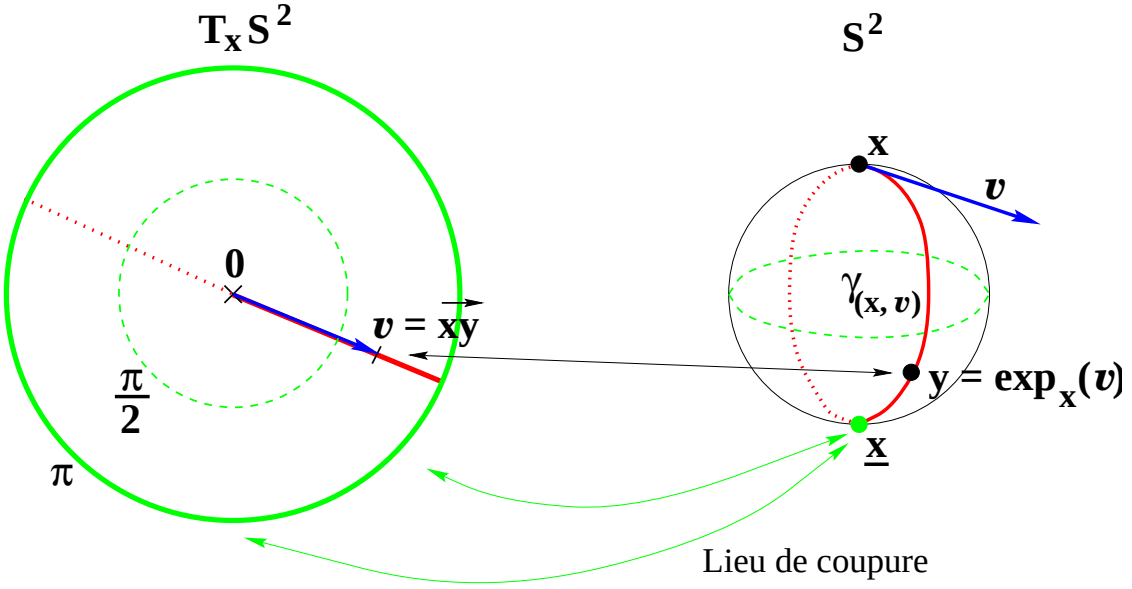


FIG. 3.6 – *Carte exponentielle et lieu de coupure sur la sphère $\mathcal{S}_2$. La géodésique $\gamma_{(x,v)}(t)$ part du point* x *avec le vecteur tangent* v*. Cette géodésique est un grand cercle sur la sphère et une droite passant par* 0 *dans la carte exponentielle. Le lieu de coupure de* x *est constitué de son point antipodal* $\underline{x}$*: on peut y arriver depuis* x *par n'importe quel arc de grand cercle de longueur* π*. Le lieux de coupure tangentiel est donc le cercle de rayon* π *sur lequel tous les points sont identifiés à* $\underline{x}$*. Si l'on considère maintenant que la sphère $\mathcal{S}_2$ est une représentation de l'espace projectif $\mathcal{P}_2$, le lieux de coupure de* $(x, \underline{x})$ *est l'équateur de ces deux points sur la sphère (en pointillés), et le lieux de coupure tangentiel dans la carte exponentielle est le cercle de rayon* $\pi/2$ *sur lequel les points antipodaux sont identifiés.*

domaine est connexe et étoilé, et qui couvre toute la variété modulo l'identification de certains points sur le bord (le lieu de coupure tangentiel) :

$$\begin{array}{ccc} \mathcal{D}(\mathrm{x}) \in \mathbb{R}^n & \longleftrightarrow & \mathcal{M} - C(\mathrm{x}) \\ \overrightarrow{\mathrm{xy}} = \log_x(\mathrm{y}) & \longleftrightarrow & \mathrm{y} = \exp_x(\overrightarrow{\mathrm{xy}}) \end{array}$$

Exemples: Sur la sphère $\mathcal{S}_n$ (de centre 0 et de rayon 1) avec la métrique riemannienne canonique (induite par l'espace euclidien ambiant $\mathbb{R}^{n+1}$), les géodésiques sont les grands cercles, et le lieu de coupure d'un point x est son point antipodal $-\mathrm{x}$. La carte exponentielle est obtenue en faisant rouler la sphère sur l'espace tangent et le domaine de définition maximal est donc la boule ouverte $\mathcal{D} = \mathcal{B}_n(\pi)$. Le bord de ce domaine est la sphère $\partial\mathcal{D} = \mathcal{S}_{n-1}(\pi)$ où tous les points sont identifiés (et représentent le point antipodal).

Si nous considérons l'espace projectif réel $\mathcal{P}_n$ obtenu en identifiant les points antipodaux de la sphère $\mathcal{S}_n$ (que l'on suppose de rayon 1), les géodésiques sont une fois de plus les grands cercles sur cette sphère, mais le lieu de coupure du point $\{\mathrm{x}, -\mathrm{x}\}$ est cette fois-ci l'espace projectif $\mathcal{P}_{n-1}$ obtenu en identifiant les points antipodaux de « l'équateur » de x ou $-\mathrm{x}$. Le domaine de définition maximal est alors la boule ouverte $\mathcal{D} = \mathcal{B}_n(\frac{\pi}{2})$, et le lieu de coupure tangentiel est la sphère $\partial\mathcal{D} = \mathcal{S}_{n-1}(\frac{\pi}{2})$ où les points antipodaux sont identifiés.

3.5.2.7 Symétrie des cartes exponentielles

Soit γ la géodésique joignant le point $\mathrm{y} \in \mathcal{M}$ au point z en un temps unitaire (on suppose que z n'est pas sur le lieu de coupure de y). Si l'on note $\partial_{\overrightarrow{\mathrm{yz}}} = \log_{\mathrm{y}}(\mathrm{z})$ le vecteur tangent en $t = 0$, la géodésique est décrite par $\gamma(t) = \exp_{\mathrm{y}}(t.\partial_{\overrightarrow{\mathrm{yz}}})$. En fait, on peut observer que le vecteur tangent à cette géodésique est constant le long du trajet, et on a en particulier $\partial_\gamma(1) = \partial_\gamma(0) = \partial_{\overrightarrow{\mathrm{yz}}}$.

Considérons maintenant la géodésique $\delta(t) = \gamma(1-t)$, c'est la géodésique qui joint z à y en un temps unitaire et elle a donc pour vecteur tangent (en $t = 0$ et ailleurs) $\partial_{\overrightarrow{\mathrm{zy}}}$. Par ailleurs, on a $\partial_\delta(t) = \frac{d\gamma(1-t)}{dt} = -\partial_\gamma(1-t)$. On obtient donc la formule de symétrie suivante :

$$\partial_{\overrightarrow{\mathrm{yz}}} = \log_{\mathrm{y}}(\mathrm{z}) = \log_{\mathrm{z}}(\mathrm{y}) = -\partial_{\overrightarrow{\mathrm{zy}}} \tag{3.21}$$

Cette équation permettra non seulement de simplifier nombre de formules par la suite, mais aussi de montrer la symétrie de certaines définitions ou, plus prosaïquement, de vérifier la précision numérique de l'implémentation des opérations de base.

3.5.3 Métrique riemannienne invariante sur un groupe de Lie connexe

3.5.3.1 Translation à gauche et à droite

Dans un groupe de Lie, nous avons deux moyens canoniques de comparer les espaces tangents en différents points qui sont donnés par la composition à gauche ou à droite par un élément fixe. Dans le vocabulaire traditionnel des groupes de transformations, ces applications portent de nom de **translations**.

Les translations à gauche et à droite de transformation fixe g sont les applications L_{g} et R_{g} définies par

$$L_{\mathrm{g}}(\mathrm{f}) = \mathrm{g} \circ \mathrm{f} \qquad \text{et} \qquad R_{\mathrm{g}}(\mathrm{f}) = \mathrm{f} \circ \mathrm{g}$$

On peut facilement composer et inverser les translations :

$$(L_{\mathrm{g}})^{(\text{-}1)} = L_{\mathrm{g}^{(\text{-}1)}} \qquad \text{et} \qquad L_{\mathrm{f}_1}(L_{\mathrm{f}_2}(\mathrm{f})) = \mathrm{f}_1 \circ \mathrm{f}_2 \circ \mathrm{f} = L_{\mathrm{f}_1 \circ \mathrm{f}_2}(\mathrm{f})$$

Les formules sont similaires pour la translation à droite.

On peut différencier ces applications pour obtenir les applications linéaires L^*_{g} et R^*_{g} de l'espace tangent $T_{\mathrm{f}}\mathcal{G}$ vers les espaces tangents $T_{(\mathrm{g}\circ\mathrm{f})}\mathcal{G}$ et $T_{(\mathrm{f}\circ\mathrm{g})}\mathcal{G}$. En effet, si l'on considère la transformation f comme un point mobile $\mathrm{f}(t)$ le long d'une courbe, alors le vecteur vitesse $\partial_{\mathrm{f}} = \frac{\mathrm{df}}{dt}$ de l'espace tangent $T_{\mathrm{f}}\mathcal{G}$ est transformé par la translation à gauche en le vecteur vitesse $\partial_{\mathrm{g}\circ\mathrm{f}} = \frac{d(\mathrm{g}\circ\mathrm{f})(t)}{dt} = \frac{\partial(\mathrm{g}\circ\mathrm{f})}{\partial \mathrm{f}}.\partial_{\mathrm{f}}$ de l'espace tangent $T_{(\mathrm{g}\circ\mathrm{f})}\mathcal{G}$.

On peut donc écrire ces différentielles comme des opérateurs linéaires qui sont fonction de f :

$$L^*_g(\mathrm{f}) = \frac{\partial(\mathrm{g} \circ \mathrm{f})}{\partial \mathrm{f}} \qquad \text{et} \qquad R^*_{\mathrm{g}}(\mathrm{f}) = \frac{\partial(\mathrm{f} \circ \mathrm{g})}{\partial \mathrm{f}}$$

3.5.3.2 Métrique riemannienne invariante à gauche et à droite

Les translations à gauche et à droite constituent donc deux moyens canoniques d'identifier les espaces vectoriels tangents en différents points du groupe. Ainsi, si $\partial_{v_f} \in T_{\mathrm{f}}\mathcal{M}$ est un vecteur tangent de $T_{\mathrm{f}}\mathcal{M}$, $\partial_v = (L^*_{\mathrm{f}}(\,\mathrm{Id}))^{(\text{-}1)}.\partial_{v_f}$ est un vecteur tangent à l'origine. En reprenant les notations de la section (3.3) utilisées pour la mesure de Haar, on écrira cette équation dans une carte quelconque :

$$v = J_L(f)^{(\text{-}1)}.v_f \qquad \text{où} \qquad J_L(f) = \left.\frac{\partial(f \circ e)}{\partial e}\right|_{e=Id}$$

Si l'on choisit maintenant une produit scalaire quelconque $\langle . \mid . \rangle$ sur l'espace tangent à l'origine, caractérisé par la matrice symétrique Q dans notre carte, on peut le transporter en tout point du groupe grâce à la translation à gauche (il en serait de même à droite) : si ∂_{x_1} et ∂_{x_2} sont deux vecteurs de l'espace tangent $T_{\mathrm{f}}\mathcal{M}$ en f, leur transport à l'origine est assuré par l'équation ci-dessus et on peut définir le produit scalaire sur $T_{\mathrm{f}}\mathcal{M}$ (exprimé dans notre carte) par

$$\langle x_1 \mid x_2 \rangle_{\mathrm{f}} = \left\langle J_L(f)^{(\text{-}1)}.x_1 \;\middle|\; J_L(f)^{(\text{-}1)}.x_2 \right\rangle$$

ce qui se traduit sur les matrices symétriques correspondantes par

$$Q(f) = J_L(f)^{(\text{-T})}.Q.J_L(f)^{(\text{-}1)} \tag{3.22}$$

3.5.3.3 Mesure riemannienne invariante

L'élément de volume associé à cette métrique riemannienne est

$$d\mathcal{G}(f) = \sqrt{|Q(f)|}.df = \lambda.\frac{df}{|J_L(f)|}$$

ce qui est justement le mesure de Haar invariante à gauche déterminée à la section (3.3), où le facteur d'échelle $\lambda = \sqrt{|Q|}$ s'interprète comme la mesure de la déformation de l'élément de volume unitaire à l'origine de la métrique par rapport à la représentation minimale que l'on utilise.

3.5.3.4 Détermination des géodésiques

Par définition, notre métrique est invariante à gauche et les géodésiques sont donc globalement invariantes par ces translations (i.e. une géodésique reste une géodésique). Plus précisément, si $\gamma_{(\mathrm{f},\partial_v)}$ est une géodésique partant de f avec le vecteur tangent ∂_v, alors $\mathrm{g} \circ \gamma_{(\mathrm{f},\partial_v)} = \gamma_{(\mathrm{g}\circ\mathrm{f},L_{\mathrm{g}}^*(\mathrm{f}).\partial_v)}$ est aussi une géodésique. On peut donc se contenter de déterminer les géodésiques partant de l'identité et calculer ensuite les géodésiques partant de n'importe quel point par translation à gauche.

Au passage, on pourra noter que si le groupe est compact (comme par exemple le groupe des rotations), alors il existe une métrique invariante à droite et à gauche et que les géodésiques de cette métrique partant de l'identité sont les sous-groupes à un paramètre. On peut alors utiliser l'équation suivante pour déterminer les géodésiques partant de l'origine :

$$\forall (s,t) \in \mathbb{R}^2, \qquad \gamma(s+t) = \gamma(s) \circ \gamma(t) = \gamma(t) \circ \gamma(s)$$

Cette équation est souvent plus facile à résoudre que le système d'équations différentielles du second degré mais n'est valide que pour un groupe compact et est généralement fausse pour une groupe seulement localement compact (comme les transformations rigides par exemple).

On suppose à partir de maintenant que l'on a déterminé les géodésiques pour la métrique invariante à gauche, et que le groupe est **géodésiquement complet**.

3.5.3.5 Carte principale

On appelle carte principale la représentation exponentielle à l'identité. Pour simplifier les notations, on notera $\log = \log_{Id}$ et $\exp = \exp_{Id}$ les applications faisant passer du groupe $\mathcal{G}$ à l'espace tangent $T\mathcal{M}$ en l'identité et $\mathcal{D}$ le domaine dans cet espace limité par le lieu de coupure tangentiel.

En utilisant la base de $T\mathcal{M}$ induite par la carte f dans laquelle on a déterminé les géodésiques, et en omettant pour simplifier les indices de référence à l'identité, la carte principale est donnée par :

$$\begin{array}{rcl} \mathcal{D} \in \mathbb{R}^n & \longleftrightarrow & \mathcal{G} - C \\ \vec{\mathrm{f}} = \log(\mathrm{f}) & \longleftrightarrow & \mathrm{f} = \exp(\vec{\mathrm{f}}) = \gamma_{(\,Id,\vec{\mathrm{f}})}(1) \end{array} \tag{3.23}$$

Pour éviter d'avoir à utiliser plusieurs cartes dans l'implémentation informatique, nous étendrons cette correspondance à tout le groupe $\mathcal{G}$ en identifiant au besoin certains points sur le lieu de coupure tangentiel.

Puisque cette carte couvre (presque) toute la variété, on peut y calculer sans problème la composition et l'inversion de transformations :

$$\vec{\mathrm{f}}_1 \circ \vec{\mathrm{f}}_2 \equiv \log(\mathrm{f}_1 \circ \mathrm{f}_2) = \log(\exp(\vec{\mathrm{f}}_1) \circ \exp(\vec{\mathrm{f}}_2))$$

$$\vec{\mathrm{f}}^{(\text{-}1)} \equiv \log(\mathrm{f}^{(\text{-}1)}) = \log(\exp(\vec{\mathrm{f}})^{(\text{-}1)})$$

ainsi que les différentielles de ces fonctions de base. En particulier, on peut calculer l'expression de la métrique dans cette carte :

$$Q(\vec{\mathrm{f}}) = J_L(\vec{\mathrm{f}})^{(\text{-T})}.Q.J_L(\vec{\mathrm{f}})^{(\text{-}1)} \qquad \text{avec} \qquad J_L(\vec{\mathrm{f}}) = \left.\frac{\partial(\vec{\mathrm{f}} \circ \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=\,Id=0}$$

Notons que nous avons obtenu *toutes* les distances riemanniennes invariantes à gauche sur le groupe $\mathcal{G}$, paramétrées par le choix de la matrice symétrique Q. Comme celle-ci est de plus définie positive (ses valeurs propres sont strictement positives), on peut la diagonaliser sous la forme $Q = R^{\mathrm{T}}.\Lambda^2.R$, où la matrice diagonale Λ rassemble les racines carrées des valeurs propres. La matrice carrée $A = \Lambda.R$ est alors appelée racine carrée de Q et permet de se ramener sans perte d'information au cas de la métrique euclidienne canonique : si $\vec{\mathrm{f}}' = A.\vec{\mathrm{f}}$ et que l'on retaille le domaine de définition en conséquence, toutes les propriété de la carte exponentielle sont conservées mais on a maintenant $Q' = Id$.

3.5.3.6 Cartes exponentielles

Soit $\gamma_{(\,Id,\partial_{\vec{\mathrm{f}}})}$ une géodésique. On sait que $\gamma(0) = \mathrm{Id}$ et $\gamma(1) = \mathrm{f}$. Si nous translatons maintenant cette géodésique par g, nous obtenons :

$$\delta = \mathrm{g} \circ \gamma_{(\,\mathrm{Id},\partial_{\vec{\mathrm{f}}})} = \gamma_{(\mathrm{g},L^*_{\mathrm{g}}(\,\mathrm{Id}).\partial_{\vec{\mathrm{f}}})}$$

avec $\delta(0) = \mathrm{g}$ et $\delta(1) = \mathrm{g} \circ \mathrm{f}$. En reprenant la définition de la fonction exponentielle, on obtient donc $\exp_{\mathrm{g}}(L^*_{\mathrm{g}}(\,\mathrm{Id}).\partial_{\vec{\mathrm{f}}}) = \mathrm{g} \circ \exp(\partial_{\vec{\mathrm{f}}})$. En notant $v = J_L(\vec{\mathrm{g}}).\vec{\mathrm{f}}$, ceci s'exprime *dans la base induite par la carte principale*[9], par :

$$\exp_{\vec{\mathrm{g}}}(v) = \mathrm{g} \circ \exp(J_L(\vec{\mathrm{g}})^{(\text{-}1)}.v) \tag{3.24}$$

En prenant maintenant $v = J_L(\vec{\mathrm{g}}).(\vec{\mathrm{g}}^{(\text{-}1)} \circ \vec{\mathrm{f}})$, on obtient :

$$\overrightarrow{\mathrm{gf}} = \log_{\vec{\mathrm{g}}}(\mathrm{f}) = J_L(\vec{\mathrm{g}}).(\vec{\mathrm{g}}^{(\text{-}1)} \circ \vec{\mathrm{f}}) \tag{3.25}$$

9. Il faut faire bien attention que les bases induites sur les espaces tangents par la carte f d'origine et la carte principale $\vec{\mathrm{f}}$ ne se correspondent *a priori* qu'à l'identité. Les expressions obtenues pour les cartes exponentielles en d'autres points sont donc différentes si l'on utilise la base induite par l'une ou l'autre carte.

3.5.3.7 Distance

La distance entre deux transformations f et g est la longueur de la géodésique minimisante entre ces deux points. Par définition de l'application logarithmique, une telle géodésique part de f avec le vecteur tangent $\overrightarrow{\mathrm{fg}} = \log_{\mathrm{f}}(\mathrm{g})$ (ou de manière symétrique à partir du point g), et la distance est la longueur de ce vecteur évalué avec la métrique au point f :

$$\mathrm{dist}(\mathrm{f},\mathrm{g})^2 = \|\log_{\mathrm{f}}(\mathrm{g})\|^2_{\vec{\mathrm{f}}} = \overrightarrow{\mathrm{fg}}^{\mathrm{T}}.Q(\vec{\mathrm{f}}).\overrightarrow{\mathrm{fg}}$$

En combinant avec l'équation de la métrique (3.22) et l'expression (3.25), cela se traduit *dans la carte principale* par :

$$\mathrm{dist}(\mathrm{f},\mathrm{g})^2 = (\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}})^{\mathrm{T}}.Q.(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}}) = \|\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}}\|^2_{\mathrm{Id}} \tag{3.26}$$

La définition de notre « norme » (ou distance à l'identité) N_L de la section (3.4.3) est donc :

$$N_L(\mathrm{f})^2 = \vec{\mathrm{f}}^{\mathrm{T}}.Q.\vec{\mathrm{f}} = \|\vec{\mathrm{f}}\|^2_{\mathrm{Id}} \tag{3.27}$$

3.5.3.8 Identités remarquables

Théorème 3.1 *Dans la carte principale, on a les relations suivantes :*

$$\vec{\mathrm{f}}^{(-1)} = -J_L(\vec{\mathrm{f}})^{(-1)}.\vec{\mathrm{f}} = -J_L(\vec{\mathrm{f}}^{(-1)}).\vec{\mathrm{f}} \tag{3.28}$$

$$\left(\frac{\partial \vec{\mathrm{f}}^{(-1)}}{\partial \vec{\mathrm{f}}}\right)^T.Q.\vec{\mathrm{f}}^{(-1)} = Q.\vec{\mathrm{f}} \tag{3.29}$$

Ces relations s'expriment de manière plus générale par :

$$(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}}) = -J(\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}})^{(-1)}.(\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}}) = -J_L(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}}).(\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}}) \tag{3.30}$$

$$\left(\frac{\partial(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}})}{\partial \vec{\mathrm{f}}}\right)^T.Q.(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}}) = \left(\frac{\partial(\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}})}{\partial \vec{\mathrm{f}}}\right)^T.Q.(\vec{\mathrm{g}}^{(-1)} \circ \vec{\mathrm{f}}) \tag{3.31}$$

Preuve : La géodésique $\gamma_{(\,Id,\vec{\mathrm{f}})} = \exp(t.\vec{\mathrm{f}})$ va de l'identité au point f en un temps unité et s'exprime dans la carte principale par $\gamma_o(t) = t.\vec{\mathrm{f}}$. Considérons la la géodésique inverse: $\delta(t) = \gamma(1-t)$. Elle s'exprime dans la carte principale par $\delta_o(t) = (1-t).\vec{\mathrm{f}}$ et va de $\delta(0) = \mathrm{f}$ à l'identité $\delta(1) = \mathrm{Id}$ et a une longueur unité. Elle a donc un vecteur tangent $\dot{\delta}_o$ constant et égal à $\mathrm{f}\,\overrightarrow{\mathrm{Id}} = J_L(\vec{\mathrm{f}}).\vec{\mathrm{f}}^{(-1)}$ dans la carte principale. Par ailleurs, le calcul direct à partir de l'équation de la géodésique montre que $\dot{\delta}_o = -\vec{\mathrm{f}}$. On obtient donc la première parie de l'identité remarquable (3.28) :

$$J_L(\vec{\mathrm{f}}).\vec{\mathrm{f}}^{(-1)} = -\vec{\mathrm{f}}$$

Observons également que la géodésique $\alpha(t) = \vec{\mathrm{f}}^{(-1)} \circ \gamma(1-t)$ va de l'origine au point $\vec{\mathrm{f}}^{(-1)}$ en un temps unitaire. Elle a donc pour vecteur tangent dans la carte principale $\dot{\alpha} = \vec{\mathrm{f}}^{(-1)}$. Or, le calcul direct donne :

$$\dot{\alpha}(1) = \left.\frac{d(\vec{\mathrm{f}}^{(-1)} \circ ((1-t).\vec{\mathrm{f}}))}{dt}\right|_{t=1} = -\left.\frac{\partial(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=\vec{o}}.\vec{\mathrm{f}} = -J_L(\vec{\mathrm{f}}^{(-1)}).\vec{\mathrm{f}}$$

On obtient donc au final la seconde partie de (3.28)

$$\vec{\mathrm{f}}^{(-1)} = -J_L(\vec{\mathrm{f}})^{(-1)}.\vec{\mathrm{f}} = -J_L(\vec{\mathrm{f}}^{(-1)}).\vec{\mathrm{f}}$$

En remplaçant maintenant la transformation générique $\vec{\mathrm{f}}$ par $(\vec{\mathrm{f}}^{(-1)} \circ \vec{\mathrm{g}})$ (qui est bien l'expression de la transfo $(\mathrm{f}^{(-1)} \circ \mathrm{g})$ dans la carte principale), on obtient la relation (3.30).
Une seconde identité remarquable provient de la distance invariante: $\mathrm{dist}(\mathrm{f},\mathrm{Id}) = \mathrm{dist}(\mathrm{Id},\mathrm{f}^{(-1)})$, ce qui s'exprime dans la carte principale par

$$\|\vec{\mathrm{f}}\|^2_{Id} = \vec{\mathrm{f}}^{\mathrm{T}}.Q.\vec{\mathrm{f}} = \vec{\mathrm{f}}^{(-\mathrm{T})}.Q.\vec{\mathrm{f}}^{(-1)} = \|\vec{\mathrm{f}}^{(-1)}\|^2_{Id}$$

La dérivation de cette égalité donne :

$$\frac{\partial \|\vec{\mathrm{f}}\|^2_{Id}}{\partial \vec{\mathrm{f}}} = 2.\vec{\mathrm{f}}^{\mathrm{T}}.Q = 2.\vec{\mathrm{f}}^{(-\mathrm{T})}.Q.\frac{\partial \vec{\mathrm{f}}^{(-1)}}{\partial \vec{\mathrm{f}}}$$

c'est-à-dire la relation (3.29). Notons qu'en décomposant le jacobien, on peut écrire la différentielle de l'inversion comme la composée de translations à droite et à gauche :

$$\frac{\partial(\vec{\mathrm{f}}^{(-1)})}{\partial \vec{\mathrm{f}}} = -J_L(\vec{\mathrm{f}}^{(-1)}).J_R(\vec{\mathrm{f}})^{(-1)} = -J_R(\vec{\mathrm{f}}^{(-1)}).J_L(\vec{\mathrm{f}})^{(-1)}$$

De la même façon, la dérivation de

$$\mathrm{dist}(\mathrm{f},\mathrm{g}) = \mathrm{dist}(\mathrm{f}^{(-1)} \circ \mathrm{g},\ \mathrm{Id}) = \mathrm{dist}(\ \mathrm{Id}, \mathrm{g}^{(-1)} \circ \mathrm{f})$$

dans la carte principale donne l'identité (3.31).

■

Il est important de noter que les résultats obtenus dans cette section concernant la distance et les identités remarquables ne sont valables que dans la carte principale et sont en général **faux** dans une autre représentation du groupe.

3.5.4 Métrique riemannienne invariante sur une variété homogène connexe

Dans cette section, la représentation utilisée pour le groupe n'a guère d'importance. Nous écrirons donc directement les transformations en temps qu'éléments du groupe.

3.5.4.1 Métrique riemannienne invariante

Pour comparer les espaces tangents à $T\mathcal{M}$ l'origine et $T_{\mathrm{x}}\mathcal{M}$ au point x, nous avons cette fois-ci tout un ensemble de transformations $\mathrm{f_x} \in \mathcal{F}_{\mathrm{x}}$. En choisissant une **fonction de placement** $\mathrm{f_x}$, on peut transformer le vecteur $\partial_v \in T\mathcal{M}$ de l'espace tangent à l'origine en un vecteur $\partial_{v_{\mathrm{x}}} \in T_{\mathrm{x}}\mathcal{M}$ tangent au point x. Dans un système de coordonnées locales, ceci s'exprime par :

$$v_x = J(\mathrm{f}_x).v \qquad \text{avec} \qquad J(\mathrm{f}_x) = \left.\frac{\partial(\mathrm{f}_x \star e)}{\partial e}\right|_{e=o}$$

Cette formule peut être utilisée dans l'autre sens pour définir sur chaque espace tangent $T_{\mathrm{x}}\mathcal{M}$ un produit scalaire par translation du produit scalaire à l'origine :

$$\langle y_1 \mid y_2 \rangle_x = \left\langle J(f_x)^{(-1)}.y_1 \;\middle|\; J(f_x)^{(-1)}.y_2 \right\rangle \qquad \text{avec} \qquad J(f_x) = \left.\frac{\partial(f_x \star e)}{\partial e}\right|_{e=\mathrm{o}}$$

ce qui se traduit sur l'expression locale de la métrique par

$$Q(x) = J(\mathrm{f}_x)^{(-\mathrm{T})}.Q.J(\mathrm{f}_x)^{(-1)} \tag{3.32}$$

Une condition nécessaire pour avoir une métrique invariante est évidemment la stabilité de cette expression par rapport au choix de la fonction de placement $\mathrm{f_x}$, ce qui se réduit à l'invariance de la métrique à l'origine Q par l'action du groupe d'isotropie $\mathcal{H}$:

$$\forall \mathrm{h} \in \mathcal{H} \qquad J(\mathrm{h})^{\mathrm{T}}.Q.J(\mathrm{h}) = Q \tag{3.33}$$

Le produit scalaire $\langle . \mid . \rangle_x$ est dans ce cas une fonction continue de x : notre collection de produits scalaires est donc une métrique invariante et la condition ci-dessus est suffisante.

3.5.4.2 Mesure riemannienne invariante

S'il existe une distance invariante, l'élément de volume infinitésimal associé est donné par :

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|Q(x)|}.dx = \lambda.\frac{dx}{|J(\mathrm{f}_x)|} \qquad \text{avec} \qquad \lambda = \sqrt{|Q|} \qquad \text{et} \qquad \mathrm{f}_x \in \mathcal{F}_x$$

Cette mesure est justement la mesure invariante déterminée à la section (3.3.4). Notons de plus qu'en prenant le déterminant de l'équation d'existence (3.33) on obtient la condition $|J(\mathrm{h})| = 1$, ce qui était la condition d'existence pour une mesure invariante.

Cependant nous avons ici une contrainte plus faible : l'existence d'une distance invariante implique celle d'une mesure invariante, mais il peut exister une mesure invariante sans qu'il n'y ait de distance invariante.

3.5.4.3 Carte principale

Comme dans le cas des groupes de Lie, on suppose que nous avons pu déterminer l'expression des géodésiques à partir d'une carte x et vérifier que la variété est géodésiquement complète. On définit alors la carte principale comme la représentation exponentielle de la variété à l'origine :

$$\begin{array}{rcl} \mathcal{D} \in \mathbb{R}^n & \longleftrightarrow & \mathcal{M} - C \\ \vec{\mathrm{y}} = \log(\mathrm{y}) & \longleftrightarrow & \mathrm{y} = \exp(\vec{\mathrm{y}}) = \gamma_{(o,\vec{\mathrm{y}})}(1) \end{array} \tag{3.34}$$

où $\mathcal{D}$ est le domaine délimité par le lieu de coupure tangentiel $\partial\mathcal{D} = \mathcal{C}(o)$. Pour des raisons d'implémentation pratique, on étendra ici aussi cette correspondance à toute la variété, en identifiant au besoin des points sur le lieu de coupure tangentiel. On peut alors définir l'action d'une transformation dans la carte principale :

$$\mathrm{f} \star \vec{\mathrm{x}} \equiv \log(\mathrm{f} \star \mathrm{x}) = \log(\mathrm{f} \star \exp(\vec{\mathrm{x}}))$$

ainsi que la différentielle de cette fonction. En particulier, on peut calculer l'expression de la métrique dans cette carte :

$$Q(\vec{\mathrm{x}}) = J(\mathrm{f}_{\vec{\mathrm{x}}})^{(-\mathrm{T})}.Q.J(\mathrm{f}_{\vec{\mathrm{x}}})^{(-1)} \qquad \text{avec} \qquad J(\mathrm{f}_{\vec{\mathrm{x}}}) = \left.\frac{\partial(\mathrm{f}_{\vec{\mathrm{x}}} \circ \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=o=0}$$

3.5.4.4 Distance invariante et action du groupe d'isotropie

Comme les géodésiques partant de l'origine (les droites passant par zéro) sont transformées par une transformation h du groupe d'isotropie $\mathcal{H}$ en géodésiques partant de l'origine, l'action d'une telle transformation est linéaire et s'exprime dans la carte principale par :

$$\mathrm{h} \star \vec{\mathrm{x}} = J(\mathrm{h}).\vec{\mathrm{x}}$$

Par ailleurs, l'ensemble de toutes les métriques invariantes sur $\mathcal{M}$ est paramétré par les matrices symétriques définies positives vérifiant l'équation (3.33) :

$$\forall \mathrm{h} \in \mathcal{H} \qquad\qquad J(\mathrm{h})^{\mathrm{T}}.Q.J(\mathrm{h}) = Q$$

où $J(\mathrm{h})$ est ici exprimé dans la carte principale.

On peut diagonaliser une telle matrice Q sous la forme $Q = U^{\mathrm{T}}.\Lambda^2.U$, où la matrice diagonale Λ rassemble les racines carrées des valeurs propres. En notant $A = \Lambda.U$ et en faisant le changement

de repère $\vec{y} = A.\vec{x}$, on obtient une nouvelle carte exponentielle à l'origine (de domaine $\mathcal{D}_y = A.\mathcal{D}_x$) avec la métrique canonique.

La condition d'invariance dans cette nouvelle carte est maintenant $J_y(\mathrm{h})^{\mathrm{T}}.J_y(\mathrm{h}) = Id$, ce qui implique que $J_y(\mathrm{h}) = R_{\mathrm{h}}$ soit une matrice orthogonale (une matrice de rotation ou rotation plus symétrie). Le groupe d'isotropie $\mathcal{H}$ est donc un sous-groupe du groupe orthogonal $\mathcal{O}_n$. Si de plus il est connexe, alors le déterminant ne peut pas passer continûment de $+1$ (le déterminant de l'identité) à -1 et on a dans ce cas $\mathcal{H} \subset \mathcal{SO}_n$ car $Id \in \mathcal{H}$.

Théorème 3.2 (Orthogonalité du groupe d'isotropie)
S'il existe une métrique invariante sur $\mathcal{M}$, alors l'action du groupe d'isotropie $\mathcal{H}$ exprimée dans la carte principale rectifiée $\vec{y}$ (avec $\vec{y} = A.\vec{x}$ où A est une racine carrée de $Q = A^T.A$) est orthogonale :

$$\forall \mathrm{h} \in \mathcal{H}, \quad \exists R_{\mathrm{h}} \in \mathcal{O}_n \qquad / \qquad \forall \mathrm{y} \in \mathcal{M} \quad : \quad \mathrm{h} \star \vec{y} = R_{\mathrm{h}}.\vec{y}$$

Si $\mathcal{H}$ est de plus connexe, alors l'action est une pure rotation n-D.

3.5.4.5 Cartes exponentielles

Avec l'invariance des géodésiques par l'action des transformations, et en choisissant une fonction de placement $\mathrm{f_x} \in \mathcal{F}_{\mathrm{x}}$, on peut facilement exprimer la carte exponentielle au point x. Si $\overrightarrow{\mathrm{xy}}$ est un vecteur tangent de $T_{\mathrm{x}}\mathcal{M}$ exprimé dans la base induite par la carte principale, on a :

$$\exp_{\vec{\mathrm{x}}}(\overrightarrow{\mathrm{xy}}) = \mathrm{f}_{\vec{\mathrm{x}}} \star \exp(J(f_{\vec{\mathrm{x}}})^{(-1)}.\vec{\mathrm{y}})$$

On obtient alors la fonction inverse par

$$\overrightarrow{\mathrm{xy}} = \log_{\vec{\mathrm{x}}}(\mathrm{y}) = J(\mathrm{f}_{\vec{\mathrm{x}}}).(\mathrm{f}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) \tag{3.35}$$

Ces fonctions sont indépendantes de la fonction de placement $\mathrm{f_x}$ choisie. En effet, si $\mathrm{f'_x}$ est une autre fonction de placement, alors il existe une pour chaque primitive x une transformation $\mathrm{h_x}$ telle que : $\mathrm{f'_x} = \mathrm{f_x} \circ \mathrm{h_x}$ (car $\mathrm{f'_x}$ et $\mathrm{f_x}$ sont deux éléments de $\mathcal{F}_{\mathrm{x}}$). On a donc : $J(\mathrm{f'}_{\vec{\mathrm{x}}}) = J(\mathrm{f}_{\vec{\mathrm{x}}}).J(\mathrm{h}_{\vec{\mathrm{x}}})$. Comme l'action du groupe d'isotropie sur la carte principale est linéaire, on a $\mathrm{f}_{\vec{\mathrm{x}}}^{\prime(-1)} \star \vec{\mathrm{y}} = J(\mathrm{h}_{\vec{\mathrm{x}}})^{(-1)}.(\mathrm{f_x} \star \vec{\mathrm{y}})$ et donc :

$$\overrightarrow{\mathrm{xy}} = \log_{\vec{\mathrm{x}}}(\mathrm{y}) = J(\mathrm{f}_{\vec{\mathrm{x}}}).(\mathrm{f}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) = J(\mathrm{f'}_{\vec{\mathrm{x}}}).(\mathrm{f'}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})$$

L'invariance de l'exponentielle est similaire.

3.5.4.6 Distance

Exactement comme pour le groupe de Lie, la distance s'exprime dans la carte principale par

$$\mathrm{dist}(\mathrm{x},\mathrm{y})^2 = \| \log_{\vec{\mathrm{x}}}(\mathrm{y}) \|^2_{\vec{\mathrm{x}}} = \overrightarrow{\mathrm{xy}}^{\mathrm{T}}.Q(\vec{\mathrm{x}}).\overrightarrow{\mathrm{xy}} = (\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})^{\mathrm{T}}.Q.(\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) \tag{3.36}$$

La définition de notre norme N, ou distance à l'origine, de la section (3.4.2) est donc *dans la carte principale* :

$$N(\mathrm{x})^2 = \vec{\mathrm{x}}^{\mathrm{T}}.Q.\vec{\mathrm{x}} = \|\vec{\mathrm{x}}\|^2_o \tag{3.37}$$

L'ensemble des distances invariantes sur la variété est cette fois-ci paramétré par les matrices symétriques définies positives Q *vérifiant l'équation* (3.33), et comme pour les transformations on peut se ramener sans perte d'information au cas de la métrique euclidienne canonique $Q' = Id$ par le changement de carte $\vec{y} = A.\vec{x}$

3.5.4.7 Identités remarquables

Théorème 3.3 *Dans la carte principale, on a les relations suivantes :*

$$f_{\vec{x}}^{(-1)} \star \vec{o} = -J(f_{\vec{x}})^{(-1)}.\vec{x} = -J(f_{\vec{x}}^{(-1)}).\vec{x} \tag{3.38}$$

$$\left(\frac{\partial(f_{\vec{z}}^{(-1)} \star \vec{o})}{\partial \vec{z}}\right)^{T}.Q.(f_{\vec{z}}^{(-1)} \star \vec{o}) = Q.\vec{z} \tag{3.39}$$

Ces relations s'expriment de manière plus générale par :

$$(f_{\vec{x}}^{(-1)} \star \vec{y}) = -\frac{\partial(f_{\vec{x}}^{(-1)} \star \vec{y})}{\partial \vec{y}}.J(f_{\vec{y}}).(f_{\vec{y}}^{(-1)} \star \vec{x}) = -J(f_{\vec{x}})^{(-1)}.\left(\frac{\partial(f_{\vec{y}}^{(-1)} \star \vec{x})}{\partial \vec{x}}\right)^{(-1)}.(f_{\vec{y}}^{(-1)} \star \vec{x}) \tag{3.40}$$

$$\overrightarrow{xy} = -J(f_{\vec{x}}).\frac{\partial(f_{\vec{x}}^{(-1)} \star \vec{y})}{\partial \vec{y}}.\overrightarrow{yx} = -\left(\frac{\partial(f_{\vec{y}}^{(-1)} \star \vec{x})}{\partial \vec{x}}\right)^{(-1)}.J(f_{\vec{y}})^{(-1)}.\overrightarrow{yx} \tag{3.41}$$

$$\left(\frac{\partial(f_{\vec{y}}^{(-1)} \star \vec{x})}{\partial \vec{y}}\right)^{T}.Q.(f_{\vec{y}}^{(-1)} \star \vec{x}) = \left(\frac{\partial(f_{\vec{x}}^{(-1)} \star \vec{y})}{\partial \vec{y}}\right)^{T}.Q.(f_{\vec{x}}^{(-1)} \star \vec{y}) \tag{3.42}$$

Preuve : La géodésique $\gamma_{(o,\vec{x})} = \exp(t.\vec{x})$ va de l'origine au point x en un temps unité et s'exprime dans la carte principale par $\gamma_o(t) = t.\vec{x}$. Considérons la géodésique inverse: $\delta(t) = \gamma(1-t)$. Elle s'exprime dans la carte principale par $\delta_o(t) = (1-t).\vec{x}$ et va de $\delta(0) = x$ à l'origine $\delta(1) = o$ et a une longueur unité. Elle a donc un vecteur tangent $\dot{\delta}_o$ constant et égal à $\overrightarrow{xo} = J(f_{\vec{x}}).(f_{\vec{x}}^{(-1)} \star o)$ dans la carte principale. Par ailleurs, la dérivation de son équation lui attribue un vecteur tangent égal à $-\vec{x}$. On obtient donc la première partie de l'identité remarquable (3.38) :

$$\overrightarrow{xo} = J(f_{\vec{x}}).(f_{\vec{x}}^{(-1)} \star o) = -\vec{x}$$

Observons également que la géodésique $\alpha(t) = f_{\vec{x}}^{(-1)} \star \gamma(1-t)$ va de l'origine au point $f_{\vec{x}}^{(-1)} \star \vec{o}$ en un temps unitaire. Elle a donc un vecteur tangent $\dot{\alpha}$ dans la carte principale constant et égal à $f_{\vec{x}}^{(-1)} \star \vec{o}$. Or, le calcul direct donne :

$$\dot{\alpha}(1) = \left.\frac{d(f_{\vec{x}}^{(-1)} \circ ((1-t).\vec{x}))}{dt}\right|_{t=1} = -\left.\frac{\partial(f_{\vec{x}}^{(-1)} \circ \vec{e})}{\partial \vec{e}}\right|_{\vec{e}=\vec{o}}.\vec{x} = -J_L(f_{\vec{x}}^{(-1)}).\vec{x}$$

On obtient donc au final la seconde partie de (3.38)

$$\vec{fx}^{(-1)} = -J_L(f_{\vec{x}})^{(-1)}.\vec{x} = -J_L(f_{\vec{x}}^{(-1)}).\vec{x}$$

Considérons maintenant la primitive $\vec{z} = f_{\vec{x}}^{(-1)} \star \vec{y}$. On veut exprimer la relation $\overrightarrow{zo} = -\vec{z}$ directement en fonction de $\vec{x}$ et $\vec{y}$. La fonction de placement étant choisie sur toute la variété, il existe $h \in \mathcal{H}$ tel que $f_{\vec{z}} = f_{\vec{x}}^{(-1)} \circ f_{\vec{y}} \circ h$. On a donc :

$$J(f_{\vec{z}}) = \frac{\partial((f_{\vec{x}}^{(-1)} \circ f_{\vec{y}} \circ h) \star e)}{\partial e} = \frac{\partial(f_{\vec{x}}^{(-1)} \star \vec{y})}{\partial \vec{y}}.J(f_{\vec{y}}).J(h)$$

$$J(f_{\vec{z}}^{(-1)}) = \frac{\partial((h^{(-1)} \circ f_{\vec{y}}^{(-1)} \circ f_{\vec{x}}) \star e)}{\partial e} = J(h)^{(-1)}.\frac{\partial(f_{\vec{y}}^{(-1)} \star \vec{x})}{\partial \vec{x}}.J(f_{\vec{x}})$$

D'un autre côté, puisque l'action de h est affine dans la carte principale, on a :

$$f_{\vec{z}}^{(-1)} \star o = h^{(-1)} \star (f_{\vec{y}}^{(-1)} \star \vec{x}) = J(h)^{(-1)}.(f_{\vec{y}}^{(-1)} \star \vec{x})$$

On a donc

$$-\vec{z} = J(f_{\vec{z}}).(f_{\vec{z}}^{(-1)} \star o) = \frac{\partial(f_{\vec{x}}^{(-1)} \star \vec{y})}{\partial \vec{y}}.J(f_{\vec{y}}).(f_{\vec{y}}^{(-1)} \star \vec{x})$$

$$-\vec{z} = J(f_{\vec{z}}^{(-1)})^{(-1)}.(f_{\vec{z}}^{(-1)} \star o) = J(f_{\vec{x}})^{(-1)}.\left(\frac{\partial(f_{\vec{y}}^{(-1)} \star \vec{x})}{\partial \vec{x}}\right)^{(-1)}.(f_{\vec{y}}^{(-1)} \star \vec{x})$$

ce qui donne en reportant dans $\overrightarrow{zo} = -\vec{z}$ les identités remarquables (3.40) et (3.41).

La distance invariante nous fournit cette fois-ci dans la carte principale :

$$\vec{z}^{\mathrm{T}}.Q.\vec{z} = \mathrm{dist}(z, o) = \mathrm{dist}(o, f_z^{(-1)} \star o) = (f_{\vec{z}}^{(-1)} \star \vec{o})^{\mathrm{T}}.Q.(f_{\vec{z}}^{(-1)} \star \vec{o})$$

La dérivation de cette égalité nous donne l'identité remarquable (3.39). De même la dérivation de

$$\mathrm{dist}(x, y) = \mathrm{dist}(f_x^{(-1)} \star y, o) = \mathrm{dist}(\mathrm{Id}, f_y^{(-1)} \star x)$$

par rapport à $\vec{y}$ dans la carte principale donne l'identité (3.42). ■

Ces identités remarquables vont non seulement nous permettre de simplifier les calculs symboliques, mais aussi de les vérifier et de tester la précision numérique de leur implémentation. Nous avons ainsi vérifié avec notre bibliothèque 100000 fois ces relations pour chaque type de primitive (repères, repères semi et non-orientés, points, voir chapitre 7), en tirant les primitives x et y de manière uniforme dans un volume 512x512x512, et sans que la différence ne dépasse 10^{-10}. Sur ces 100000 essais, seulement 20 en moyenne ont une erreur qui dépasse 10^{-12}.

3.6 Résumé

Nous avons formalisé dans ce chapitre les primitives géométriques comme des variétés différentielles sur lesquelles agissent un groupe de Lie. Cette formulation met en avant 3 opérations fondamentales : la composition et l'inversion au sein du groupe, et l'action du groupe sur la variété. En factorisant les caractéristiques invariantes de nos primitives, on peut considérer notre variété comme le produit de la variété des invariants unaires, sur laquelle le groupe n'agit pas, et une variété homogène, entièrement soumise à l'action du groupe.

On développe alors un ensemble d'outils mathématiques pour travailler sur cette variété homogène *en accord avec l'action du groupe*. On détermine ainsi les conditions d'existence et l'expression d'une mesure uniforme ou invariante sur le groupe et la variété, puis les caractéristiques d'une distance invariante. Ce n'est toutefois qu'en formalisant cette contrainte en terme de géométrie riemannienne que nous pouvons développer les conditions d'existence et l'expression d'une telle distance. Ces développements nous fournissent au passage une représentation très spécifique que nous appelons *carte principale*, et que l'on peut considérer comme le développement de la variété le long de ses géodésiques dans l'espace vectoriel tangent à l'origine. Cette représentation permet une expression simple de la distance invariante et prendra une importance capitale au chapitre suivant pour caractériser l'espérance d'une primitive aléatoire et permettre la définition d'une matrice de covariance. Un résumé des formules importantes en pratique sera fourni par le chapitre 6.

Chapitre 4

Probabilités sur les primitives géométriques

« Ainsi, joignant la rigueur des démonstrations de la science à l'incertitude du hasard, et conciliant ces choses en apparence contraires, elle peut, tirant son nom des deux, s'arroger à bon droit ce titre stupéfiant :
La Géométrie du Hasard. *»*
B. Pascal, Adresse à l'Académie Parisienne.

La mesure invariante du chapitre précédent nous permet de résoudre des problèmes impliquant une distribution *uniforme* des primitives. Pour pouvoir gérer correctement les erreurs de mesure sur nos primitives géométriques, nous devons plutôt utiliser des distribution centrées autour de la primitive exacte. Nous introduisons donc tout d'abord la notion de **primitive aléatoire**, dont nous pouvons caractériser la distribution par une densité de probabilité basée sur la mesure invariante. Nous examinons également à la section (4.1) la propagation de ces densités par les opérations de base (composition, inversion d'une transformation et action du groupe sur la variété).

Nous avons vu qu'on obtient une approximation efficace de la densité en ne conservant que la moyenne et la covariance. Le problème que nous attaquerons à la section (4.2) sera de définir la notion de moyenne sur une variété, indépendante de la représentation utilisée et de préférence cohérente avec l'action du groupe. Un second problème, envisagé à la section (4.3) est la caractérisation et l'obtention de la moyenne ainsi définie. La section suivante sera consacrée à la définition de la matrice de covariance, ainsi qu'à sa propagation, et la dernière section montrera très rapidement comment on peut utiliser toute cette artillerie pour gérer plusieurs primitives ou transformations aléatoires simultanément.

4.1 Densité de probabilité d'une primitive aléatoire

« Comment oser parler des lois du hasard?
Le hasard n'est-il pas l'antithèse de toute loi? »
J. Bertrand, Calcul des probabilités, 1907

4.1.1 Primitive aléatoire

Si nous reprenons la structure du chapitre 2, nous n'avons pas besoin de toucher à la notion d'espace probabilisé, de tribu d'événements et de mesure de probabilité. En bref, la section (2.2.1) reste d'actualité.

Maintenant, au lieu de considérer que nous mesurons des *variables* ou des *vecteurs* qui sont fonction des événements de l'espace probabilisé d'origine Ω, nous nous intéressons à des mesures à valeur dans une variété $\mathcal{M}$ (rappelons que les groupes que nous utilisons sont aussi des variétés). Soit $\omega \in \Omega$ un événement élémentaire. On appelle **primitive aléatoire** une fonction (borélienne) $\mathbf{x} = \mathrm{x}(\omega)$ de Ω dans $\mathcal{M}$. Comme dans le cas vectoriel, on peut faire abstraction de l'espace d'origine Ω et travailler avec la mesure de probabilité induite sur l'espace de mesure, en l'occurrence la variété $\mathcal{M}$.

On supposera par défaut dans la suite de cette section que $\mathcal{M}$ est homogène et possède une mesure invariante notée $d\mathcal{M}$. Dans le cas d'un groupe de Lie $\mathcal{G}$, on le supposera localement compact, auquel cas il possède toujours une mesure invariante à gauche $d_L\mathcal{G}$.

4.1.2 Définition de la densité de probabilité

On note $\mathcal{A}$ la tribu borélienne de la variété $\mathcal{M}$ (engendrée par la classe des ouverts). On dit que la primitive aléatoire $\mathbf{x}$ possède ou suit une densité de probabilité p (fonction réelle positive intégrable [1]) si :

$$\forall \mathcal{X} \in \mathcal{A}, \qquad \Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \tag{4.1}$$

La contrainte de normalisation s'exprime bien sûr par

$$\Pr(\mathcal{M}) = \int_{\mathcal{M}} p(\mathrm{y}).d\mathcal{M}(\mathrm{y}) = 1$$

On notera par la suite $p_{\mathbf{x}}$ la densité de la primitive aléatoire $\mathbf{x}$. Un exemple simple de densité est la **densité uniforme** sur un ensemble $\mathcal{X}$ borné :

$$p_{\mathcal{X}}(\mathrm{y}) = \frac{1}{\int_{\mathcal{X}} d\mathcal{M}} \mathbf{1}_{\mathcal{X}}(\mathrm{y}) = \frac{\mathbf{1}_{\mathcal{X}}(\mathrm{y})}{\mathcal{V}(\mathcal{X})} \tag{4.2}$$

où $\mathcal{V}(\mathcal{X})$ peut être considéré comme le volume de l'ensemble $\mathcal{X}$.

Les densités se définissent de manière similaire pour un groupe de transformation $\mathcal{G}$. On peut cependant utiliser la mesure invariante *à gauche* où *à droite*, qui sont différentes si le groupe n'est pas uni-modulaire. Comme le groupe agit à droite, nous supposons dans la suite que la densité est définie en utilisant la mesure invariante *à gauche*.

1. Les fonctions à support compact sont en particulier intégrables.

4.1.3 Densité dans une représentation

Soit $p_{\mathbf{x}}$ la densité d'une primitive aléatoire $\mathbf{x}$. Si $x = \pi(\mathrm{x})$ est une représentation de la variété (définie presque partout pour simplifier), on obtient un vecteur aléatoire $\boldsymbol{x} = \pi(\mathbf{x})$, dont la densité $\rho_{\boldsymbol{x}}$ est maintenant définie par la mesure de Lebesgue dx à la place de la mesure invariante $d\mathcal{M}(x)$. En incorporant l'expression de la mesure invariante obtenue à l'équation (3.10), la probabilité pour que $\mathbf{x}$ soit dans un ensemble $\mathcal{X}$ est

$$\Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) = \int_{\pi(\mathcal{X})} \frac{p_{\mathbf{x}}(y)}{|J(f_y)|}.dy = \int_{\pi(\mathcal{X})} \rho_{\boldsymbol{x}}(y).dy = \Pr(\boldsymbol{x} \in \pi(\mathcal{X}))$$

où $J(f_y)$ est le jacobien de l'action de $f_y \in \mathcal{F}_y$ sur l'origine. Par définition, la densité du *vecteur aléatoire* x est donc (en se limitant au domaine de définition de la carte : $y \in \mathcal{D}$) :

$$\rho_{\boldsymbol{x}}(y) = \frac{p_{\mathbf{x}}(y)}{|J(f_y)|} \tag{4.3}$$

Notons que la densité $p_{\mathbf{x}}$ que nous avons définie pour une primitive aléatoire $\mathbf{x}$ ne dépend pas de la carte utilisée pour la variété, ce qui n'est absolument pas le cas de la densité $\rho_{\boldsymbol{x}}$ du vecteur aléatoire $\boldsymbol{x}$ la représentant : si $x_1 = \pi_1(\mathrm{x})$ et $x_2 = \pi_2(\mathrm{x})$ sont deux représentations différentes de la variété, les jacobiens sont généralement différents et on a :

$$\rho_{\boldsymbol{x}_1}(\pi_1(\mathrm{x})) \neq \rho_{\boldsymbol{x}_2}(\pi_2(\mathrm{x}))$$

4.1.4 Propagation d'une densité

Nous examinons ici le cas d'une seule primitive ou transformation aléatoire soumise aux opérations géométriques de base.

Théorème 4.1 (Action d'une transformation fixe sur une primitive aléatoire)
Soit $\mathbf{x}$ *une primitive aléatoire de densité* $p_{\mathbf{x}}$ *et* $\mathrm{f} \in \mathcal{G}$ *une transformation fixe. Alors la densité de la primitive aléatoire* $\mathrm{f} \star \mathbf{x}$ *est*

$$p_{(\mathrm{f}\star\mathbf{x})}(\mathrm{y}) = p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y}) \tag{4.4}$$

Preuve : La probabilité que $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ soit dans un ensemble $\mathcal{Y}$ est

$$P(\mathrm{f} \star \mathbf{x} \in \mathcal{Y}) = P(\mathbf{x} \in \mathrm{f}^{(-1)} \star \mathcal{Y}) = \int_{(\mathrm{f}^{(-1)} \star \mathcal{Y})} p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z})$$

ce qui donne avec le changement de variable $\mathrm{z} = \mathrm{f}^{(-1)} \star \mathrm{y}$, et puisque $d\mathcal{M}$ est la mesure invariante,

$$P(\mathrm{f} \star \mathbf{x} \in \mathcal{Y}) = \int_{\mathcal{Y}} p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y}).d\mathcal{M}(\mathrm{y})$$

On obtient l'équation (4.4) avec la définition de la densité. ■

Rappelons que le module $\Delta\mathcal{G}(\mathrm{g})$ du groupe quantifie par différence entre la mesure invariante à droite et la mesure invariante à gauche : $d_L\mathcal{G}(\mathrm{g}) = \Delta\mathcal{G}(\mathrm{g}).d_R\mathcal{G}(\mathrm{g})$ (voir section 3.3.2). A moins que le groupe ne soit uni-modulaire, la composition d'une transformation aléatoire par une transformation fixe conduit à des densités différentes suivant que l'on compose à gauche ou à droite. Nos densités sont effectivement définies par la mesure invariante à gauche, qui n'est généralement pas invariante à droite.

Théorème 4.2 (Translation d'une transformation aléatoire)
Soit $\mathbf{f}$ *une transformation aléatoire de densité* $p_{\mathbf{f}}$*, et* f_o *une transformation fixe. La densité de* $\mathbf{g}_l = \mathrm{f}_o \circ \mathbf{f}$*, la translation à gauche de* $\mathbf{f}$*, est obtenue (avec la même preuve que précédemment) par*

$$p_{(\mathrm{f}_o\circ\mathbf{f})}(\mathrm{g}) = p_{\mathbf{f}}(\mathrm{f}_o^{(-1)} \circ \mathrm{g}) \tag{4.5}$$

La densité de la translation à droite $\mathbf{g_r} = \mathbf{f} \circ \mathrm{f}_o$ *est quant à elle :*

$$p_{(\mathbf{f} \circ \mathrm{f}_o)}(\mathrm{g}) = \frac{\Delta\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)})}{\Delta\mathcal{G}(\mathrm{g})} . p_{\mathbf{f}}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) \tag{4.6}$$

Preuve : La probabilité que $\mathbf{g_r}$ soit dans un ensemble $\mathcal{Y} \subset \mathcal{G}$ est :

$$\Pr\left((\mathbf{f} \circ \mathrm{f}_o) \in \mathcal{Y}\right) = \Pr(\mathbf{f} \in (\mathcal{Y} \circ \mathrm{f}_o^{(-1)})) = \int_{(\mathcal{Y} \circ \mathrm{f}_o^{(-1)})} p_{\mathbf{f}}(\mathrm{f}') . d_L\mathcal{G}(\mathrm{f}')$$

Avec le changement de variable $\mathrm{f}' = \mathrm{g} \circ \mathrm{f}_o^{(-1)}$, on obtient

$$\begin{aligned} \Pr((\mathbf{f} \circ \mathrm{f}_o) \in \mathcal{Y}) &= \int_{\mathcal{Y}} p_{\mathbf{f}}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . d_L\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) \\ &= \int_{\mathcal{Y}} \Delta\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . p_{\mathbf{f}}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . d_R\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) \end{aligned}$$

Maintenant, puisque $d_R\mathcal{G}$ est la mesure invariante à droite, on peut simplifier et revenir à la mesure invariante à gauche :

$$\begin{aligned} \Pr((\mathbf{f} \circ \mathrm{f}_o) \in \mathcal{Y}) &= \int_{\mathcal{Y}} \Delta\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . p_{\mathbf{f}}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . d_R\mathcal{G}(\mathrm{g}) \\ &= \int_{\mathcal{Y}} \frac{\Delta\mathcal{G}(\mathrm{g} \circ \mathrm{f}_o^{(-1)})}{\Delta\mathcal{G}(\mathrm{g})} . p_{\mathbf{f}}(\mathrm{g} \circ \mathrm{f}_o^{(-1)}) . d_L\mathcal{G}(\mathrm{g}) \end{aligned}$$

On conclut avec la définition de la densité. ■

Théorème 4.3 (Action d'une transformation aléatoire sur l'origine)
Soit $\mathbf{f}$ *une transformation aléatoire de densité* $p_{\mathbf{f}}$*. Son action sur l'origine* o *détermine une primitive aléatoire* $\mathbf{y} = \mathbf{f} \star \mathrm{o}$ *de densité*

$$p_{(\mathbf{f} \star \mathrm{o})}(\mathrm{y}) = \int_{\mathcal{H}} p_{\mathbf{f}}(\mathrm{f}_\mathrm{y} \circ \mathrm{h}) . d\mathcal{H}(\mathrm{h}) \tag{4.7}$$

Preuve : La probabilité que $\mathbf{x} = \mathbf{f} \star \mathrm{o}$ soit dans un ensemble $\mathcal{X} \in \mathcal{M}$ est donnée par

$$\Pr(\mathbf{f} \star \mathrm{o} \in \mathcal{X}) = \int_{\mathcal{F}_\mathcal{X}} p_{\mathbf{f}}(\mathrm{g}) . d\mathcal{G}(\mathrm{g}) \qquad \text{avec} \qquad \mathcal{F}_\mathcal{X} = \{\mathrm{g} \in \mathcal{G} \,/\, \mathrm{g} \star \mathrm{o} \in \mathcal{X}\}$$

Si f_x est un représentant de $\mathcal{F}_\mathrm{x}$, l'ensemble des transformations $\mathcal{F}_\mathcal{X}$ est donc $\mathcal{F}_\mathcal{X} = \{\mathrm{f}_\mathrm{x} \circ \mathrm{h} / \mathrm{x} \in \mathcal{X}, \mathrm{h} \in \mathcal{H}\}$.
La probabilité ci-dessus s'écrit donc

$$\Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X},\mathcal{H}} p_{\mathbf{f}}(\mathrm{f}_\mathrm{x} \circ \mathrm{h}) . d\mathcal{M}(\mathrm{x}) . d\mathcal{H}(\mathrm{h}) = \int_{\mathcal{X}} \left(\int_{\mathcal{H}} p_{\mathbf{f}}(\mathrm{f}_\mathrm{x} \circ \mathrm{h}) . d\mathcal{H}(\mathrm{h}) \right) . d\mathcal{M}(\mathrm{x})$$

L'équation (4.7) suit. ■

Si l'on veut maintenant appliquer la transformation aléatoire à une primitive x différente de l'origine, on peut choisir $\mathrm{f}_\mathrm{x} \in \mathcal{F}_\mathrm{x}$ et écrire : $\mathbf{y} = \mathbf{f} \star x = (\mathbf{f} \circ \mathrm{f}_\mathrm{x}) \star \mathrm{o}$

Théorème 4.4 (Action d'une transformation aléatoire sur une primitive fixe)
Soit $\mathbf{f}$ *une transformation aléatoire de densité* $p_{\mathbf{f}}$*. Son action sur la primitive* x *détermine une primitive aléatoire* $\mathbf{y} = \mathbf{f} \star \mathrm{x}$ *de densité*

$$p_{(\mathbf{f} \star x)}(y) = \int_{\mathcal{H}} p_{(\mathbf{f} \circ f_x)}(f_y \circ h) . d\mathcal{H}(h) = \int_{\mathcal{H}} \frac{\Delta\mathcal{G}(f_y \circ h \circ f_x^{(-1)})}{\Delta\mathcal{G}(f_y \circ h)} . p_{\mathbf{f}}(f_y \circ h \circ f_x^{(-1)}) . d\mathcal{H}(h) \tag{4.8}$$

4.1.5 Algèbre des densités des primitives et transformations aléatoires

Nous avons vu comment la densité se propageait si une transformation ou une primitive était aléatoire. La question est maintenant de propager les densités si tous les éléments sur lesquels on

agit sont aléatoires. Ceci nous donnera un ensemble cohérent d'opérations, au moins au niveau mathématique, pour travailler avec les primitives et les transformations aléatoires.

Théorème 4.5 Composition de transformation aléatoires
Soit $\mathbf{f}_1$ *et* $\mathbf{f}_2$ *deux transformation aléatoires de densité* $p_{\mathbf{f}_1}$ *et* $p_{\mathbf{f}_2}$. *La densité de la transformation composée est*

$$p_{(\mathbf{f}_1 \circ \mathbf{f}_2)}(\mathrm{f}) = \int_{\mathcal{G}} p_{\mathbf{f}_1}(\mathrm{g}).p_{\mathbf{f}_2}(\mathrm{g}^{(-1)} \circ \mathrm{f}).d_L\mathcal{G}(\mathrm{g}) \tag{4.9}$$

Preuve : Soit $\mathcal{F} \subset \mathcal{G}$ un ensemble de transformations. L'ensemble $\mathcal{A}$ des couples $(\mathrm{g}_1, \mathrm{g}_2)$ tels que $\mathrm{g}_1 \circ \mathrm{g}_2 \in \mathcal{F}$ peut s'écrire de deux façons :

$$(1) \qquad \mathcal{A} = \left\{ (\mathrm{g}_1, \mathrm{g}_2) \,/\, \mathrm{g}_1 \in \mathcal{G} \,,\, \mathrm{g}_2 \in \mathrm{g}_1^{(-1)} \circ \mathcal{F} \right\}$$
$$(2) \qquad \mathcal{A} = \left\{ (\mathrm{g}_1, \mathrm{g}_2) \,/\, \mathrm{g}_2 \in \mathcal{G} \,,\, \mathrm{g}_1 \in \mathcal{F} \circ \mathrm{g}_2^{(-1)} \right\}$$

En utilisant la première formulation, la probabilité pour que $\mathbf{f} = \mathbf{f}_1 \circ \mathbf{f}_2$ soit dans l'ensemble $\mathcal{F}$ est

$$\begin{aligned}
\Pr(\mathbf{f}_1 \circ \mathbf{f}_2 \in \mathcal{F}) &= \int_{\mathcal{G}} \left(\int_{\mathrm{g}_1^{(-1)} \circ \mathcal{F}} p_{\mathbf{f}_2}(\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2) \right) p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1) \\
&= \int_{\mathcal{G}} \left(\int_{\mathcal{F}} p_{\mathbf{f}_2}(\mathrm{g}_1^{(-1)} \circ \mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2) \right) p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1) \\
&= \int_{\mathcal{F}} \left(\int_{\mathcal{G}} p_{\mathbf{f}_2}(\mathrm{g}_1^{(-1)} \circ \mathrm{g}_2).p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1) \right) d_L\mathcal{G}(\mathrm{g}_2)
\end{aligned}$$

Nous avons donc obtenu la densité recherchée. En utilisant la seconde formulation de l'ensemble $\mathcal{A}$, on a :

$$\begin{aligned}
\Pr(\mathbf{f}_1 \circ \mathbf{f}_2 \in \mathcal{F}) &= \int_{\mathcal{G}} \left(\int_{\mathcal{F} \circ \mathrm{g}_2^{(-1)}} p_{\mathbf{f}_1}(\mathrm{g}_1).d_L\mathcal{G}(\mathrm{g}_1) \right) p_{\mathbf{f}_2}(\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2) \\
&= \int_{\mathcal{G}} \left(\int_{\mathcal{F}} p_{\mathbf{f}_1}(\mathrm{g} \circ \mathrm{g}_2^{(-1)}).d_L\mathcal{G}(\mathrm{g} \circ \mathrm{g}_2^{(-1)}) \right) p_{\mathbf{f}_2}(\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2) \\
&= \int_{\mathcal{G}} \left(\int_{\mathcal{F}} \frac{\Delta\mathcal{G}(\mathrm{f} \circ \mathrm{g}^{(-1)})}{\Delta\mathcal{G}(\mathrm{f})} p_{\mathbf{f}_1}(\mathrm{g} \circ \mathrm{g}_2^{(-1)}).d_L\mathcal{G}(\mathrm{g}) \right) p_{\mathbf{f}_2}(\mathrm{g}_2).d_L\mathcal{G}(\mathrm{g}_2)
\end{aligned}$$

Ceci donne la densité

$$p_{(\mathbf{f}_1 \circ \mathbf{f}_2)}(\mathrm{f}) = \int_{\mathcal{G}} \frac{\Delta\mathcal{G}(\mathrm{f} \circ \mathrm{g}^{(-1)})}{\Delta\mathcal{G}(\mathrm{f})} . p_{\mathbf{f}_1}(\mathrm{f} \circ \mathrm{g}^{(-1)}).p_{\mathbf{f}_2}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})$$

Cependant, on doit noter que le changement de variable $\mathrm{g}_1 = g \circ \mathrm{g}_2^{(-1)}$ nous ramène à la première forme de la densité. ■

Analogie avec le produit de convolution On doit noter l'analogie remarquable de la formule (4.9) avec le produit de convolution classique obtenu pour l'addition de deux vecteurs aléatoires de $\mathbb{R}^n$. Rappelons la formule (2.11) : la densité du vecteur aléatoire $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ est donné par le produit de convolution

$$p_{(\boldsymbol{x}+\boldsymbol{y})}(z) = p_{\boldsymbol{x}} \circledast p_{\boldsymbol{y}}(z) = \int_{\mathbb{R}^n} p_{\boldsymbol{x}}(t).p_{\boldsymbol{y}}(z-t).dt$$

Cette formule peut être trouvée comme un cas particulier de l'équation (4.9) dérivée ci-dessus : considérons le groupe des translations de $\mathbb{R}^n$. On trouve que la mesure invariante à gauche (et à droite dans ce cas) est $d_L\mathcal{G}(f) = df$ (f est ici un vecteur de translation), et comme l'inversion et la composition s'écrivent $g^{(-1)} = -g$ et $f \circ g = f + g$, on obtient :

$$p_{(\boldsymbol{f}_1+\boldsymbol{f}_2)}(f) = \int_{\mathbb{R}^n} p_{\boldsymbol{f}_1}(g).p_{\boldsymbol{f}_2}(f-g).dg$$

Théorème 4.6 (Inversion d'une transformation aléatoire)
Soit $\mathbf{f}$ *une transformation aléatoire de densité* $p_{\mathbf{f}}$. *La densité de la transformation inverse s'exprime*

en utilisant le module du groupe :

$$p_{\mathbf{f}^{(-1)}}(\mathrm{g}) = \Delta\mathcal{G}(\mathrm{g}^{(-1)}).p_{\mathbf{f}}(\mathrm{g}^{(-1)}) \tag{4.10}$$

Preuve : La première étape est de déterminer la relation entre $d_L\mathcal{G}(\mathrm{g}^{(-1)})$ et $d_L\mathcal{G}(\mathrm{g})$. En supposant que l'on soit dans un système de coordonnées locales, on a $d(g^{(-1)}) = \left|\frac{\partial g^{(-1)}}{\partial g}\right|.dg$, et donc :

$$d_L\mathcal{G}(g^{(-1)}) = \left|\frac{\partial g^{(-1)}}{\partial g}\right|.\frac{|J_L(g)|}{|J_L(g^{(-1)})|}.d_L\mathcal{G}(g)$$

D'un autre côté, on a :

$$J_R(g^{(-1)}) = \left.\frac{\partial(e\circ g^{(-1)})}{\partial e}\right|_{e=Id} = \left.\frac{\partial(e\circ g^{(-1)})}{\partial(g\circ e^{(-1)})}\right|_{e=Id}.\left.\frac{\partial(g\circ e^{(-1)})}{\partial e^{(-1)}}\right|_{e^{(-1)}=Id}.\left.\frac{\partial(e^{(-1)})}{\partial e}\right|_{e=Id}$$

et donc

$$J_R(g^{(-1)}) = \frac{\partial g^{(-1)}}{\partial g}.J_L(g).\left.\frac{\partial(e^{(-1)})}{\partial e}\right|_{e=Id}$$

Mais le dernier termes est la matrice identité, et la première formule se simplifie donc en

$$d_L\mathcal{G}(g^{(-1)}) = \frac{|J_R(g^{(-1)})|}{|J_L(g^{(-1)})|}.d_L\mathcal{G}(g) \qquad \Longleftrightarrow \qquad d_L\mathcal{G}(\mathrm{g}^{(-1)}) = \Delta\mathcal{G}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g})$$

Maintenant, soit $\mathbf{f}$ une transformation aléatoire de densité $p_{\mathbf{f}}$, et $\mathcal{F}$ un ensemble de transformations :

$$\begin{aligned}\Pr(\mathbf{f}^{(-1)}\in\mathcal{F}) &= \Pr(\mathbf{f}\in\mathcal{F}^{(-1)}) = \int_{\mathcal{F}^{(-1)}} p_{\mathbf{f}}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})\\ &= \int_{\mathcal{F}} p_{\mathbf{f}}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g}^{(-1)}) = \int_{\mathcal{F}} \Delta\mathcal{G}(\mathrm{g}^{(-1)})p_{\mathbf{f}}(\mathrm{g}^{(-1)}).d_L\mathcal{G}(\mathrm{g})\end{aligned}$$

Ceci donne la densité proposée. ■

Théorème 4.7 (Action d'une transformation aléatoire sur une primitive aléatoire) *Soit* $\mathbf{f}$ *une transformation aléatoire de densité* $p_{\mathbf{f}}$ *et* $\mathbf{x}$ *une primitive aléatoire de densité* $p_{\mathbf{x}}$. *La densité de la primitive aléatoire* $\mathbf{y} = \mathbf{f}\star\mathbf{x}$ *est*

$$p_{\mathbf{f}\star\mathbf{x}}(\mathrm{y}) = \int_{\mathcal{G}} p_{\mathbf{f}}(\mathrm{g}).p_{\mathbf{x}}(\mathrm{g}^{(-1)}\star\mathrm{y}).d_L\mathcal{G}(\mathrm{g}) \tag{4.11}$$

On peut encore noter la similarité avec un produit de convolution. L'analogie serait plutôt cette fois-ci l'action d'un vecteur de translation aléatoire à un point aléatoire.

Preuve : Soit $\mathcal{X}$ un ensemble de primitives. L'ensemble $\mathcal{A}$ des couples (f,x) tels que $\mathrm{f}\star\mathrm{x}\in\mathcal{X}$ peut s'écrire

$$\mathcal{A} = \left\{(\mathrm{f},\mathrm{x}) \,/\, \mathrm{f}\in\mathcal{G}\ ,\ \mathrm{x}\in\mathrm{f}^{(-1)}\star\mathcal{X}\right\}$$

et la probabilité que $\mathbf{y} = \mathrm{f}\star\mathrm{x}$ soit dans $\mathcal{X}$ est donc

$$\begin{aligned}\Pr(\mathbf{f}\star\mathbf{x}\in\mathcal{X}) &= \int_{\mathcal{G}}\left(\int_{\mathrm{g}^{(-1)}\star\mathcal{X}} p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y})\right).p_{\mathbf{f}}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})\\ &= \int_{\mathcal{G}}\left(\int_{\mathcal{X}} p_{\mathbf{x}}(\mathrm{g}^{(-1)}\star\mathrm{z}).d\mathcal{M}(\mathrm{z})\right)p_{\mathbf{f}}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})\\ &= \int_{\mathcal{X}}\left(\int_{\mathcal{G}} p_{\mathbf{x}}(\mathrm{g}^{(-1)}\star\mathrm{z}).p_{\mathbf{f}}(\mathrm{g}).d_L\mathcal{G}(\mathrm{g})\right).d\mathcal{M}(\mathrm{z})\end{aligned}$$

La densité de la primitive aléatoire $\mathbf{y} = \mathbf{f}\star\mathbf{x}$ est donc donnée par l'équation (4.11). ■

4.1.6 Espérance d'une fonction réelle (observable)

Soit $\varphi(\mathrm{x})$ une fonction à valeur réelle sur la variété $\mathcal{M}$, et $\mathbf{x}$ une primitive aléatoire de densité $p_{\mathbf{x}}$. Alors $\varphi(\mathbf{x})$ est une variable aléatoire réelle dont on peut calculer l'espérance. On note :

$$\mathbf{E}\left[\,\varphi(\mathbf{x})\,\right] = \mathbf{E}_{\mathbf{x}}\left[\,\varphi\,\right] = \int_{\mathcal{M}} \varphi(\mathrm{y}).p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \tag{4.12}$$

Cette notion de l'espérance correspond à celle que l'on a définie sur les variables et vecteurs aléatoires, mais n'est pas ici extensible à la définition d'une valeur moyenne de la distribution. Notons que l'opérateur d'espérance ainsi défini est toujours linéaire : si λ_1 et λ_2 sont deux scalaires et φ_1 et φ_2 deux fonctions à valeur dans $\mathbb{R}$, on a

$$\mathbf{E}_{\mathbf{x}}[\,\lambda_1.\varphi_1 + \lambda_2.\varphi_2\,] = \lambda_1.\mathbf{E}_{\mathbf{x}}[\,\varphi_1\,] + \lambda_2.\mathbf{E}_{\mathbf{x}}[\,\varphi_2\,]$$

Une propriété intéressante concerne le devenir de cette espérance lorsque l'on transforme le système par une transformation fixe f :

$$\begin{aligned}\mathbf{E}_{(\mathrm{f}\star\mathbf{x})}[\,\varphi\,] &= \textstyle\int_{\mathcal{M}} \varphi(\mathrm{y}).p_{(\mathrm{f}\star\mathbf{x})}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) = \int_{\mathcal{M}} \varphi(\mathrm{y}).p_{\mathbf{x}}((\mathrm{f}^{(-1)}\star\mathrm{y}).d\mathcal{M}(\mathrm{y}) \\ &= \textstyle\int_{\mathcal{M}} \varphi(\mathrm{f}\star\mathrm{z}).p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z})\end{aligned}$$

où l'on a utilisé le changement de variable $\mathrm{z} = \mathrm{f}^{(-1)}\star\mathrm{y}$. Notons φ_{f} la fonction translatée : $\varphi_{\mathrm{f}}(\mathrm{x}) = \varphi(\mathrm{f}\star\mathrm{x})$. Alors, on a

$$\mathbf{E}_{(\mathrm{f}\star\mathbf{x})}[\,\varphi\,] = \mathbf{E}_{\mathbf{x}}[\,\varphi_{\mathrm{f}}\,]$$

En particulier, si φ est une fonction invariante (ce qui ne peut être le cas sur une variété homogène), alors son espérance est aussi invariante.

L'espérance d'une fonction sur le groupe de transformation est bien évidemment similaire, et la propriété ci-dessus est toujours valable *pour la translation à gauche.*

4.1.7 Discussion

Nous avons développé dans cette section une théorie des probabilités sur les primitives homogènes et les transformations aléatoires fondée sur des densités relatives à la mesure invariante (à gauche pour le groupe). Ces densités sont intrinsèques à la variété (i.e. ne dépendent pas de la carte choisie) mais peuvent aisément être reliées à la densité classique du *vecteur aléatoire* représentant la primitive aléatoire dans n'importe quelle représentation.

Nous avons ensuite étendu les opérations de base (composition, inversion et action) aux primitives aléatoires en calculant la propagation des densités. D'un point de vue mathématique, nous avons donc obtenu une algèbre cohérente pour gérer les primitives et les transformations « déterministes » et aléatoires.

Dans le cas où il n'existerait pas de mesure invariante sur la variété, on peut encore définir la densité de probabilité à partir d'une mesure « dite uniforme » sur la variété, mais on doit gérer les changements de mesures induits par les transformations dans les formules de propagation. Les formules en causes sont calculables, bien que nettement plus complexes et n'offrent plus du tout la simplicité du schéma présenté ici.

D'un point de vue pratique, il nous reste maintenant à approximer ces densités par des valeurs significatives, comme nous avons « simplifié » la densité des vecteurs aléatoires par leur moyenne et leur matrice de covariance.

4.2 Primitive et transformation moyenne

Nous nous intéressons ici à la notion de **moyenne** d'une primitive aléatoire, en préférant ce terme à celui d'**espérance** (même si nous l'emploierons aussi) pour mieux différencier la notion d'élément milieu d'une distribution de celle d'espérance d'une fonction réelle.

Rappelons l'espérance ou la valeur moyenne d'un vecteur aléatoire $\boldsymbol{x}$ de densité $p_{\boldsymbol{x}}$:

$$\overline{x} = \mathbf{E}[\,\boldsymbol{x}\,] = \int_{\mathcal{D}} y.p_{\boldsymbol{x}}(y).dy$$

Comme d'un point de vue pratique nous aurons rarement des densités mais plus souvent un ensemble de mesures ou une population $\{x_i\}$ provenant de cette densité, on rappelle également la moyenne empirique

$$\overline{x} = \mathbf{E}\,[\,\{x_i\}\,] = \frac{1}{n}\sum_i x_i$$

qui est en fait un estimateur statistique de la moyenne ayant de bonnes propriétés.

La généralisation de ces formules du cas vectoriel au cas d'une variété n'est pas triviale, comme nous l'avons vu avec le « paradoxe » de la section (2.3.3). En particulier, le résultat de l'intégrale ou de la somme n'est pas obligatoirement dans le domaine de définition : une somme de matrices de rotations n'a pas de raison d'être orthogonale, particulièrement pour de grandes déviations. De plus, cet opérateur d'espérance ne commute pas (en général) avec un changement de repère ou l'application d'une transformation. Ceci signifie dans le cas d'objets euclidiens que notre moyenne dépend du repère de l'espace choisi, ce qui est inacceptable.

Il existe de nombreuses solutions *ad hoc* pour éviter chacun des problème indépendemment et la plupart du temps pour des cas particuliers. L'idée principale de ces heuristiques et de « centrer » le domaine de définition de la représentation de la variété utilisée avant de faire la moyenne. Si cela est simple dans le cas du cercle, cela devient tout de suite plus problématique pour des primitives comme les droites 3D ou les repères, où la variété est bien plus complexe. Remarquons cependant que si l'on a pu centrer notre représentation, alors la moyenne est justement au centre et le problème est résolu. L'espérance ou la moyenne au sens de Fréchet est un formalisme bien posé qui réalise cette idée : la centralité d'une primitive est basée sur sa distance par rapport à une distribution ou d'autres mesures, et la primitive moyenne est celle qui optimise cette « centralité ». Par contre, comme on parle d'optimisation, on perd en général l'unicité de la solution : il peut y avoir plusieurs primitives moyennes.

4.2.1 Espérance ou moyenne de Fréchet

Soit $\boldsymbol{x}$ un vecteur aléatoire de $\mathbb{R}^n$. Fréchet a observé dans (Fréchet, 1944; Fréchet, 1948) que la variance $\sigma^2_{\boldsymbol{x}}(y) = \mathbf{E}\,[\ \mathrm{dist}(\boldsymbol{x}, y)^2\ ]$ est minimisée pour la valeur moyenne $\overline{x} = \mathbf{E}\,[\,\boldsymbol{x}\,]$. Le point important pour qui permet de généraliser cette formulation est que l'espérance d'une fonction réelle (mesurable) est toujours bien définie (voir section 4.1.6).

On considère donc une distance sur la variété $\mathcal{M}$. Dans notre cas, on choisira la distance invariante si elle existe ou la distance invariante à gauche sur le groupe $\mathcal{G}$. Soit $\mathbf{x}$ une primitive aléatoire de densité $p_{\mathbf{x}}$. L'espérance de la distance au carré entre cette primitive aléatoire et une primitive fixe y est définie par :

$$\sigma^2_{\mathbf{x}}(\mathrm{y}) = \mathbf{E}\,[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^2\ ] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) \tag{4.13}$$

Si la variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ est finie pour toute primitive y (ce qui est en particulier vérifié si la densité a un support compact), nous appelons **primitive moyenne ou espérée** toute primitive $\bar{\mathrm{x}}$ minimisant cette variance. On note $\mathbb{E}\,[\,\mathbf{x}\,]$ l'ensemble des primitives moyennes. On a donc :

$$\mathbb{E}\,[\,\mathbf{x}\,] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\,[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^2\ ]\right) \tag{4.14}$$

S'il existe au moins une primitive moyenne $\bar{\mathrm{x}}$, on appelle **variance** la valeur minimale $\sigma^2_{\mathbf{x}} = \sigma^2_{\mathbf{x}}(\bar{\mathrm{x}})$ et **écart-type** la racine carrée de cette valeur.

De la même façon, on définit la **moyenne empirique ou discrète** d'un ensemble de mesures $\mathrm{x}_1, \ldots \mathrm{x}_n$ par la version discrète :

$$\mathbb{E}\left[\,\{\mathrm{x}_i\}\,\right] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\,\{\,\mathrm{dist}(\mathrm{y},\mathrm{x}_i)^2\}\,\right]\right) = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\frac{1}{n}\sum_i \mathrm{dist}(\mathrm{y},\mathrm{x}_i)^2\right) \tag{4.15}$$

et s'il existe au moins une primitive moyenne $\bar{\mathrm{x}}$, on appelle écart-type empirique (ou RMS pour Root Mean Square) la valeur $s = \sqrt{\frac{1}{n}\sum_i \mathrm{dist}(\bar{\mathrm{x}},\mathrm{x}_i)^2}$.

On peut définir ainsi d'autres types de valeurs centrales : on appelle déviation moyenne à l'ordre α la valeur

$$\sigma_{\mathbf{x},\alpha}(\mathrm{y}) = (\mathbf{E}\left[\,\mathrm{dist}(\mathrm{y},\mathbf{x})^\alpha\,\right])^{1/\alpha} = \left(\int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^\alpha . p_{\mathbf{x}}(\mathrm{z}) . d\mathcal{M}(\mathrm{z})\right)^{1/\alpha} \tag{4.16}$$

Si cette fonction est bornée sur $\mathcal{M}$, on appelle **primitive centrale à l'ordre** α toute primitive $\bar{\mathrm{x}}_\alpha$ la minimisant. A titre d'exemple, on obtient les modes de la densité pour $\alpha = 0$ (les primitives pour lesquelles la densité est maximale), la médiane pour $\alpha = 1$, comme nous l'avons observé à la section (2.2.2.5), et le « barycentre » du support de la densité (qui doit être un compact dans ce cas) pour $\alpha \to \infty$. La définition de ces valeurs centrales s'applique sans problème dans le cas discret d'un ensemble de mesures de primitives, excepté sans doute pour les modes ($\alpha = 0$) et $\alpha \to \infty$ où l'ensemble des primitives espérées est respectivement $\mathbb{E}^0\left[\,\{\mathrm{x}_i\}\,\right] = \mathcal{M}$ et $\mathbb{E}^\infty\left[\,\{\mathrm{x}_i\}\,\right] = \{\mathrm{x}_i\}$, ce qui n'apporte pas grand chose.

Notons que la moyenne de Fréchet est définie ainsi dans tout espace métrique et donc en particulier dans toute variété riemannienne, indépendamment des hypothèses d'invariance que l'on considère ici. En particulier, tout ce que nous venons de dire se transpose littéralement dans le cas d'une moyenne sur le groupe $\mathcal{G}$, avec la convention que nous utilisons la mesure et la distance invariante à gauche.

4.2.2 Existence et unicité : espérance de Karcher

Il est évident que, comme notre moyenne est le résultat d'une minimisation, l'existence de la moyenne n'est pas assuré (le minimum global n'est pas forcément atteint), et le résultat est de toute façon un ensemble et non plus un seul élément (il peut y avoir plusieurs minimums). En ce sens, on se rapproche du comportement classique de certaines valeurs centrales de vecteurs aléatoires. C'est le cas des modes, par exemple : on peut tout à fait envisager des distributions multimodales qui représentent une variable centrée autour de plusieurs valeurs. Cependant, l'espérance de Fréchet ne permet pas de définir tous les modes, même dans le cas vectoriel : on ne conserve que le (ou les) modes d'intensité maximale.

Pour assouplir cette contrainte, (Karcher, 1977) propose de considérer les minimums locaux de la variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ (équation 4.13) et non plus simplement les minimums globaux. Comme il y a bien sûr plus de minimums locaux que globaux, l'ensemble des moyennes au sens de Fréchet est un sous-ensemble de celles de Karcher. Notons au passage que le fait d'utiliser un minimum local permet de caractériser les solutions en utilisant simplement les dérivées d'ordre deux au point considéré.

En utilisant cette définition étendue, (Karcher, 1977) et (Kendall, 1990) ont pu établir des conditions sur la variété et la distribution pour garantir l'existence et l'unicité de la moyenne. L'expression de ces conditions nécessite l'introduction de quelques notions complémentaires de géométrie différentielle dont nous donnons ici un aperçu très simplifié.

Courbure riemannienne d'une variété Gauss a montré que la courbure (gaussienne) d'une surface peut s'exprimer en fonction de la métrique de cette surface. Riemann a utilisé cette propriété pour généraliser la notion de courbure aux variétés (riemanniennes) de la façon suivante : en un point x de la variété, on choisit un sous-espace vectoriel $\mathcal{V} \subset T_{\mathrm{x}}\mathcal{M}$ de dimension 2 dans l'espace tangent en ce point. L'espace engendré par les géodésiques de $\mathcal{M}$ démarrant du point x avec un vecteur tangent inclus dans $\mathcal{V}$ est une sous-variété de $\mathcal{M}$ possédant la métrique induite. C'est donc une surface dont on peut déterminer la courbure de Gauss : on appelle **courbure sectionnelle** ou **courbure de Riemann** $\kappa(\mathrm{x}, \mathcal{V})$ cette quantité intrinsèque. Il est clair que dans le cas d'un espace vectoriel, la courbure est toujours nulle. Dans les cas simples, la « surface » de la variété est toujours du même côté du plan tangent, et la courbure est toujours positive. C'est le cas de la sphère ou de $\mathcal{SO}_3$. Par contre, pour les transformations rigides, on rajoute les translations (donc un espace vectoriel de courbure nulle), et la variété est simplement dite de **courbure non-négative**. On peut envisager des variétés de courbure toujours négative, comme le paraboloïde hyperbolique $z = x^2 - y^2$ (la selle de cheval).

Boule géodésique régulière Une boule $\mathcal{B}(\mathrm{x}, r)$ est l'ensemble des point $\mathrm{y} \in \mathcal{M}$ dont la distance à la primitive x est strictement inférieure au rayon r. La boule est dite géodésique si elle ne rencontre pas le lieu de coupure du centre x, c'est-à-dire s'il existe une unique géodésique joignant le centre à n'importe quel point de la boule. Soit κ le maximum de la courbure riemannienne dans la boule. La boule est dite **régulière** si son rayon vérifie l'inégalité $2.r.\sqrt{\kappa} < \pi$.

Par exemple, sur la sphère $\mathcal{S}_2$ de rayon 1, la courbure est constante et égale à 1, et une boule géodésique est régulière si $r < \pi/2$. Elle peut donc presque couvrir un hémisphère, mais elle ne peut jamais inclure un point de l'équateur. Dans une variété de courbure négative, une boule géodésique régulière peut couvrir l'espace entier (d'après le théorème d'Hadamard, une telle variété (si elle est connexe) est d'ailleurs difféomorphe à $\mathbb{R}^n$).

Nous pouvons maintenant revenir à notre problème.

Théorème 4.8 (Existence et unicité de la moyenne de Karcher)
Soit $\mathbf{x}$ *une primitive aléatoire de densité* $p_{\mathbf{x}}$.

- **(Kendall, 1990)** *Si le support de* $p_{\mathbf{x}}$ *est inclus dans une boule géodésique régulière* $\mathcal{B}(\mathrm{y}, r)$, *alors il existe une et une seule moyenne de Karcher de* $\mathbf{x}$ *sur cette boule.*
- **(Karcher, 1977)** *Si le support de* $p_{\mathbf{x}}$ *est inclus dans une boule géodésique régulière* $\mathcal{B}(\mathrm{y}, r)$ *et que la boule de rayon double* $\mathcal{B}(\mathrm{y}, 2.r)$ *est encore géodésique et régulière, alors la variance* $\sigma^2_{\mathbf{x}}(\mathrm{z})$ *est une fonction convexe de* z *et a un et un seul point critique sur* $\mathcal{B}(\mathrm{y}, r)$, *nécessairement la moyenne de Karcher.*

Ces conditions sont relativement contraignantes, mais assurent cependant un comportement cohérent de notre moyenne pour des distributions localisées.

4.2.3 Autres définitions possibles de l'espérance

Si la moyenne de Fréchet nous convient très bien car elle présente de bonnes propriétés pour réaliser l'optimisation, il existe des travaux proposant d'autres définitions de la notion de moyenne ou de barycentre dans une variété. Nous les signalons ici par souci de complétude et par l'intérêt mathématique qu'elle peuvent présenter, mais elles semblent peu applicables d'un point de vue pratique.

Une autre propriété qui peut également servir de point de départ pour une généralisation de l'espérance est la suivante (Doss, 1949) : si $\boldsymbol{x}$ est une variable aléatoire réelle, alors le seul nombre

réel $\bar{x}$ vérifiant

$$\forall y \in \mathbb{R} \qquad |y - \bar{x}| \leq \mathbf{E}\,[\,|\boldsymbol{x} - \bar{x}|\,]$$

est égal à la moyenne de $\boldsymbol{x}$: $\bar{x} = \mathbf{E}\,[\,\boldsymbol{x}\,]$.

Si l'on se place maintenant dans un espace métrique, on peut définir la **moyenne au sens de Doss** comme l'ensemble des éléments $\bar{\mathrm{x}} \in \mathcal{M}$ de la variétés vérifiant :

$$\forall \mathrm{y} \in \mathcal{M} \qquad \mathrm{dist}(\mathrm{y}, \bar{\mathrm{x}}) \leq \mathbf{E}\,[\ \mathrm{dist}(\mathbf{x}, \bar{\mathrm{x}})\,]$$

(Herer, 1986; Herer, 1988) montre que cette définition comprend entre autre l'espérance classique dans un espace de Banach (mais comprend éventuellement d'autres points en plus), et développe une espérance conditionnelle basée sur cette définition.

Une définition similaire mais ne faisant pas intervenir de propriétés métriques est utilisée dans (Emery et Mokobodzki, 1991) et reprise dans (Arnaudon, 1994; Arnaudon, 1995) en utilisant les fonction convexes sur la variété. Rappelons qu'une fonction sur $\mathcal{M}$ à valeur dans $\mathbb{R}$ est convexe si sa restriction à toute géodésique (considérée comme fonction de $\mathbb{R}$ dans $\mathbb{R}$) est convexe. Le **barycentre convexe** d'une primitive aléatoire $\mathbf{x}$ de densité $p_{\mathbf{x}}$ est l'ensemble $\mathbb{B}(\mathbf{x})$ des primitives $\mathrm{y} \in \mathcal{M}$ telles que $\alpha(\mathrm{y}) \leq \mathbf{E}\,[\,\alpha(\mathbf{x})\,]$ *pour toute fonction réelle α convexe et bornée sur un voisinage du support de $p_{\mathbf{x}}$*.

Cette définition semble d'un intérêt restreint dans notre cas puisque sur les variétés compactes, comme la sphère où la variété $\mathcal{SO}_3$ des rotations, les géodésiques se ferment sur elles-mêmes et les seules fonctions convexes sont les fonctions constantes. Toute variable aléatoire dont la distribution a pour support la variété entière, par exemple si la densité ne s'annule pas, a donc pour barycentre la variété toute entière.

Cependant, dans le cas où le support de la distribution est inclus dans un ouvert fortement convexe[2] $\mathcal{U}$, Emery montre que les **barycentres exponentiels**, définis comme les points critiques de la variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ (équation 4.13), sont un sous-ensemble du barycentre $\mathbb{B}(\mathbf{x})$. Les minimums locaux et globaux étant en particulier des points critiques, les barycentres exponentiels incluent donc les moyennes de Karcher et de Fréchet.

(Picard, 1994) réalise une synthèse très bien construite de toutes ces notions de moyenne et montre que toute définition d'un barycentre est reliée à un connecteur, qui détermine lui-même une connexion (et donc éventuellement une métrique). Un propriété très intéressante de cette formulation est que la distance entre deux barycentres (de définitions différentes) pour la même primitive aléatoire est de l'ordre de $O(\sigma^2_{\mathbf{x}})$. Pour les primitives aléatoires suffisamment centrées, toutes les valeurs centrales sont donc proches.

4.2.4 Propagation de la moyenne

Comme nous avons calculé la propagation des densités, il nous serait fort utile de connaître la propagation de la moyenne au sein des opérations de base. Notons que les propriétés « d'invariance » ou de cohérence de la moyenne avec l'action des transformations sont dues à l'utilisation d'une distance invariante et ne sont donc plus valables s'il n'existe pas de distance invariante.

Théorème 4.9 (Action d'une transformation fixe sur une primitive aléatoire)

$$\mathbb{E}\,[\,\mathrm{g} \star \mathbf{x}\,] = \mathrm{g} \star \mathbb{E}\,[\,\mathbf{x}\,] \qquad \textit{et} \qquad \mathbb{E}\,[\,\{\mathrm{g} \star \mathrm{x}_i\}\,] = \mathrm{g} \star \mathbb{E}\,[\,\{x_i\}\,] \tag{4.17}$$

Ce résultat tient également pour l'ensemble des primitives centrales d'ordre quelconques.

2. On utilise ici *fortement convexe* au sens où deux points quelconques de $\mathcal{U}$ sont reliés par une unique géodésique minimisante incluse dans $\mathcal{U}$ et dépendant de façon C^∞ des deux points. Il existe alors en tout point une carte exponentielle couvrant $\mathcal{U}$ et ne rencontrant pas le lieu de coupure.

Preuve : Soit $\mathbf{z} = \mathrm{g} \star \mathbf{x}$ la primitive aléatoire obtenue par l'action de la transformation fixe g sur la primitive aléatoire $\mathbf{x}$. On a :

$$\sigma^2_{\mathbf{z}}(\mathrm{y}) = \mathbf{E}\left[\ \mathrm{dist}(\mathrm{g} \star \mathbf{x}, \mathrm{y})^2\ \right] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y}, \mathrm{g} \star \mathrm{x})^2 . p_{\mathbf{x}}(\mathrm{x}) . d\mathcal{M}(\mathrm{x})$$

Grâce à l'invariance de la distance, on obtient

$$\sigma^2_{\mathbf{z}}(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{g}^{(-1)} \star \mathrm{y}, \mathrm{x})^2 . p_{\mathbf{x}}(\mathrm{x}) . d\mathcal{M}(\mathrm{x}) = \sigma^2_{\mathbf{x}}(\mathrm{g}^{(-1)} \star \mathrm{y})$$

La primitive $\bar{\mathrm{x}}$ minimise donc $\sigma^2_{\mathbf{x}}(\mathrm{x})$ si et seulement si $\bar{\mathrm{z}} = \mathrm{g} \star \bar{\mathrm{x}}$ minimise $\sigma^2_{\mathbf{z}}(\mathrm{z})$, ce qui se réécrit $\mathbb{E}[\,\mathbf{z}\,] = \mathrm{g} \star \mathbb{E}[\,\mathbf{x}\,]$. De plus, les variances sont égales : $\sigma_{\mathbf{z}} = \sigma_{\mathbf{x}}$.
La même argumentation tient pour la stabilité de la primitive centrale d'ordre quelconque, ainsi que pour la moyenne empirique comme nous l'avons remarqué à la section (3.4.1). ■

Théorème 4.10 (Translation à gauche d'une transformation aléatoire)

$$\mathbb{E}[\,\mathrm{g} \circ \mathbf{f}\,] = \mathrm{g} \circ \mathbb{E}[\,\mathbf{f}\,] \qquad \textit{et} \qquad \mathbb{E}[\,\{\mathrm{g} \circ \mathrm{f}_i\}\,] = \mathrm{g} \circ \mathbb{E}[\,\{f_i\}\,] \tag{4.18}$$

Ce résultat tient également pour l'ensemble des primitives centrales d'ordre quelconques.

Preuve : Remplacer dans la preuve précédente l'action ($\star$) par la composition ($\circ$) et une primitive x par une transformation f. ■

Question ouverte 4.1 (Inversion d'une transformation aléatoire)
Quelles sont les relations entre ces ensembles moyens ?

$$\mathbb{E}\left[\ \mathbf{g}^{(-1)}\ \right] \stackrel{?}{=} \mathbb{E}[\,\mathbf{g}\,]^{(-1)} \qquad \textit{et} \qquad \mathbb{E}\left[\ \{\mathrm{g}_i^{(-1)}\}\ \right] \stackrel{?}{=} \mathbb{E}[\,\{\mathrm{g}_i\}\,]^{(-1)}$$

L'écriture des variances nous donne :

$$\sigma^2_{\mathbf{g}^{(-1)}}(\mathrm{f}) = \mathbb{E}\left[\ \mathrm{dist}(\mathbf{g}^{(-1)}, \mathrm{f})^2\ \right] = \int_{\mathcal{G}} \mathrm{dist}(\mathrm{g}^{(-1)}, \mathrm{f})^2 . p_{\mathbf{g}}(\mathrm{g}) . d_L\mathcal{G}(\mathrm{g})$$

mais il n'est pas évident d'en conclure quelque chose.

Question ouverte 4.2 (Translation à droite d'une transformation aléatoire)
Quelles sont les relations entre ces ensembles moyens ?

$$\mathbb{E}[\,\mathbf{g} \circ \mathrm{f}\,] \stackrel{?}{=} \mathbb{E}[\,\mathbf{g}\,] \circ \mathrm{f} \qquad \textit{et} \qquad \mathbb{E}[\,\{\mathrm{g}_i\} \circ \mathrm{f}\,] \stackrel{?}{=} \mathbb{E}[\,\{\mathrm{g}_i\}\,] \circ \mathrm{f}$$

L'écriture des variances nous donne (z est ici une transformation) :

$$\sigma^2_{(\mathbf{g} \circ \mathrm{f})}(\mathrm{z}) = \mathbb{E}\left[\ \mathrm{dist}(\mathbf{g} \circ \mathrm{f}, \mathrm{z})^2\ \right] = \int_{\mathcal{G}} \mathrm{dist}(\mathrm{g} \circ \mathrm{f}, \mathrm{z})^2 . p_{\mathbf{g}}(\mathrm{g}) . d_L\mathcal{G}(\mathrm{g})$$

mais cette fois-ci, la distance n'a pas de raison d'être invariante à droite et nous ne pouvons rien conclure.

Question ouverte 4.3 (Action d'une transformation aléatoire)
Quelles sont les relations entre ces ensembles moyens ?

$$\mathbb{E}[\,\mathbf{f} \star \mathrm{x}\,] \stackrel{?}{=} \mathbb{E}[\,\mathbf{f}\,] \star \mathrm{x} \qquad \textit{et} \qquad \mathbb{E}[\,\mathbf{f} \star \mathbf{x}\,] \stackrel{?}{=} \mathbb{E}[\,\mathbf{f}\,] \star \mathbb{E}[\,\mathbf{x}\,]$$

Si la primitive x est déterministe, la variance s'écrit :

$$\sigma^2_{(\mathbf{f}\star\mathrm{x})}(\mathrm{z}) = \int_{\mathcal{G}} \mathrm{dist}(\mathrm{f}\star\mathrm{x},\mathrm{z})^2.p_{\mathbf{f}}(\mathrm{f}).d_L\mathcal{G}(\mathrm{f})$$

et si elle est aléatoire :

$$\sigma^2_{(\mathbf{f}\star\mathbf{x})}(\mathrm{z}) = \int_{\mathcal{G},\mathcal{M}} \mathrm{dist}(\mathrm{f}\star\mathrm{x},\mathrm{z})^2.p_{\mathbf{f}}(\mathrm{f}).p_{\mathbf{x}}(\mathrm{x}).d_L\mathcal{G}(\mathrm{f}).d\mathcal{M}(\mathrm{x})$$

ce qui se peut encore s'écrire

$$\sigma^2_{(\mathbf{f}\star\mathbf{x})}(\mathrm{z}) = \int_{\mathcal{G}} \sigma^2_{\mathbf{x}}(\mathrm{f}^{(-1)}\star\mathrm{z}).p_{\mathbf{f}}(\mathrm{f}).d_L\mathcal{G}(\mathrm{f}) = \int_{\mathcal{M}} \sigma^2_{(\mathbf{f}\star\mathrm{x})}.p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x})$$

Une fois de plus, on ne peut pas conclure grand chose.

4.3 Propriétés et obtention des primitives moyennes

En reprenant l'idée de barycentre exponentiel de Emery, nous allons pouvoir caractériser les moyennes de Karcher d'une primitive aléatoire comme étant, entre autres, des points critiques de la variance. L'idée classique pour faire cela est de dire que la dérivée s'annule en ces points. Dans le cas d'une variété, cela se traduit par l'annulation de l'espérance vectorielle classique dans la représentation exponentielle en ce point. Dans le cas, plus intéressant pour nous, d'une variété homogène géodésiquement complète pour une métrique invariante donnée, nous n'utilisons que la **carte principale** définie à la section (3.5.4.3) comme le log de la variété à l'origine, mais nous pouvons transporter toutes les cartes exponentielles en ce point grâce à des transformations adaptés. Ceci permet d'écrire les résultats de manière synthétique dans une seule représentation, et mène directement à un algorithme itératif pour la détermination de la primitive moyenne.

Nous présentons les résultats pour le cas de la variété, mais il est clair qu'ils s'appliquent de la même façon au groupe de transformation $\mathcal{G}$ muni de la distance invariante à gauche. Il faut alors évidemment exprimer les transformations dans la carte principale : il suffit de remplacer la fonction de placement $f_{\vec{\mathrm{x}}}$ par $\vec{\mathrm{f}}$, $J(f_{\vec{\mathrm{x}}})$ par $J_L\vec{\mathrm{f}}$ et $d\mathcal{M}$ par $d_L\mathcal{G}$ pour obtenir les propriétés recherchées sur le groupe. Notons que si l'on ne s'intéresse qu'à la variété, on peut utiliser n'importe quelle représentation pour le groupe de transformation, du moment que son action sur la variété peut être exprimée dans la carte principale de celle-ci : $\vec{\mathrm{y}} = f\star\vec{\mathrm{x}}$ doit être calculable.

4.3.1 Caractérisation d'une valeur moyenne

Les optimums de la variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ sont caractérisés par une dérivée nulle par rapport à la primitive y. Nous pourrions développer cette condition directement dans la carte principale et simplifier les équations en utilisant les identités remarquables du théorème (3.3), mais il est plus élégant et plus rapide de suivre (Emery et Mokobodzki, 1991) et de demander un **gradient** nul.

4.3.1.1 Gradient d'une fonction à valeur réelle

Soit α une fonction à valeur réelle. Le gradient en un point est la direction dans laquelle cette fonction croît le plus vite. Sur une variété $\mathcal{M}$, on défini le gradient au point x d'une fonction α par :

$$\forall\ \partial_v \in T_{\mathrm{x}}\mathcal{M} \qquad \langle \nabla\alpha \mid \partial_v \rangle_{\mathrm{x}} = \partial_v\alpha$$

Cette définition correspond au gradient classique dans $\mathbb{R}^n$ même dans le cas d'une base non orthonormée et s'étend bien sûr en tout point de la variété. Le gradient est alors un champs de vecteur.

Si l'on exprime cette relation dans une carte x, on a : $\partial_v \alpha = \frac{\partial \alpha(x)}{\partial x}.v$ et $\langle \nabla\alpha \mid v \rangle_x = \nabla\alpha^{\mathrm{T}}.Q(x).v$. L'égalité de ces deux termes pour tout vecteur v entraîne donc la relation :

$$\frac{\partial \alpha}{\partial x}^{\mathrm{T}} = Q(x).\nabla\alpha \tag{4.19}$$

4.3.1.2 Gradient du carré de la distance

Observons tout d'abord que la distance entre deux primitives x et y est la longueur de la géodésique entre les points, et donc la norme du vecteur tangent $\partial_{\overrightarrow{\mathrm{xy}}} \in T_{\mathrm{x}}\mathcal{M}$, en accord avec la formule (3.36) : $\mathrm{dist}(\mathrm{x},\mathrm{y})^2 = \|\partial_{\overrightarrow{\mathrm{xy}}}\|^2$ où $\|.\|$ est la norme canonique de $\mathbb{R}^n$. La primitive x étant fixée, le gradient de $\alpha_{\mathrm{x}}(\mathrm{y}) = \mathrm{dist}(\mathrm{x},\mathrm{y})^2 = \|\partial_{\overrightarrow{\mathrm{xy}}}\|^2$ est alors classique :

$$\nabla\alpha_{\mathrm{x}}(\mathrm{y}) = 2.\partial_{\overrightarrow{\mathrm{xy}}} = -2.\partial_{\overrightarrow{\mathrm{yx}}}$$

La seconde expression est plus adaptée et l'on obtient en terme de dérivées partielles dans la carte principale :

$$\frac{\partial(\,\mathrm{dist}(\vec{\mathrm{x}},\vec{\mathrm{y}})^2)}{\partial\vec{\mathrm{y}}}^{\mathrm{T}} = -2.Q(\vec{\mathrm{y}}).\overrightarrow{\mathrm{yx}} = -2.J(f_{\vec{\mathrm{y}}})^{(\text{-T})}.(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}}) \tag{4.20}$$

4.3.1.3 Hessienne du carré de la distance

On peut aller plus loin et calculer les dérivées secondes : on a tout d'abord la dérivée croisée exacte (le second terme est obtenu par symétrie) :

$$\frac{\partial^2(\,\mathrm{dist}(\vec{\mathrm{x}},\vec{\mathrm{y}})^2)}{\partial\vec{\mathrm{x}}\,\partial\vec{\mathrm{y}}} = -2.J(f_{\vec{\mathrm{y}}})^{(\text{-T})}.\frac{\partial(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}})}{\partial\vec{\mathrm{x}}} = -2.J(f_{\vec{\mathrm{x}}})^{(\text{-T})}.\frac{\partial(f_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})}{\partial\vec{\mathrm{y}}} \tag{4.21}$$

Pour la matrice hessienne $H_{\vec{\mathrm{y}}}$, observons tout d'abord qu'en utilisant l'identité remarquable (3.40), on a :

$$J(f_{\vec{\mathrm{y}}})^{(\text{-T})}.(f_{\vec{\mathrm{x}}}^{(-1)} \star y) = -Q(\vec{\mathrm{y}}).\left(\frac{\partial(f_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})}{\partial\vec{\mathrm{y}}}\right)^{(-1)}.(f_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})$$

En négligeant les termes en $\ddot{y}.y$ devant les termes en $\dot{y}^2$ dans la dérivation (voir par exemple (Gill et al., 1981, section 4.7)), on obtient :

$$H_{\vec{\mathrm{y}}} = \frac{\partial^2(\,\mathrm{dist}(\vec{\mathrm{x}},\vec{\mathrm{y}})^2)}{\partial\vec{\mathrm{y}}^2} \simeq 2.Q(\vec{\mathrm{y}}).\left(\frac{\partial(f_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})}{\partial\vec{\mathrm{y}}}\right)^{(-1)}.\left(\frac{\partial(f_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})}{\partial\vec{\mathrm{y}}}\right) \simeq 2.Q(\vec{\mathrm{y}}) \tag{4.22}$$

Notons que ce résultat est également la matrice hessienne exacte du carré d'une distance euclidienne, même si le repère n'est pas euclidien ($Q \neq Id$).

4.3.1.4 Optimums de la variance

On considère maintenant une primitive aléatoire $\mathbf{x}$ et sa variance par rapport à un point fixe y

$$\sigma^2_{\mathbf{x}}(\mathrm{y}) = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y},\mathrm{z})^2.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z})$$

Les points critiques $\bar{\mathrm{x}}$ ont un gradient nul :

$$\left.\frac{\partial(\sigma^2_{\mathbf{x}}(\vec{\mathrm{y}}))}{\partial\vec{\mathrm{y}}}\right|^{\mathrm{T}}_{\vec{\mathrm{y}}=\bar{\mathrm{x}}} = 0 \qquad \Longleftrightarrow \qquad \nabla\sigma^2_{\mathbf{x}}(\bar{\mathrm{x}}) = 0 = -2.\int_{\mathcal{M}} \log_{\bar{\mathrm{x}}}(\mathrm{z}).p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z})$$

Notons que cette dernière intégrale a bien un sens puisque l'on intègre la fonction $\overrightarrow{\bar{\mathrm{x}}\mathrm{z}} = \log_{\bar{\mathrm{x}}}(\mathrm{z})$ à valeurs dans *l'espace vectoriel* $T_{\bar{\mathrm{x}}}\mathcal{M}$. En utilisant la carte principale, on obtient

$$\mathbf{E}\left[\,\log_{\bar{\mathrm{x}}}(\mathbf{x})\,\right] = J(f_{\bar{\mathrm{x}}}).\int_{\mathcal{M}} (f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathrm{z}}).p_{\mathbf{x}}(\vec{\mathrm{z}}).d\mathcal{M}(\vec{\mathrm{z}}) = 0$$

Considérons maintenant la primitive aléatoire $\mathbf{e} = \mathrm{f}^{(-1)}_{\bar{\mathrm{x}}} \star \mathbf{x}$ ou de manière équivalent, $\mathbf{x} = \mathrm{f}_{\bar{\mathrm{x}}} \star \mathbf{e}$. Alors, d'après le théorème (4.1), les densités sont reliées par $p_{\mathbf{x}}(z) = p_{\mathbf{e}}(\mathrm{f}^{(-1)}_{\bar{\mathrm{x}}} \star z)$, et la formule ci-dessus se simplifie en

$$\int_{\mathcal{M}} (f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathrm{z}}).p_{\mathbf{e}}(f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathrm{z}}).d\mathcal{M}(\vec{\mathrm{z}}) = 0$$

et comme la mesure est invariante, le changement de variable $\vec{\mathrm{y}} = f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathrm{z}}$ donne :

$$\int_{\mathcal{D}} \vec{\mathrm{y}}.p_{\mathbf{e}}(\vec{\mathrm{y}}).\frac{d\vec{\mathrm{y}}}{|J_L(f_{\vec{\mathrm{y}}})|} = 0 \qquad \Longleftrightarrow \qquad \mathbf{E}\left[\,\vec{\mathbf{e}}\,\right] = \int_{\mathcal{D}} \vec{\mathrm{y}}.\rho_{\vec{\mathbf{e}}}(\vec{\mathrm{y}}).d\vec{\mathrm{y}} = 0$$

en notant $\vec{\mathbf{e}}$ le *vecteur aléatoire* représentant la primitive aléatoire $\mathbf{e} = \mathrm{f}^{(-1)}_{\bar{\mathrm{x}}} \star \mathbf{x}$ dans la *carte principale*. Nous avons donc obtenu une caractérisation des barycentres exponentiels :

Théorème 4.11 (Caractérisation des primitives moyennes)
Une condition nécessaire (mais pas suffisante) pour qu'une primitive $\bar{\mathrm{x}}$ *soit une moyenne de Fréchet ou de Karcher de la primitive aléatoire* $\mathbf{x}$ *est que le vecteur aléatoire* $\vec{\mathbf{e}}$ *représentant la primitive aléatoire* $\mathbf{e} = \mathrm{f}^{(-1)}_{\bar{\mathrm{x}}} \star \mathbf{x}$ *dans la carte principale ait une espérance nulle au sens vectoriel classique.*

$$\bar{\mathrm{x}} \in \mathbb{E}\left[\,\mathbf{x}\,\right] \quad \Longrightarrow \quad \mathbf{E}\left[\,\vec{\mathbf{e}}\,\right] = \mathbf{E}\left[\,f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathbf{x}}\,\right] = \int_{\mathcal{D}} \vec{\mathrm{y}}.\rho_{\vec{\mathbf{e}}}(\vec{\mathrm{y}}).d\vec{\mathrm{y}} = 0 \tag{4.23}$$

La caractérisation est similaire pour le cas discret ou empirique :

$$\bar{\mathrm{x}} \in \mathbb{E}\left[\,\{\mathrm{x}_i\}\,\right] \quad \Longrightarrow \quad \mathbf{E}\left[\,\{\vec{\mathrm{e}}_i\}\,\right] = \mathbf{E}\left[\,\{f^{(-1)}_{\bar{\mathrm{x}}} \star \vec{\mathrm{x}}_i\}\,\right] = \frac{1}{n}\sum_i \vec{\mathrm{e}}_i = 0 \tag{4.24}$$

De manière plus générale, la condition nécessaire $\mathbf{E}\left[\,\log_{\bar{\mathrm{x}}}(\mathbf{x})\,\right] = 0$ exprime le fait que l'espérance de Fréchet ou de Karcher correspond à l'espérance classique dans la carte exponentielle centrée au point moyen $\bar{\mathrm{x}}$. L'utilisation de distances invariantes et d'une fonction de placement $\mathrm{f}_{\mathrm{x}} \in \mathcal{F}_{\mathrm{x}}$ nous permet de translater ces propriétés pour les exprimer à l'origine en n'utilisant qu'une seule carte : la carte principale. C'est un atout majeur pour la gestion informatique de ces variétés.

4.3.1.5 Exemples

Si l'on considère le groupe des translations, ou bien les points soumis aux translations ou même aux transformations rigides, on choisit comme fonction de placement la translation de l'origine au point et on a $f^{(-1)}_x = -x$. La caractérisation d'un point moyen $\bar{x}$ est :

$$\mathbf{E}\left[\,-\bar{x} + \boldsymbol{x}\,\right] = 0$$

qui n'est qu'une formulation alternative de la formule bien connue $\bar{x} = \mathbf{E}[\boldsymbol{x}]$.

Pour des rotations, la carte principale est le vecteur rotation pourvu d'un angle inférieur à π, et le vecteur rotation $\bar{r}$ correspondant à la rotation espérée satisfait :

$$\mathbf{E}\left[\,\bar{r}^{(-1)} \circ \boldsymbol{r}\,\right] = 0$$

Si le vecteur rotation espéré $\bar{r}$ est suffisamment centré dans la carte principale (i.e. si l'angle de la rotation est faible) et si la distribution est suffisamment rassemblée autour de cette valeur (il est souhaitable par exemple que $\bar{r} + \delta r$ ne dépasse pas les frontières du domaine de définition pour la plus grosse partie de la distribution), alors un développement limité au premier ordre dans l'intégrale définissant l'espérance donne

$$\bar{r}^{(-1)} \circ r = r - \bar{r} + O(\|\bar{r}\|^2) + O(\|\bar{r} - r\|^2)$$

et on obtient donc une espérance classique proche de l'espérance de Fréchet : $\mathbf{E}[\,\boldsymbol{r}\,] \simeq \bar{r}$.

Il faut cependant faire attention car ces condition ne sont pas aisément vérifiables en pratique, et plus on s'en éloigne, plus la différence est sensible : voir par exemple la figure (4.1). De plus, il n'est pas évident que ce genre d'approximation soit valide avec d'autres représentations.

4.3.2 Un algorithme pour obtenir la moyenne

Un algorithme classique pour déterminer un minimum est la descente de gradient. Comme nous avons vu que l'on pouvait calculer simplement le gradient du carré de la distance et que, de plus, on a un moyen canonique de revenir d'un plan tangent à un point de la variété (l'application exponentielle), cet algorithme itératif paraît tout à fait adapté.

Supposons que l'on ait, à un instant t, une estimation $\bar{\mathrm{x}}_t$ de la moyenne de la primitive aléatoire $\mathbf{x}$. Le gradient de la variance en ce point est :

$$\nabla\sigma_t^2 = \nabla\sigma^2_{\mathbf{x}}(\bar{\mathrm{x}}) = -2.\mathbf{E}\left[\,\log_{\bar{\mathrm{x}}_t}(\mathbf{x})\,\right] \quad \in \quad T_{\bar{\mathrm{x}}_t}\mathcal{M}$$

La dérivée est donc

$$\delta\sigma_t^2 = \frac{\partial {\sigma_t^2}^{\mathrm{T}}}{\partial \bar{\mathrm{x}}} = Q(\bar{\mathrm{x}}_t).\nabla\sigma_t^2 = -2.J(f_{\bar{\mathrm{x}}_t})^{(-\mathrm{T})}.\mathbf{E}\left[\,f^{(-1)}_{\bar{\mathrm{x}}_t} \star \vec{\mathbf{x}}\,\right]$$

et la matrice hessienne :

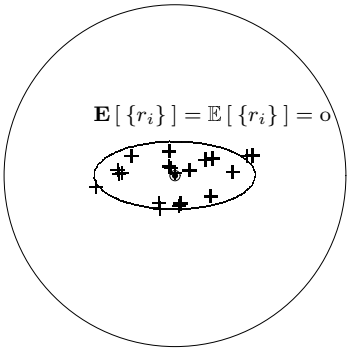
$$H_{\sigma_t^2} = \int_{\mathcal{M}} \left.\frac{\partial^2 \operatorname{dist}(\vec{\mathrm{y}}, \vec{\mathrm{z}})^2}{\partial \vec{\mathrm{y}}^2}\right|_{\mathrm{y}=\bar{\mathrm{x}}} .p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) = 2.Q(\bar{\mathrm{x}})$$

L'idée est donc d'avancer depuis $\bar{\mathrm{x}}_t$ en suivant la géodésique partant dans la direction opposée à $\delta\sigma_t^2$. Pour savoir de combien il faut avancer, on utilise l'approximation au second ordre qui consiste à pondérer ce vecteur par l'inverse de la matrice hessienne. On avance donc du vecteur :

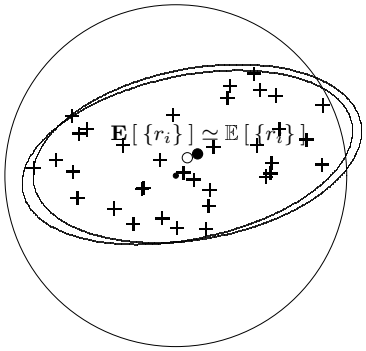
$$\overrightarrow{\delta\mathrm{x}_t} = -H^{(-1)}_{\sigma^2}.\delta\sigma^2 = J(f_{\bar{\mathrm{x}}_t}).\mathbf{E}\left[\,f^{(-1)}_{\bar{\mathrm{x}}_t} \star \vec{\mathbf{x}}\,\right] = \mathbf{E}\left[\,\log_{\bar{\mathrm{x}}_t}(\mathbf{x})\,\right]$$

Comme ce vecteur est donné dans l'espace tangent en $\bar{\mathrm{x}}_t$, il suffit d'utiliser l'exponentielle en ce point :

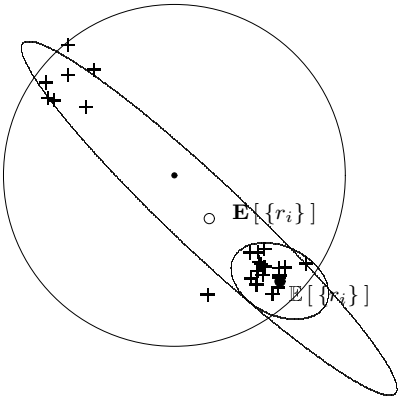
$$\bar{\mathrm{x}}_{t+1} = \exp_{\bar{\mathrm{x}}_t}\left(\overrightarrow{\delta\mathrm{x}_t}\right) = f_{\bar{\mathrm{x}}_t} \star \exp\left(J(f_{\bar{\mathrm{x}}_t})^{(-1)}.\overrightarrow{\delta\mathrm{x}_t}\right) = f_{\bar{\mathrm{x}}_t} \star \exp\left(\mathbf{E}\left[\,f^{(-1)}_{\bar{\mathrm{x}}_t} \star \vec{\mathbf{x}}\,\right]\right) \tag{4.25}$$



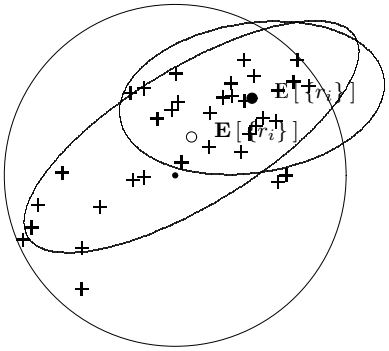
(a) L'espérance standard et celle de Fréchet sont identiques à l'origine.



(b) Quand l'espérance de Fréchet est près de l'origine, l'espérance classique peut être considérée comme une approximation au premier ordre.



(c) Près de la frontière du domaine, les deux espérances diffèrent significativement. Cependant, avec ce bruit relativement faible, il serait encore possible de détecter et prendre en compte cet effet de frontière.



(d) Avec un niveau de bruit plus élevé, il n'est plus possible de deviner la séparation permettant d'éviter l'effet de frontière.

FIG. 4.1 – *Comportement des rotations espérées et des matrices de covariance associées : projection dans le plan (r_x, r_y) des* **vecteurs rotation** *mesurés, de leur espérance classique et de Fréchet et des ellipsoïdes d'incertitude correspondant à $\chi^2 = 15$. Le cercle représente la frontière de la carte principale ($\theta = \|r\| = \pi$). Rappelons que lorsqu'on traverse cette frontière pour sortir de la carte principale au point $r = \pi.n$, on rentre en fait dans la carte par le point symétrique $r' = \pi.(-n)$. Notons également que la bonne façon de visualiser les matrices de covariance serait de ramener la rotation espérée au centre comme sur la figure (4.1(a)) pour que la représentation corresponde à la carte exponentielle en ce point.*

Si l'on considère maintenant le cas de la moyenne discrète ou empirique, nettement plus intéressant d'un point de vue pratique, le vecteur $\overrightarrow{\delta \mathrm{x}_t}$ est donné par

$$\overrightarrow{\delta \mathrm{x}_t} = \mathbf{E}\left[\ \{\log_{\bar{\mathrm{x}}_t}(\mathrm{x}_i)\}\ \right] = \frac{1}{n}\sum_i J(f_{\bar{\mathrm{x}}_t}).\left(f_{\bar{\mathrm{x}}_t}^{(-1)} \star \vec{\mathrm{x}}_i\right)$$

On avance donc de la moitié de ce vecteur dans $T_{\bar{\mathrm{x}}}\mathcal{M}$ et on revient à la variété, ce qui s'exprime aussi dans la carte principale par

$$\bar{\mathrm{x}}_{t+1} = \exp_{\bar{\mathrm{x}}_t}(\overrightarrow{\delta \mathrm{x}_t}) = f_{\bar{\mathrm{x}}_t} \star \exp\left(J(f_{\bar{\mathrm{x}}_t})^{(-1)}.\overrightarrow{\delta \mathrm{x}_t}\right)$$

Ceci qui se simplifie pour donner la formule d'évolution suivante *dans la carte principale* :

$$\bar{\mathrm{x}}_{t+1} = f_{\bar{\mathrm{x}}_t} \star \left(\frac{1}{n}\sum_i \left(f_{\bar{\mathrm{x}}_t}^{(-1)} \star \vec{\mathrm{x}}_i\right)\right) \tag{4.26}$$

Notons que dans le cas où l'action du groupe est linéaire ou affine *dans la carte principale*, l'action de la transformation est distributive vis-à-vis de l'addition, et l'évolution se simplifie en

$$x_{t+1} = \frac{1}{n}\sum_i x_i$$

On retrouve donc bien le barycentre ou moyenne classique dans un espace vectoriel. Qui plus est, on converge dans ce cas en une seule étape et de manière exacte. Cet algorithme est donc tout à fait adapté dans une optique « orientée objet » où l'on veut implémenter des algorithmes génériques qui fonctionnent sur toutes les primitives de la même façon. Les seules opérations qui diffèrent suivant le type de primitive considéré sont ici l'action du groupe sur les primitives (dans la carte principale) et la fonction de placement $f_{\bar{\mathrm{x}}}$.

Un avantage important de cette descente de gradient par rapport à d'autres algorithmes similaires est que nous n'avons aucun problème de projection du plan tangent vers la variété, grâce à la relation canonique qui relie les deux : l'exponentielle. De plus, même si le gradient nous fait sortir du domaine de définition de la carte, l'application exponentielle est définie sur le plan tangent tout entier (parce que la variété est supposée géodésiquement complète) et on peut donc toujours revenir sans problème. Dans la pratique, le problème ne se pose même pas puisque les domaines de définition des cartes principales que nous envisagerons sont convexes, comme le disque $\mathcal{B}_1(\pi)$ pour un vecteur unitaire ou la boule $\mathcal{B}_3(\pi)$ pour le vecteur rotation, et le barycentre est assuré de rester dans le domaine.

Un dernier point important pour finir cet algorithme itératif est de trouver un point de départ de préférence proche de la solution. On peut choisir l'une des primitives x_i au hasard, ou bien attribuer à chaque primitive sa distance moyenne (ou médiane pour être robuste) par rapport aux autres primitives et choisir celle qui minimise ce critère de centralité initiale. Cette méthode de choix du point de départ peut être randomisée de manière assez efficace (Huber, 1981; Rousseeuw et Leroy, 1987).

Le problème de l'unicité de la solution peut être résolu en répétant l'algorithme à partir de plusieurs primitives différentes pour vérifier l'obtention d'un minimum unique. Quand on connaît la courbure de la variété (si par exemple elle est constante ou majorée par κ), on peut aussi utiliser a posteriori le théorème (4.8) d'existence et d'unicité. En effet, dans la carte principale, toute boule centrée en zéro et ne rencontrant pas le lieu de coupure est une boule géodésique de centre l'origine et de même rayon sur la variété. Lorsqu'on a obtenu une moyenne $\bar{\mathrm{x}}$, on peut translater une boule

géodésique centrée sur ce point en une boule géodésique centrée à l'origine en considérant les résidus $\vec{\mathrm{e}}_i = f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i$. Si le maximum r de la norme de ces résidus est suffisamment faible :

$$r = \max_i \|\vec{\mathrm{e}}_i\| = \max_i \ \mathrm{dist}(\bar{\mathrm{x}}, \mathrm{x}_i) < \frac{\pi}{\sqrt{\kappa}}$$

la boule géodésique de centre $\bar{\mathrm{x}}$ et de rayon r est régulière et contient le support de la distribution discrète : il n'existe donc qu'un seul minimum, nécessairement celui qu'on a déjà trouvé.

4.3.2.1 Exemple : le repère moyen

Prenons comme exemple un ensemble de mesures de repères $f_i = (r_i, x_i)$. En utilisant le vecteur rotation et le vecteur translation comme représentation, on est dans la carte exponentielle mais pour simplifier les formules, on oubliera la flèche sur ces vecteurs. On cherche donc le repère moyen $\bar{f} = (\bar{r}, \bar{x})$ au sens de Karcher. La distance invariante (ou invariante à gauche si l'on considère que ce sont des transformations) est donne en section (7.4.1.1) :

$$\mathrm{dist}(\bar{f}, f_i) = N_\lambda(\bar{f}^{(-1)} \circ f_i) = N_\lambda(f_i^{(-1)} \circ \bar{f})$$

avec

$$N_\lambda(r,x)^2 = \lambda.\|r\|^2 + \|x\|^2 \qquad \text{et} \qquad \bar{f}^{(-1)} \circ f_i = \left(\bar{r}^{(-1)} \circ r_i \ ; \ \bar{r}^{(-1)} \star (x_i - \bar{x})\right)$$

On peut appliquer directement la formule (4.26) qui se simplifie en :

$$\bar{f}_{t+1} = \left(\bar{r}_t \circ \left(\frac{1}{n} \sum_i (\bar{r}_t^{(-1)} \circ r_i) \right) \ ; \ \frac{1}{n} \sum_i x_i \right)$$

La position moyenne $\bar{x}$ est donc le barycentre des positions, comme on pouvait s'y attendre intuitivement, et la minimisation pour la rotation est indépendante. La séparation entre la rotation et la translation (ou le trièdre et la position) était en effet visible directement sur le critère des moindres carrés :

$$s^2_\lambda(\bar{f}) = \lambda^2 \sum_i \|\bar{r}^{(-1)} \circ r_i\|^2 + \sum_i \|x_i - \bar{x}\|^2$$

Notons encore que la solution obtenue est indépendante du paramètre λ qui nous sert à comparer les angles de rotation avec les distances euclidiennes de points, ce qui est satisfaisant.

Nous avons également proposé dans (Pennec et Thirion, 1995) et nous verrons au chapitre (8.3) des méthodes similaires, mais utilisant les informations de deuxième ordre (les matrices de covariances).

4.4 Matrice de covariance

Nous avons obtenu jusqu'ici un équivalent de l'espérance, ou plutôt de la valeur centrale pour les primitives aléatoires, et un indice de dispersion : la variance par rapport à un point. Pour aller plus loin, observons que la matrice de covariance d'un vecteur aléatoire $\boldsymbol{x}$ par rapport à un point y représente la dispersion *directionnelle* du vecteur « différentiel » $\boldsymbol{x} - y$. En reprenant les notations vectorielles classiques, on noterait ce vecteur : $\overrightarrow{y\boldsymbol{x}}$, et la définition de la matrice de covariance du point aléatoire $\boldsymbol{x}$ par rapport à un point fixe y serait :

$$\Sigma_{\boldsymbol{x}\boldsymbol{x}}(y) = \mathbf{E}\left[\ \overrightarrow{y\boldsymbol{x}}.\overrightarrow{y\boldsymbol{x}}^{\mathrm{T}}\ \right] = \int_{\mathbb{R}^n} (\overrightarrow{yx}).(\overrightarrow{yx})^{\mathrm{T}}.p_{\boldsymbol{x}}(x).dx$$

En fait, c'est vraiment la notion de bipoint développée à l'origine par Grassman qui convient pour l'extension de la notion de vecteur aux variétés différentielles : si x et y sont deux point de la variété $\mathcal{M}$ (supposée suivre les hypothèses usuelles), on peut interpréter le vecteur $\overrightarrow{\mathrm{yx}} = \log_{\mathrm{y}}(\mathrm{x})$ définit précédemment comme un représentant du « bipoint » (y,x). Par contre, on ne peut plus faire abstraction du point de départ pour constituer un vecteur comme une classe d'équivalence des bipoints. Observons également que le vecteur $\overrightarrow{\mathrm{yx}}$ est en général contraint de rester dans le domaine de définition $\mathcal{D}(\mathrm{y})$ de la carte logarithmique en y, délimité par le lieu de coupure tangentiel $\mathcal{C}(\mathrm{y})$.

Pour en revenir à la matrice de covariance, remarquons que l'on peut exprimer la densité de probabilité du *vecteur aléatoire* $\overrightarrow{\mathrm{y}\boldsymbol{x}}$ dans la carte $\log_{\mathrm{y}}$ grâce à l'équation (4.3) ou utiliser la mesure invariante et comme le domaine de définition $\mathcal{D}(\mathrm{y})$ est étoilé par rapport à l'origine (et que le lieu de coupure est de mesure nulle), la définition exhibée ci-dessus a encore un sens :

$$\Sigma_{\mathbf{xx}}(\mathrm{y}) = \mathbf{E}\left[\, \overrightarrow{\mathrm{yx}}.\overrightarrow{\mathrm{yx}}^{\mathrm{T}} \,\right] = \int_{\mathcal{D}(\mathrm{y})} (\overrightarrow{\mathrm{yx}}).(\overrightarrow{\mathrm{yx}})^{\mathrm{T}}.p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x}) \tag{4.27}$$

Avec nos hypothèses, on peut toujours se ramener à la carte principale pour obtenir :

$$\Sigma_{\mathbf{xx}}(\mathrm{y}) = J(f_{\vec{\mathrm{y}}}).\mathbf{E}\left[\, (f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathbf{x}}).(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathbf{x}})^{\mathrm{T}} \,\right].J(f_{\vec{\mathrm{y}}})^{\mathrm{T}}$$

En fait, on s'intéresse en général à la matrice de covariance *centrée*, c'est-à-dire relative à la moyenne :

Définition 4.1 *Soit* $\mathbf{x}$ *une primitive aléatoire sur la variété* $\mathcal{M}$ *possédant les propriétés habituelles, et* $\bar{\mathrm{x}} \in \mathbb{E}[\,\mathbf{x}\,]$ *une primitive moyenne que l'on suppose unique pour simplifier les notations (dans le cas contraire, il faut conserver la moyenne de référence). On note* $\Sigma_{\mathbf{xx}}$ *et on appelle matrice de covariance de* $\mathbf{x}$ *la matrice de covariance par rapport à cette moyenne :*

$$\Sigma_{\mathbf{xx}} = \Sigma_{\mathbf{xx}}(\bar{\mathrm{x}}) = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{T} \,\right] = \int_{\mathcal{D}(\bar{\mathrm{x}})} (\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}).(\overrightarrow{\bar{\mathrm{x}}\mathrm{x}})^{T}.p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x})$$

En utilisant la carte principale et la primitive aléatoire $\mathbf{e} = \mathrm{f}_{\bar{\mathrm{x}}}^{(-1)} \star \mathbf{x}$ *que l'on peut interpréter comme l'erreur autour de l'origine, on a :*

$$\Sigma_{\mathbf{xx}} = J(f_{\bar{x}}).\Sigma_{\mathbf{ee}}.J(f_{\bar{x}})^{T} \qquad \text{où} \qquad \Sigma_{\mathbf{ee}} = \mathbf{E}\left[\, \vec{\mathbf{e}}.\vec{\mathbf{e}}^{T} \,\right] = \int_{\mathcal{D}} \vec{\mathrm{y}}.\vec{\mathrm{y}}^{T}.\rho_{\vec{\mathbf{e}}}(y).dy$$

La covariance empirique est définie de la même façon en utilisant la version discrète de l'espérance : si $\bar{\mathrm{x}}$ est la moyenne de l'ensemble de primitive $\{\mathrm{x}_i\}$, la covariance est

$$\Sigma_{\mathbf{xx}} = J(f_{\bar{\mathrm{x}}}).\left(\frac{1}{n}\sum_i (f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i).(f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i)^{\mathrm{T}}\right).J(f_{\bar{\mathrm{x}}})^{\mathrm{T}} \tag{4.28}$$

Notons que, comme dans le cas d'un vecteur aléatoire, la trace de la matrice de covariance est égale à la variance :

$$\mathrm{Tr}(\Sigma_{\mathbf{xx}}) = \mathbf{E}\left[\, \mathrm{Tr}(\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}}) \,\right] = \mathbf{E}\left[\, \mathrm{dist}(\bar{\mathrm{x}},\mathbf{x}) \,\right] = \sigma_{\mathbf{x}}^2$$

Ceci est bien sûr valide pour la covariance et la variance relatives à un point fixe autre que la moyenne.

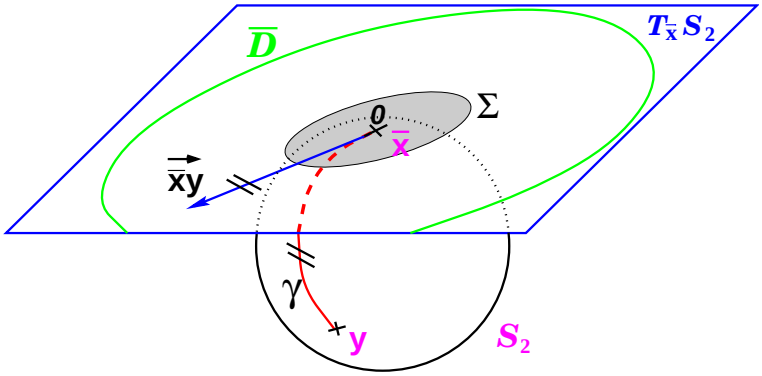


FIG. 4.2 – *La covariance est définie dans le plan tangent au point moyen comme la matrice de covariance classique du « vecteur déviation » :* $\Sigma_{\mathbf{xx}} = \mathbf{E}\left[\overrightarrow{\bar{x}\mathbf{x}} \,.\, \overrightarrow{\bar{x}\mathbf{x}}^T\right]$.

4.4.1 Approximation d'une primitive aléatoire

D'un point de vue informatique, nous avons maintenant suffisamment de paramètres pour approximer une primitive aléatoire $\mathbf{x}$: si $\bar{x}$ est la moyenne (supposée unique pour simplifier) et $\Sigma_{\mathbf{xx}}$ la covariance associée, on écrira comme dans le cas d'un vecteur aléatoire :

$$\mathbf{x} \sim (\bar{x}, \Sigma_{\mathbf{xx}})$$

Dans le cas où il y a plusieurs moyennes qui réalisent le minimum de la variance (global au sens de Fréchet ou local au sens de Karcher), les matrices de covariances n'ont aucune raison générique d'être les mêmes. On doit donc gérer l'ensemble des valeurs moyennes et des covariances associées. En pratique, il est rare que l'on ait à le faire.

4.4.2 Algèbre des primitives et transformations déterministes et aléatoires

Comme nous avons calculé la propagation des densités et des moyennes, nous nous attachons ici à la propagation des matrices de covariances dans les fonctions de base. Notons encore une fois que les propriétés « d'invariance » ou de cohérence avec l'action des transformations sont dues à l'utilisation d'une distance invariante et ne sont donc pas forcément valables s'il n'existe pas de distance invariante.

Théorème 4.12 (Action d'une transformation fixe sur une primitive aléatoire)
Soit $\mathbf{x} \sim (\bar{x}, \Sigma_{\mathbf{xx}})$ *une primitive aléatoire. L'action d'une transformation fixe est:*

$$\mathbf{y} = \mathrm{f} \star \mathbf{x} \sim (\mathrm{f} \star \bar{x}\,,\; J.\Sigma_{\mathbf{xx}}.J^T) \qquad \textit{avec} \qquad J = \left.\frac{\partial(\mathrm{f} \star \vec{x})}{\partial \vec{x}}\right|_{\vec{x}=\bar{x}} \tag{4.29}$$

Preuve : Puisque $\bar{y} = f \star \bar{x}$, il existe une transformation $h \in \mathcal{H}$ telle que $f_y = f \circ f_x \circ h$. Il suffit alors d'écrire dans la définition de $\Sigma_{\mathbf{yy}}$ le jacobien sous la forme composée :

$$J(f_{\bar{y}}) = \left.\frac{\partial(f_{\bar{y}} \star \vec{e})}{\partial \vec{e}}\right|_{\vec{e}=0} = \left.\frac{\partial((f \circ f_{\bar{x}} \circ h) \star \vec{e})}{\partial \vec{e}}\right|_{\vec{e}=0} = \left.\frac{\partial(\mathrm{f} \star \vec{x})}{\partial \vec{x}}\right|_{\vec{x}=\bar{x}} . \left.\frac{\partial(f_{\bar{x}} \star \vec{e})}{\partial \vec{e}}\right|_{\vec{e}=0} .J(h) = J.J(f_{\bar{x}}).J(h)$$

et $f_{\bar{y}}^{(-1)} \star \vec{y} = J(h)^{(-1)}.(f_{\bar{x}}^{(-1)} \star \vec{x})$ puisque l'action de $\mathcal{H}$ est linéaire dans la carte principale. En reportant dans la définition, on trouve le résultat. ■

Théorème 4.13 (Translation à gauche d'une transformation aléatoire)
Soit $\mathbf{f}_1 \sim (\bar{f}_1, \Sigma_{\mathbf{f_1 f_1}})$ *une transformation aléatoire. La translation à gauche par une transformation*

fixe est :

$$\mathbf{f}_2 = \mathrm{g} \circ \mathbf{f}_1 \sim \left(\mathrm{g} \circ \bar{\mathrm{f}}_1 \, , \; J.\Sigma_{\mathbf{f_1 f_1}}.J^T\right) \qquad \textit{avec} \qquad J = \left.\frac{\partial(\vec{\mathrm{g}} \circ \vec{\mathrm{f}})}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}_1} \tag{4.30}$$

Preuve : Remplacer dans la preuve précédente l'action ($\star$) par la composition ($\circ$) et une primitive x par une transformation f. Enlever également toute allusion à h puisque $\mathcal{H} = \{\,\mathrm{Id}\}$. ■

Les formules sont bien entendu aussi valides pour la propagation de la version empirique (ou statistique) de la matrice de covariance. Notons que dans ces deux cas, les formules de propagation de la moyenne et de la covariance sont **exactes**, comme dans le cas affine pour les vecteurs aléatoires (section 2.2.4.2).

Pour la propagation dans les autres opérations, nous n'avons pas été capable de déterminer dans quelles conditions les formules suivantes sont exactes. Nous nous contenterons donc de les justifier par l'approximation au 1$^{\mathrm{er}}$ ordre du théorème (2.1).

- **(Inversion d'une transformation aléatoire)**
  Soit $\mathbf{f} \sim (\bar{\mathrm{f}}, \Sigma_{\mathbf{ff}})$ une transformation aléatoire. La transformation inverse est donnée au 1$^{\mathrm{er}}$ ordre par aléatoire. La translation à gauche par une transformation fixe est :

$$\mathbf{f}^{(-1)} \sim \left(\bar{\mathrm{f}}^{(-1)} \, , \; J.\Sigma_{\mathbf{ff}}.J^{\mathrm{T}}\right) \qquad \text{avec} \qquad J = \left.\frac{\partial \vec{\mathrm{f}}^{(-1)}}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}} \tag{4.31}$$

- **(Translation à droite d'une transformation aléatoire)**
  Soit $\mathbf{f}_1 \sim (\bar{\mathrm{f}}_1, \Sigma_{\mathbf{f_1 f_1}})$ une transformation aléatoire. La translation à droite par une transformation fixe est :

$$\mathbf{f}_2 = \mathbf{f}_1 \circ \mathrm{g} \sim \left(\bar{\mathrm{f}}_1 \circ \mathrm{g} \, , \; J.\Sigma_{\mathbf{f_1 f_1}}.J^{\mathrm{T}}\right) \qquad \text{avec} \qquad J = \left.\frac{\partial(\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}_1} \tag{4.32}$$

- **(Action d'une transformation aléatoire sur une primitive fixe)**
  Soit $\mathbf{f} \sim (\bar{\mathrm{f}}, \Sigma_{\mathbf{ff}})$ une transformation aléatoire. L'action sur une primitive fixe est :

$$\mathbf{y} = \mathbf{f} \star \mathrm{x} \sim \left(\bar{\mathrm{f}} \star \mathrm{x} \, , \; J.\Sigma_{\mathbf{ff}}.J^{\mathrm{T}}\right) \qquad \text{avec} \qquad J = \left.\frac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}} \tag{4.33}$$

- **(Action d'une transformation aléatoire sur une primitive aléatoire)**
  Soit $\mathbf{f} \sim (\bar{\mathrm{f}}, \Sigma_{\mathbf{ff}})$ une transformation aléatoire et $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ une primitive aléatoire. L'action de la transformation sur la primitive produit une nouvelle primitive aléatoire :

$$\mathbf{y} = \mathbf{f} \star \mathbf{x} \sim \left(\bar{\mathrm{f}} \star \bar{\mathrm{x}} \, , \; J_{\vec{\mathrm{f}}}.\Sigma_{\mathbf{ff}}.J_{\vec{\mathrm{f}}}^{\mathrm{T}} + J_{\vec{\mathrm{x}}}.\Sigma_{\mathbf{xx}}.J_{\vec{\mathrm{x}}}^{\mathrm{T}}\right) \tag{4.34}$$

  avec

$$J_{\vec{\mathrm{f}}} = \left.\frac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\bar{\mathrm{f}}} \qquad \text{avec} \qquad J_{\vec{\mathrm{x}}} = \left.\frac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}}\right|_{\vec{\mathrm{x}}=\bar{\mathrm{x}}}$$

- **(Composition de deux transformations aléatoires)**
  Soit $\mathbf{f_1} \sim (\bar{\mathrm{f}}_1, \Sigma_{\mathbf{f_1 f_1}})$ et $\mathbf{f_2} \sim (\bar{\mathrm{f}}_2, \Sigma_{\mathbf{f_2 f_2}})$ deux transformations aléatoires leur composition est :

$$\mathbf{g} = \mathbf{f_2} \circ \mathbf{f_1} \sim \left(\bar{\mathrm{f}}_2 \circ \bar{\mathrm{f}}_1 \, , \; J_{\vec{\mathrm{f}}_2}.\Sigma_{\mathbf{f_2 f_2}}.J_{\vec{\mathrm{f}}_2}^{\mathrm{T}} + J_{\vec{\mathrm{f}}_1}.\Sigma_{\mathbf{f_1 f_1}}.J_{\vec{\mathrm{f}}_1}^{\mathrm{T}}\right) \tag{4.35}$$

avec

$$J_{\vec{f}_2} = \left.\frac{\partial(\vec{f}_2 \circ \vec{f}_1)}{\partial \vec{f}_2}\right|_{\vec{f}_2=\bar{\vec{f}}_2} \qquad \text{avec} \qquad J_{\vec{f}_1} = \left.\frac{\partial(\vec{f}_2 \circ \vec{f}_1)}{\partial \vec{f}_1}\right|_{\vec{f}_1=\bar{\vec{f}}_1}$$

4.5 Plusieurs primitives aléatoires

On considère maintenant que l'on a plusieurs mesures simultanées provenant de la même expérience aléatoire. Nous développons rapidement ici le cas de deux primitives aléatoires $\mathbf{x}$ et $\mathbf{y}$ à valeurs dans deux variétés $\mathcal{M}$ et $\mathcal{N}$ de dimensions quelconques. Ce paragraphe se généralise aisément à un nombre quelconque (mais fini) de primitives aléatoires.

On range ces deux mesures dans une primitive $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ appartenant à la variété produit $\mathcal{M} \times \mathcal{N}$. La mesure et la métrique sur cette variété produit est simplement le produit des mesures et des métriques et si les groupes agissant sur $\mathcal{M}$ et $\mathcal{N}$ sont notés $\mathcal{G}_1$ et $\mathcal{G}_2$, cette mesure et cette métrique sont invariantes pour le groupe produit (direct) $\mathcal{G}_1 \times \mathcal{G}_2$. On peut donc définir simplement la densité $p_{\mathbf{z}}(\mathrm{z}) = p_{(\mathbf{x},\mathbf{y})}(\mathrm{x}, \mathrm{y})$ (relativement à la mesure $d\mathcal{M}.d\mathcal{N}$). Comme dans le cas des vecteurs aléatoires, les densités marginales des primitives aléatoires $\mathbf{x}$ et $\mathbf{y}$ sont obtenues par intégration partielle :

$$p_{\mathbf{x}}(\mathrm{x}) = \int_{\mathcal{N}} p_{(\mathbf{x},\mathbf{y})}(\mathrm{x}, \mathrm{y}).d\mathcal{N}(\mathrm{y}) \qquad \text{et} \qquad p_{\mathbf{y}}(\mathrm{y}) = \int_{\mathcal{M}} p_{(\mathbf{x},\mathbf{y})}(\mathrm{x}, \mathrm{y}).d\mathcal{M}(\mathrm{x})$$

La moyenne de la mesure conjointe est

$$\bar{\mathrm{z}} \in \mathbb{E}[\,\mathbf{z}\,] = \mathbb{E}[\,\mathbf{x}\,] \times \mathbb{E}[\,\mathbf{y}\,] \quad \Longleftrightarrow \quad \bar{\mathrm{x}} \in \mathbb{E}[\,\mathbf{x}\,] \quad \text{et} \quad \bar{\mathrm{y}} \in \mathbb{E}[\,\mathbf{y}\,]$$

et, en supposant pour simplifier qu'il n'y a qu'une seule moyenne, la covariance est donnée par

$$\Sigma_{\mathbf{zz}} = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$$

où $\Sigma_{\mathbf{xy}} = \Sigma_{\mathbf{yx}}^{\mathrm{T}}$ est la matrice de covariance croisée

$$\Sigma_{\mathbf{xy}} = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{y}}\mathbf{y}}^{\mathrm{T}} \,\right]$$

où les vecteurs signifient que l'on utilise la carte logarithmique propre à chacune des variétés.

Mesures indépendantes Notons que la densité conjointe $p_{\mathbf{z}}(\mathrm{z})$ se factorise en

$$p_{(\mathbf{x},\mathbf{y})}(\mathrm{x}, \mathrm{y}) = p_{\mathbf{x}}(\mathrm{x}).p_{\mathbf{y}}(\mathrm{y})$$

si et seulement si les primitives aléatoires $\mathbf{x}$ et $\mathbf{y}$ sont indépendants, ce qui sera souvent supposé en statistiques. Il est aisé de voir que l'intégrale de la covariance croisée se factorise également dans ce cas en $\Sigma_{\mathbf{xy}} = \mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{x}}\mathbf{x}} \,\right].\mathbf{E}\left[\, \overrightarrow{\bar{\mathrm{y}}\mathbf{y}}^{\mathrm{T}} \,\right]$ et grâce à la caractérisation de la moyenne, chacune de ces intégrales est nulle : $\Sigma_{\mathbf{xy}} = 0$.

4.6 Discussion sur les aspects probabilistes

La donnée d'une mesure invariante sur le groupe ou la variété nous permet de définir sans aucun problème les densités de probabilité associées à des primitives aléatoires. De plus, l'invariance de cette mesure permet de calculer simplement la propagation de ces densités pour les opérations de bases sur les primitives et leurs transformations : composition, inversion et action. En ce qui concerne l'espérance, on ne peut par contre définir proprement que celle d'une observable, c'est-à-dire d'une fonction réelle, ou par extension vectorielle.

La définition d'une valeur moyenne pour une primitive aléatoire est donc nettement plus complexe que dans le cas vectoriel et nécessite une formulation variationelle basée sur la distance : l'espérance au sens de Fréchet ou de Karcher. Comme la moyenne est maintenant le résultat d'un problème de minimisation, l'existence et l'unicité ne sont pas assurés, sauf sous certaines conditions (théorème 4.8). Pour définir une matrice de covariance, nous devons faire intervenir la carte exponentielle au point moyen : la primitive aléatoire est alors représentée vectoriellement et est de plus centrée dans un ouvert étoilé et symétrique. La généralisation de la matrice de covariance classique ne pose alors pas de problème. De manière plus générale, on pourrait concevoir une matrice de covariance associée à d'autres définitions de la moyenne en utilisant la notion de connecteur introduite dans (Picard, 1994), dont la carte exponentielle n'est qu'un exemple.

L'utilisation de la distance invariante nous permet d'obtenir des formules de propagation exactes et cohérentes avec l'action d'une transformation déterministe (ce qui équivaut à changer de repère). Nous avons donc obtenu la stabilité recherché pour cette définition. Par contre, nous n'avons pas été capable de déterminer des formules exactes en ce qui concerne les autres opérations de base, et en particulier la translation à droite. Comme dans le cas des densités il a fallu faire intervenir le module du groupe (le rapport entre la mesure invariante à gauche et la mesure invariante à droite), on peut se demander s'il ne faut pas faire intervenir ici un rapport entre l'invariance à gauche et à droite de la distance. Lorsque le groupe est compact, cela ne pose pas de problème puisqu'il existe une distance bi-invariante (Spivak, 1979; Carmo, 1992), par exemple pour les rotations ou la sphère unitaire. Ceci est en général faux lorsque le groupe est seulement localement compact, par exemple pour les transformations rigides. D'un autre côté, les propriétés de stabilité peuvent être conservées avec une contrainte plus faible que l'invariance : si

$$\forall (\mathrm{x}, \mathrm{y}) \in \mathcal{M}^2 \qquad \mathrm{dist}(\mathrm{f} \star \mathrm{x}, \mathrm{f} \star \mathrm{y}) = \alpha(\mathrm{f}).\, \mathrm{dist}(\mathrm{x}, \mathrm{y})$$

alors les variances minimums ne sont pas conservées mais les primitives qui les réalisent le sont. On pourrait peut être rechercher une distance sur le groupe ayant ce type de propriété avec des fonction différentes α_L et α_R pour la composition à gauche et à droite, et en déduire des propriétés supplémentaires pour la propagation des moments dans les opérations de base.

Chapitre 5

Aspects statistiques

« Ne croyez qu'en les statistiques
que vous avez vous même falsifiées. »
Berlin-Est, Novembre 1989

Pour pouvoir appliquer les développements théoriques du chapitre précédent à des problèmes statistiques, et en particulier modéliser le bruit de mesure, nous avons besoin de supposer que plusieurs mesures en des point différents de la variété sont « identiques ». La question que nous nous poserons à la section (5.1) est : qu'est-ce qu'une distribution identique et indépendante ? Ceci conduira aux modèles de bruits homogènes et isotropes.

Pour d'autres applications statistiques, la connaissance de la moyenne et de la covariance n'est pas suffisante : il nous faut supposer une densité de probabilité. Comme dans le cas vectoriel, on peut définir l'information d'une distribution et chercher la densité (de moyenne et de covariance fixée) qui minimise cette information. Nous définirons ainsi à la section (5.2) l'équivalent de la distribution gaussienne sur notre variété.

Afin de conclure notre ensemble d'opérations sur les primitives et les transformations aléatoires, il nous reste encore à définir dans la section (5.3) une distance statistique entre primitives déterministes et aléatoires ou entre primitives aléatoires : c'est la distance de Mahalanobis. En utilisant la loi gaussienne précédemment déterminée, on peut alors envisager un test similaire au χ^2 pour rejeter les mesures aberrantes.

Pour finir ce chapitre, nous résumerons les résultats importants des chapitres précédents, en essayant de dégager clairement la démarche mathématique à suivre pour obtenir les opérations de base sur les primitives, et les algorithmes génériques sur les primitives géométriques qui s'expriment alors à partir de ces opérations.

5.1 Modèles de bruit

Pourquoi doit-on supposer que les bruits de mesures sur différentes primitives sont « identiques », ou plutôt suivent une loi similaire ? On pourrait se passer de cette hypothèse si l'on pouvait répéter un grand nombre de fois la mesures de chaque primitive géométrique. Dans la réalité, c'est très

difficile, en particulier en imagerie médicale : on ne peut pas faire passer 50 fois un même patient au scanner ou à l'IRM, d'autant plus qu'il faudrait conserver les mêmes paramètres et le même positionnement. Il nous faut donc supposer que l'ensemble des primitives que l'on mesure dans une acquisition est corrompue par le même type de bruit, ce qui est par ailleurs sans doute justifié au regard de la variation des paramètres entre divers acquisitions. Nous avons vu à la section (2.3.5) qu'un bruit additif identique sur les représentations des primitives ne convenait pas, ce qui devrait, en ce point du manuscrit, sembler évident.

Pour comprendre quelle modélisation de l'erreur nous devons adopter, examinons tout d'abord ce qu'est une mesure : soit x une primitive exacte. A cause du bruit, sa mesure est corrompue et nous ne pouvons mesurer que la primitive aléatoire $\mathbf{x}$ avec la densité de probabilité $p_{\mathbf{x}}(\mathrm{y})$. Cette densité est en général différente pour chaque point exact x de la variété. La description du processus d'erreur pour toute la variété est donc un champ de primitives aléatoires $\mathbf{x}(\mathrm{x})$ de densité $p_{\mathbf{x}(\mathrm{x})}(\mathrm{y})$: sachant que l'on doit mesurer x, nous observerons une réalisation $\hat{\mathrm{x}}$ de la primitive aléatoire $\mathbf{x}(\mathrm{x})$ avec la probabilité $p_{\mathbf{x}(\mathrm{x})}(\hat{\mathrm{x}})$. Un processus de bruit est entièrement décrit par un tel champ.

5.1.1 Processus homogènes

L'idée de base pour définir une distribution identique en différents points de la variété est que, dans le cadre géométrique, on ne peut comparer des « objets » que par l'intermédiaire de transformations : une primitive aléatoire $\mathbf{x}$ est donc similaire en terme de bruit à une primitive aléatoire $\mathbf{y}$ s'il existe une transformation de $\mathcal{G}$ qui permet d'identifier leur densités. Le processus de bruit tout entier sur la variété $\mathcal{M}$ est dit **homogène** si le bruit de mesure de tout point x de $\mathcal{M}$ est similaire au bruit de mesure à l'origine (et donc similaire au bruit de mesure de tout autre points y).

5.1.1.1 Exemple sur les points dans le plan

Nous avons représenté dans la figure (5.1) les ellipsoïdes d'incertitude du bruit de mesure de différents points dans le plan. En utilisant l'approximation au second ordre de la distribution (que l'on suppose centrée), la densité est caractérisée par la matrice de covariance et cet ellipsoïde représente toute l'information dont nous disposons : deux bruits sont similaires si leurs ellipsoïdes d'incertitudes sont congruents, c'est-à-dire superposables au moyen d'une transformation rigide. Dans l'exemple de la figure (5.1), les bruits de mesures des X_i sont similaires, mais les bruits sur Y_1 et Y_2 sont différents de tous les autres.

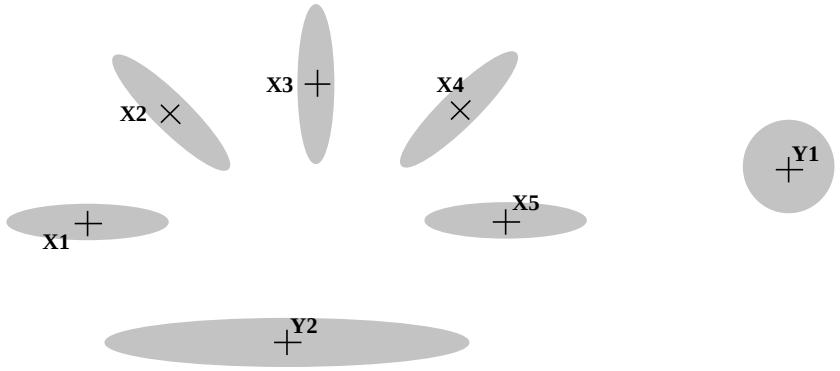


FIG. 5.1 – *Les distributions du bruit sur les points X_i sont similaires (pour les transformation rigides) mais les bruits sur les points Y_1 et Y_2 n'ont de similitude avec aucun autre.*

5.1.1.2 Formalisation mathématique

Soit $p_{\mathbf{e}}$ la densité d'une primitive aléatoire $\mathbf{e} = \mathbf{x}(\mathrm{o})$ mesurant l'origine o et $\mathrm{f_x} \in \mathcal{F}_\mathrm{x}$ une transformation qui amène l'origine au point x. La densité de probabilité de $\mathbf{x} = \mathrm{f_x} \star \mathbf{e}$ (mesurant $\mathrm{x} = \mathrm{f_x} \star \mathrm{o}$) est donnée par l'équation (4.4) : $p_{(\mathrm{f_x} \star \mathbf{e})}(\mathrm{y}) = p_{\mathbf{e}}(\mathrm{f_x^{(-1)}} \star \mathrm{y})$. La distribution du bruit $p_{\mathbf{x}(\mathrm{x})}$ au point x est donc similaire à la densité du bruit à l'origine s'il existe une transformation $\mathrm{f_x} \in \mathcal{F}_\mathrm{x}$ telle que $p_{\mathbf{x}(\mathrm{x})}(\mathrm{y}) = p_{(\mathrm{f_x} \star \mathbf{e})}(\mathrm{y})$.

Définition 5.1 Bruit homogène
Un processus de bruit $p_{\mathbf{x}(\mathrm{x})}$ sur une variété $\mathcal{M}$ est homogène si

$$\forall \mathrm{x} \in \mathcal{M}\ ,\ \exists \mathrm{f_x} \in \mathcal{F}_\mathrm{x} \quad \textit{tel que} \quad p_{\mathbf{x}(\mathrm{x})}(\mathrm{y}) = p_{\mathbf{e}}\left(\mathrm{f_x^{(-1)}} \star \mathrm{y}\right)$$

En revenant au niveau des primitives aléatoires, on peut dire que

Théorème 5.1 *Un processus de bruit homogène $\mathbf{x}(\mathrm{x})$ sur $\mathcal{M}$ est entièrement déterminé par la primitive aléatoire $\mathbf{e}$ qui mesure l'origine, de densité $p_{\mathbf{e}}$, et par une fonction de placement $\mathrm{f_x} : \mathrm{x} \in \mathcal{M} \mapsto \mathrm{f_x} \in \mathcal{F}_\mathrm{x}$ qui « met en place » l'erreur de mesure du point* x.

5.1.1.3 Exemple avec les repères et les transformations rigides

Dans ce cas particulier, les primitives et les transformations sont identifiées, les cosets sont réduits à des singletons et la fonction de placement $\mathrm{f_g}$ est l'identité ($\mathrm{f_g} = \mathrm{g}$). On peut donc écrire un processus de bruit homogène sur les repères comme suit : soit $\mathbf{e}$ la primitive aléatoire « mesure de l'origine ». Alors la primitive aléatoire mesurant f est $\mathbf{f} = \mathrm{f} \circ \mathbf{e}$. C'est le modèle de bruit « compositif » que nous avons proposé dans (Pennec et Thirion, 1995).

5.1.1.4 Choix de la fonction de placement

Les repères sont un cas très spécifiques puisque le groupe d'isotropie $\mathcal{H}$ est réduit à l'identité. Pour les autres types de primitives, les cosets ne sont plus des singletons et de multiples choix sont alors possibles pour la fonction de placement $\mathrm{f_x}$. Une contrainte raisonnable est la continuité (ou la différentiabilité) de la densité $p_{\mathbf{x}(\mathrm{x})}(\mathrm{y})$ par rapport aux primitives x et y. Cependant, ce n'est souvent pas suffisant pour obtenir une fonction de placement unique.

Prenons encore une fois l'exemple des points et approchons la densité au second ordre par la matrice de covariance [1] : le bruit additif standard est obtenu avec la translation comme fonction de placement ($\mathrm{f}_x = (0, x)$) et n'importe quelle matrice de covariance $\Sigma_{\boldsymbol{ee}} = \mathbf{E}\left[\, \boldsymbol{e}.\boldsymbol{e}^{\mathrm{T}} \,\right]$ pour la mesure de l'origine. On a alors le processus de bruit $\boldsymbol{x}(x) = \mathrm{f}_x \star \boldsymbol{e} = x + \boldsymbol{e}$. Cependant, on peut concevoir d'autres fonctions de placement continues, qui amènent à considérer des modèles de bruit étranges mais possibles : nous en avons exhibé un exemple dans la figure (5.2).

Une autre idée pour restreindre le choix de la fonction de placement est d'imposer une contrainte plus forte basée sur l'invariance : c'est ce qui amène aux modèles de bruit isotropes.

5.1.2 Modèles de bruit isotropes ou invariants

« Isotrope » est un adjectif habituellement utilisé pour les modèles de bruit sur les points qui sont invariants par rotation (et translation). Nous généralisons cette notion aux variété homogènes en imposant que le processus de bruit soit globalement invariant sur la variété par l'action de n'importe

1. L'approximation par la matrice de covariance ne simplifie que la densité et n'interfère aucunement avec le choix de la fonction de placement.

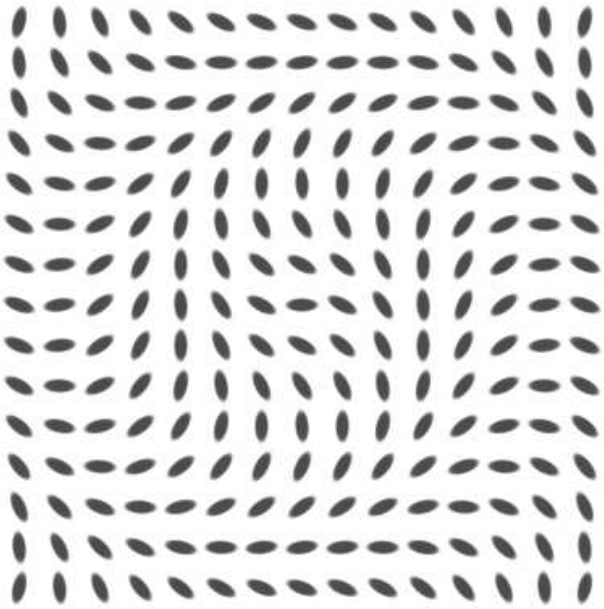


FIG. 5.2 – *Un modèle de bruit homogène dans le plan (représenté par les ellipsoïdes d'incertitude) induit par la fonction de placement* $\mathrm{f}_x = (R(x), x)$ *où* $R(x)$ *est la rotation 2D d'angle* $\theta = \|x\|\ [2\pi]$.

quelle transformation du groupe. Cela s'exprime par l'invariance du champ de densités $p_{\mathbf{x}(\mathrm{x})}$ par une transformation f quelconque :

$$p_{\mathbf{x}(\mathrm{f}\star\mathrm{x})}(\mathrm{f}\star\mathrm{y}) = p_{\mathbf{x}(\mathrm{x})}(\mathrm{y}) = p_{\mathbf{x}(\mathrm{x})}(\mathrm{f}^{(-1)}\star(\mathrm{f}\star\mathrm{y})) = p_{\mathrm{f}\star\mathbf{x}(\mathrm{x})}(\mathrm{f}\star\mathrm{y})$$

En prenant $\mathrm{f} = \mathrm{f_z} \in \mathcal{F}_\mathrm{z}$ et $\mathrm{x} = \mathrm{o}$, on obtient la condition :

$$\forall \mathrm{f_z} \in \mathcal{F}_\mathrm{z} \qquad \mathbf{x}(\mathrm{z}) = \mathbf{x}(\mathrm{f_z}\star\mathrm{o}) = \mathrm{f_z}\star\mathbf{x}(\mathrm{o}) = \mathrm{f_z}\star\mathbf{e}$$

En particulier, on a pour l'origine :

$$\forall \mathrm{h} \in \mathcal{H} \quad p_\mathbf{e}(\mathrm{h}\star\mathrm{y}) = p_\mathbf{e}(\mathrm{y})$$

Au niveau des primitives aléatoires, cela se résume par

Définition 5.2 Bruit isotrope
Un bruit isotrope est un bruit homogène dont l'erreur de mesure de l'origine $\mathbf{e}$ *est invariante sous l'action du groupe d'isotropie* $\mathcal{H}$ *:* $\mathrm{h}\star\mathbf{e} = \mathbf{e}$

Notons qu'un bruit isotrope est indépendant de la fonction de placement $\mathrm{f_x}$ choisie. La seule caractéristique d'un tel bruit est la densité de probabilité de mesurer e lorsqu'on cherche à mesurer l'origine, soumise à la condition d'invariance ci-dessus.

5.1.2.1 Exemple sur les points

Dans ce cas, $\mathrm{h}\star\boldsymbol{e} = R.\boldsymbol{e}$ où R est une matrice de rotation quelconque, et on a donc $p_{\boldsymbol{e}}(y) = p_{\boldsymbol{e}}(R\star y)$.Cela signifie que le bruit (autour de l'origine $x = 0$) est invariant par rotation et donc isotrope au sens classique. On peut aller de l'origine au point x par n'importe quelle transformation $\mathrm{f}_x \in \mathcal{F}_x$ et en particulier avec la translation de vecteur x. La mesure $\boldsymbol{x}$ de x est donc : $\boldsymbol{x} = x + \boldsymbol{e}$, ce qui est un bruit additif. Cependant, l'addition représente ici la composition des translations.

Si nous nous restreignons à l'approximation de la densité au second ordre (la matrice de covariance), la propriété d'invariance implique que $\Sigma_{\boldsymbol{ee}} = \mathbf{E}\,[\,\boldsymbol{e}.\boldsymbol{e}^{\mathrm{T}}\,] = \sigma^2\,Id$, ce qui est le bruit isotrope classique sur les points.

5.1.3 Choix du modèle de bruit

En l'absence de connaissance sur la formation de l'image et le processus d'extraction des primitives, le bruit le plus simple que l'on puisse utiliser est le bruit isotrope. Dans ce cas, la densité $p_{\mathbf{e}}$ est contrainte à être invariante sous l'action du groupe d'isotropie $\mathcal{H}$. Si l'on veut capturer une anisotropie, nous devons utiliser un modèle de de bruit homogène et choisir une fonction de placement. En vision par ordinateur, la plupart des primitives sont extraites d'images qui sont des tableaux réguliers de pixels (ou de voxels en 3D) et il semble raisonnable de choisir si c'est possible un modèle de bruit invariant par *translation*. Dans le cas général, il peut être possible de supposer l'invariance du bruit par un sous-groupe du groupe de transformation.

Dans le cas où la variété est identifiée avec le groupe, comme pour les points en translation ou les repères avec les transformations rigides, les bruit homogènes sont automatiquement isotropes car la fonction de placement est unique et il n'y a pas de contrainte sur la matrice de covariance à l'origine.

5.1.4 Relation avec le bruit additif

Il est intéressant de noter que pour les rotations 2D (avec la représentation angulaire) ou pour les translations (i.e. les points), la composition correspond à l'addition (éventuellement modulo $2.\pi$). Dans ces cas, les bruits homogènes et additifs sont identiques, ce qui induit une intuition fausse pour les cas plus complexes. Plus généralement, le modèle de bruit additif correspond au modèle homogène dès que l'action du groupe de transformation sur les primitives (ou sur lui-même) est linéaire *dans la représentation considérée*, auquel cas l'espace tangent correspond à la variété. Il existe alors une matrice F pour chaque transformation f telle que $\mathrm{f} \star x = F.x$ (ou $f \circ g = F.g$) et le jacobien $\frac{\partial \mathrm{f} \star x}{\partial x} = F$ est donc indépendant de la position x de la primitive.

L'adéquation du modèle de bruit homogène (ou « compositif ») sur les primitives de type repère sera montrée à la section (9.3.3) avec l'analyse du bruit de mesure estimé sur les points extrémaux.

5.2 Distribution « gaussienne »

Les développements de cette section ne doivent pas être considérés comme l'unique façon de généraliser la loi gaussienne aux variétés ni comme la meilleure façon de le faire, mais plutôt comme l'une des nombreuses pistes qui se présentent. En effet, il est difficile de trouver parmi les très nombreuses propriétés de la gaussienne celles qui pourront servir efficacement de base pour la généralisation.

L'approche que nous présentons ici est basée sur la minimisation de l'information et permet de définir dans quelques cas simples les équations explicites de la « gaussienne » sur une variété. Cela donne une bonne idée de ce qu'elle peut être sur des variétés plus complexes et montre que l'on peut approximer cette distribution par la gaussienne usuelle pour des covariances faibles.

5.2.1 Information et loi uniforme

Puisque nous pouvons intégrer sans problème une fonction à valeur réelle, nous pouvons étendre la définition de l'entropie à une primitive aléatoire :

$$\mathbf{H}[\,\mathbf{x}\,] = -\mathbf{E}[\,\log(p_{\mathbf{x}}(\mathbf{x}))\,] = -\int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{x})).p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x}) \tag{5.1}$$

ou de manière équivalente l'information :

$$\mathbf{I}[\,\mathbf{x}\,] = -\mathbf{H}[\,\mathbf{x}\,] = \int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{x})).p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x})$$

Cette définition est bien raisonnable puisque la densité $p_{\mathcal{U}}$ qui minimise l'information si l'on sait seulement que la mesure est dans le compact $\mathcal{U}$ est la densité uniforme sur ce compact :

$$p_{\mathcal{U}}(\mathrm{x}) = \mathbf{1}_{\mathcal{U}}(\mathrm{x}) \Big/ \int_{\mathcal{U}} d\mathcal{M}(\mathrm{y})$$

Si l'on veut exprimer l'information de la primitive aléatoire dans une carte, la densité est selon l'équation (4.3) : $\rho_{\boldsymbol{x}}(y) = p_{\mathbf{x}}(y)/\,|J(f_y)|$. En tenant compte de l'expression de la mesure invariante, l'information s'écrit donc

$$\mathbf{I}[\,\mathbf{x}\,] = \int_{\mathcal{D}} \log(\rho_{\boldsymbol{x}}(x)).\rho_{\boldsymbol{x}}(x).dx + \int_{\mathcal{D}} \log(|J(f_x)|).\rho_{\boldsymbol{x}}(x).dx$$

On peut interpréter le premier terme comme l'information « classique » que l'on aurait pu définir hâtivement directement dans la carte, mais le second terme nous montre qu'une fois de plus, cette définition aurait dépendu de la carte choisie. En effet, ce terme exprime le fait que la mesure uniforme n'est pas la mesure de Lebesgue dans la carte et mesure en quelque sorte la localisation de l'information de notre distribution par rapport à la mesure uniforme sur la variété.

Une propriété tout à fait centrale de notre notion d'information découle de l'utilisation de la mesure invariante :

Théorème 5.2 (Invariance de l'information)
Soit $\mathbf{x}$ *une primitive aléatoire et* f *une transformation déterministe. Alors l'information de la primitive aléatoire* $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ *est égale à celle de* $\mathbf{x}$ *:*

$$\mathbf{I}[\,\mathrm{f} \star \mathbf{x}\,] = \mathbf{I}[\,\mathbf{x}\,]$$

Preuve : La densité de $\mathbf{y}$ est donnée par le théorème (4.4) :

$$p_{\mathbf{y}}(\mathrm{y}) = p_{(\mathrm{f}\star\mathbf{x})}(\mathrm{y}) = p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y})$$

En reportant dans la définition de l'information, on a :

$$\mathbf{I}[\,\mathbf{y}\,] = \int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y})).p_{\mathbf{x}}(\mathrm{f}^{(-1)} \star \mathrm{y}).d\mathcal{M}(\mathrm{y})$$

En faisant le changement de variable $\mathrm{y} = \mathrm{f} \star \mathrm{x}$, et vu que la mesure est invariante, on obtient

$$\mathbf{I}[\,\mathbf{y}\,] = \int_{\mathcal{M}} \log(p_{\mathbf{x}}(\mathrm{x})).p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x}) = \mathbf{I}[\,\mathbf{x}\,]$$

■

5.2.2 Loi gaussienne

Si l'on suppose maintenant que l'on connaît la moyenne et la covariance d'une primitive aléatoire : $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma)$, on peut chercher comme dans le cas réel ou vectoriel la densité qui minimise l'information conditionnelle

$$\mathbf{I}\left[\,\mathbf{x} \,|\, \mathbb{E}\left[\,\mathrm{x}\,\right] = \bar{\mathrm{x}},\ \Sigma_{\mathbf{xx}} = \Sigma\,\right]$$

Avant de chercher à former le lagrangien, observons que comme l'information est invariante par action d'une transformation fixe, on peut rechercher la densité de la primitive $\mathbf{e} = \mathrm{f}_{\mathrm{x}}^{(-1)} \circ \mathbf{x}$ minimisant l'information associée, ce qui signifie que l'on peut toujours se ramener à la recherche d'une densité de moyenne nulle[2]. Ceci permet de simplifier grandement les calculs et de ne considérer que des distributions qui seront centrées dans la carte principale.

On considère donc pour la dérivation de la densité que $\bar{\mathrm{x}} = \mathrm{o}$. Rappelons que les contraintes à vérifier sont les suivantes, en utilisant pour simplifier la densité ρ correspondant à p dans la carte principale :

- C'est une densité de probabilité normalisée : $\quad p(\mathrm{x}) \geq 0 \quad$ et $\quad \displaystyle\int_{\mathcal{D}} \rho(\vec{\mathrm{x}}).d\vec{\mathrm{x}} = 1,$
- de moyenne unique l'origine, donc caractérisée par : $\quad \mathbf{E}\left[\,\vec{\boldsymbol{x}}\,\right] = \displaystyle\int_{\mathcal{D}} \vec{\mathrm{x}}.\rho(\vec{\mathrm{x}}).d\vec{\mathrm{x}} = 0,$
- et de covariance Σ : $\quad \mathbf{E}\left[\,\vec{\boldsymbol{x}}.\vec{\boldsymbol{x}}^{\mathrm{T}}\,\right] = \displaystyle\int_{\mathcal{D}} \vec{\mathrm{x}}.\vec{\mathrm{x}}^{\mathrm{T}}.p(\vec{\mathrm{x}}).d\vec{\mathrm{x}} = \Sigma_{\mathbf{xx}}$

Vu les contraintes, nous allons écrire le lagrangien dans la carte principale et non pas directement dans la variété. Il y a donc des contraintes à rajouter sur le bord du domaine : la valeur de la densité ρ doit être identique pour les points du lieu de coupure tangentiel $\mathcal{C} = \partial\mathcal{D}$ qui correspondent au même point physiquement sur la variété : si $\mathrm{x} \in C(\mathrm{o})$ est un point du lieu de coupure sur la variété, l'ensemble des points correspondants sur le bord de la carte principale (le lieu de coupure tangentiel) est constitué d'au moins deux points (il y a au moins deux géodésiques minimisantes qui se rencontrent en x), et les conditions aux limites s'écrivent donc :

$$\forall \vec{\mathrm{x}} \in \log(\mathrm{x}) \qquad p(\mathrm{x}) = \mathrm{Cte} = \rho(\vec{\mathrm{x}}).|J(f_{\vec{\mathrm{x}}})|$$

En pratique, on peut envisager deux cas :

- il y a deux géodésiques seulement qui se rencontrent en ce point du lieu de coupure, et en ce cas elle partent avec des vecteurs tangent opposés, ce qui veut dire que les conditions aux limites sont du style $\rho(\vec{\mathrm{x}}) = \rho(-\vec{\mathrm{x}})$. Dans ce cas, nous pouvons oublier d'inclure la contrainte : elle sera automatiquement vérifiée grâce à la symétrie de la matrice de covariance. Nous verrons un exemple avec le cercle $\mathcal{S}_1$, mais c'est aussi le cas des rotations $\mathcal{S}O_3$ et plus généralement des espaces projectifs.
- Il y a une infinité de géodésiques qui se rencontrent en ce point. C'est le cas par exemple sur la sphère $\mathcal{S}_2$ où toutes les géodésiques partant de l'origine (le pôle nord) se rencontrent à une distance de π au pôle sud : la valeur de la densité ρ doit être identique sur tout le cercle de rayon π autour du domaine de définition.

  Nous conjecturons que les conditions aux limites de ce type sont automatiquement vérifiées *dans le cas d'une matrice de covariance isotrope*, mais il faut absolument les inclure dans le lagrangien dans le cas général.

2. Nous ne considérons ici que le cas où la moyenne existe et est unique.

Le second cas nécessiterait une étude plus poussée et sans doute plus rigoureuse. Nous ne développons ici que le premier cas (on ne gère pas de conditions aux limites), ce qui permet de donner une idée générale de la distribution gaussienne sur une variété. Dans ce qui suit, log et exp représentent le logarithme et l'exponentielle classique sur $\mathbb{R}$. Le lagrangien s'écrit donc

$$\Lambda(\rho) = \int_{\mathcal{D}} \left(\rho.\log(\rho) + \rho.\log(|J_{\vec{\mathrm{x}}}|) + \alpha.\rho + \beta^{\mathrm{T}}.\vec{\mathrm{x}}.\rho + \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2}.\rho \right) .d\vec{\mathrm{x}}$$

Le lagrangien est stationnaire si la dérivée est nulle :

$$\frac{\partial \Lambda(\rho)}{\partial \rho} = 0 = \log(\rho) + 1 + \log(|J_{\vec{\mathrm{x}}}|) + \alpha + \beta^{\mathrm{T}}.\vec{\mathrm{x}} + \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2}$$

ce qui donne :

$$\rho(\vec{\mathrm{x}}) = \frac{\exp(-\alpha - 1)}{|J_{\vec{\mathrm{x}}}|}.\exp\left(-\beta^{\mathrm{T}}.\vec{\mathrm{x}} - \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2} \right)$$

En reportant dans les contraintes, on trouve

- Normalisation : $\exp(1+\alpha) = \displaystyle\int_{\mathcal{D}} \exp\left(-\beta^{\mathrm{T}}.\vec{\mathrm{x}} - \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2} \right).\frac{d\vec{\mathrm{x}}}{|J_{\vec{\mathrm{x}}}|}$
- Moyenne nulle : $\displaystyle\int_{\mathcal{D}} \vec{\mathrm{x}}.\exp\left(-\beta^{\mathrm{T}}.\vec{\mathrm{x}} - \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2} \right).\frac{d\vec{\mathrm{x}}}{|J_{\vec{\mathrm{x}}}|} = 0$
- Covariance fixée : $\displaystyle\int_{\mathcal{D}} (\vec{\mathrm{x}}.\vec{\mathrm{x}}^{\mathrm{T}} - \Sigma).\exp\left(-\beta^{\mathrm{T}}.\vec{\mathrm{x}} - \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma.\vec{\mathrm{x}}}{2} \right).\frac{d\vec{\mathrm{x}}}{|J_{\vec{\mathrm{x}}}|} = 0$

En fait, comme le domaine de définition $\mathcal{D}$ est symétrique par rapport à l'origine, on trouve que $\beta = 0$ convient pour assurer une moyenne nulle. En simplifiant les équations et en revenant à la densité sur la variété, on obtient donc :

$$p_{\mathbf{x}}(\mathrm{x}) = k.\exp\left(-\frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\vec{\mathrm{x}}}{2} \right) \qquad \text{avec} \qquad k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\vec{\mathrm{x}}}{2} \right).d\mathcal{M}(\mathrm{x})$$

et la matrice symétrique de paramètres $\Gamma_{\mathbf{x}}$ est solution de l'équation :

$$\int_{\mathcal{M}} \vec{\mathrm{x}}.\vec{\mathrm{x}}^{\mathrm{T}}.\exp\left(-\frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\vec{\mathrm{x}}}{2} \right).d\mathcal{M}(\mathrm{x}) = k^{(-1)}.\Sigma_{\mathbf{xx}}$$

A partir de ces équations, on peut calculer l'information de cette distribution : comme $\log(p_{\mathbf{x}}(\mathrm{x})) = \log(k) - \frac{\vec{\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\vec{\mathrm{x}}}{2}$, on a :

$$\mathbf{I}\left[\, p_{\mathbf{x}} \,\right] = \log(k) - \frac{1}{2}.\int_{\mathcal{M}} \vec{\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}^{(-1)}.\vec{\mathrm{x}}.p_{\mathbf{x}}(\mathrm{x}).d\mathcal{M}(\mathrm{x})$$

En notant que $x^{\mathrm{T}}.V.y = \mathrm{Tr}(V.y.x^{\mathrm{T}})$, on peut sortir la matrice Γ de l'intégrale, et celle-ci se simplifie comme dans le cas vectoriel pour donner la matrice de covariance. On a au final :

$$\mathbf{I}\left[\, p_{\mathbf{x}} \,\right] = \log(k) - \frac{1}{2}.\mathrm{Tr}\left(\Gamma_{\mathbf{x}}.\Sigma_{\mathbf{xx}}\right)$$

Soit maintenant une primitive aléatoire $\mathbf{x}$ de moyenne $\bar{\mathrm{x}}$ quelconque et de covariance $\Sigma_{\mathbf{xx}}$. La primitive aléatoire $\mathbf{e} = f_{\bar{\mathrm{x}}}^{(-1)} \star \mathbf{x}$ est de moyenne nulle et de covariance $\Sigma_{\mathbf{ee}} = J(f_{\bar{\mathrm{x}}})^{(-1)}.\Sigma_{\mathbf{xx}}.J(f_{\bar{\mathrm{x}}})^{(-\mathrm{T})}$. Grâce à l'invariance de l'information, la densité minimisante sachant la moyenne $\bar{\mathrm{x}}$ et la covariance $\Sigma_{\mathbf{xx}}$ est donnée par la translation inverse $\mathbf{x} = f_{\bar{\mathrm{x}}} \star \mathbf{e}$ où la primitive $\mathbf{e}$ est de densité gaussienne de

moyenne nulle et de covariance $\Sigma_{\mathbf{ee}}$. La densité gaussienne de moyenne $\bar{\mathrm{x}}$ et de covariance $\Sigma_{\mathbf{xx}}$ est donc :

$$p_{\mathbf{x}}(\mathrm{x}) = p_{\mathbf{e}}(f_{\bar{\mathrm{x}}}^{(-1)} \star \mathrm{x}) = k.\exp\left(-\frac{(f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}})^{\mathrm{T}}.\Gamma_{\mathbf{e}}.(f_{\bar{\mathrm{x}}}^{(-1)} \star \mathrm{x})}{2}\right)$$

En utilisant la carte exponentielle en $\bar{\mathrm{x}}$, et en notant $\Gamma_{\mathbf{x}} = J(f_{\bar{\mathrm{x}}})^{(-\mathrm{T})}.\Gamma_{\mathbf{e}}.J(f_{\bar{\mathrm{x}}})^{(-1)}$ on peut écrire

$$p_{\mathbf{x}}(\mathrm{x}) = k.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}}{2}\right) \qquad \text{avec} \qquad k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}}{2}\right).d\mathcal{M}(\mathrm{x})$$

Observons par ailleurs que l'équation qui relie $\Sigma_{\mathbf{ee}}$ et $\Gamma_{\mathbf{e}}$ peut se aussi réécrire dans la carte exponentielle en $\bar{\mathrm{x}}$ pour relier directement $\Sigma_{\mathbf{xx}}$ et $\Gamma_{\mathbf{x}}$:

$$k^{(-1)}.\Sigma_{\mathbf{xx}} = \int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}^{\mathrm{T}}.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{x}}}{2}\right).d\mathcal{M}(\mathrm{x})$$

L'information étant invariante, cette distribution a la même information que la distribution centrée à l'origine, et comme à la fois la constante de normalisation k et $\mathrm{Tr}\,(\Gamma_{\mathbf{x}}.\Sigma_{\mathbf{xx}})$ sont invariants, la formule ne change pas. On peut résumer les résultats ainsi :

Théorème 5.3 (Distribution gaussienne)
On appelle distribution gaussienne sur la variété $\mathcal{M}$ (vérifiant les propriétés usuelle plus la « simplicité » du lieu de coupure) la densité minimisant l'information connaissant la moyenne et la covariance. La loi gaussienne $N_{(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})}(\mathrm{y})$ de moyenne $\bar{\mathrm{x}}$ et de covariance $\Sigma_{\mathbf{xx}}$ sur la variété $\mathcal{M}$ (vérifiant les propriétés usuelle plus la « simplicité » du lieu de coupure) est

$$N_{(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})}(\mathrm{y}) = k.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right) \tag{5.2}$$

où la constante de normalisation est

$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{5.3}$$

et la matrice symétrique de paramètres $\Gamma_{\mathbf{x}}$ est solution de l'équation :

$$k^{(-1)}.\Sigma_{\mathbf{xx}} = \int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{5.4}$$

Les paramètres de la loi sont $\bar{\mathrm{x}}$ et $\Gamma_{\mathbf{x}}$, et son information est donnée par :

$$\mathbf{I}\left[\, N_{(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})} \,\right] = \log(k) - \frac{1}{2}.Tr\,(\Gamma_{\mathbf{x}}.\Sigma_{\mathbf{xx}}) \tag{5.5}$$

A partir du paramètre Γ, on peut calculer, au moins de manière numérique, la covariance de la primitive aléatoire. La détermination inverse est plus dure et nécessite la résolution des équations couplées (5.3) et (5.4).

Notons au passage qui si $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ est une primitive aléatoire de densité gaussienne $N_{(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})}$, alors $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ est encore une primitive aléatoire gaussienne $N_{(\bar{\mathrm{y}},\Gamma_{\mathbf{y}})}$ de moyenne $\bar{\mathrm{y}} = \mathrm{f} \star \bar{\mathrm{x}}$ et de paramètre

$$\Gamma_{\mathbf{y}} = J^{(-\mathrm{T})}.\Gamma_{\mathbf{x}}.J^{(-1)} \qquad \text{où} \qquad J = \left.\frac{\partial(f \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}}\right|_{\vec{\mathrm{x}}=\bar{\mathrm{x}}}$$

alors que sa covariance est $\Sigma_{\mathbf{yy}} = J.\Sigma_{\mathbf{xx}}.J^{\mathrm{T}}$. On pressent ici une relation du type $\Sigma = \Gamma^{(-1)}$ que nous allons préciser dans les exemples ci-dessous.

5.2.3 Exemple 1 : le cas vectoriel

On considère bien sûr les transformations rigides. La mesure invariante se simplifie dans ce cas à $d\mathcal{M}(x) = dx$ et la carte principale est de plus identifiée à l'espace vectoriel d'origine. L'intégration des équations pour la normalisation et la covariance est alors classique, et, en utilisant le changement de variable $y = A.x$ dans l'intégrale de la covariance où A est une racine carrée de la matrice symétrique Γ ($\Gamma = A^{\mathrm{T}}.A$), on obtient :

$$k^{(-1)} = \int_{\mathbb{R}^n} \exp\left(-\frac{x^{\mathrm{T}}.\Gamma.x}{2}\right).dx = |A|^{(-1)}.\int_{\mathbb{R}^n} \exp(-\frac{\|y\|^2}{2}).dy$$

Cette dernière intégrale se calcule aisément en coordonnées polaires dans $\mathbb{R}^n$: si $r = \|y\|$ est le rayon et Ω l'angle solide, on a $dy = r^{n-1}.dr.d\Omega$ et donc :

$$\begin{aligned} k^{(-1)} &= |A|^{(-1)}.\left(\int_{\mathcal{S}_{n-1}} d\Omega\right).\left(\int_0^\infty \exp(-r^2/2).r^{n-1}.dr\right) \\ &= |A|^{(-1)}.\left(\frac{2.\pi^{n/2}}{\mathbf{\Gamma}\left(\frac{n}{2}\right)}\right).\left(\mathbf{\Gamma}\left(\frac{n}{2}\right).2^{(n-2)/2}\right) = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|\Gamma|}} \end{aligned}$$

où $\mathbf{\Gamma}$ est la fonction Gamma ou fonction Eulérienne de deuxième espèce. Pour la covariance, rappelons que $\Sigma = \int_{\mathbb{R}^n} x.x^{\mathrm{T}}.k.\exp\left(-\frac{x^{\mathrm{T}}.\Gamma.x}{2}\right).dx$. En utilisant le changement de variable $y = A.x$, l'intégrale se simplifie en

$$A.\Sigma.A^{\mathrm{T}} = |A|^{(-1)}.k.\int_{\mathbb{R}^n} y.y^{\mathrm{T}}.\exp(-\|y\|^2/2).dy$$

Les composantes hors diagonales sont nulles car on intègre des fonction anti-symétriques, et on a pour la i$^{\mathrm{e}}$ composante diagonale :

$$\begin{aligned} [A.\Sigma.A^{\mathrm{T}}]_{ii} &= |A|^{(-1)}.k.\left(\int_{\mathbb{R}} y_i^2.\exp(y_i^2/2).dy_i\right)\left(\prod_{j\neq i}\int_{\mathbb{R}} \exp(y_j^2/2).dy_j\right) \\ &= |A|^{(-1)}.\left(\frac{|A|}{(2\pi)^{\frac{n}{2}}}\right).\left(\sqrt{2.\pi}\right).\left((2.\pi)^{\frac{n-1}{2}}\right) = 1 \end{aligned}$$

On a donc $A.\Sigma.A^{\mathrm{T}} = Id$, ce qui se traduit en $\Gamma = \Sigma^{(-1)}$. La densité est donc bien la densité gaussienne classique :

$$N_{(0,\Gamma)}(x) = k.\exp\left(-\frac{x^{\mathrm{T}}.\Gamma.x}{2}\right) = \frac{1}{(2.\pi)^{n/2}.\sqrt{|\Sigma|}}.\exp\left(-\frac{x^{\mathrm{T}}.\Sigma^{(-1)}.x}{2}\right)$$

5.2.4 Exemple 2 : le cercle

En considérant le cercle dans le plan $\mathbb{R}^2$, un groupe de transformation approprié est le groupe des rotations, et la représentation exponentielle pour la distance invariante est l'angle $\theta \in \mathcal{D} =]-\pi;\pi[$ du point du cercle avec, par exemple, l'axe des x. La mesure invariante est alors simplement $d\theta$. Si l'on considère maintenant un cercle de rayon r et que l'on conserve la métrique du plan, la représentation exponentielle devient $x = r.\theta$ et le domaine est $\mathcal{D} =]-a;a[$, où $a = \pi.r$. La mesure invariante est alors $dx = r.d\theta$ et le facteur de normalisation s'écrit :

$$k^{(-1)} = \int_{-a}^{a} \exp\left(-\frac{\gamma.x^2}{2}\right).dx = \sqrt{\frac{2\pi}{\gamma}}.\mathrm{erf}\left(\sqrt{\frac{\gamma}{2}}.a\right)$$

où erf est la fonction d'erreur $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}.\int_0^x \exp(-t^2).dt$. La densité est alors

$$N_{(0,\gamma)}(x) = \sqrt{\frac{\gamma}{2\pi}}.\frac{\exp\left(-\frac{\gamma.x^2}{2}\right)}{\mathrm{erf}\left(\sqrt{\frac{\gamma}{2}}.a\right)} = k.\exp\left(-\frac{\gamma.x^2}{2}\right)$$

ce qui est en fait une gaussienne classique tronquée. Ceci se ressent dans la liaison entre la variance σ^2 et le paramètre γ par l'introduction d'un biais dans la relation $\sigma^2 = 1/\gamma$. L'intégration de l'équation de variance donne en effet :

$$\sigma^2 = \int_{-a}^{a} x^2.k.\exp\left(-\frac{\gamma.x^2}{2}\right).dx = \frac{1}{\gamma}\left(1 - 2.a.k.\exp\left(-\frac{\gamma.a^2}{2}\right)\right)$$

Pour fixer les idées, il est intéressant de regarder quelques propriétés aux limites : si le rayon tend vers l'infini, le cercle devient la droite réelle $\mathbb{R}$ et on obtient $\sigma^2 = 1/\gamma$ et la densité gaussienne unidimensionnelle classique, comme on pouvait s'y attendre.

Un autre cas limite est intéressant : comme le cercle est compact et borné, la variance ne peut pas devenir infinie comme dans le cas réel. En effet, si l'on fait tendre le paramètre γ vers 0 (ce qui correspond dans le cas vectoriel à faire tendre la variance vers l'infini), un développement limité de σ^2 donne $\sigma^2 = a^2/3 + O(\gamma)$. L'étalement maximal de la gaussienne est donc :

$$\sigma_0^2 = \lim_{\gamma \to 0} \sigma^2 = \frac{a^2}{3} \qquad \text{pour une densité de} \qquad N_{(0,0)}(x) = \frac{1}{2.a}$$

On retrouve donc la densité uniforme sur le cercle comme la densité gaussienne de paramètre 0. A l'opposé, si γ tend vers l'infini, on obtient le Dirac à l'origine (voir figure 5.3).

5.2.5 Approximation pour Σ faible

Pour finir, remarquons quand même que si la distribution gaussienne est « suffisamment centrée[3] », alors la mesure invariante dans la carte exponentielle autour de la primitive $\bar{\mathrm{x}}$ est à peu de choses près la mesure de Lebesgue :

$$d\mathcal{M}(\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}) = \sqrt{|Q(\overrightarrow{\bar{\mathrm{x}}\mathrm{y}})|}.d\overrightarrow{\bar{\mathrm{x}}\mathrm{y}} \simeq \sqrt{|Q(\vec{\mathrm{x}})|}.d\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}$$

Notons que la base induite par la carte principale sur l'espace tangent au point $\bar{\mathrm{x}}$ n'est pas orthonormée et nous sommes obligés de conserver la métrique pour remédier à cela. L'autre solution consisterait à se ramener à l'origine. La partie de l'exponentielle qui serait en dehors du domaine, comme les contraintes éventuelles sur les bords, deviennent aussi négligeables et on est alors dans le cas « quasi-vectoriel » où la distribution est donnée par

$$N_{(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})}(\mathrm{y}) \simeq k.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}.\Gamma_{\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right)$$

avec

$$k^{(\text{-}1)} \simeq (2.\pi)^{n/2}.\sqrt{|Q(\vec{\mathrm{x}}).\Sigma_{\mathbf{xx}}|} \qquad \text{et} \qquad \Gamma_{\mathbf{x}} \simeq \Sigma_{\mathbf{xx}}^{(\text{-}1)}$$

Il serait intéressant de développer plus avant cette approximation, et en particulier d'en préciser l'ordre par rapport à la variance σ^2 et sans doute également du rayon d'injection et de la courbure maximale.

3. Ceci dépend a priori de Σ mais aussi de la courbure de la variété qui intervient dans l'expression de la mesure invariante et de l'extension du domaine de définition de la carte exponentielle.

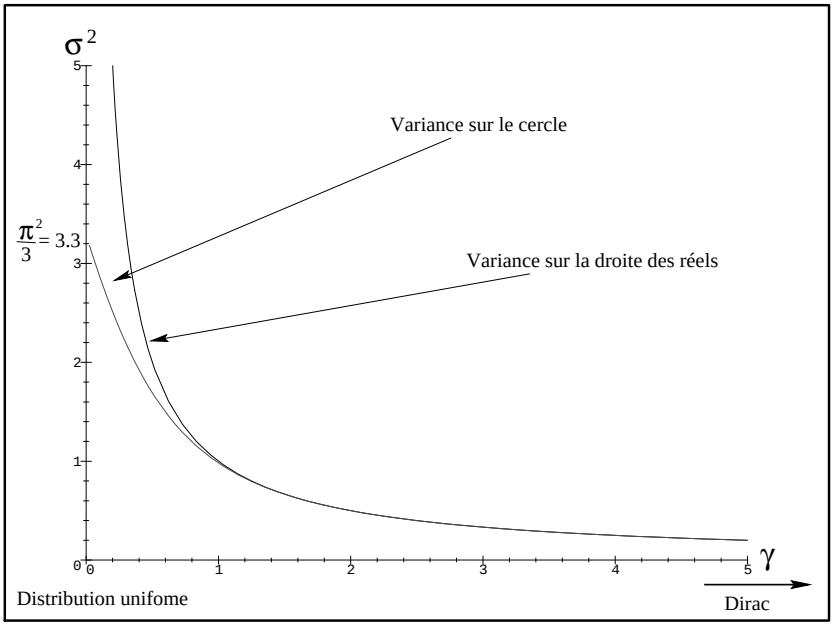


FIG. 5.3 – *Variance en fonction du paramètre γ de la gaussienne sur le cercle de rayon 1 et dans $\mathbb{R}$. Cette variance tend vers $\sigma_0^2 = \pi^2/3$ pour la distribution uniforme sur le cercle ($\gamma = 0$) alors qu'elle tend vers l'infini pour la mesure uniforme sur $\mathbb{R}$. Il est intéressant de noter que pour une faible variance (ici $\gamma > 1$), une bonne valeur approchée de la relation entre la variance sur le cercle et le paramètre γ est $\sigma^2 \simeq 1/\gamma$, c'est-à-dire identique au cas réel.*

Retour à l'information : approximation pratique Supposons que l'on ait estimé, à partir du même ensemble de données, les matrices de covariances à l'origine de deux modèles de bruits différents. Pour savoir lequel est le plus adapté, il paraît raisonnable de choisir le modèle de bruit le plus informatif. Pour cela, on associe à chaque modèle de bruit l'information minimale parmi les distribution ayant cette moyenne et covariance.

Le problème qui se pose est évidemment de calculer l'information de la distribution gaussienne, étant donné qu'on ne sait pas en général en calculer les paramètres. Afin d'obtenir quand même un critère pratique, nous nous placerons dans le cas d'une covariance Σ faible développé ci-dessus. La formule (5.5) se réduit alors à

$$\mathbf{I}\,[\,N(\bar{\mathrm{x}},\Gamma_{\mathbf{x}})\,] = -\frac{n}{2}\,(1+\log(2.\pi)) - \frac{1}{2}\log\left(\det(Q(\vec{\mathrm{x}}).\Sigma)\right)$$

En pratique, on associera donc à une primitive aléatoire $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ l'information :

$$\mathbf{I}\,[\,\mathbf{x}\,] = -\frac{n}{2}\,(1+\log(2.\pi)) + \log\left(|J(\mathrm{f}_{\vec{\mathrm{x}}})|\right) - \frac{1}{2}.\log\left(|\Sigma|\right) \tag{5.6}$$

5.2.6 Discussion

L'approche « minimisation de l'information » pour la définition de la gaussienne sur une variété semble prometteuse et permet d'obtenir une famille de distributions allant du Dirac, la distribution ponctuelle exacte, à la distribution uniforme (ou la mesure uniforme si l'espace n'est pas compact). Cependant, le lien entre le paramètre Γ et la covariance Σ de la distribution gaussienne est loin d'être

aussi simple que dans le cas vectoriel et les équations ne sont pas toujours intégrables formellement, même si l'on pourrait concevoir des algorithmes pour le faire numériquement.

Ainsi, dans le cas des rotations, la mesure invariante est donnée pour le vecteur rotation (la carte principale) par l'équation (3.7) : $d\mathcal{G}(r) = \frac{4\sin^2(\theta/2)}{\theta^2}\,dr$ et la densité gaussienne centrée à l'origine de paramètre Γ est $N_{(0,\Gamma)}(r) = k.\exp(-r^{\mathrm{T}}.\Gamma.r/2)$. En passant en coordonnées polaires ($r = \theta.n$), on peut exprimer la constante de normalisation par :

$$k^{(-1)} = 4\int_{n\in\mathcal{S}_2}\int_{\theta=0}^{\pi}\sin^2(\theta/2).\exp\left(-\theta^2.\frac{n^{\mathrm{T}}.\Gamma.n}{2}\right).d\theta.dn$$

et la covariance est :

$$\Sigma = 4.k.\int_{n\in\mathcal{S}_2} n.n^{\mathrm{T}}.\int_{\theta=0}^{\pi}\theta^2.\sin^2(\theta/2).\exp\left(-\theta^2.\frac{n^{\mathrm{T}}.\Gamma.n}{2}\right).d\theta.dn$$

Or les intégrales $\int_{\theta=0}^{\pi}\sin^2(a.x).e^{-a.x^2}.dx$ et $\int_{\theta=0}^{\pi}x^2.\sin^2(a.x).e^{-a.x^2}.dx$ ne sont pas calculable formellement à notre connaissance. On ne peut donc pas développer plus la notion de gaussienne sans une étude approfondie des algorithmes numériques qui nous permettraient de calculer k et Σ en fonction de Γ et de façon réciproque k et Γ en fonction Σ.

Un second problème qui demanderait également une étude approfondie concerne la cas où plus de deux géodésiques minimisantes se rencontrent sur le lieu de coupure, comme par exemple sur le sphère $\mathcal{S}_2$. Nous conjecturons qu'il n'y a pas de problème si la distribution est isotrope, mais il faut absolument introduire des contraintes sur le bord du domaine en cas d'anisotropie et la distribution ne devrait plus être de « forme exponentielle » dans ce cas. Il serait également intéressant de faire le lien avec la théorie des statistiques directionnelles (voir par exemple (Bingham, 1974; Jupp et Mardia, 1989; Kent, 1992; Mardia, 1995)). Il semblerait d'ailleurs qu'un certain nombre de distributions examinées dans ces travaux puissent être exprimées dans le formalisme variationnel développé ici grâce à l'utilisation de *connecteurs* invariants (au sens de (Picard, 1994)) différents du connecteur métrique invariant (la carte exponentielle) sur lequel nous avons basé notre théorie.

5.3 Distance de Mahalanobis

5.3.1 Définition de la distance simple

Le problème que l'on se pose maintenant est de comparer la mesure $\hat{\mathrm{y}}$ d'une primitive avec la distribution d'une primitive aléatoire $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$, par exemple pour savoir si la mesure peut décemment provenir de cette distribution. Dans le cadre statistique, on suppose plutôt que l'on a une mesure bruitée $\mathbf{x} \sim (\hat{\mathrm{x}}, \Sigma_{\mathbf{xx}})$: $\hat{\mathrm{x}}$ est alors notre mesure et $\Sigma_{\mathbf{xx}}$ représente l'incertitude que l'on a estimée sur cette mesure. En supposant que le bruit est centré, on veut savoir si cette primitive aléatoire peut être une mesure de la primitive exacte y.

Dans le cas d'une distribution gaussienne dans un espace vectoriel, c'est le test du χ^2 qui permet de répondre à cette question. Ce test mesure la probabilité de la distance de Mahalanobis $\mu^2 = (\hat{x} - \bar{x})^{\mathrm{T}}.\Sigma_{\boldsymbol{xx}}^{(-1)}.(\hat{x} - \bar{x})$ en supposant que $\hat{x}$ soit une réalisation de $\boldsymbol{x}$. Si la probabilité est trop faible (i.e. μ^2 est trop grand), on rejette l'hypothèse. Cette définition de la distance de Mahalanobis se généralise aisément au cas des variétés :

Définition 5.3 (Distance de Mahalanobis simple)
On appelle distance de Mahalanobis simple entre une primitive aléatoire $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ *et une primitive fixe* y *la valeur :*

$$\mu^2(\mathbf{x}, \mathrm{y}) = \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{T}.\Sigma_{\mathbf{xx}}^{(-1)}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}} \tag{5.7}$$

On peut évidement exprimer cette distance dans la carte principale : si l'on note $\mathbf{e} = \mathrm{f}_{\bar{\mathrm{x}}}^{(-1)} \star \mathbf{x}$ l'erreur ramenée à l'origine et $z = \mathrm{f}_{\bar{\mathrm{x}}}^{(-1)} \star y$ l'erreur de mesure également ramenée à l'origine, on a :

$$\mu^2(\mathbf{x}, \mathrm{y}) = (f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}})^{\mathrm{T}}.J(f_{\bar{\mathrm{x}}})^{\mathrm{T}}.\Sigma_{\mathbf{xx}}^{(-1)}.J(f_{\bar{\mathrm{x}}}).(f_{\bar{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}) = \vec{\mathrm{z}}^{\mathrm{T}}.\Sigma_{\mathbf{ee}}^{(-1)}.\vec{\mathrm{z}} = \mu^2(\mathbf{e}, \mathrm{z})$$

5.3.2 Propriétés

Soit f une transformation déterministe et $\mathbf{x'} = \mathrm{f} \star \mathbf{x}$ et $\mathrm{y}' = \mathrm{f} \star \mathrm{y}$ les transformées de la primitive aléatoires $\mathbf{x}$ et de la primitive fixe y. Alors on a :

$$\overrightarrow{\bar{\mathrm{x}}'\mathrm{y}'} = J.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}} \qquad \text{et} \qquad \Sigma_{\mathbf{x'x'}} = J.\Sigma_{\mathbf{xx}}.J^{\mathrm{T}} \qquad \text{avec} \qquad J = \left.\frac{\partial(f \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}}\right|_{\vec{\mathrm{x}}=\vec{x}}$$

Il suffit donc d'écrire la définition de $\mu^2(\mathbf{x'}, \mathrm{y}')$ pour voir que cette expression est égale à la distance de Mahalanobis avant transformation :

$$\mu^2(\mathrm{f} \star \mathbf{x}, \mathrm{f} \star \mathrm{y}) = \mu^2(\mathbf{x}, \mathrm{y}) \tag{5.8}$$

Ceci nous permet également de nous ramener à la carte principale avec une complexité algorithmique plus faible que précédemment grâce à la transformation $\mathrm{f}_{\mathrm{y}}^{(-1)}$:

$$\mu^2(\mathbf{x}, \mathrm{y}) = \mu^2(\mathrm{f}_{\mathrm{y}}^{(-1)} \star \mathbf{x}, \mathrm{o}) = \vec{e}^{\mathrm{T}}.\Sigma_{\mathbf{ee}}^{(-1)}.\vec{e} \qquad \text{où} \qquad \mathbf{e} = \mathrm{f}_{\mathrm{y}}^{(-1)} \star \mathbf{x} \tag{5.9}$$

Enfin, une dernière propriété qui nous sera utile concerne la distance de Mahalanobis moyenne. La parallèle est à faire avec la variance : $\sigma_{\mathbf{x}}^2 = \mathbf{E}\left[\ \mathrm{dist}(\mathbf{x}, \bar{\mathrm{x}})^2\ \right]$, mais en utilisant ici la distance de Mahalanobis :

$$\mathbf{E}\left[\ \mu^2(\mathbf{x}, \bar{\mathrm{x}})\ \right] = \int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}.\Sigma_{\mathbf{xx}}^{(-1)}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}.p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) = \mathrm{Tr}\left(\Sigma_{\mathbf{xx}}^{(-1)}.\int_{\mathcal{M}} \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}.p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y})\right)$$

où le dernier terme est obtenu en utilisant l'identité $z^{\mathrm{T}}.A.z = \mathrm{Tr}(A.z.z^{\mathrm{T}})$. L'intégrale est donc égale à la matrice de covariance et comme $\mathrm{Tr}(\Sigma_{\mathbf{xx}}^{(-1)}.\Sigma_{\mathbf{xx}}) = \mathrm{Tr}(\ Id_n) = n$, on obtient l'identité suivante

$$\mathbf{E}\left[\ \mu^2(\mathbf{x}, \bar{\mathrm{x}})\ \right] = n \tag{5.10}$$

Cette identité est très utile pour différents tests statistiques. Nous l'utiliserons en particulier au chapitre 9 pour valider a posteriori l'incertitude estimée sur le recalage de deux ensembles de primitives.

5.3.3 Autre définition

En partant de la densité gaussienne de moyenne $\bar{\mathrm{x}}$ et de covariance $\Sigma_{\mathbf{xx}}$:

$$N_{(\bar{\mathrm{x}},\Gamma)}(\mathrm{y}) = k.\exp\left(-\frac{\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}.\Gamma.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}}{2}\right)$$

on pourrait envisager une définition de la distance de Mahalanobis comme l'entropie « normalisée » de la primitive fixe y par rapport à la gaussienne :

$$\mu'^2(\mathbf{x}, \mathrm{y}) = \log(k) - 2.\log(N_{(\bar{\mathrm{x}},\Gamma)}(\mathrm{y})) = \overrightarrow{\bar{\mathrm{x}}\mathrm{y}}^{\mathrm{T}}.\Gamma.\overrightarrow{\bar{\mathrm{x}}\mathrm{y}}$$

Cette définition, si elle peut présenter de nombreux avantages, a l'inconvénient majeur de reposer sur une famille de distributions fixée (la gaussienne) et n'est donc pas adaptée si on connaît la distribution exacte de la primitive aléatoire $\mathbf{x}$ et non plus seulement sa moyenne et sa covariance. De plus, d'un point de vue complexité algorithmique, elle repose sur le paramètre Γ de la gaussienne que nous avons vu difficilement calculable.

Cependant, pour une distribution quasi-gaussienne suffisamment centrée, on a la relation $\Sigma_{\mathbf{xx}}^{(-1)} \simeq \Gamma$ et les deux définitions sont donc équivalentes. Dans le cas vectoriel, ces deux définitions sont évidemment identiques.

5.3.4 Test du χ^2

Connaissant la loi de $\mathbf{x}$ (par exemple gaussienne) et en supposant que $\hat{\mathrm{y}}$ en soit une réalisation, on pourrait penser à calculer la distribution explicite de $\mu^2(\mathbf{x}, \hat{\mathrm{y}})$ et généraliser ainsi le test du χ^2 aux variétés. En pratique, ce type de calcul s'avère difficile, complexe et surtout spécifique à chaque type de primitive. Étant donné qu'on ne sait déjà pas mener les calculs des paramètres de la gaussienne sur les rotations, il paraît illusoire de vouloir dériver la distribution du χ^2 associée. De manière plus générale, il serait intéressant d'explorer ces distributions mais le travail nécessaire dépasse le cadre de ce manuscrit.

Dans la pratique, on supposera que les distributions sont quasi-gaussiennes et suffisamment centrées pour que l'on soit dans le cadre « quasi-vectoriel » : la distribution de $\mu^2(\mathbf{x}, \mathrm{y})$ est alors quasiment un χ^2_n *si* y *est bien une réalisation de* $\mathbf{x}$, où le nombre de degrés de liberté n est la dimension de la variété. Rappelons pour mémoire que la densité de probabilité du χ^2 à n degrés de libertés est donnée par

$$p_{\chi^2_n}(u) = \frac{1}{2.\mathbf{\Gamma}\left(\frac{n}{2}\right)}.\left(\frac{u}{2}\right).\exp\left(-\frac{u}{2}\right) \tag{5.11}$$

Rappelons également que la fonction $\mathbf{\Gamma}$ est calculable par la propriété de récurrence : $\mathbf{\Gamma}(x+1) = x.\mathbf{\Gamma}(x)$ avec $\mathbf{\Gamma}(1) = 1$ et $\mathbf{\Gamma}(\frac{1}{2}) = \sqrt{\pi}$. On obtient donc

$$\mathbf{\Gamma}(\frac{n}{2}) = (k-1)! \quad \text{si} \quad n = 2.k \qquad \text{et} \qquad \mathbf{\Gamma}(\frac{n}{2}) = \sqrt{\pi}.\prod_{i=0}^{k-1}(i+\frac{1}{2}) \quad \text{si} \quad n = 2.k+1$$

5.3.5 Distance de Mahalanobis entre primitives aléatoires

5.3.6 Définition théorique

On suppose ici que l'on a deux mesures bruitées $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ et $\mathbf{y} \sim (\bar{\mathrm{y}}, \Sigma_{\mathbf{yy}})$, et l'on voudrait tester si ces deux primitives aléatoires peuvent être les mesures d'une même primitive exacte z. Si c'est le cas, il existe une primitive z telle que $\mu^2(\mathbf{x}, \mathrm{z})$ et $\mu^2(\mathbf{y}, \mathrm{z})$ soient simultanément faibles, et la primitive fixe la mieux placée pour remplir ces conditions est évidemment celle qui minimise la somme de ces deux distances de Mahalanobis. La définition de la distance de Mahalanobis entre deux distributions serait donc

$$\mu^2(\mathbf{x}, \mathbf{y}) = \min_{\mathrm{z}}\left(\mu^2(\mathbf{x}, \mathrm{z}) + \mu^2(\mathbf{y}, \mathrm{z})\right) = \min_{\mathrm{z}}\left(\overrightarrow{\bar{\mathrm{x}}\mathrm{z}}^{\mathrm{T}}.\Sigma_{\mathbf{xx}}^{(-1)}.\overrightarrow{\bar{\mathrm{x}}\mathrm{z}} + \overrightarrow{\bar{\mathrm{y}}\mathrm{z}}^{\mathrm{T}}.\Sigma_{\mathbf{yy}}^{(-1)}.\overrightarrow{\bar{\mathrm{y}}\mathrm{z}}\right) \tag{5.12}$$

L'inconvénient majeur de cette définition est qu'elle nécessite une minimisation. Or, dans un algorithme de reconnaissance, cette distance sert non seulement de test de décision, par exemple pour accepter un appariement ou le rejeter comme aberrant, mais aussi de critère de classification de ces appariements. Il est donc souhaitable que son calcul soit rapide. Cette constatation nous amène à considérer une autre définition, sans doute plus ad hoc, mais qui ne nécessite pas de minimisation.

5.3.7 Définition pratique

Définition 5.4 (Distance de Mahalanobis double)
On appelle distance de Mahalanobis entre deux primitives aléatoires $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ *et* $\mathbf{y} \sim (\bar{\mathrm{y}}, \Sigma_{\mathbf{yy}})$ *la distance de Mahalanobis simple entre la primitive aléatoire* $\mathbf{z} = f_{\mathbf{y}}^{(-1)} \star \mathbf{x}$ *et l'origine :*

$$\nu^2(\mathbf{x}, \mathbf{y}) = \mu^2(f_{\mathbf{y}}^{(-1)} \star \mathbf{x}, \mathrm{o}) \tag{5.13}$$

En utilisant les équations de propagation de la moyenne et de la covariance (section 4.4.2) et en notant :

$$J_{\vec{\mathrm{x}}} = \frac{\partial(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}} \qquad J_{\vec{\mathrm{y}}} = \frac{\partial(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{y}}} = \frac{\partial(f_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}})}{\partial f_{\vec{\mathrm{y}}}^{(-1)}} . \left. \frac{\partial f^{(-1)}}{\partial f} \right|_{f=f_{\vec{\mathrm{y}}}} . \frac{\partial f_{\vec{\mathrm{y}}}}{\partial \vec{\mathrm{y}}}$$

les jacobiens estimés en $(\bar{\mathrm{x}}, \bar{\mathrm{y}})$, la primitive aléatoire $\mathbf{z} = f_{\mathbf{y}} \star \mathbf{x}$ est caractérisée par une moyenne de $\bar{\mathrm{z}} = f_{\bar{\mathrm{y}}} \star \bar{\mathrm{x}}$ et une covariance de $\Sigma_{\mathbf{zz}} = J_{\vec{\mathrm{x}}} . \Sigma_{\mathbf{xx}} . J_{\vec{\mathrm{x}}}^{\mathrm{T}} + J_{\vec{\mathrm{y}}} . \Sigma_{\mathbf{yy}} . J_{\vec{\mathrm{y}}}^{\mathrm{T}}$ et la distance de Mahalanobis est

$$\nu^2(\mathbf{x}, \mathbf{y}) = \bar{\mathrm{z}}^{\mathrm{T}} . \Sigma_{\mathbf{zz}}^{(-1)} . \bar{\mathrm{z}} = (f_{\bar{\mathrm{y}}}^{(-1)} \star \bar{\mathrm{x}})^{\mathrm{T}} . (J_{\vec{\mathrm{x}}} . \Sigma_{\mathbf{xx}} . J_{\vec{\mathrm{x}}}^{\mathrm{T}} + J_{\vec{\mathrm{y}}} . \Sigma_{\mathbf{yy}} . J_{\vec{\mathrm{y}}}^{\mathrm{T}})^{(-1)} . (f_{\bar{\mathrm{y}}}^{(-1)} \star \bar{\mathrm{x}})$$

Théorème 5.4 (Invariance de la distance de Mahalanobis)
Soient $\mathbf{x}$ *et* $\mathbf{y}$ *deux primitives aléatoires et* f *une transformation déterministe. Alors :*

$$\nu^2(\mathrm{f} \star \mathbf{x}, \mathrm{f} \star \mathbf{y}) = \nu^2(\mathbf{x}, \mathbf{y}) \tag{5.14}$$

Preuve :
Notons $\mathrm{x}' = f \star \mathrm{x}$ et $\mathrm{y}' = f \star \mathrm{y}$. Il existe donc $h \in \mathcal{H}$ tel que $f_{\mathrm{y}'} \circ h = f \circ f_{\mathrm{y}}$, ce qui donne : $f_{\mathrm{y}'}^{(-1)} = h \circ f_{\mathrm{y}}^{(-1)} \circ f^{(-1)}$. Soit maintenant $\mathrm{z} = f_{\mathrm{y}}^{(-1)} \star \mathrm{x}$. On obtient donc $\mathrm{z}' = f_{\mathrm{y}'}^{(-1)} \star \mathrm{x}' = h \star \mathrm{z}$, ce qui s'exprime dans la carte principale pour la primitive moyenne par

$$\bar{\mathrm{z}'} = J(h_{\bar{\mathrm{z}}}) . \bar{\mathrm{z}}$$

puisque l'action est $h_{\bar{\mathrm{z}}}$ est linéaire dans cette carte.
Pour calculer la matrice de covariance de $\mathbf{z'}$, on utilise justement le jacobien $\frac{\partial \vec{\mathrm{z}'}}{\partial \vec{\mathrm{z}}}$ estimé à la valeur moyenne. On a donc : $\Sigma_{\mathbf{z'z'}} = J(h_{\bar{\mathrm{z}}}) . \Sigma_{\mathbf{zz}} . J(h_{\bar{\mathrm{z}}})^{\mathrm{T}}$. En reportant dans la définition de la distance de Mahalanobis, on trouve :

$$\nu^2(\mathrm{f} \star \mathbf{x}, \mathrm{f} \star \mathbf{y}) = \bar{\mathrm{z}'}^{\mathrm{T}} . \Sigma_{\mathbf{z'z'}}^{(-1)} . \bar{\mathrm{z}'} = \bar{\mathrm{z}}^{\mathrm{T}} . \Sigma_{\mathbf{zz}}^{(-1)} . \bar{\mathrm{z}} = \nu^2(\mathbf{x}, \mathbf{y})$$

■

5.3.8 Équivalence des définitions dans le cas vectoriel

Théorème 5.5 (Distance de Mahalanobis double et minimum)
Soient $\boldsymbol{x} \sim (x, \Sigma_{\boldsymbol{xx}})$ *et* $\boldsymbol{y} \sim (y, \Sigma_{\boldsymbol{yy}})$ *deux vecteurs aléatoires. On note* $\Lambda = (\Sigma_{\boldsymbol{xx}} + \Sigma_{\boldsymbol{yy}})^{(-1)}$. *Le minimum sur* $z \in \mathbb{R}^n$ *de* $\mu^2(\boldsymbol{x}, z) + \mu^2(\boldsymbol{y}, z)$ *est obtenu pour le vecteur*

$$z = (\, Id - \Sigma_{\boldsymbol{xx}} . \Lambda) . x + (\, Id - \Sigma_{\boldsymbol{yy}} . \Lambda) . y \tag{5.15}$$

avec la valeur :

$$\mu^2(\boldsymbol{x}, \boldsymbol{y}) = \min_z \left(\mu^2(\boldsymbol{x}, z) + \mu^2(\boldsymbol{y}, z) \right) = (x - y)^T . \Lambda . (x - y) = \nu^2(\boldsymbol{x}, \boldsymbol{y}) \tag{5.16}$$

Notons que cette équivalence et la formulation exacte du résultat sont à la base des équations du filtre de Kalman. Pour développer la preuve, nous aurons besoin du lemme d'inversion, dont on pourra trouver une démonstration dans (Maybeck, 1979, p. 213).

Lemme 5.1 (Lemme d'inversion)
Soit deux matrices symétriques définies positives W *et* S *de dimension* m *et* n, *et une matrice* M *de dimension* $n \times m$. *Alors :*

$$(M^T . W^{(-1)} . M + S^{(-1)})^{(-1)} = S - S . M^T (W + M . S . M^T)^{(-1)} . M . S \tag{5.17}$$

Preuve : (théorème 5.5)
Le minimum dans $C(z) = \mu^2(\boldsymbol{x}, z) + \mu^2(\boldsymbol{y}, z)$ est caractérisé par une dérivée nulle par rapport à z :

$$\frac{\partial C(z)}{\partial z} = 2\left((x-z)^{\mathrm{T}}.\Sigma_{\boldsymbol{xx}}^{(-1)} + (y-z)^{\mathrm{T}}.\Sigma_{\boldsymbol{yy}}^{(-1)}\right) = 0$$

Le minimum est donc obtenu pour

$$z = (\Sigma_{\boldsymbol{xx}}^{(-1)} + \Sigma_{\boldsymbol{yy}}^{(-1)})^{(-1)}.(\Sigma_{\boldsymbol{xx}}^{(-1)}.x + \Sigma_{\boldsymbol{yy}}^{(-1)}.y)$$

avec une matrice hessienne définie positive : $H_z = \Sigma_{\boldsymbol{xx}}^{(-1)} + \Sigma_{\boldsymbol{yy}}^{(-1)}$. On a donc bien un minimum unique. Utilisons le lemme d'inversion pour « simplifier » z : on a $M = Id$, $W = \Sigma_{\boldsymbol{xx}}$ et $S = \Sigma_{\boldsymbol{yy}}$ (et réciproquement). En notant $\Lambda = (\Sigma_{\boldsymbol{xx}} + \Sigma_{\boldsymbol{yy}})^{(-1)}$ et en factorisant, on obtient :

$$(\Sigma_{\boldsymbol{xx}}^{(-1)} + \Sigma_{\boldsymbol{yy}}^{(-1)})^{(-1)} = \{\, Id - \Sigma_{\boldsymbol{yy}}.\Lambda\}\,.\Sigma_{\boldsymbol{yy}} = \{\, Id - \Sigma_{\boldsymbol{xx}}.\Lambda\}\,.\Sigma_{\boldsymbol{xx}}$$

En reportant dans z, on trouve :

$$z = (\, Id - \Sigma_{\boldsymbol{xx}}.\Lambda).x + (\, Id - \Sigma_{\boldsymbol{yy}}.\Lambda).y$$

Reportant maintenant cette valeur dans $\mu^2(\boldsymbol{x}, z)$. Comme $\Sigma_{\boldsymbol{xx}}^{(-1)}.(z-x) = \Sigma_{\boldsymbol{xx}}^{(-1)}(\, Id - \Sigma_{\boldsymbol{yy}}.\Lambda).y - \Lambda.x$, on a :

$$\begin{aligned}\mu^2(\boldsymbol{x}, z) = &\quad (z-x)^{\mathrm{T}}.\Sigma_{\boldsymbol{xx}}^{(-1)}.(z-x)\\ = &\quad x^{\mathrm{T}}.\Lambda.\Sigma_{\boldsymbol{xx}}.\Lambda.x - 2.y^{\mathrm{T}}.(\, Id - \Lambda.\Sigma_{\boldsymbol{yy}}).\Lambda.x\\ &+ y^{\mathrm{T}}.(\, Id - \Lambda.\Sigma_{\boldsymbol{yy}}).\Sigma_{\boldsymbol{xx}}^{(-1)}.(\, Id - \Sigma_{\boldsymbol{yy}}.\Lambda).y\end{aligned}$$

Pour simplifier le terme en $y^{\mathrm{T}}.y$, posons $S = \Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda$, et $W = -M$ avec $M = M^{\mathrm{T}} = S^{(-1)} - \Lambda^{(-1)}$. Nous allons réutiliser le lemme d'inversion. Pour cela, observons que

$$\begin{aligned}M.S &= Id - \Lambda^{(-1)}.S = Id - \Sigma_{\boldsymbol{yy}}.\Lambda\\ M.S.M^{\mathrm{T}} + W &= (M.S - Id).M = -\Lambda^{(-1)}.S.M\\ &= \Lambda^{(-1)} - \Sigma_{\boldsymbol{yy}} = -\Sigma_{\boldsymbol{xx}}\end{aligned}$$

Le terme en $y^{\mathrm{T}}.y$ se simplifie donc grâce au lemme d'inversion en

$$\begin{aligned}(\, Id - \Lambda.\Sigma_{\boldsymbol{yy}}).\Sigma_{\boldsymbol{xx}}^{(-1)}.(\, Id - \Sigma_{\boldsymbol{yy}}.\Lambda) &= S.M^{\mathrm{T}}.(M.S.M^{\mathrm{T}} + W)^{(-1)}.M.S\\ &= S - (S^{(-1)} + M^{\mathrm{T}}.W^{(-1)}.M)^{(-1)}\\ &= S - (S^{(-1)} - M)^{(-1)}\\ &= \Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda - \Lambda\end{aligned}$$

ce qui simplifie $\mu^2(\boldsymbol{x}, z)$ en :

$$\mu^2(\boldsymbol{x}, z) = -x^{\mathrm{T}}.\Lambda.\Sigma_{\boldsymbol{xx}}.\Lambda.x - 2.y^{\mathrm{T}}.(\, Id - \Lambda.\Sigma_{\boldsymbol{yy}}).\Lambda.x + y^{\mathrm{T}}.(\Lambda - \Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda).y$$

L'addition des distances de Mahalanobis par rapport à $\boldsymbol{x}$ et $\boldsymbol{y}$ donne donc :

$$\begin{aligned}\mu^2(\boldsymbol{x}, z) + \mu^2(\boldsymbol{y}, z) = &\quad x^{\mathrm{T}}(\Lambda.\Sigma_{\boldsymbol{xx}}.\Lambda + \Lambda - \Lambda.\Sigma_{\boldsymbol{xx}}.\Lambda).x\\ &+ y^{\mathrm{T}}(\Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda + \Lambda - \Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda).y\\ &- 2.y^{\mathrm{T}}(2.\Lambda - \Lambda.\Sigma_{\boldsymbol{yy}}.\Lambda + -\Lambda.\Sigma_{\boldsymbol{xx}}.\Lambda).x\end{aligned}$$

Comme $(\Sigma_{\boldsymbol{xx}} + \Sigma_{\boldsymbol{yy}}) = \Lambda^{(-1)}$, on trouve finalement :

$$\mu^2(\boldsymbol{x}, z) + \mu^2(\boldsymbol{y}, z) = (x-y)^{\mathrm{T}}.\Lambda.(x-y) = \nu^2(\boldsymbol{x}, \boldsymbol{y})$$

■

5.3.9 Discussion sur la distance de Mahalanobis

Nous avons défini une distance statistique entre une mesure y d'une primitive et une primitive aléatoire $\mathbf{x}$ qui généralise, sur la base de la matrice de covariance, la distance de Mahalanobis classique. De plus, elle est invariante par l'action d'une transformation fixe et suit, pour une covariance suffisamment faible, une loi du χ^2 à n degrés de liberté (n étant la dimension de la variété).

L'extension de cette définition à une distance entre deux primitives pose cependant plus de problèmes et nécessiterait sans doute une étude plus poussée. Nous avons tout d'abord proposé une définition basée sur la minimisation :

$$\mu^2(\mathbf{x}, \mathbf{y}) = \min_{\mathrm{z}}\left(\mu^2(\mathbf{x}, \mathrm{z}) + \mu^2(\mathbf{y}, \mathrm{z})\right)$$

qui généralise correctement la distance de Mahalanobis simple de l'équation (5.9) si l'on considère que la distance de Mahalanobis entre deux primitives déterministes différentes est infinie partout et

que la distance entre deux primitives déterministes égales est nulle. Cette définition est évidemment invariante par l'action d'une transformation fixe et symétrique mais elle est difficile à implémenter et augmente considérablement les temps de calculs à cause de la définition implicite.

Nous avons donc proposé une autre définition qui se calcule directement et correspond à la première dans le cas vectoriel. Elle est également invariante mais elle n'est généralement pas symétrique :

$$\mu^2(f_{\mathbf{y}}^{(-1)} \star \mathbf{x}, \mathrm{o}) = \nu^2(\mathbf{x}, \mathbf{y}) \neq \nu^2(\mathbf{y}, \mathbf{x}) = \mu^2(f_{\mathbf{x}}^{(-1)} \star \mathbf{y}, \mathrm{o})$$

Cependant, nous n'avons observé lors de nos expériences que des différences relatives de l'ordre de 2% entre $\mu^2(\mathbf{x}, \mathbf{y})$ et $\mu^2(\mathbf{y}, \mathbf{x})$. De plus, l'équivalence de cette définition avec la définition par minimisation dans le cas vectoriel nous permet de considérer que nous avons ainsi obtenu une approximation de $\nu^2(\mathbf{x}, \mathbf{y})$. Cette valeur approchée apparaît en pratique largement suffisante pour classer des appariements et rejeter des mesures aberrantes.

Par contre, la définition par minimisation ouvre la voie à des algorithmes de calcul de la moyenne incluant des informations du second ordre : voir à ce sujet les divers algorithmes du chapitre 8.

5.4 Conclusion

On peut considérer que nous avons conclu avec les aspects statistiques notre théorie de l'incertitude sur les primitives géométriques. L'étude des modèles de bruits nous a permis de caractériser les bruits homogènes et isotropes qui seront à la base de nos applications statistiques, et nous avons pu ébaucher une étude des distributions équivalentes à la gaussienne sur les variété homogènes grâce à la minimisation de l'information. Ceci nous permet en particulier de justifier le choix d'une distribution quasi-gaussienne *dans la carte exponentielle au point* $\bar{\mathrm{x}}$ comme distribution de référence pour la primitive aléatoire $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$, et de justifier le test du χ^2 sur la distance de Mahalanobis simple. L'approximation de la distance de Mahalanobis entre deux primitive probabilistes prendra son importance dans la seconde partie de ce manuscrit, pour la fusion de primitives ou l'estimation de la transformation entre deux objets (le recalage).

Chapitre 6

Synthèse : implémentation pratique

Nous avons développé jusqu'ici un ensemble d'outils mathématiques pour gérer l'incertitude sur les primitives géométriques. Nous présentons dans cette section un résumé pratique des formules importantes et surtout de la démarche à suivre pour pouvoir appliquer cette théorie.

La première section rappelle les étapes mathématiques qui conduisent à l'obtention des géodésiques et de la carte principale. Nous présentons ensuite les quelques opérations de base à implémenter pour chaque type de primitive à partir desquelles nous pouvons construire la plupart des algorithmes sur les primitives, de manière totalement indépendante du type de primitive considéré. Cette façon « orientée objet » de travailler avec les primitives géométriques est à la base de la bibliothèque C qui implémente la théorie développée jusqu'ici pour les repères, les repères semi-orientés et non orientés et les points. Nous développerons les calculs nécessaires à ces primitives et les applications dans la seconde partie de ce manuscrit (voir chapitre 7).

6.1 Phase mathématique : géodésiques et carte principale

On s'intéresse donc ici à une variété de primitives géométriques $\mathcal{M}$ soumises à l'action d'un groupe de transformation $\mathcal{G}$. Nous nous focaliserons dans un premier temps sur le groupe de transformation, en supposant qu'il soit défini de manière indépendante des primitives, par exemples par son action sur un espace euclidien. C'est le cas du groupe des transformations rigides que nous considérons dans ce manuscrit, mais on pourrait penser à d'autres groupes comme les similitudes, les transformations affines ou les transformations projectives.

6.1.1 Le groupe de transformation $\mathcal{G}$

Le premier travail est de définir les éléments du groupes, par exemple comme une sous-variété d'un espace vectoriel euclidien. Les transformations affines en n-D sont ainsi définies par une matrice $n \times n$ A et un vecteur de translation t, c'est-à-dire $n.(n+1)$ éléments réels. Cependant, tous les points de cet espace $\mathbb{R}^{n.(n+1)}$ ne représentent pas un élément du groupe puisque l'on a la contrainte $\det(A) \neq 0$ pour que la matrice soit inversible. De même, les transformations rigides sont représentables par une matrice $n \times n$ R, vérifiant les contraintes $R.R^{\mathrm{T}} = R^{\mathrm{T}}.R = Id$ et $\det(R) = +1$, et un vecteur de translation $t \in \mathbb{R}^n$. Cette façon de définir l'ensemble des transformations permet de montrer

aisément qu'on a bien une variété différentielle puisqu'il suffit pour cela que les contraintes soient non dégénérées (de jacobien non nul). On note habituellement f et g des points de cette variété $\mathcal{G}$. Pour que l'on puisse travailler sur le groupe, il faut encore un élément neutre que l'on appellera identité (Id) et l'expression des opérations de composition ($f \circ g$) et d'inversion ($f^{(-1)}$).

Nous supposerons de plus, pour pouvoir obtenir une distance entre tous les éléments du groupe, que le groupe est connexe. Cela revient à dire que deux transformations quelconques peuvent être reliés par une suite de transformations continues. Cette hypothèse est en général justifiée si l'on travaille en vrai 3D. On devrait ainsi restreindre le groupe des transformations affines à celles de déterminant positif (la feuille qui contient l'identité) car on ne peut pas passer continûment d'un déterminant positif à un déterminant négatif sans passer par une matrice de déterminant nul, qui n'est plus une transformation puisqu'elle n'est pas inversible. Cela correspondrait par exemple à replier $\mathbb{R}^3$ tout entier dans un plan, une droite ou un seul point, ce qui semble peu réaliste.

Le cas d'un groupe connexe compact C'est un cas un peu spécial puisqu'il existe alors une métrique invariante à droite et à gauche en même temps, et que les géodésiques partant de l'identité pour cette métrique sont les sous-groupes à un paramètre. Une propriété intéressante est que toutes les transformations d'un sous groupe à un paramètre commutent. La démarche est alors de caractériser la commutation de deux transformation ($f \circ g = g \circ f$), puis de trouver l'expression des courbes $\gamma(t)$ sur le groupe satisfaisant :

$$\forall (s,t) \in \mathbb{R}^2, \qquad \gamma(s+t) = \gamma(s) \circ \gamma(t) = \gamma(t) \circ \gamma(s)$$

Puisque ces sous-groupes sont définis sur $\mathbb{R}$ tout entier et que le groupe est supposé connexe, le groupe est géodésiquement complet. De plus, on obtient les géodésiques partant d'un point f par la translation à gauche $f \circ \gamma(t)$ (ou à droite $\gamma(t) \circ f$ puisque la métrique est bi-invariante).

Il ne reste plus alors qu'à caractériser l'espace tangent à l'identité (i.e. trouver un espace vectoriel de dimension n qui lui soit difféomorphe), calculer le vecteur tangent v en $t = 0$ à une telle courbe, et réécrire l'équation de la courbe γ_v en fonction de ce vecteur tangent. L'exponentielle est alors définie par :

$$\exp(v) = \gamma_v(1)$$

La suite est alors identique au cas général.

Cas général : groupe connexe localement compact La démarche est ici un peu plus compliquée puisque l'on doit passer par une carte pour déterminer les géodésiques. On suppose donc que l'on a une représentation minimale du groupe que l'on notera f et dont le domaine est un ouvert autour de l'identité. Par une translation appropriée dans la carte, on peut se ramener à une carte centrée en l'identité, c'est-à-dire telle que l'identité soit représentée par le vecteur nul. On détermine alors l'expression de la composition $f \circ g$ dans cette carte (pour des transformations suffisamment proches des l'identité), ainsi que les dérivées de cette opération et en particulier le jacobien de la translation à gauche, qui permet de construire la représentation locale de la métrique invariante à gauche.

$$Q(x) = J_L(f)^{(-T)}.Q.J_L(f)^{(-1)} \qquad \text{avec} \qquad J_L(f) = \left.\frac{\partial(f \circ e)}{\partial e}\right|_{e=Id}$$

Notons qu'on peut choisir sans problème la métrique à l'identité $Q = Id$ puisque les géodésiques sont globalement invariantes au choix de cette métrique. De plus, on pourra toujours changer après-coup cette métrique Q en appliquant un changement de variable affine dans la carte principale.

Le plus difficile est alors à faire : résoudre le système d'équations différentielles du second ordre suivant pour trouver l'équation des géodésiques.

$$\frac{d^2\gamma_i}{dt^2} + \sum_{j,k=1}^{n} \Gamma^i_{j,k}.\frac{d\gamma_j}{dt}.\frac{d\gamma_k}{dt} = 0$$

On rappelle que les symboles de Christoffel sont donnés par

$$\Gamma^i_{j,k} = \frac{1}{2}\sum_{m=1}^{n} q^{im}\left(\frac{\partial q_{mj}}{\partial x_k} + \frac{\partial q_{mk}}{\partial x_j} - \frac{\partial q_{jk}}{\partial x_m}\right)$$

où $q^{ij} = [Q(f)^{(-1)}]_{ij}$. On pourra cependant se contenter des géodésiques partant de l'origine, c'est-à-dire de zéro dans notre carte centrée. S'il existe toujours une solution, il n'est pas dit qu'elle soit explicite et exprimable avec nos fonctions usuelles. Cependant, on supposera que c'est le cas ici.

Une fois qu'on a l'équation locale des géodésiques autour de l'identité dans notre carte, on détermine leur équation directement dans le groupe $\mathcal{G}$ et on vérifie qu'elles s'étendent sans singularité à $\mathbb{R}$ tout entier et que le groupe est ainsi géodésiquement complet.

On calcule alors le vecteur tangent v à chaque géodésique dans la carte locale ou, comme dans le cas du groupe compact, dans un espace vectoriel difféomorphe à l'espace tangent à l'identité, et en réécrit l'équation de la courbe γ_v en fonction de ce vecteur. L'exponentielle est alors définie par :

$$\exp(v) = \gamma_v(1)$$

On inverse alors cette expression pour obtenir l'ensemble $\mathcal{V}(\mathrm{f})$ des vecteurs v satisfaisant $\exp(v) = \mathrm{f}$, puis on détermine le lieu de coupure, limitant le domaine de définition. Pour cela, il suffit de se souvenir qu'il n'existe qu'une seule géodésique minimisante à l'intérieur du domaine et éventuellement plusieurs sur le bord (le lien de coupure). On peut donc calculer la « norme » de chaque primitive f :

$$N_L(\mathrm{f}) = \mathrm{dist}(\mathrm{f}, \mathrm{Id}) = \min_{v\in\mathcal{V}(\mathrm{f})}(\|v\|)$$

et réduire $\mathcal{V}(\mathrm{f})$ à l'ensemble des vecteurs de norme minimale :

$$\mathcal{V}_{\min}(\mathrm{f}) = \arg\min_{v\in\mathcal{V}(\mathrm{f})}(\|v\|)$$

On a alors obtenu la carte principale. Si $\mathcal{V}_{\min}(\mathrm{f})$ n'est pas réduit à un seul élément, alors f fait partie du lieu de coupure de l'identité et $\mathcal{V}_{\min}(\mathrm{f})$ est l'ensemble des vecteurs à identifier sur le lieu de coupure tangentiel, et si $\mathcal{V}_{\min}(\mathrm{f}) = \{\vec{\mathrm{f}}\}$ est réduit à un seul élément, alors on est en général à l'intérieur du domaine de définition $\mathcal{D}$ (bien qu'occasionnellement f puisse appartenir au lieu de coupure).

6.1.2 La variété $\mathcal{M}$

Comme on a défini le groupe de transformation, il nous faut définir la variété, par exemple comme une sous-variété d'un espace vectoriel euclidien. On notera $\mathrm{x} \in \mathcal{M}$ un tel élément. On représentera ainsi un point extrémal par un repère semi-orienté, c'est-à-dire une position $x \in \mathbb{R}^3$, un vecteur unitaire n normal à la surface en ce point, et les deux directions principales unitaires $(\pm t_1, \pm t_2)$ orthogonales entre elles et au vecteur normal. Si les contraintes ne sont pas dégénérées, alors on a une structure de variété différentielle (ce qui est le cas par exemple pour $\|n\| = 1$ ou $\langle t_1 \mid t_2 \rangle = 0$).

L'étape suivante est de définir l'action d'une transformation f (quelle qu'en soit sa représentation) sur une primitive x. On notera : $\mathrm{y} = \mathrm{f} \star \mathrm{x}$. Il est bien évident que le résultat de cette action doit être une primitive de $\mathcal{M}$. Il faudra au besoin étendre l'ensemble des primitives utilisées pour satisfaire cette contrainte. On utilisera par exemple des trièdres quelconques (mais non dégénérés) et non plus simplement des trièdres orthonormés si l'on utilise des transformations affines au lieu des transformations rigides.

Pour que notre variété soit homogène, il nous faut alors séparer la partie invariante des primitives de la partie qui « bouge » avec le groupe. On ne s'intéresse dans la suite qu'à la partie homogène, mais les invariants (dits unaires puisqu'ils ne font intervenir qu'une seule primitive) peuvent être conservés et utilisés à d'autres fins telles que la mise en correspondance. Remarquons au passage que si notre variété est homogène pour un groupe connexe, elle est automatiquement connexe.

On choisit alors une origine que l'on appelle o, et on détermine le groupe d'isotropie en ce point :

$$\mathcal{H} = \{\mathrm{h} \in \mathcal{G} \ / \ \mathrm{h} \star \mathrm{o} = \mathrm{o}\}$$

puis le coset d'un point $\mathrm{x} \in \mathcal{M}$:

$$\mathcal{F}_\mathrm{x} = \{\mathrm{g} \in \mathcal{G} \ / \ \mathrm{g} \star \mathrm{o} = \mathrm{x}\}$$

et enfin on se choisit une fonction de placement $\mathrm{f_x}$, de préférence la plus simple possible, mais qui pourra éventuellement servir à définir un modèle de bruit homogène « standard ». La translation de vecteur x est ainsi tout a fait adaptée comme fonction de placement pour les points soumis aux transformations rigides. De manière générale, un choix généralement agréable est la transformation (ou l'une des transformations) de $\mathcal{F}_\mathrm{x}$ qui minimise la distance à l'identité :

$$\mathrm{f_x} \in \arg\min_{\mathrm{f} \in \mathcal{F}_\mathrm{x}} (\ \mathrm{dist}(\mathrm{f},\ \mathrm{Id}))$$

On a ainsi choisi la façon la plus rapide d'aller de l'origine o au point x[1]. Remarquons également que dans le cas d'un groupe d'isotropie discret et fini, on a ainsi obtenu la carte principale (section 3.4.4).

La démarche pour obtenir les géodésiques dans le cas général est très similaire au cas du groupe général : on suppose que l'on a une représentation minimale de la variété $\mathcal{M}$ centrée à l'origine et on détermine l'expression de l'action du groupe $y = \mathrm{f} \star x$ dans cette carte (pour x suffisamment proche de l'origine et f suffisamment proche de l'identité), et le jacobien de la translation de l'origine :

$$J(\mathrm{f}) = \left.\frac{\partial(\mathrm{f} \star x)}{\partial x}\right|_{x=o}$$

On peut alors enfin savoir s'il existe une métrique invariante par $\mathcal{G}$ sur $\mathcal{M}$: il existe alors une matrice symétrique définie positive Q vérifiant

$$\forall \mathrm{h} \in \mathcal{H} \qquad J(\mathrm{h})^{\mathrm{T}}.Q.J(\mathrm{h}) = Q$$

En supposant qu'une telle matrice existe, on construit alors la représentation locale de la métrique invariante $Q(x) = J(\mathrm{f}_x)^{(\text{-T})}.Q.J(\mathrm{f}_x)^{(\text{-1})}$ et on cherche à résoudre le système d'équations différentielles du second ordre définissant les géodésiques dans la carte locale. On peut se contenter des géodésiques partant de l'origine, les géodésiques partant de x étant obtenues grâce à la translation de transformation $\mathrm{f_x}$. On exprime alors l'équation des géodésiques directement dans la variété $\mathcal{M}$ et on vérifie qu'elles s'étendent sans singularité à $\mathbb{R}$ tout entier et que la variété est ainsi géodésiquement complète. La détermination de la carte principale est alors calquée sur le cas du groupe.

1. Il serait d'ailleurs intéressant de savoir dans quelles conditions, avec ce choix de $\mathrm{f_x}$, la courbe $\gamma(t) = (t.\vec{\mathrm{f}}_\mathrm{x}) \star \mathrm{o}$ est une géodésique de $\mathcal{M}$ pour la métrique invariante, et quel est le lien avec l'existence d'une métrique invariante. Cette propriété est vérifiée pour les primitives que nous utilisons dans ce manuscrit et pourrait s'avérer un moyen efficace pour déterminer les géodésiques de $\mathcal{M}$ si on a déterminé celles de $\mathcal{G}$.

6.2 Implémentation des opérations atomiques

Nous appelons opérations atomiques les opérations qui sont spécifiques à chaque type de primitive. Nous baserons tous nos algorithmes (qu'ils soient bas ou haut niveau) sur ces quelques opérations atomiques, ce qui nous permettra à tout moment de modéliser un problème avec un nouveau type de primitive en ayant à notre disposition tous les algorithmes déjà développés : il suffira pour cela de mener les calculs mathématiques nécessaires à l'implémentation des opérations atomiques. Certaines de ces opérations et cette façon de voir les choses étaient déjà pressenties dans (Smith et Cheeseman, 1987; Smith et al., 1988), mais ces travaux passent totalement sous silence les problèmes liés à la représentation choisie pour les primitives et les singularités que cela peut occasionner (par exemple avec les angles d'Euler).

Revenons pour l'instant à un type de primitive et un groupe de transformation : on suppose que l'on a déterminé leur cartes principales. Pour simplifier les expressions, on suppose de plus que ces cartes sont exprimées dans une base orthonormée pour la métrique (i.e. $Q = Id$).

Le premier type d'opérations important, que cela soit pour le groupe ou pour la variété, est la traduction entre les représentations classiques et la carte principale. Cela permet de relier sans problème la bibliothèque qui implémente cette théorie avec le reste du monde et constitue en quelque sorte son interface avec d'autres programmes. Nous aurons ainsi besoin de traduire un vecteur rotation en un quaternion unitaire ou en une matrice de rotation, et vice-versa. A l'intérieur de la bibliothèque, on ne travaillera cependant qu'avec les cartes principales.

6.2.1 Opérations atomiques sur le groupe

Il s'agit pour le groupe de la composition, l'inversion et leur jacobiens :

- **Composition :** $\vec{\mathrm{f}} \circ \vec{\mathrm{g}}$; $\dfrac{\partial(\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}{\partial \vec{\mathrm{f}}}$; $\dfrac{\partial(\vec{\mathrm{f}} \circ \vec{\mathrm{g}})}{\partial \vec{\mathrm{g}}}$
- **Inversion :** $\vec{\mathrm{f}}^{(-1)}$; $\dfrac{\partial \vec{\mathrm{f}}^{(-1)}}{\partial \vec{\mathrm{f}}}$

On pourra également envisager d'implémenter le jacobien des translations à gauche et à droite pour des raisons d'efficacité algorithmique.

- **Translation de l'identité :** $J_L(\vec{\mathrm{f}}) = \left.\dfrac{\partial(\vec{\mathrm{f}} \circ \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=0}$; $J_R(\vec{\mathrm{f}}) = \left.\dfrac{\partial(\vec{\mathrm{e}} \circ \vec{\mathrm{f}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=0}$

Enfin, nous utiliserons le filtrage de Kalman étendu pour certains problèmes d'estimation, et la mise à jour de l'état peut faire sortir celui-ci du domaine de définition de notre carte principale. Cela ne pose pas vraiment de problème théorique puisque l'exponentielle est définie sur tout l'espace tangent, mais plutôt un problème pratique puisque toutes nos opérations sont conçues pour travailler avec des valeurs comprises dans le domaine de la carte. Nous avons donc besoin d'une dernière opération qui consiste à ramener un vecteur $\vec{\mathrm{f}}'$ représentant une primitive f à sa valeur $\vec{\mathrm{f}}$ dans le domaine (ou à l'une de ses valeurs sur le lieu de coupure), en empruntant bien sûr la géodésique. Le jacobien de cette opération sera nécessaire pour pouvoir réajuster la matrice de covariance.

- **Domaine :** $\vec{\mathrm{f}}(\vec{\mathrm{f}}')$; $\dfrac{\partial \vec{\mathrm{f}}(\vec{\mathrm{f}}')}{\partial \vec{\mathrm{f}}'}$

6.2.2 Opérations atomiques sur la variété

En ce qui concerne la variété des primitives, nous avons l'action du groupe comme analogue de la composition et la fonction de placement ($\mathrm{f}_{\mathrm{x}}^{(-1)} \star \mathrm{o}$ pouvant être considéré comme un analogue de

l'inverse). Les jacobiens sont bien sûr nécessaires.

- **Action :** $\vec{\mathrm{f}} \star \vec{\mathrm{x}}$; $\dfrac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{f}}}$; $\dfrac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}}$
- **Fonction de placement :** $\vec{\mathrm{f}}_{\vec{\mathrm{x}}}$; $\dfrac{\partial \vec{\mathrm{f}}_{\vec{x}}}{\partial \vec{\mathrm{x}}}$

De même que pour le groupe, on peut envisager d'implémenter directement le jacobien de la translation de l'origine :

- **Translation de l'origine :** $J(\vec{\mathrm{f}}_{\vec{\mathrm{x}}}) = \left.\dfrac{\partial(\vec{\mathrm{f}}_{\vec{\mathrm{x}}} \star \vec{\mathrm{e}})}{\partial \vec{\mathrm{e}}}\right|_{\vec{\mathrm{e}}=0}$

et nous utiliserons le filtre de Kalman, donc nous aurons besoin de nous ramener dans le domaine :

- **Domaine :** $\vec{\mathrm{x}}(\vec{\mathrm{x}}')$; $\dfrac{\partial \vec{\mathrm{x}}(\vec{\mathrm{x}}')}{\partial \vec{\mathrm{x}}'}$

Dans notre bibliothèque, nous avons implémenté pour toutes ces opérations une version dite simple, sans les jacobiens, et une version calculant l'opération et ses jacobiens. Ceci permet de réduire au minimum le nombre de calculs de jacobiens et de multiplications matricielles, et ainsi de diminuer les temps de calcul (au détriment, mais dans une faible part, du volume du code).

6.3 Opérations de base sur les primitives probabilistes

Ayant construit et implémenté les fonction atomiques pour les transformations et tous les types de primitives qui nous intéressent, on peut s'abstraire du type de primitive utilisé et implémenter une fois pour toutes les opérations qui suivent. Le jour où l'on voudra rajouter une nouvelle primitive, il suffira de programmer ses fonctions atomiques et toutes les opérations et algorithmes décrits à partir de maintenant fonctionneront automatiquement.

Le groupe de transformation $\mathcal{G}$ étant une variété, il est clair qu'un grand nombre des opérations de base s'applique également au groupe. Nous développons donc dans la suite les opérations de base sur la variété, en ne les traduisant sur le groupe que si les différences ou les simplifications sont significatives.

6.3.1 Primitives probabilistes

Soit $\mathbf{x}$ une primitive aléatoire. Nous supposerons que sa moyenne de Fréchet

$$\mathbb{E}[\,\mathbf{x}\,] = \arg\min_{\mathrm{y}\in\mathcal{M}} \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y},\mathbf{x})^2\ \right]\right)$$

est unique et a pour expression $\mathbb{E}[\,\vec{\mathbf{x}}\,] = \vec{x}$ dans la carte principale. Sa matrice de covariance est donnée dans cette même carte par :

$$\Sigma_{\mathbf{xx}} = \mathbf{E}\left[\ \overrightarrow{\mathrm{x}\mathrm{x}}.\overrightarrow{\mathrm{x}\mathrm{x}}^{\mathrm{T}}\ \right] = J(\vec{\mathrm{f}}_{\vec{x}}).\mathbf{E}\left[\ (\vec{\mathrm{f}}_{\vec{x}}^{(-1)} \star \vec{\mathbf{x}}).(\vec{\mathrm{f}}_{\vec{x}}^{(-1)} \star \vec{\mathbf{x}})^{\mathrm{T}}\ \right].J(\vec{\mathrm{f}}_{\vec{x}})^{\mathrm{T}}$$

Ces paramètres sont en général suffisants pour gérer de manière informatique la primitive aléatoire et on définit donc un « objet géométrique aléatoire » (ou probabiliste) avec l'approximation :

$$\mathbf{x} \sim (\vec{x}, \Sigma_{\mathbf{xx}})$$

Une primitive probabiliste sera donc pour nous un objet composé d'un vecteur de dimension n (la dimension de la variété), d'une matrice de covariance $n \times n$, et d'un type indiquant de quelle

primitive il s'agit. A chaque type de primitive est associé son ensemble d'opérations atomiques. Pour être précis, on devrait subordonner le type de primitive au groupe de transformation que l'on considère, mais nous ne considérons dans notre implémentation que les transformations rigides. Il faut également noter que les transformations sont des objets d'une classe différente puisque certaines opérations différent.

Note sur l'implémentation Dans cette optique, une primitive exacte ou déterministe a simplement une covariance nulle. Elle peut donc être gérée exactement comme une primitive probabiliste, mais pour éviter des calculs de jacobiens longs et inutiles, on pourra rajouter dans la structure informatique de l'objet un drapeau indiquant si l'objet est déterministe ou probabiliste et ne faire les calculs reliés à la matrice de covariance que dans ce dernier cas.

Notons également que la matrice de covariance est symétrique. Si la variété est de dimension n, elle n'a donc que $n.(n+1)/2$ composantes libres. Afin d'éviter des calculs inutiles et de rester cohérent, il est souhaitable d'utiliser un codage de la matrice de covariance ne conservant que $n.(n+1)/2$ composantes indépendantes et de réécrire toutes les opérations sur la matrice de covariance directement dans ce codage. A titre d'information, les principales opérations sur les matrices symétriques (et surtout celles qui sont les plus coûteuses en temps de calcul) sont l'inversion, la diagonalisation et surtout « l'action » d'un jacobien : $\Lambda = J.\Sigma.J^{\mathrm{T}}$. Une optimisation de ces opérations dans le codage symétrique est donc de la plus haute importance pour obtenir des algorithmes haut niveau rapides sur les primitives aléatoires. Par ailleurs, la minimisation du nombre des opérations à effectuer (ici dans le sens de multiplication et addition de réels) apporte également un gain de précision numérique.

6.3.2 Opérations géométriques sur les primitives probabilistes

La première opération géométrique, c'est la distance entre deux primitives de même type. On ne précise pas ici les matrices de covariance puisqu'elles n'interviennent pas.

– **Distance** $\quad (\vec{\mathrm{x}} \;\; ; \;\; \vec{\mathrm{y}}) \qquad \longmapsto \qquad \mathrm{dist}(\mathrm{x},\mathrm{y}) = \sqrt{\left(\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}\right)^{\mathrm{T}} . \left(\vec{\mathrm{f}}_{\vec{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{y}}\right)}$

Ensuite, nous avons besoin de faire agir une transformation probabiliste sur une primitive probabiliste et, pour simplifier l'expression des algorithmes de plus haut niveau, de la « fonction de placement probabiliste » :

– **Action d'une transformation**

$$\left(\mathbf{f} \sim (\vec{f}, \Sigma_{\mathbf{ff}}) \quad ; \quad \mathbf{x} \sim (\vec{\bar{x}}, \Sigma_{\mathbf{xx}})\right) \qquad \longmapsto \qquad \mathbf{y} = \mathbf{f} \star \mathbf{x} \sim \left(\vec{f} \circ \vec{\bar{x}} \;,\; \Sigma_{\mathbf{yy}}\right)$$

avec

$$\Sigma_{\mathbf{yy}} = J_{\vec{\mathrm{f}}}.\Sigma_{\mathbf{ff}}.J_{\vec{\mathrm{f}}}^{\mathrm{T}} + J_{\vec{\mathrm{x}}}.\Sigma_{\mathbf{xx}}.J_{\vec{\mathrm{x}}}^{\mathrm{T}} \qquad \text{où} \quad J_{\vec{\mathrm{f}}} = \left.\frac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}}=\vec{f}} \quad \text{et} \quad J_{\vec{\mathrm{x}}} = \left.\frac{\partial(\vec{\mathrm{f}} \star \vec{\mathrm{x}})}{\partial \vec{\mathrm{x}}}\right|_{\vec{\mathrm{x}}=\vec{\bar{x}}}$$

– **Fonction de placement probabiliste**

$$\mathbf{x} \sim (\vec{\bar{x}}, \Sigma_{\mathbf{xx}}) \qquad \longmapsto \qquad \mathrm{f}_{\mathbf{x}} \sim \left(\vec{\mathrm{f}}_{\vec{\bar{x}}} \quad ; \quad J.\Sigma_{\mathbf{xx}}.J^{\mathrm{T}}\right) \qquad \text{avec} \qquad J = \left.\frac{\partial \vec{\mathrm{f}}_{\vec{\mathrm{x}}}}{\partial \vec{\mathrm{x}}}\right|_{\vec{\mathrm{x}}=\vec{\bar{x}}}$$

Pour implémenter la distance de Mahalanobis, nous procédons en deux étapes : une première fonction pour la distance de Mahalanobis simple par rapport à l'origine, et une seconde fonction pour la distance de Mahalanobis double approchée.

– **Distance de Mahalanobis à l'origine** $\quad \mathbf{z} \sim (\vec{\bar{z}}, \Sigma_{\mathbf{zz}}) \qquad \longmapsto \qquad \mu^2(\mathbf{z}, \mathrm{o}) = \vec{\bar{z}}^{\mathrm{T}}.\Sigma_{\mathbf{zz}}^{(-1)}.\vec{\bar{z}}$

– **Distance de Mahalanobis double**

$$\left(\mathbf{x} \sim (\vec{\hat{x}}, \Sigma_{\mathbf{xx}}) \quad ; \quad \mathbf{y} \sim (\vec{\hat{y}}, \Sigma_{\mathbf{yy}})\right) \qquad \longmapsto \qquad \nu^2(\mathbf{x}, \mathbf{y}) = \mu^2(\mathrm{f}_{\mathbf{y}}^{(-1)} \star \mathbf{x}, \mathrm{o})$$

En prenant comme convention que, s'il y a une primitive déterministe, c'est y (on inverse au besoin les arguments dans la procédure), la distance de Mahalanobis double implémente également la distance simple.

6.3.3 Opérations géométriques sur les transformations probabilistes

Toute ces opérations sont très similaires pour le groupe, excepté que l'action doit être remplacé par la composition et l'inverse de la fonction de placement par l'inversion :

– **Composition**

$$\left(\mathbf{f_1} \sim (\vec{\hat{f}}_1, \Sigma_{\mathbf{f_1 f_1}}) \quad ; \quad \mathbf{f_2} \sim (\vec{\hat{f}}_2, \Sigma_{\mathbf{f_2 f_2}})\right) \qquad \longmapsto \qquad \mathbf{g} = \mathbf{f_2} \circ \mathbf{f_1} \sim \left(\vec{\hat{f}}_2 \circ \vec{\hat{f}}_1 \, , \, \Sigma_{\mathbf{gg}}\right)$$

avec

$$\Sigma_{\mathbf{gg}} = J_{\vec{\mathrm{f}}_2}.\Sigma_{\mathbf{f_2 f_2}}.J_{\vec{\mathrm{f}}_2}^{\mathrm{T}} + J_{\vec{\mathrm{f}}_1}.\Sigma_{\mathbf{f_1 f_1}}.J_{\vec{\mathrm{f}}_1}^{\mathrm{T}} \qquad \text{où} \quad J_{\vec{\mathrm{f}}_2} = \left.\frac{\partial(\vec{\mathrm{f}}_2 \circ \vec{\mathrm{f}}_1)}{\partial \vec{\mathrm{f}}_2}\right|_{\vec{\mathrm{f}}_2 = \vec{\hat{f}}_2} \quad \text{et} \quad J_{\vec{\mathrm{f}}_1} = \left.\frac{\partial(\vec{\mathrm{f}}_2 \circ \vec{\mathrm{f}}_1)}{\partial \vec{\mathrm{f}}_1}\right|_{\vec{\mathrm{f}}_1 = \vec{\hat{f}}_1}$$

– **Inversion**

$$\left(\mathbf{f} \sim (\vec{\hat{f}}, \Sigma_{\mathbf{ff}})\right) \qquad \longmapsto \qquad \mathbf{f}^{(-1)} \sim \left(\vec{\hat{f}}^{(-1)} \, , \, J.\Sigma_{\mathbf{ff}}.J^{\mathrm{T}}\right) \qquad \text{avec} \qquad J = \left.\frac{\partial \vec{\mathrm{f}}^{(-1)}}{\partial \vec{\mathrm{f}}}\right|_{\vec{\mathrm{f}} = \vec{\hat{f}}}$$

Il est immédiat d'en déduire les procédures pour les calculs de distance.

6.3.4 Opérations statistiques sur les primitives

D'un point de vue statistique, nous avons plutôt une mesure $\hat{\mathrm{x}}$ et un modèle de bruit sur cette mesure, estimé par ailleurs. Un modèles de bruit isotrope étant représenté par une matrice de covariance Σ à l'origine vérifiant $J(\vec{\mathrm{h}}).\Sigma.J(\vec{\mathrm{h}})^{\mathrm{T}} = \Sigma$, on considérera notre mesure comme une primitive aléatoire

$$\mathbf{x} \sim \left(\vec{\hat{x}}, \Sigma_{\mathbf{xx}}\right) \qquad \text{avec} \qquad \Sigma_{\mathbf{xx}} = J(\vec{\mathrm{f}}_{\vec{\hat{x}}}).\Sigma.J(\vec{\mathrm{f}}_{\vec{\hat{x}}})^{\mathrm{T}}$$

où $\vec{\mathrm{f}}_{\vec{\mathrm{x}}}$ est notre fonction de placement par défaut (elle n'a pas ici d'importance).

Dans le cas d'un bruit homogène, il n'y a plus de contraintes sur la covariance à l'origine mais la fonction de placement est un paramètre du bruit. Dans notre implémentation, on se contente de la fonction de placement par défaut (d'où l'intérêt de bien la choisir au départ), mais on pourrait concevoir la fonction plus générique suivante :

– **Modélisation statistique** $\left(\vec{\hat{x}} \quad ; \quad \Sigma \quad ; \quad \vec{\mathrm{f}}_{\vec{\mathrm{x}}}\right) \qquad \longmapsto \qquad \mathbf{x} \sim \left(\vec{\hat{x}}, J(\vec{\mathrm{f}}_{\vec{\hat{x}}}).\Sigma.J(\vec{\mathrm{f}}_{\vec{\hat{x}}})^{\mathrm{T}}\right)$

A l'inverse, il est souvent utile de pouvoir interpréter le bruit sur une primitive aléatoire résultant d'une expérience, ne serait-ce que pour pouvoir estimer le modèle ou le niveau de bruit. Il est impossible d'estimer la fonction de placement avec une seule mesure et même très difficile de l'estimer avec plusieurs mesures (supposées distribuées identiquement et indépendamment). On la suppose donc connue (ou devinée) et la covariance à l'origine est donnée par la fonction :

– **Mesure du bruit** $\left(\mathbf{x} \sim (\vec{\hat{x}}, \Sigma_{\mathbf{xx}}) \quad ; \quad \vec{\mathrm{f}}_{\vec{\mathrm{x}}}\right) \qquad \longmapsto \qquad \Sigma = J(\vec{\mathrm{f}}_{\vec{\hat{x}}})^{(-1)}.\Sigma_{\mathbf{xx}}.J(\vec{\mathrm{f}}_{\vec{\hat{x}}})^{(-\mathrm{T})}$

L'action du groupe d'isotropie sur la carte principale étant une rotation (puisque l'on considère que $Q = Id$), le choix d'une autre fonction de placement n'occasionne en fait qu'une rotation de la covariance Σ à l'origine.

Deux autres opérations statistiques sont encore intéressantes à implémenter pour pouvoir juger de l'incertitude d'une primitive aléatoire et comparer par exemple différents modèles de bruit. Il s'agit de la variance $\sigma^2_{\mathbf{x}}$ et de l'information (approchée) de la distribution gaussienne associée :

- **Variance** $\quad \mathbf{x} \sim (\vec{\bar{x}}, \Sigma_{\mathbf{xx}}) \quad \longmapsto \quad \sigma^2_{\mathbf{x}} = \mathrm{Tr}\left(J(\vec{\mathrm{f}}_{\vec{\bar{x}}})^{(\text{-}1)}.\Sigma_{\mathbf{xx}}.J(\vec{\mathrm{f}}_{\vec{\bar{x}}})^{(\text{-}\mathrm{T})}\right)$
- **Information**

$$(\vec{\bar{x}}, \Sigma_{\mathbf{xx}}) \quad \longmapsto \quad \mathbf{I}[\,\mathbf{x}\,] = -\frac{n}{2}(1 + \log(2.\pi)) + \log\left(|J(\vec{\mathrm{f}}_{\vec{\mathrm{x}}})|\right) - \frac{1}{2}.\log(|\Sigma|)$$

6.3.5 Opérations statistiques sur les transformations

Le groupe d'isotropie étant réduit à l'identité, nous n'avons pas ici de problème de fonction de placement et tous les bruits homogènes sont isotropes. La modélisation et l'interprétation d'une transformation probabiliste sont donc simplifiées en :

- **Modélisation statistique** $\quad \left(\vec{f} \quad ; \quad \Sigma\right) \quad \longmapsto \quad \mathbf{f} \sim \left(\vec{f}, J_L(\vec{f}).\Sigma.J_L(\vec{f})^{\mathrm{T}}\right)$
- **Mesure du bruit** $\quad \left(\mathbf{f} \sim (\vec{f}, \Sigma_{\mathbf{ff}})\right) \quad \longmapsto \quad \Sigma = J_L(\vec{f})^{(\text{-}1)}.\Sigma_{\mathbf{ff}}.J(\vec{f})^{(\text{-}\mathrm{T})}$

La mesure de la variance et de l'information se déduisent de manière similaire.

6.4 Quelques algorithmes moyen niveau

6.4.1 Primitive moyenne et covariance

Supposons que l'on ait une série de m mesures x_i que l'on considère comme des réalisations indépendantes d'une même primitive aléatoire $\mathbf{y} \sim (\bar{\mathrm{y}}, \Sigma_{\mathbf{yy}})$. On peut déterminer une estimation de la primitive moyenne grâce à l'algorithme de descente de gradient développé à la section (4.3.2) :

- Initialiser l'estimation avec une mesure quelconque, disons la première : $\vec{\bar{y}}_0 = \vec{\mathrm{x}}_1$ (l'indice de l'estimation se réfère à l'itération tandis que celui des mesures se réfère au numéro de la mesure).
- Calculer l'estimation au temps $t+1$:

$$\vec{\bar{y}}_{t+1} = \vec{\mathrm{f}}_{\vec{\bar{y}}_t} \star \left(\frac{1}{m}\sum_i \left(\vec{\mathrm{f}}^{(\text{-}1)}_{\vec{\bar{y}}_t} \star \vec{\mathrm{x}}_i\right)\right)$$

- Un critère d'arrêt cohérent est une distance entre deux estimations suffisamment faible : $\mathrm{dist}(\vec{\bar{y}}_t, \vec{\bar{y}}_{t+1}) < \varepsilon$.

On peut alors estimer la covariance sur la primitive aléatoire $\mathbf{y}$ qui a généré les mesures par

$$\Sigma_{\mathbf{yy}} = \frac{1}{m-1}.J(\vec{\mathrm{f}}_{\vec{\bar{y}}}).\left(\sum_{i=1}^{m}(\vec{\mathrm{f}}^{(\text{-}1)}_{\vec{\bar{y}}} \star \vec{\mathrm{x}}_i).(\vec{\mathrm{f}}^{(\text{-}1)}_{\vec{\bar{y}}} \star \vec{\mathrm{x}}_i)^{\mathrm{T}}\right).J(\vec{\mathrm{f}}_{\vec{\bar{y}}})^{\mathrm{T}}$$

La normalisation par $\frac{1}{m-1}$ au lieu de $\frac{1}{m}$ est classique en statistiques et compense le biais introduit l'utilisation d'une *estimée* de la valeur moyenne au lieu de la valeur moyenne exacte (voir par exemple (Anderson, 1958; Bard, 1974) ou (Koch, 1988)).

Il faut bien faire attention que la covariance $\Sigma_{\mathbf{yy}}$ que nous venons de calculer n'est pas représentative de l'incertitude sur la primitive moyenne estimée, mais caractérise le processus qui a donné lieu aux mesures. Cet algorithme sera repris et implémenté au chapitre 8 en compagnie d'algorithmes de fusion et de recalage du même type.

6.4.2 Génération de primitives aléatoires

« The generation of random numbers is too important to be left to chance. »
Robert R. Coveyou, Oak Ridge National Laboratory

Pour pouvoir faire des expériences avec des données synthétiques, nous avons deux problèmes de génération aléatoire. Supposons pour l'instant que nous ayons un ensemble de primitives données $\{\vec{\mathrm{x}}_i\}$ représentant un objet dans une image. Afin de simuler le processus d'acquisition de l'image, on applique éventuellement une transformation f à cet objet pour obtenir les primitives (exactes) $\{\vec{\mathrm{y}}_i\}$ dans la seconde image et on veut bruiter chaque primitive indépendemment selon un certain modèle de bruit, le bruit gaussien étant le plus adapté puisqu'il minimise l'information en ne connaissant que la moyenne et la covariance.

Le second problème apparaît lorsque l'on veut simuler des faux positifs, c'est-à-dire des primitives qui sont là « par hasard » ou s'affranchir de la forme de notre objet et partir d'un « objet aléatoire ». Le modèle aléatoire le plus adapté est alors la distribution uniforme *dans l'image*.

Génération aléatoire gaussienne En supposant que notre modèle de bruit soit homogène et centré, il est caractérisé par une fonction de placement $\mathrm{f_x}$ et une covariance à l'origine Σ. Le modèle de bruit pour la primitive i est donc : $\vec{\mathbf{y}}_i = \vec{\mathrm{f}}_{\mathrm{y}_i} \star \vec{\mathbf{e}}_i$ où tous les $\vec{\mathbf{e}}_i$ sont des bruits indépendants de moyenne nulle et de covariance Σ dans la carte principale. La gaussienne classique étant une bonne approximation de la distribution minimisant l'information, le problème se ramène donc au tirage de vecteurs aléatoires gaussiens. L'algorithme est donc le suivant pour chaque primitive (on oublie ici l'indice) :

- Tirer un vecteur aléatoire $\vec{\mathrm{e}}$ suivant la loi gaussienne approchée $N_{(0,\Sigma^{(-1)})}$.
- Vérifier si ce vecteur est dans le domaine de la carte principale (en utilisant l'opération atomique « domaine »). S'il est en dehors en retirer un autre.
- Le « vecteur » $\vec{\mathrm{e}}$ représente l'erreur de mesure sur l'origine. Il ne reste plus qu'à le placer comme un erreur de mesure sur la primitive y : $\hat{\vec{y}} = \vec{\mathrm{f}}_{\vec{\mathrm{y}}} \star \vec{\mathrm{e}}$.

De manière plus générale, on peut ainsi générer des primitives aléatoires ayant la loi gaussienne approchée $N_{(\mathrm{z},\Sigma^{(-1)}_{\mathbf{zz}})}$: il suffit de se ramener à l'origine pour travailler dans la carte principale et d'appliquer l'algorithme précédent. Si $\vec{\mathrm{f}}_{\vec{z}}$ est la fonction de placement utilisée au point z, la covariance à utiliser dans l'algorithme précédent est :

$$\Sigma = J(\vec{\mathrm{f}}_{\vec{z}})^{(-1)}.\Sigma_{\mathbf{zz}}.J(\vec{\mathrm{f}}_{\vec{z}})^{(-\mathrm{T})}$$

Il ne reste qu'à préciser comment on peut tirer un vecteur aléatoire $\boldsymbol{x}$ de loi $N_{(0,\Sigma^{(-1)})}$:

- Diagonaliser la covariance : $\Sigma = R^{\mathrm{T}}.\Lambda.R$, où R est une matrice de rotation n-D et Λ est la matrice diagonale des valeurs propres.

- Comme Σ est positive, les valeurs propres sont positives et on peut en prendre une racine carrée : $\Lambda = D^2$. On note $A = D.R$ la « racine carrée » de Σ. Le vecteur aléatoire $\boldsymbol{y} = A^{(-1)}.\boldsymbol{x}$ doit alors suivre la loi normale réduite à n dimensions $N_{(0,\, Id_n)}$.
- Tirer les n composantes du vecteur aléatoire $\boldsymbol{y}$ indépendamment selon la loi normale unidimensionnelle $N_{(0,1)}$ pour obtenir $\hat{y}$.
- Retourner le vecteur $\hat{x} = A.\hat{y}$.

Rappelons qu'on peut obtenir un tirage $\boldsymbol{a}$ de loi normale $N_{(0,1)}$ à partir de deux tirages $\boldsymbol{u}_1$ et $\boldsymbol{u}_2$ uniformes sur $[0, 1]$ par la formule :

$$\boldsymbol{a} = \sqrt{-2.\ln(\boldsymbol{u}_1)}.\cos(2.\pi.\boldsymbol{u}_2)$$

Génération aléatoire uniforme Le problème du tirage uniforme est plus complexe et nécessiterait théoriquement une procédure spécifique pour chaque type de primitive. Comme nous nous intéressons ici à des primitives 3D définies à partir de points de l'image 3D (comme les les repères), nous nous contentons d'une procédure de tirage uniforme sur les transformations rigides 3D en suivant ce protocole : on tire la translation comme un point 3D uniforme dans l'image et une rotation 3D uniforme. On applique ensuite cette transformation aléatoire uniforme $\vec{\mathrm{f}}$ à la primitive origine pour obtenir la primitive aléatoire $\vec{\mathrm{y}} = \vec{\mathrm{f}} \star \vec{\mathrm{o}}$. L'idée générale pour voir que l'on obtient une distribution uniforme est de considérer la mesure $d\mathcal{G}$ sur le groupe comme $d\mathcal{G} = d\mathcal{M}.d\mathcal{H}$ (puisque nous avons une mesure invariante). Comme les intégrales sont effectuées sur des compacts (l'image est finie), on obtient bien une primitive aléatoire uniforme dans l'image.

Cette procédure de génération aléatoire uniforme est un peu *ad hoc* et demanderait une étude plus poussée pour pouvoir être exprimée génériquement dans le même cadre que le reste de nos opérations. Néanmoins, elle fournit dans notre cas les résultats nécessaires à des expérimentations statistiques synthétiques fiables.

6.5 Conclusion

Nous avons montré dans la synthèse comment la théorie développée dans cette première partie du manuscrit pouvait être appliquée et implémentée en machine dans une structure orientée objet générique, ne dépendant pas du type de primitive considéré. Nous suivrons ainsi au chapitre 7 cette méthodologie pour dériver les développements mathématiques nécessaires à l'expression de la carte principale et des quelques opérations atomiques associées aux primitives qui nous intéressent : repères, repères semi et non-orientés, points, points.

Cette structure nous permettra ensuite de concevoir relativement simplement d'autres algorithmes de moyen niveau et de haut niveau sur les primitives, par exemple concernant le recalage et sa validation aux chapitres 8, 9 et 12, ou encore la reconnaissance aux chapitres 10 et 11.

Nous pensons que cette structure orientée objet pourrait être étendue aux variétés homogènes ne possédant pas de métrique invariante mais simplement une connexion invariante moyennant quelques modifications des opérations atomiques. Ceci permettrait alors l'utilisation directe des algorithmes de haut niveau pour ces nouveaux types de primitives.

Deuxième partie

Reconnaissance, Recalage et Applications

Chapitre 7

Primitives rigides en 3D

« Mathematicians have long since regarded it as demeaning to work on problems related to elementary geometry in two or three dimensions, in spite of the fact that it it precisely this sort of mathematics which is of practical value. »
B. Grünbaum and G. C. Shephard,
Handbook of Applicable Mathematics.

La théorie de l'incertitude sur les variétés que nous avons développée jusqu'ici nécessite l'expression de la carte exponentielle et des opérations atomiques associées. Nous étudierons tout d'abord dans ce chapitre les rotations vectorielles de $\mathbb{R}^3$. Cette étude ne suit pas exactement et dépasse même le cadre de la synthèse développée au chapitre précédent (elle a été écrite antérieurement), mais elle peut ainsi permettre une autre approche des techniques génériques présentées dans la partie théorique.

A partir du vecteur rotation et de ses opérations atomiques, nous pourrons développer relativement simplement les calculs nécessaires aux transformations rigides, donc aux repères. Nous envisagerons alors le cas des repères semi et non orientés, et nous conclurons par les points.

7.1 Rotations vectorielles de $\mathbb{R}^3$

La littérature concernant les rotations 3D est abondante et diversifiée. D'un point de vue mathématique, nous ne citerons que (Altmann, 1986), (Kanatani, 1990, chap. 3 & 6) et (Faugeras, 1993) qui fournissent des synthèses assez complètes. On pourra également consulter la littérature sur les quaternions indiquée à la section (7.2), qui a aussi influencé l'écriture de cette section, ainsi que la littérature sur les représentations des rotations évoquée à la section (7.3).

7.1.1 Definition

Let $\{i, j, k\}$ be a right handed orthonormal basis of the euclidean space $\mathbb{R}^3$, and $\mathcal{B} = \{e_1, e_2, e_3\}$ be a set of three vectors. The linear mapping from $\{i, j, k\}$ to $\mathcal{B}$ is given by the matrix

$$R = [e_1, e_2, e_3] = \begin{bmatrix} e_{11} & e_{21} & e_{31} \\ e_{12} & e_{22} & e_{32} \\ e_{13} & e_{23} & e_{33} \end{bmatrix} \tag{7.1}$$

If we now want $\mathcal{B}$ to be an orthonormal basis, the matrix R verifies

$$R.R^{\mathrm{T}} = R^{\mathrm{T}}.R = I_3 \quad \text{and} \quad \det(R) = +1 \tag{7.2}$$

which implies $\det(R) = \pm 1$ and gives rise to the group $O(3)$ of orthogonal matrices. If we want $\mathcal{B}$ to be also right handed, we have to impose $\det(R) = 1$. Such matrices are called *rotations*. Each right handed orthonormal basis can then be represented by a unique rotation and conversely each rotation maps $\{i, j, k\}$ onto a unique right handed orthonormal basis.

Rotations can also be seen as a subset of linear maps of $\mathbb{R}^3$. They correspond in this case to their usual interpretation as transformations of $\mathbb{R}^3$: they are positive isometries (maps conserving orientation and dot product): $\langle R.x \mid R.y \rangle = \langle x \mid y \rangle$. In particular, they conserve the length of a vector: $||R.x|| = ||x||$. With the composition law, and I_3 as identity, rotations forms a non-commutative group denoted by $\mathcal{SO}_3$ (*3D rotation group*).

The three main operations on rotations are:

- the composition of R_1 and R_2: $R = R_2.R_1$ (beware of the order),
- the inverse of R: $R^{(-1)} = R^{\mathrm{T}}$,
- and the application of R to a vector x: $y = R.x$.

7.1.2 Geometric parameters: axis and angle

Let R be a 3-D rotation matrix. It is characterized by its axis n (unit vector) and its angle θ. The relationship between these two representations are given by the Rodrigues formula (see for instance (Kanatani, 1993)).

$$R = Id + \sin\theta.S_n + (1 - \cos\theta).S_n^2 = \cos\theta.\, Id + \sin\theta.S_n + (1 - \cos\theta).n.n^{\mathrm{T}} \tag{7.3}$$

The matrix S_n is the skew matrix corresponding to (left) cross product: for all vector v we have $S_n.v = n \times v$. If the coordinates of n are (n_x, n_y, n_z), the matrix S_n is

$$S_n = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix} \tag{7.4}$$

and we have the relation $S_n^2 = n.n^{\mathrm{T}} - Id$, used to derive the second part of equation (7.3). Note that S_n uniquely determines the vector n. Conversely, let $\mathrm{Tr}(R)$ be the trace of R; the parameters are:

$$\theta = \arccos\left(\frac{\mathrm{Tr}(R) - 1}{2}\right) \quad \text{and} \quad S_n = \frac{R - R^{\mathrm{T}}}{2.\sin\theta} \tag{7.5}$$

The last equation is valid only when $\theta \in]0; \pi[$. Indeed, for $\theta = 0$ (i.e. identity) the rotation axis n is not determined, and $\sin(\theta) = 0$ for reflections (i.e. when $\theta = \pi$).

7.1.2.1 R close to identity: θ is small

Since the axis n is not defined for identity, there is a singularity and a numerical instability around it. However, we can compute the rotation vector with a Taylor expansion:

$$S_r = \theta\, S_n = \frac{\theta}{2\sin\theta}.(R - R^{\mathrm{T}}) = \frac{1}{2}.\left(1 + \frac{\theta^2}{6}\right).(R - R^{\mathrm{T}}) + O\left(\theta^4\right)$$

7.1.2.2 R close to a reflection: $\pi - \theta$ is small

The axis is this time well defined, but we have to use another equation. From Rodrigues formula, we get $R + R^{\mathrm{T}} - 2.\,Id = 2.(1 - \cos\theta).S_n^2$, and since $S_n^2 = n.n^{\mathrm{T}} - \,Id$, we have

$$n.n^{\mathrm{T}} = \,Id + \frac{1}{2.(1-\cos\theta)}.\,(R + R^{\mathrm{T}} - 2.\,Id)$$

Let $\varrho = 1/(1 - \cos\theta)$; taking diagonal terms gives

$$n_i^2 = 1 + \varrho.(R_{i,i} - 1) \qquad \Rightarrow \qquad n_i = \varepsilon_i\sqrt{1 + \varrho.(R_{i,i} - 1)}$$

The off diagonal terms are used to determine the signs ε_i: considering that the sign of n_1 is $\varepsilon \in \{-1; +1\}$, we can compute that

$$\mathrm{sign}(n_k) = \varepsilon.\mathrm{sign}(R_{1,k} + R_{k,1})$$

If we have an exact reflection, the sign ε does not matter since rotating clockwise or counter-clockwise gives the same result, but for a quasi-reflection, this sign is important. In this case, the vector $w = 2.\sin\theta.n$ is very small but not identically null: it can be computed without numerical instabilities with $S_w = R - R^{\mathrm{T}}$. Since $\theta < \pi$, the largest component w_k in absolute value of this vector must have the same sign as the corresponding component n_k of vector n.

7.1.3 Differential properties

The orthogonal group $\mathcal{O}_3$ is defined as a subspace of $\mathbb{R}^{3\times3}$ by equation (7.2) which give rise to 6 independent scalar equations since $R.R^{\mathrm{T}}$ is symmetric. Hence $\mathcal{O}_3$ is a differential manifold of dimension 3. Taking into account the constraint $\det(R) = 1$ amounts to keep the component of identity. Since the composition and inversion maps are infinitely differentiable, $\mathcal{SO}_3$ is moreover a Lie group.

7.1.3.1 Tangent space

Let $R(t)$ be a curve on $\mathcal{SO}_3$: constraint (7.2) is differentiated into

$$\frac{dR}{dt}.R^{\mathrm{T}} + \left(\frac{dR}{dt}.R^{\mathrm{T}}\right)^{\mathrm{T}} = 0 \qquad \text{or} \qquad R^{\mathrm{T}}.\frac{dR}{dt} + \left(R^{\mathrm{T}}.\frac{dR}{dt}\right)^{\mathrm{T}} = 0$$

which means that $\dfrac{dR}{dt}.R^{\mathrm{T}}$ and $R^{\mathrm{T}}.\dfrac{dR}{dt}$ are skew matrices. Hence

$$\exists\, \omega_r, \omega_l \in \mathbb{R}^3 \qquad \text{such as} \qquad \frac{dR}{dt} = S_{\omega_r}.R = R.S_{\omega_l} \tag{7.6}$$

In particular, if $R = I_3$, the derivative is S_ω:

Théorème 7.1 *the tangent space $T\mathcal{SO}_3$ of $\mathcal{SO}_3$ at identity is the vectorial space of skew matrices, isomorphic to $\mathbb{R}^3$. The tangent space $T_R\mathcal{SO}_3$ at R is given by*

$$T_R\mathcal{SO}_3 = \{S_\omega.R/\omega \in \mathbb{R}^3\} = \{R.S_\omega/\omega \in \mathbb{R}^3\}$$

Two important and canonical maps on a Lie group are very useful for studying the tangent space: these are the left and right translations. In the case of $\mathcal{SO}_3$ they are the left and right composition by a fixed rotation R_0

$$\begin{array}{rccc} & \mathcal{SO}_3 & \longrightarrow & \mathcal{SO}_3 \\ L_{R_0}: & R & \longmapsto & L_{R_0}(R) = R_0.R \\ R_{R_0}: & R & \longmapsto & R_{R_0}(R) = R.R_0 \end{array} \tag{7.7}$$

Their differentials realize canonical isomorphisms between the tangent spaces of $\mathcal{SO}_3$ at different points. The differential of a function φ maps the tangent vector of any curve $\gamma(t)$ to the tangent vector of the curve $\varphi(\gamma(t))$. Let X be a vector of $T_R\mathcal{SO}_3$ and $\gamma(t)$ a curve on $\mathcal{SO}_3$ defined around R with $\gamma(0) = R$ and having the tangent vector $\frac{d\gamma}{dt} = X$. The left translation of γ is the curve $\gamma_1(t) = R_0.\gamma(t)$ and its tangent vector at 0 is $Y = \frac{d\gamma_1}{dt} = R_0.X$. The differential $L^*_{R_0}$ of L_{R_0} is then for any R in $\mathcal{SO}_3$

$$\begin{array}{rccc} & T_R\mathcal{SO}_3 & \longrightarrow & T_{R_0.R}\mathcal{SO}_3 \\ L^*_{R_0}: & X & \longmapsto & L^*_{R_0}(X) = R_0.X \end{array}$$

The differential $R^*_{R_0}$ is defined in the same way.

In particular, if $R = I_3$, we have two canonical isomorphisms between the tangent space at identity $T\mathcal{SO}_3$ and the tangent space $T_R\mathcal{SO}_3$ at any point R (since the formulations of $L^*_{R_0}$ and $R^*_{R_0}$ and independent of R, we denote their restriction to $R = I_3$ by the same names).

$$\begin{array}{rccc} & T\mathcal{SO}_3 & \longrightarrow & T_R\mathcal{SO}_3 \\ L^*_R: & S_\omega & \longmapsto & R.S_\omega \\ R^*_R: & S_\omega & \longmapsto & S_\omega.R \end{array} \tag{7.8}$$

Differential properties of rotations are then completely reducible to differential properties around identity in $T\mathcal{SO}_3$.

7.1.3.2 Structure of $T\mathcal{SO}_3$: the Lie algebra of $\mathcal{SO}_3$

We already know that $T\mathcal{SO}_3$ is the vectorial space of skew symmetric matrices, identifiable with $\mathbb{R}^3$. If we now consider that $(\mathbb{R}^3, +, \times)$ is an algebra ($\times$ is the cross product), we can induce an algebra on $T\mathcal{SO}_3$. Let $X = S_{\omega_1}$ and $Y = S_{\omega_2}$ be two vectors of $T\mathcal{SO}_3$, then the vector $Z = S_{(\omega_1 \times \omega_2)}$ belongs to $T\mathcal{SO}_3$ and is called the bracket of X and Y:

$$Z = S_{(\omega_1 \times \omega_2)} = S_{\omega_1}.S_{\omega_2} - S_{\omega_2}.S_{\omega_1} = X.Y - Y.X = [X, Y]$$

Hence $(T\mathcal{SO}_3, +, [.,.])$ is an algebra. It is in fact the Lie algebra of the group $\mathcal{SO}_3$. The next step is to give a metric to the group: consider the following euclidean dot products on matrices and on $\mathbb{R}^3$:

$$\langle M_1 \mid M_2 \rangle_{\mathbb{R}^{n\times m}} = \mathrm{Tr}(M_1^{\mathrm{T}}.M_2) = \mathrm{Tr}(M_2.M_1^{\mathrm{T}}) \qquad \text{and} \qquad \langle x \mid y \rangle_{\mathbb{R}^3} = x^{\mathrm{T}}.y = \mathrm{Tr}(x.y^{\mathrm{T}})$$

They induces on $T\mathcal{SO}_3$ the dot product

$$\langle S_{\omega_1} \mid S_{\omega_2} \rangle_{T\mathcal{SO}_3} \stackrel{def}{=} \frac{1}{2}.\mathrm{Tr}(S_{\omega_1}^{\mathrm{T}}.S_{\omega_2}) = \langle \omega_1 \mid \omega_2 \rangle_{\mathbb{R}^3}$$

The dot product on matrices can also be used on any tangent space : let $X = S_{\omega_l}.R = R.S_{\omega_r}$ and $Y = S_{\omega'_l}.R = R.S_{\omega'_r}$ be two vectors of $T_R\mathcal{SO}_3$. Their dot products is

$$\langle X \mid Y \rangle_{T_R\mathcal{SO}_3} = \left\langle R.S_{\omega_r} \;\middle|\; R.S_{\omega'_r} \right\rangle_{T_R\mathcal{SO}_3} = \frac{1}{2}.\mathrm{Tr}(S_{\omega_r}^{\mathrm{T}}.R^{\mathrm{T}}.R.S_{\omega'_r}) = \left\langle S_{\omega_r} \;\middle|\; S_{\omega'_r} \right\rangle_{T\mathcal{SO}_3}$$

or equivalently

$$\langle X \mid Y \rangle_{T_R\mathcal{SO}_3} = \left\langle S_{\omega_l}.R \;\middle|\; S_{\omega'_l}.R \right\rangle_{T_R\mathcal{SO}_3} = \frac{1}{2}.\mathrm{Tr}(S_{\omega'_l}.R.R^{\mathrm{T}}.S_{\omega_l}^{\mathrm{T}}) = \left\langle S_{\omega_l} \;\middle|\; S_{\omega'_l} \right\rangle_{T\mathcal{SO}_3}$$

Hence the metric on $T_R\mathcal{SO}_3$ is invariant by left or right translation from $T\mathcal{SO}_3$. Such a metric is called a bi-invariant Riemannian metric.

Théorème 7.2 *The tangent space of $\mathcal{SO}_3$ at identity $T\mathcal{SO}_3$ is the vectorial space of skew symmetric matrices. With the Lie bracket $[.,.]$ and the euclidean metric $\langle X \mid Y \rangle = \frac{1}{2}.Tr(X^T.Y)$, it forms a metric algebra which is canonically isomorphic to $(\mathbb{R}^3, +, \times, \langle . \mid . \rangle_{\mathbb{R}^3})$.*

The tangent space $T_R\mathcal{SO}_3$ at R is transported from $T\mathcal{SO}_3$ with the left or right translation by equation (7.8).

7.1.4 Exponential map

7.1.4.1 Integral curves

Let $R_X(s)$ be a one parameter subgroup of $\mathcal{SO}_3$ (homomorphism from $(\mathbb{R}, +)$ to $(\mathcal{SO}_3, .)$. This is a continuous curve which is also a subgroup of $\mathcal{SO}_3$. By definition and since $(\mathbb{R}, +)$ is commutative

$$R_X(s+t) = R_X(s).R_X(t) = R_X(t).R_X(s)$$

This means in particular that $R_X(t)$ and $R_X(s)$ commute: they have the same rotation axis n. $R_X(s)$ is thus a rotation of axis n and angle $\theta(s)$. We will denote by $\mathcal{R}(n, \theta)$ such a rotation. Reporting this in the definition, we find $\theta(s+t) = \theta(s) + \theta(t)$ and hence $\theta(s) = \lambda.s$ with some $\lambda \in \mathbb{R}$. Computing the derivatives, we find

$$\left.\frac{dR_X}{ds}\right|_0 = X = \lambda.S_n \in T\mathcal{SO}_3 \qquad \text{and} \qquad \left.\frac{dR_X}{ds}\right|_s = X.R_X(s) = R_X(s).X \in T_{R_X}\mathcal{SO}_3$$

We established that to each one-parameter subgroup $R_X(s) = \mathcal{R}(n, \lambda.s)$ of $\mathcal{SO}_3$ corresponds a unique vector $X = \lambda.S_n$ of $T\mathcal{SO}_3$. The converse is also true and $R_X(s)$ is called the integral curve of X. It is to be noted that X and $\lambda.X$ generate the same integral curve with proportional parameterizations.

This is a particular case of a more general theorem for Lie groups which state that there is a one to one correspondence between one parameter subgroups of the Lie group and one dimensional sub-algebras of its Lie algebra.

7.1.4.2 Exponential map

Let $X = \theta.S_n$ with $\|n\| = 1$ be a vector of $T\mathcal{SO}_3$ and $R_X(s) = \mathcal{R}(n, \theta.s)$ the integral curve of X. The exponential map[1] is defined as the map from $T\mathcal{SO}_3$ to $\mathcal{SO}_3$ which assigns $R_X(1)$ to X. Hence

$$\exp(X) = \exp(\theta.S_n) = \mathcal{R}(n, \theta)$$

1. The name "exponential map" will be justified with the metric properties: integral curves are geodesics.

Another exponential map, the matrix exponential, can be defined for X as the limit of the series $Id + X/1! + X^2/2! + \ldots$. Since X is a skew matrix, $X^3 = -\theta^2.X$ and the series reduced to

$$e^X = I_3 + \frac{\sin\theta}{\theta}.X + \frac{1-\cos\theta}{\theta^2}.X^2 = I_3 + \sin\theta.S_n + (1-\cos\theta).S_n^2 = \mathcal{R}(n,\theta)$$

Hence both exponentials turn out to be the same and using the isomorphism between $\mathbb{R}^3$ and $T\mathcal{SO}_3$ we can define the exponential of the **rotation vector** $r = \theta.n$:

$$\mathcal{R}(r) = \exp(S_r) = \mathcal{R}(n,\theta) \qquad \text{with} \qquad \theta = \|r\| \qquad \text{and} \qquad n = \frac{r}{\theta}$$

The exponential map is a sort of "development" of $\mathcal{SO}_3$ onto its tangent space $T\mathcal{SO}_3$: each one-dimensional subspace $\mathbb{R}.S_n$ is mapped on its integral curve $\mathcal{R}(n,\mathbb{R})$ and we will see that the length along these curves are conserved.

7.1.5 Metric properties

7.1.5.1 Definitions

Let $\gamma : s \in [a,b] \subset \mathbb{R} \mapsto \gamma(s) \in \mathcal{SO}_3$ be a piecewise curve on $\mathcal{SO}_3$ and $\dot\gamma$ the tangent vector of γ at s. The *length* of the curve γ is defined by

$$\mathcal{L}(\gamma) = \int_a^b \sqrt{\langle \dot\gamma(s) \mid \dot\gamma(s) \rangle}.ds \tag{7.9}$$

Let now Γ be the set of curves joining rotations R_1 and R_2. The map

$$\rho : \begin{array}{ccc} \mathcal{SO}_3 \times \mathcal{SO}_3 & \longrightarrow & \mathbb{R} \\ (R_1, R_2) & \longmapsto & \inf_{\gamma\in\Gamma} \mathcal{L}(\gamma) \end{array} \tag{7.10}$$

is the canonical metric on $\mathcal{SO}_3$. The curves γ minimizing the criterion $\mathcal{L}(\gamma)$ are called geodesics. Since the metric comes from a bi-invariant metric on $T\mathcal{SO}_3$, the length criterion (7.9) is invariant by left and right translations and finding geodesics amounts to find geodesics starting from identity.

7.1.5.2 Geodesics and metric on $\mathcal{SO}_3$

For a Lie group with a bi-invariant Riemannian metric, it turns out that geodesics starting from identity, one-parameter subgroups and integral curves (starting also from identity) are three different approaches for the same curves (Spivak, 1979, chap.10). Let $\gamma_X(s) = \exp(\lambda.s.S_n)$ be such a curve. Its derivative at identity is $\dot\gamma_X(0) = X = \lambda.S_n$ and $\dot\gamma_X(s) = X.\gamma_X(s) = \gamma_X(s).X$ elsewhere. Hence

$$\langle \dot\gamma(s) \mid \dot\gamma(s) \rangle = \langle X \mid X \rangle = \lambda^2.\langle n \mid n \rangle = \lambda^2$$

and the distance from identity to $\mathcal{R}(n,\theta) = \gamma_X(\theta/\lambda)$ is

$$\rho(I_3, \mathcal{R}(n,\theta)) = \int_0^{\theta/\lambda} \sqrt{\lambda^2}.ds = \theta$$

With an arc-length parameterization, we obtain $\gamma_{S_n}(\theta) = \mathcal{R}(n,\theta) = \exp(\theta.S_n)$. These curves are 2π-periodic and $\gamma_{S_n}(\theta) = \gamma_{(-S_n)}(-\theta)$. Hence shortest paths (or minimizing geodesics) are uniquely defined for $\theta < \pi$ and doubly defined for $\theta = \pi$.

Théorème 7.3 *The canonical metric on $\mathcal{SO}_3$ is given by*

$$\rho(R_1,R_2) = \rho(I_3, R_1^T.R_2) = \theta(R_1^T.R_2) = \arccos\left(\frac{(Tr(R_1^T.R_2) - 1)}{2}\right)$$

Geodesics of $\mathcal{SO}_3$ starting from identity are the curves $\theta \mapsto \mathcal{R}(n,\theta) = \exp(\theta.S_n)$. Shortest paths from identity to the non reflection rotation $R = \exp(\theta.S_n)$ $(0 \leq \theta < \pi)$ are given by

$$t \in [0,\theta] \mapsto \exp(t.S_n)$$

Shortest paths from identity to reflection $R = \exp(\pi S_n)$ are doubly defined by the above formula $(\theta = \pi)$ with n and $-n$ as unit vectors. Other geodesics are obtained by left or right translation.

7.2 Quaternions

« The invention of the calculus of quaternions is a step towards the knowledge of quantities related to space which can only be compared for its importance, with the invention of triple coordinates by Descartes. The ideas of this calculus, as distinguished from its operations and symbols, are fitted to be one of the greatest use in all parts of science. »
J.C. Maxwell, Proceedings of the London Mathematical Society 3, 1869

Avant de définir proprement la carte principale et d'explorer les propriétés du vecteur rotation, il n'est pas inutile de s'intéresser à une autre représentation des rotations fort utilisée : les quaternions unitaires. Nous introduirons tout d'abord les propriétés générales de l'algèbre des quaternions et examinerons ensuite le lien entre rotation et quaternion unitaire. Cette section nous donnera des jacobiens relativement simples (en particulier pour la composition des rotations) que nous utiliserons postérieurement avec le vecteur rotation. Nous finirons par le calcul de la mesure uniforme, ce qui nous fournira un algorithme très simple pour obtenir des rotations aléatoires uniformes.

Cette section est principalement inspirée de (Le Borgne, 1987) pour la partie différentielle et de (Casteljau, 1987) pour les quaternions matriciels (on pourra également voir (Walker et Shao, 1991) et (Horaud et Monga, 1993)). Les avantages calculatoirs des quaternions pour représenter les rotations ont été explorés dans (Reyes-Avila, 1990; Reyes-Avila, 1991; Funda et Paul, 1988) et surtout (Capolsini et al., 1993) pour leur implémentation symbolique. On pourra bien sûr se référer à (Altmann, 1986) pour un point de vue plus mathématique.

7.2.1 Definitions

Introduced by Hamilton around 1843, the 4 dimensional algebra of quaternions $\mathbb{Q}$ is a generalization of complex numbers. The basic set is $\mathbb{R}^4$ with the canonical basis $\{1,i,j,k\}$. The addition is the canonical addition of $\mathbb{R}^4$ and the product $*$ is defined by distributivity and the following formulas

- $i^2 = j^2 = k^2 = -1$
- $i*j = -j*i = k \qquad j*k = -k*j = i \qquad k*i = -i*k = j$

The quaternion algebra has hence a structure of skew field (non commutative division ring). Another equivalent definition is the set of complex matrices

$$q = \begin{bmatrix} a & b \\ -\overline{b} & \overline{a} \end{bmatrix} = \begin{bmatrix} t+ix & y+iz \\ -y+iz & t-ix \end{bmatrix}$$

with the matrix addition and product ($\overline{a}$ is here the complex conjugate of a).

7.2.1.1 Vectorial quaternions

From the first definition, it is easy to see that $\mathbb{Q}$ has a subfield isomorphic to $\mathbb{R}$ (basis 1) and a vectorial subspace of basis $\{i, j, k\}$ identified with $\mathbb{R}^3$. This subsets are called **real and pure quaternions**. Each quaternion q can be uniquely written as the sum of a real and a pure quaternion and can then be considered equivalently as a pair $q = (a, v)$ or a vector $q = (a, v^{\mathrm{T}})^{\mathrm{T}}$, where $a \in \mathbb{R}$ is the real part and $v \in \mathbb{R}^3$ the pure part. For simpler notations, we will use $q = (a, v)$ to denote the vectorial form of the quaternion. The operations defined on quaternions to form the algebra are then :

- Addition : $(a_1, v_1) + (a_2, v_2) = (a_1 + a_2, v_1 + v_2)$
- Internal multiplication : $(a_1, v_1) * (a_2, v_2) = (a_1.a_2 - \langle v_1 \mid v_2 \rangle, v_1 \times v_2 + a_1 v_2 + a_2 v_1)$ where "$\times$" and "$\langle . \mid . \rangle$" are the usual cross and dot products on $\mathbb{R}^3$.

As for complex numbers, we define the conjugate quaternion

- $\overline{(a, v)} = (a, -v)$

and the canonical dot product of $\mathbb{R}^4$ can be written

- $\langle q_1 \mid q_2 \rangle = \frac{1}{2}.(\overline{q}_1 * q_2 + \overline{q}_2 * q_1) = a_1.a_2 + \langle v_1 \mid v_2 \rangle$

The norm is then

- $|q|^2 = \langle q \mid q \rangle = \|q\|^2_{\mathbb{Q}} = \overline{q} * q = a^2 + \|v\|^2_{\mathbb{R}^3} = \|q\|^2_{\mathbb{R}^4}$

which is compatible with the product: $|q_1 * q_2| = |q_1|.|q_2|$. This allows to write very simply the inverse quaternion :

- $q^{(-1)} = \dfrac{\overline{q}}{|q|^2}$ for $q \neq 0$.

7.2.1.2 Commutation

Let $q_1 = (a_1, v_1)$ and $q_2 = (a_2, v_2)$ be two quaternions. Their product is usually non commutative. The commutator of q_1 and q_2 is defined by

$$[q_1, q_2] = q_1 * q_2 - q_2 * q_1 = (0, 2.v_1 \times v_2)$$

This means that two quaternions commutes if and only if their pure part are collinear. It shall be noted that the commutator is always a pure quaternion.

7.2.1.3 Pure quaternions

As said above, the set of quaternions $(0, x)$ with $x \in \mathbb{R}^3$ is trivially identified with $\mathbb{R}^3$ itself. Let x and y be two vectors (elements of $\mathbb{R}^3$). Their quaternion product is

$$x * y = (-\langle x \mid y \rangle, x \times y)$$

and their cross product is then

$$x \times y = \frac{1}{2}[x, y]$$

7.2.1.4 Derivation

Let $p(t)$ and $q(t)$ be curves on $\mathbb{Q}$. The derivation of the product is easily proved to be the non commutative derivation

$$\frac{d}{dt}(p*q) = \frac{dp}{dt}*q + p*\frac{dq}{dt}$$

and the derivative of the conjugate is the conjugate derivative

$$\frac{d\overline{q}}{dt} = \overline{\left(\frac{dq}{dt}\right)}$$

The derivatives of most expressions easily follows, as we can treat variable quaternions as usual, but without commutativity.

7.2.1.5 Matricial quaternions and anti-quaternions

Let $q = (a, v)$ be a quaternion considered as a vector of $\mathbb{R}^4$. We define the matricial quaternion Q_q as the 4x4 matrix verifying for any other quaternion p in vector form:

$$Q_q.p = q*p$$

Writing the matrix by part, we have

$$Q_q = a.I_4 + \begin{bmatrix} 0 & -v^{\mathrm{T}} \\ v & S_v \end{bmatrix} \tag{7.11}$$

It can be easily verified that $Q_{q_1}.Q_{q_2} = Q_{(q_1*q_2)}$. Quaternions can then be also identified with a 4 dimensional subspace of 4x4 real matrices. If we now consider quaternions q and p by row vectors (denoted by q^{T} and p^{T}), we can define another 4x4 matrix P_q verifying for any other p^{T} :

$$p^{\mathrm{T}}.P_q = (p*q)^{\mathrm{T}} \qquad \Longleftrightarrow \qquad P_q^{\mathrm{T}}.p = (p*q)$$

Writing the matrix by part,we have

$$P_q = a.I_4 + \begin{bmatrix} 0 & v^{\mathrm{T}} \\ -v & S_v \end{bmatrix} \tag{7.12}$$

We have also $P_{q_1}.P_{q_2} = P_{(q_1*q_2)}$. The matrix P_q will be called the anti-quaternion of q.

Properties of matricial quaternions

- $P_q.p = p*\overline{q}$ and $Q_q.p = q*p$
- $Q_{\overline{q}} = Q_q^{\mathrm{T}}$ and $P_{\overline{q}} = P_q^{\mathrm{T}}$
- $Q_p.P_q = P_q.Q_p$
- $Q_q.Q_q^{\mathrm{T}} = Q_q^{\mathrm{T}}.Q_q = \|q\|^2.I_4 = P_q.P_q^{\mathrm{T}} = P_q^{\mathrm{T}}.P_q$
- $\det(Q_q) = \det(P_q) = \|q\|^4$

If q be a unit quaternion ($\|q\| = 1$), we have thus $Q_q.Q_q^{\mathrm{T}} = Q_q.{}^{\mathrm{T}}Q_q = I_4$ and $\det(Q_q) = 1$, and similarly for P_q, which shows that the matrices Q_q and P_q are rotations of $\mathbb{Q} \equiv \mathbb{R}^4$.

7.2.2 Quaternions and rotations

7.2.2.1 From unit quaternions to rotations

A particular subgroup of $\mathbb{Q}$ is the sphere of unit quaternions (equivalent to $\mathcal{S}_3$). Let $q = (a, v)$ be such a quaternion. The map

$$R_q : \begin{array}{ccc} \mathbb{Q} & \longrightarrow & \mathbb{Q} \\ p & \longmapsto & q * p * \overline{q} \end{array}$$

is an inner automorphism of $\mathbb{Q}$ that conserves pure quaternions. Its restriction to $\mathbb{R}^3$ conserves orientation and the dot product: this is a rotation of $\mathbb{R}^3$.

7.2.2.2 From unit quaternions to rotations

Let x be a vector considered as a pure quaternion. The action of R_q produces the vector $y = q * x * \overline{q}$ and since $x * \overline{q} = P_q.x$, we can write $y = (Q_q.P_q).x$. Developing the product and identifying with $y = R.x$ (the action of a 3-D rotation), we find

$$Q_q.P_q = P_q.Q_q = \|q\|^2.I_4 + 2.\begin{bmatrix} 0 & 0 \\ 0 & a.S_v + S_v^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$$

Théorème 7.4 *Let $q = (a, v)$ be a unit quaternion. The associated rotation matrix is given by*

$$R = I_3 + 2.a.S_v + 2.S_v^2 \tag{7.13}$$

This formula is the equivalent of the Rodrigue formula for quaternions.

Let $R \leftrightarrow q$ denote the association between rotation matrix R and rotation quaternion q. As direct properties of this representation, we have:

- If $R_1 \leftrightarrow q_1$ and $R_2 \leftrightarrow q_2$ then $R_1.R_2 \leftrightarrow q_1 * q_2$.
- If $R \leftrightarrow q$ then $R^{(-1)} \leftrightarrow \overline{q} = q^{(-1)}$

and by definition, the application of to vector x is

- $y = R.x = q * x * \overline{q} = Q_q.P_q.x$

7.2.2.3 From rotation matrices to unit quaternions

Assume now that R is a rotation of angle θ around axis n. From Rodrigues' formula we have for the symmetric part $2.S_v^2 = (1 - \cos\theta).S_n^2$ and hence $v = \pm \sin\left(\frac{\theta}{2}\right) n$. With the unit constraint, we obtain

Théorème 7.5 *Two opposite unit quaternions are associated with the rotation of angle θ around axis n. They are given by*

$$q = \pm\left(\cos\frac{\theta}{2}\,,\ \sin\frac{\theta}{2}.n\right)$$

The space $\mathcal{SO}_3$ is then isomorphic to $\mathcal{P}^3$ the projective space obtained by identifying antipodal points on the sphere S^3.

From now on, we will call rotation quaternion q the couple $(q, -q)$ with $\|q\| = 1$, and give as value one of the two opposite quaternions.

Using the relations of section (7.1.2) about the geometric parameters of the rotation, we can directly obtain the rotation quaternion q from the rotation matrix R from the following formulas:

$$\mathrm{Tr}(R) = 1 + 2\cos\theta \qquad R - R^{\mathrm{T}} = 2.\sin\theta.S_n$$

Hence, q is defined for $\mathrm{Tr}(R) \neq -1$ by

$$q = \pm(a, v) \qquad \text{with} \qquad a = \frac{\sqrt{1+\mathrm{Tr}(R)}}{2} \qquad \text{and} \qquad S_v = \frac{1}{2.\sqrt{1+\mathrm{Tr}(R)}}.(R - R^{\mathrm{T}}) \tag{7.14}$$

If $\mathrm{Tr}(R) = -1$ we have to use the other form of Rodrigues' formula : $R = I_3 + 2.S_n^2 = 2.n.n^{\mathrm{T}} - I_3$. Assuming $n = (x, y, z)^{\mathrm{T}}$, we have in this case $2.n.n^{\mathrm{T}} = R - I_3 = [2.x.n, 2.y.n, 2.z.n]$. It is then sufficient to normalize the greater column vector (in norm) of $R - I_3$ to obtain the axis n. The rotation quaternion q is then given by $q = \pm(0, n)$.

7.2.3 Differential properties of rotation quaternions

Let $R(t) \leftrightarrow q(t)$ be a curve on $\mathcal{SO}_3$ and $\dot{R}(t)$ and $\dot{q}(t)$ the derivatives of these quantities. Remember that the rotation quaternion q is the couple $(q, -q)$. We call derivative of the rotation quaternion q the couple $(\dot{q}, -\dot{q})$ in the same sign order: the derivative representation is determinated by the choice of the quaternion. Since $|q|^2 = \overline{q} * q = q * \overline{q} = 1$, we have $\dot{q} * \overline{q} + q * \dot{\overline{q}} = 0$ and then

$$\langle \dot{q} \mid q \rangle = 0 \qquad \text{and} \qquad \dot{\overline{q}} = -\overline{q} * \dot{q} * \overline{q}$$

Let now x and y be vectors such that

$$R(t).x = y(t) = q(t) * x * \overline{q}(t)$$

The derivation of y can be done for both sides:

$$\begin{aligned} \dot{y} &= \dot{R}.x = S_{\omega_r}.R.x = \omega_r \times (q * x * \overline{q}) = \tfrac{1}{2}.(\omega_r * q * x * \overline{q} - q * x * \overline{q} * \omega_r) \\ &= \dot{q} * x * \overline{q} + q * x * \dot{\overline{q}} = \dot{q} * x * \overline{q} - q * x * \overline{q} * \dot{q} * \overline{q} \end{aligned}$$

Hence we find that $\overline{q} * \dot{y} * q = [\overline{q} * \dot{q}, x] = \frac{1}{2}.[\overline{q} * \omega_r * q, x]$ and since this is true for every $x \in \mathbb{R}^3$, we conclude with the following theorem.

Théorème 7.6 *The tangent space $T_q\mathcal{SO}_3$ of $\mathcal{SO}_3$ at rotation quaternion q is given by $\langle \dot{q} \mid q \rangle = 0$ and if $R(t) \leftrightarrow q(t)$ we have*

$$\dot{R} = S_{\omega_r}.R = R.S_{\omega_l} \quad \Longleftrightarrow \quad \dot{q} = \frac{\omega_r}{2} * q = q * \frac{\omega_l}{2}$$

From the above theorem it is clear that the left and right translations and their differentials are just left and right product by the corresponding rotation quaternion. The tangent space $T_1\mathcal{SO}_3$ of $\mathcal{SO}_3$ at rotation quaternion identity is simply the set of pure quaternions, i.e. $\mathbb{R}^3$. It is canonically link to the tangent space $T\mathcal{SO}_3$ at rotation matrix identity by

$$\begin{array}{ccc} T_1\mathcal{SO}_3 & \longleftrightarrow & T\mathcal{SO}_3 \\ \frac{\omega}{2} & \longleftrightarrow & S_\omega \end{array} \tag{7.15}$$

7.2.3.1 Lie algebra of $T_1\mathcal{SO}_3$ and dot product

From the commutation properties of section (7.2.1.2), the Lie bracket in $T\mathcal{SO}_3$ $[S_{\omega_1}, S_{\omega_2}] = S_{(\omega_1 \times \omega_2)}$ immediately transfers in $T_1\mathcal{SO}_3$ as the classical commutator

$$\left[\frac{\omega_1}{2}, \frac{\omega_2}{2}\right] = \frac{\omega_1 \times \omega_2}{2}$$

and the dot product on $T\mathcal{S}O_3$: $\langle S_{\omega_1} \mid S_{\omega_2} \rangle_{T\mathcal{S}O_3} = \langle \omega_1 \mid \omega_2 \rangle_{\mathbb{R}^3}$. is transported on $T_1\mathcal{S}O_3$ by

$$\left\langle \frac{\omega_1}{2} \;\middle|\; \frac{\omega_2}{2} \right\rangle_{T_1\mathcal{S}O_3} = \langle \omega_1 \mid \omega_2 \rangle_{\mathbb{R}^3} = 4. \left\langle \frac{\omega_1}{2} \;\middle|\; \frac{\omega_2}{2} \right\rangle_{\mathbb{Q}}$$

This metric is trivially invariant by left and right translations. Hence the dot product on the tangent space of rotation quaternions (identified with left or right translation to $T_1\mathcal{S}O_3$) is 4 times the canonical dot product of the quaternions space $\mathbb{Q}$.

7.2.3.2 Differentials of standard maps

Now that we have the basic properties of the tangent spaces, it is interesting to see how vectors are mapped between these tangent spaces by standard functions on quaternions. these maps can also be restricted to rotation quaternion. We indicate if a simplification occurs in this case. Some of these jacobians will be used in the next section for computing the jacobians of some maps on the rotation vector.

- **Left translation** by p: $q \mapsto p * q$.

$$J_l(p) = \frac{\partial(p*q)}{\partial q} = \frac{\partial(Q_p.q)}{\partial q} = Q_p$$

- **Right translation** by q: $p \mapsto p * q = P_{\overline{q}}.p = P_q^{\mathrm{T}}.p$

$$J_r(q) = \frac{\partial(p*q)}{\partial p} = \frac{\partial(P_{\overline{q}}.p)}{\partial p} = P_q^{\mathrm{T}}$$

- **Inversion**: $q = (a, v) \mapsto q^{(-1)} = \dfrac{\overline{q}}{|q|} = \dfrac{(a, -v)}{a^2 + \|v\|^2}$

$$J_{Inv}(q) = \frac{\partial q^{(-1)}}{\partial q} = \frac{\overline{J}}{\langle q \mid q \rangle} + 2.\frac{\overline{q}.q^{\mathrm{T}}}{\langle q \mid q \rangle^2} \qquad \text{with} \qquad \overline{J} = \frac{\partial \overline{q}}{\partial q} = \begin{bmatrix} 1 & 0 \\ 0 & -I_3 \end{bmatrix}$$

  In the case of rotation quaternions, we have $\|q\| = 1$, and the formula simplifies to

$$J_{Inv}(q) = -Q_q^{\mathrm{T}}.P_q$$

- **Action of a "Rotation"** q **on** x: $x \mapsto q * x * \overline{q} = Q_q.P_q.x$

$$J_R(q) = \frac{\partial(q*x*\overline{q})}{\partial x} = Q_q.P_q$$

  If q is a rotation quaternion ($\|q\| = 1$) associated with R, this jacobian simplifies to :

$$J_R(q) = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$$

  To obtain the derivative with respect to q, we have to come back to the definition of a jacobian matrix : the differential J_f of a function f is the linear function that maps to the tangent vector $\dot{q}$ of the moving point q, the tangent vector $J_f.\dot{q}$ of the moving point $f(q)$. In our case, we have :

$$J(q,x).\dot{q} = \dot{q} * x * \overline{q} + q * x * \dot{\overline{q}} = P_{(q*\overline{x})}.\dot{q} + Q_{(q*x)}.\overline{J}.\dot{q}$$

which shows that

$$J(q,x) = \frac{\partial(q * x * \overline{q})}{\partial q} = P_q.P_x^{\mathrm{T}} + Q_q.Q_x.\overline{J}$$

Developing this formula, we find if x is a vector and $q = (a, v)$:

$$J(q,x) = 2\begin{bmatrix} 0 & 0 \\ a.x + v \times x & \langle x \mid v \rangle .I_3 - a.S_x \end{bmatrix}$$

7.2.4 Exponential map

As for rotations matrices, the exponential map can be defined in $T_1\mathcal{S}O_3$ from integral curves, but also with the quaternion exponential from the formula

$$\exp(q) = \sum_{i=0}^{+\infty} \frac{q^i}{i!} = 1 + \frac{q}{1!} + \frac{q^2}{2!} + \dots$$

Since the quaternions product is not commutative, this exponential behave globally as the matrix exponential.

In our case, we are interested in exponential of vectors (from $T_1\mathcal{S}O_3$), which turn out to give unit quaternions. Indeed, let $x = (0, \theta.n)$ be a vector, then $x^2 = (-\theta^2, 0)$, $x^3 = (0, -\theta^2.n)$ and so on. The series reduces then to

$$\exp(x) = (\cos\theta\ ,\ \sin\theta.n) = \left(\cos(\|x\|)\ ,\ \sin(\|x\|).\frac{x}{\|x\|}\right)$$

Let $r = \theta.n$ be a (rotation) vector, we have then the following relation

$$R = \exp(S_r) \longleftrightarrow q = \exp\left(\frac{r}{2}\right)$$

where the first is the matrix exponential and the second the quaternion one.

7.2.5 Geodesics for rotation quaternions

Let $R_p \leftrightarrow p$ and $R_q \leftrightarrow q$ be two rotations. The transportation of the $\mathcal{S}O_3$ metric gives:

$$\rho(p,q) = \rho(R_p, R_q) = \rho(I_3, R_p^{\mathrm{T}}.R_q) = \theta$$

where θ is the angle of the rotation $R_p^{\mathrm{T}}.R_q$. On the other hand, rotation quaternion $q * \overline{p} = \pm(\cos(\theta/2), \sin(\theta/2)n)$ has a dot product with identity

$$\langle 1 \mid q * \overline{p} \rangle = \langle q \mid p \rangle = \pm\cos(\theta/2)$$

We can then distinguish two cases depending on the choice of the representations (signs) of p and q leading to a positive or a negative dot product $\langle p \mid q \rangle$. The two representations will be said compatible if the dot product is positive: the distance is $\theta = 2\arccos(\langle p \mid q \rangle)$ in this case. We can then write the canonical distance of $\mathcal{S}O_3$:

$$\rho(p,q) = 2.\arccos\left(\left|\langle p \mid q \rangle_{\mathbb{Q}}\right|\right) \tag{7.16}$$

The rotation quaternion distance is then *twice* the classic distance on the surface of S^3 (between two quaternions of the same hemisphere).

Geodesics starting from identity to the rotation quaternion $q = \exp(\theta.n/2)$ are given by the following formula for $\theta \leq \pi$

$$q(t) = \exp(t.n) = (\cos(t)\ ,\ \sin(t).n) \qquad \text{with} \qquad 0 \leq t \leq \frac{\theta}{2}\left(\leq \frac{\pi}{2}\right)$$

This is in fact the equation of a piece of circle on the (unit) sphere S^3 and in the plane $\{1, n\}$. By left and right translation, we obtain every other geodesic. If we want to express these geodesics in terms of quaternions (not rotation quaternions), we have

Théorème 7.7 *Geodesics for rotation quaternions are the pieces of circle of center $q = 0$ (on the unit sphere S^3 of $\mathbb{Q}$) with a length (in $\mathbb{Q}$) less than $\pi/2$, along with their antipodal piece of circle.*

7.2.6 Uniform density for rotations

We saw in section (3.3) how to compute the invariant measure (or uniform measure) directly or from the metric. Here is another approches that give the same result. This will give us an easy way to sample unifrom rotations.

7.2.6.1 Uniform measure over rotation quaternions

Let $\sigma(q).dq$ be the probability of a small area around q, and $\sigma_p(p*q).d(p*q)$ the probability of the same area after a global left translation of the group: they are the same and hence relate by the classical change in variables formula $\sigma_p(p*q) = \sigma(q).(\det(J_l))^{(-1)}$ where J_l is the Jacobian of the left translation. The uniform randomness is defined by requiring that σ is invariant by left translation, that is $\sigma_p = \sigma$. In the case of quaternions, this jacobian does not depend upon the application point:

$$J_l = \frac{\partial(p*q)}{\partial q} = Q_p$$

and if p is a unit quaternion, we have $\det(Q_p) = 1$. Hence $\sigma(p*q) = \sigma(q)$ and taking $q = 1$ gives the constant density $\sigma(p) = \sigma(1)$ for any p. Since the set of rotation quaternions is compact, we can obtain the uniform density by normalization (remember that q and $-q$ represent the same rotation):

$$1 = \frac{1}{2}.\int_{\mathcal{SO}_3} \sigma(q).dq = \frac{\sigma(1)}{2}.\int_{\mathcal{S}^3} dq = \sigma(1)\pi^2$$

We obtain thus the invariant (or Haar) measure on rotation quaternions:

$$\sigma(q).dq = \frac{dq}{\pi^2} \tag{7.17}$$

There are in fact two invariant measures since we could have chosen a right invariance for σ. In the $\mathcal{SO}_3$ case, left and right Haar invariant measures turn out to be the same.

7.2.6.2 Computation of a (uniform) random rotation

From formula (7.17), the problem reduces to find a uniform random vector on the sphere S^3. This can be achieved as follows.

- Choose the 4 coordinates of a quaternion p uniformly in $[-1, 1]$. The quaternion p has then a uniform probability in the hypercube.

- Reject the sample if $\|p\| > 1$ or $\|p\| < \varepsilon$ where ε is a small threshold set to avoid numerical instabilities around 0 for next step. We obtain a uniform probability in the ball with a hole $\mathcal{B}_4(0,1) - \mathcal{B}_4(0,\varepsilon)$.
- Normalize p. The quaternion $q = \dfrac{p}{\|p\|}$ obtained has a uniform probability over the sphere S^3.

7.3 Vecteur rotation

La représentation des rotations est un sujet étudié depuis longtemps (voir par exemple (Stuelpnagel, 1964)), mais qui prend une importante toute particulière en robotique (Paul, 1982; Latombe, 1991) et en vision par ordinateur (Kanatani, 1993). Nous avons vu dans la section précédente les quaternions unitaires, mais plusieurs types d'angles d'Euler sont également couramment utilisés.

Le vecteur rotation n'est guère employé en temps que représentation pratique avant (Ayache, 1989, chap. 12), bien que les paramètres axe et angle soient fort connus. Remarquons que, dans notre cas, ce sont les propriétés génériques de la carte exponentielle qui nous amènent à choisir cette représentation et que les propriétés théoriques dérivées sur les primitives probabilistes dans les chapitres précédents ne seraient généralement pas valides avec une autre représentation.

7.3.1 La carte principale

Si l'on reprend les résultats (heureusement concordants) des deux sections précédentes sur les géodésiques pour la métrique bi-invariante, on sait déjà que la carte principale est constituée du vecteur rotation $r = \theta.n$, où n est l'axe de rotation (unitaire) et θ l'angle de la rotation autour de cet axe. Comme cet angle est défini à $2.\pi$ près, tous les vecteurs $r_k = (\theta + 2.k.\pi).n$ $(k \in \mathbb{Z})$ représentent la même rotation $R = \mathcal{R}(\theta, n)$. La « norme » de cette rotation est donc

$$N(R) = \text{ dist}(R,\, Id) = \inf_{k \in \mathbb{Z}} |\theta + 2.k.\pi|$$

Le lieu de coupure (de l'identité) est constitué des rotations pouvant être atteintes par deux géodésiques de même distance : ce n'est possible que si $\theta = \pi$, ce qui correspond à un retournement que l'on peut atteindre en partant avec le vecteur tangent $r = \pi.n$ ou $r' = \pi.(-n)$.

Théorème 7.8 (Carte principale du groupe $\mathcal{SO}_3$)
La carte principale du groupe des rotations 3D est formée des vecteurs rotation $r = \theta.n$ de la boule ouverte $\mathcal{D} = \mathcal{B}(0,\pi)$. La norme est dans cette carte est : $\theta(R) = N(r) = \ dist(R, I_3) = \|r\|$.

Le lieu de coupure est constitué des retournements (rotations d'angle π), qui correspondent aux vecteurs rotation (identifiés) $r = \pi.n$ et $-r = \pi.(-n)$ sur le bord du domaine $\mathcal{C} = \mathcal{S}_3(0,\pi)$.

Notons pour mémoire que si l'on voulait obtenir un atlas, il faudrait définir au minimum trois autres cartes. En suivant (Ayache, 1989) , on pourrait garder la même représentation $r = \theta.n$ mais avec les domaines formés des demi boules ouvertes $\mathcal{B}_x = \{r \in \mathcal{B}(0,2\pi) / r_x > 0\}$ (respectivement $\mathcal{B}_y$ et $\mathcal{B}_z$) couvrant les rotations différentes de l'identité et ayant un axe orthogonal à l'axe des x (respectivement de y et des z).

Grâce à l'utilisation de la métrique invariante, nous pouvons en fait nous contenter de la carte principale, puisque l'on obtient une carte centrée en n'importe quelle rotation R en translatant (à gauche) la carte principale.

7.3.2 Relation entre le vecteur rotation et les autres représentations

7.3.2.1 Conversion entre vecteur et matrice de rotation

Le vecteur rotation $r = \theta.n$ peut être calculé à partir de la matrice de rotation en utilisant les paramètres angle et axe comme à la section (7.1.2). A l'inverse, la formule de Rodrigues nous permet de calculer directement la matrice de rotation à partir du vecteur de rotation :

$$R = Id + \frac{\sin\theta}{\theta}.S_r + \frac{(1-\cos\theta)}{\theta^2}.S_r^2 \qquad \text{avec} \qquad \theta = \|r\|$$

Pour éviter les instabilités numériques autour de $\theta = 0$, on utilisera les développements limités suivants :

$$\frac{\sin\theta}{\theta} = 1 - \frac{\theta^2}{6} + O\left(\theta^4\right) \qquad \text{et} \qquad \frac{(1-\cos\theta)}{\theta^2} = \frac{1}{2} - \frac{\theta^2}{24} + O\left(\theta^4\right)$$

7.3.2.2 Conversion entre vecteurs rotation et quaternions unitaires

Les deux fonctions suivantes permettent de passer de l'une des représentations à l'autre, et nous permettront en particulier de calculer le jacobien de la composition de deux vecteurs rotation en passant par les quaternions. Nous ne donnons ici que les formules de passage, leur jacobiens seront développés à la section (7.3.4).

Considérons un vecteur rotation $r = \theta.n$ et un quaternion rotation $q = \pm(a, v)$ représentant la même rotation R. Nous avons donc $a = \cos(\theta/2)$ et $v = \sin(\theta/2).n$. La conversion du vecteur rotation au quaternion est donc :

$$q(r) = \left(\cos(\theta/2)\,;\, \frac{\sin(\theta/2)}{\theta}.r\right) \qquad \text{où} \qquad \theta = \|r\|$$

A l'inverse, le vecteur rotation peut être obtenu à partir du quaternion par la formule suivante (la notation $\overset{\|q\|=1}{=\!=}$ signifie « égal si $\|q\| = 1$ ») :

$$r(q) = 2.\mathrm{sign}(a).\arcsin\left(\frac{\|v\|}{\sqrt{a^2+\|v\|^2}}\right).\frac{v}{\|v\|} \overset{\|q\|=1}{=\!=} 2.\mathrm{sign}(a).\frac{\arcsin(\|v\|)}{\|v\|.}v$$

7.3.3 Opérations atomiques sur le vecteur rotation

Soit $r = \theta.n$ (avec n unitaire) un vecteur rotation représentant la rotation R. Nous nous intéressons dans cette section à l'action de r sur un vecteur x ($y = r \star x$), à l'inversion $r^{(-1)}$, à la composition de deux vecteurs rotation $r = r_2 \circ r_1$ et enfin à la normalisation (opération qui ramène dans le domaine). Si les opérations ne sont pas très dures à réaliser, le calcul de leur jacobiens, lui, est plus difficile. Nous étudierons d'abord l'inversion et la normalisation, qui sont les plus faciles, puis l'action sur un vecteur et enfin la composition qui demandera de passer par les quaternions unitaires. Le lecteur pourra se référer à l'appendice (A.3) pour la différenciation des opérateurs usuels.

7.3.3.1 Inversion d'un vecteur rotation : $r^{(-1)} = -r$

Tout est déjà dans le titre, il ne reste plus qu'à écrire le jacobien :

$$J_{Inv}(r) = \frac{\partial r^{(-1)}}{\partial r} = \frac{\partial(-r)}{\partial r} = -\,Id \tag{7.18}$$

7.3.3.2 Domaine : $r = \psi.n \mapsto r' = \theta.n \qquad (|\theta| \leq \pi)$

Considérons un vecteur rotation $r = \psi.n$ où n est un vecteur unitaire mais où l'angle $\psi = \|r\|$ est quelconque. Pour optimiser l'implémentation de cette fonction, il conviendra de remarquer que si $\psi \leq \pi$, alors on est déjà dans le domaine de la carte principale ou sur sa frontière et il n'y a aucune modification à faire.

Dans le cas général, on recherche l'entier k_0 minimisant la norme du vecteur rotation. Comme φ est positif, la solution est donnée par

$$k_0 = \arg\min_{k\in\mathbb{N}} |\varphi - 2.k.\pi| = \left\lfloor \frac{\varphi}{2.\pi} + \frac{1}{2} \right\rfloor = \left\lfloor \frac{\|r\|}{2.\pi} + \frac{1}{2} \right\rfloor$$

où $\lfloor x \rfloor$ est la valeur entière immédiatement inférieure à x. On doit donc retourner le vecteur rotation suivant (si $\varphi \neq 0$) :

$$r' = (\varphi - 2.k_0.\pi).\frac{r}{\varphi} = \left(1 - \frac{2.k_0.\pi}{\|r\|}\right).r$$

et le jacobien est donné par la formule :

$$\frac{\partial r'}{\partial r} = I_3 - \frac{2.k_0.\pi}{\|r\|^3}.S_r^2$$

7.3.3.3 Action d'un vecteur rotation sur un vecteur : $r \star x = R.x$

Sachant calculer la matrice de rotation associée à r, il n'y pas de problème pour calculer $r \star x = R.x$ ni le jacobien de l'action par rapport au vecteur x :

$$\frac{\partial (r \star x)}{\partial x} = R$$

Le calcul de jacobien $\frac{\partial (r\star x)}{\partial r}$ par rapport à r est un peu plus difficile. Il sera cependant nécessaire pour les transformations rigides et pour estimer des rotations. Considérons pour cela les fonctions suivantes :

$$\alpha = \frac{\sin\theta}{\theta} \qquad \beta = \frac{(1-\cos\theta)}{\theta^2} \qquad \gamma = \frac{\dot{\alpha}}{\theta} \qquad \delta = \frac{\dot{\beta}}{\theta}$$

On peut alors écrire, la formule de Rodrigues sous la forme :

$$r \star x = x + \alpha.S_r.x + \beta.S_r^2.x$$

En prenant en compte les dérivées suivantes :

$$\frac{\partial (S_r.x)}{\partial r} = -S_x \qquad \frac{\partial (S_r^2.x)}{\partial r} = S_x.S_r - 2.S_r.S_x \qquad \frac{\partial \theta}{\partial r} = \frac{r^{\mathrm{T}}}{\theta}$$

on peut différencier $r\star x$ par la composition des jacobiens. Après factorisation, on obtient le résultat suivant (Les développements limités permettent d'éviter les instabilités numériques autour de $\theta = 0$).

Théorème 7.9 (Action du vecteur rotation r sur le vecteur x)
Soit r un vecteur rotation, et α, β, δ et γ les fonction suivantes de $\theta = \|r\|$:

$$\alpha = \tfrac{\sin\theta}{\theta} = 1 - \tfrac{\theta^2}{6} \qquad \beta = \tfrac{(1-\cos\theta)}{\theta^2} = \tfrac{1}{2} - \tfrac{\theta^2}{24} + O\left(\theta^4\right)$$

$$\gamma = \tfrac{\dot{\alpha}}{\theta} = \tfrac{(\cos\theta-\alpha)}{\theta^2} = \tfrac{1}{3} - \tfrac{\theta^2}{30} + O\left(\theta^4\right) \qquad \delta = \tfrac{\dot{\beta}}{\theta} = \tfrac{(\alpha-2\beta)}{\theta^2} = -\tfrac{1}{12} + \tfrac{\theta^2}{180} + O\left(\theta^4\right)$$

Alors on a :

$$r \star x = R.x \qquad et \qquad \frac{\partial (r \star x)}{\partial x} = R \qquad avec \qquad R = I_3 + \alpha.S_r + \beta.S_r^2 \tag{7.19}$$

$$\frac{\partial (r \star x)}{\partial r} = -S_x.\left(\gamma.r.r^T - \beta.S_r + \alpha.\, Id\right) - S_r.S_x.\left(\delta.r.r^T + 2.\beta.\, Id\right) \tag{7.20}$$

Pour optimiser ce dernier calcul, on pourra noter que $S_r.S_x = x.r^{\mathrm{T}} - \langle x \mid r \rangle .I_3 = x.r^{\mathrm{T}} - \mathrm{Tr}(x.r^{\mathrm{T}}).I_3$.

Lien avec d'autres travaux : Une autre façon de différencier l'action d'un vecteur rotation est donné dans (Ayache, 1989) : soient η et U_r la fonction et la matrice suivante :

$$\eta = \frac{(1-\alpha)}{\theta^2} = \frac{1}{6} - \frac{\theta^2}{120} + O\left(\theta^4\right) \qquad \text{and} \qquad U_r = \eta.S_r^2 + \beta.S_r + I_3$$

et du le vecteur infinitésimal $du = U_r.dr$. Alors N. Ayache a montré que, pour n'importe quel incrément infinitésimal dr du vecteur rotation r, l'incrément dR de la matrice de rotation R est donné par $dR = S_{du}.R$. Pour obtenir à partir de cela le jacobien de l'action d'un vecteur, on peut écrire :

$$(R + dR).x - R.x = dR.x = S_{du}.R = -S_{(R.x)}.du = -S_{(R.x)}.U_r.dr$$

et on obtient donc

$$\frac{\partial (r \star x)}{\partial r} = -S_{(R.x)}.U_r \tag{7.21}$$

ce qui n'est qu'une forme factorisée de l'équation (7.20). D'un point de vue applicatif, la forme précédente est plus rapide car elle nécessite moins d'opérations en machine.

7.3.4 Composition de deux vecteurs rotation

La composition est sans doute l'opération la plus complexe. On pourrait bien sûr calculer les matrices de rotation associée, les multiplier ($R = R_2.R_1$) et revenir au vecteur rotation, mais ce serait très difficile à différencier. Nous avons choisi de passer par les quaternions unitaires comme étape intermédiaire, en sachant que la dérivée du produit des quaternions est très simple.

Si r_1 et r_2 sont deux vecteurs rotation, le principe est donc de calculer les quaternions unitaires q_1 et q_2 associés, de les multiplier ($q = q_2 * q_1$), et de revenir au vecteur rotation :

$$r = r_2 \circ r_1 = r(q(r_2) * q(r_1))$$

On peut alors obtenir les jacobiens de r par rapport à r_1 et r_2 grâce aux jacobiens composés suivants :

$$\frac{\partial r}{\partial r_1} = \frac{\partial r}{\partial q}.\frac{\partial q}{\partial q_1}.\frac{\partial q_1}{\partial r_1} \qquad \text{et} \qquad \frac{\partial r}{\partial r_2} = \frac{\partial r}{\partial q}.\frac{\partial q}{\partial q_2}.\frac{\partial q_2}{\partial r_2} \tag{7.22}$$

Rappelons pour mémoire que les jacobiens de la composition des quaternions sont (voir section (7.2.3.2)):

$$\frac{\partial (q_2 * q_1)}{\partial q_1} = Q_{q_2} \qquad \text{et} \qquad \frac{\partial (q_2 * q_1)}{\partial q_2} = P_{q_1}^{\mathrm{T}}$$

Le problème est maintenant de calculer les jacobiens de la conversion entre vecteur rotation et quaternion unitaire. Nous n'écrirons pas ici de formule complète pour les jacobiens composés, mais ces multiplications matricielles ne posent aucun problème numériquement.

7.3.4.1 Du vecteur rotation r au quaternion q

Soit $q = (a, v)$ le quaternion unitaire (avec $a > 0$) associé au vecteur rotation r. A partir des formules $a = \cos(\theta/2)$ et $v = \frac{\sin(\theta/2)}{\theta}.r$, on obtient

$$\frac{\partial a}{\partial r} = -\frac{\sin(\theta/2)}{2}.\frac{r^{\mathrm{T}}}{\theta} = -\frac{v^{\mathrm{T}}}{2} \qquad \text{and} \qquad \frac{\partial v}{\partial r} = \frac{cos(\theta/2)}{2.\theta}.r.r^{\mathrm{T}} - \frac{\sin(\theta/2)}{2.\theta}.\left(\frac{S_r}{\theta}\right)^2$$

En utilisant l'identité $S_r^2 = r.r^{\mathrm{T}} - \theta^2.I_3$, on peut simplifier le résultat pour obtenir le théorème suivant.

Théorème 7.10 (Conversion vecteur rotation / quaternion)
Soit un vecteur de rotation r et κ, λ les fonctions suivantes de $\theta = \|r\|$:

$$\kappa = \frac{\sin(\theta/2)}{\theta} = \frac{1}{2} - \frac{\theta^2}{48} + O\left(\theta^4\right) \qquad \lambda = \frac{\sin(\theta/2)}{\theta^3} - \frac{\cos(\theta/2)}{2.\theta^2} = \frac{1}{24}.\left(1 - \frac{\theta^2}{40}\right) + O\left(\theta^4\right)$$

Les deux quaternions unitaires associés et les jacobiens de cette conversion sont donnés avec $\varepsilon = \pm 1$ par :

$$q(r) = \varepsilon.\left|\begin{array}{c} a \\ v \end{array}\right. \qquad avec \qquad a = \cos(\theta/2) \qquad et \qquad v = \kappa.r \tag{7.23}$$

$$\frac{\partial q}{\partial r} = \varepsilon.\left[\begin{array}{c} \frac{\partial a}{\partial r} \\ \\ \frac{\partial v}{\partial r} \end{array}\right] \qquad avec \qquad \frac{\partial a}{\partial r} = -\frac{v^T}{2} \qquad et \qquad \frac{\partial v}{\partial r} = \kappa.I_3 - \lambda.r.r^T \tag{7.24}$$

7.3.4.2 Du quaternion unitaire q au vecteur rotation r

Soit $q = (a, v)$ un quaternion et r le vecteur rotation associé au quaternion unitaire $\frac{q}{\|q\|}$. On peut obtenir r à partir de q grâce aux équations suivantes :

$$r(q) = 2.\mathrm{sign}(a).\arcsin\left(\frac{\|v\|}{\sqrt{a^2 + \|v\|^2}}\right).\frac{v}{\|v\|} \overset{\|q\|=1}{=\!=} 2.\mathrm{sign}(a).\frac{\arcsin(\|v\|)}{\|v\|}.v$$

Notons que d'autres équations seraient possibles mais conduiraient à des dérivations plus complexes. Si a est positif, on peut écrire :

$$\begin{array}{rcl} \frac{\partial r}{\partial a} & = & \frac{-2.v}{a^2+\|v\|^2} \overset{\|q\|=1}{=\!=} -2.v \\ \frac{\partial r}{\partial v} & = & \frac{2.a}{a^2+\|v\|^2}.\frac{v.v^{\mathrm{T}}}{\|v\|^2} - \frac{2}{\|v\|}\arcsin\left(\frac{\|v\|}{a^2+\|v\|^2}\right).\frac{S_v^2}{\|v\|^2} \\ & \overset{\|q\|=1}{=\!=} & 2.\left\{\frac{\arcsin(\|v\|)}{\|v\|}.I_3 + \left(a - \frac{\arcsin(\|v\|)}{\|v\|}\right).\frac{v.v^{\mathrm{T}}}{\|v\|^2}\right\} \end{array}$$

Pour a négatif, on utilise les équations ci-dessus avec $q' = -q$:

$$\begin{array}{lll} \frac{\partial r}{\partial a} & = \frac{\partial r}{\partial a'}.\frac{\partial a'}{\partial a} & = -2.v'.(-1) = -2.v \\ \frac{\partial r}{\partial v} & = \frac{\partial r}{\partial v'}.\frac{\partial v'}{\partial v} & = -2.\left\{\frac{\arcsin(\|v\|)}{\|v\|}.I_3 + \left(-a - \frac{\arcsin(\|v\|)}{\|v\|}\right).\frac{v.v^{\mathrm{T}}}{\|v\|^2}\right\} \end{array}$$

Au final, on peut regrouper les calculs pour $\|q\| = 1$ dans le théorème suivant :

Théorème 7.11 (Conversion quaternion unitaire / vecteur rotation)
Soit $q = (a, v)$ un quaternion et τ, ξ les fonctions suivantes de $\mu = \|v\|$:

$$\tau = 2.sign(a).\frac{\arcsin(\mu)}{\mu} = 2.sign(a).\left(1 + \frac{\mu^2}{6}\right) + O\left(\mu^4\right)$$
$$\xi = \frac{2.a-\tau}{\mu^2} = 2.sign(a).\frac{\mu.\sqrt{1-\mu^2}-\arcsin(\mu)}{\mu^3} = -2.sign(a)\left(\frac{2}{3} + \frac{\mu^2}{5}\right) + O\left(\mu^4\right)$$

où sign est la fonction signe avec $sign(0) = \pm 1$ indifféremment. Le vecteur rotation associé et le jacobien de cette conversion sont donnés par

$$r(q) = \tau.v \qquad et \qquad \frac{\partial r}{\partial q} = \frac{\partial r}{\partial (a, v)} = [-2.v\ ;\ \tau.I_3 + \xi.v.v^T] \tag{7.25}$$

7.3.5 Jacobien de la translation à gauche de l'identité

Les calculs concernant les opérations atomiques sont maintenant terminés : on peut faire agir, composer et inverser des vecteurs rotation, et calculer le jacobien de ces opérations. Cependant, pour une implémentation efficace, il est utile d'extraire l'expression symbolique du jacobien de la translation à gauche de l'identité. Cela nous permettra d'optimiser numériquement son calcul, mais aussi de calculer l'expression symbolique de la densité uniforme dans la carte principale.

Le principe est le même que pour la composition, mais avec des simplifications dûes à $r_1 = 0$. A partir de $r_2 \circ r_1 = r(q(r_2) * q(r_1))$, la dérivation en chaîne nous donne :

$$\left.\frac{\partial (r_2 \circ r_1)}{\partial r_1}\right|_{r_1=0} = \left.\frac{\partial r(q)}{\partial q}\right|_{q=q_2} . \left.\frac{\partial (q_2 * q_1)}{\partial q_1}\right|_{\substack{q_1=0 \\ q_2=q(r_2)}} . \left.\frac{\partial q(r_1)}{\partial r_1}\right|_{r_1=0}$$

7.3.5.1 Des vecteurs rotation aux quaternions unitaire

On a ici $r_1 = 0$ et $r_2 = \theta_2.n_2$. Les quaternions associés sont donc

$$q_1 = (1, 0) \qquad \text{et} \qquad q_2 = \left(\cos(\theta_2/2)\ ,\ \frac{\sin(\theta_2/2)}{\theta_2}.r_2\right)$$

Nous n'avons besoin que du jacobien $\frac{\partial q_1}{\partial r_1}$: en reprenant les notations de la section (7.3.4.1), on calcule aisément avec $\theta_1 = 0$ que $\kappa_1 = \frac{1}{2}$ et $\lambda_1 = \frac{1}{24}$. Il ne reste plus qu'à reporter dans l'équation (7.24) pour obtenir :

$$\left.\frac{\partial q_1}{\partial r_1}\right|_{r_1=0} = \frac{1}{2}.\begin{bmatrix} 0 \\ I_3 \end{bmatrix}$$

7.3.5.2 Composition des quaternions

Comme $q_1 = 1$, on a évidemment $q = q_2 * q_1 = q_2$. Le jacobien est simplement :

$$\frac{\partial (q_2 * q_1)}{\partial q_1} = Q_{q_2} = \cos(\theta_2/2).I_4 + \frac{\sin(\theta_2/2)}{\theta_2}.\begin{bmatrix} 0 & -r_2^{\mathrm{T}} \\ r_2 & S_{r_2} \end{bmatrix}$$

La multiplication des deux premiers jacobiens donne donc :

$$\frac{\partial (q_2 * q_1)}{\partial q_1}.\frac{\partial q_1}{\partial r_1} = \frac{\cos(\theta_2/2)}{2}.\begin{bmatrix} 0 \\ I_3 \end{bmatrix} + \frac{\sin(\theta_2/2)}{2.\theta_2}.\begin{bmatrix} -r_2^{\mathrm{T}} \\ S_{r_2} \end{bmatrix}$$

7.3.5.3 Du quaternion unitaire au vecteur rotation

Comme $q = q(r_2)$, il est bien évident que $r(q) = r_2$. Le jacobien est par contre un peu plus compliqué. Avec les notations de la section (7.3.4.2), et puisque l'on sait que $a = \cos(\theta_2/2)$ et $v = \frac{\sin(\theta_2/2)}{\theta_2}.r_2$, on a :

$$\tau = 2.\text{sign}(a).\frac{\arcsin(\|v\|)}{\|v\|} = \frac{\theta_2}{\sin(\theta_2/2)} \qquad \text{et} \qquad \xi = \frac{2.a - \tau}{\mu^2} = \frac{2.\cos(\theta_2/2) - \frac{\theta_2}{\sin(\theta_2/2)}}{\sin(\theta_2/2)^2}$$

ce qui donne en reportant dans l'équation (7.25) :

$$\frac{\partial r}{\partial q} = \left[-2.\frac{\sin(\theta_2/2)}{\theta_2}.r_2 \;\; ; \;\; \frac{\theta_2}{\sin(\theta_2/2)}.I_3 + \left(2.\cos(\theta_2/2) - \frac{\theta_2}{\sin(\theta_2/2)}\right).\frac{r_2.r_2^{\mathrm{T}}}{\theta^2} \right]$$

Il ne reste plus qu'à multiplier cette matrice avec la précédente pour obtenir le résultat.

Théorème 7.12 (Jacobien de la translation à gauche de l'identité)
Soit r un vecteur rotation de norme $\|r\| = \theta$ et φ et ω les fonctions suivantes

$$\varphi = \frac{\theta/2}{\tan(\theta/2)} = 1 - \frac{\theta^2}{12} + O(\theta^4) \qquad et \qquad \omega = \frac{1-\varphi}{\theta^2} = \frac{1}{12} + \frac{\theta^2}{720} + O(\theta^4)$$

Le jacobien de la translation à gauche de l'identité est

$$J_L(r) = \left.\frac{\partial(r \circ e)}{\partial e}\right|_{e=0} = \varphi.I_3 + \omega.r.r^T + \frac{S_r}{2} \tag{7.26}$$

D'un point de vue numérique, il faudra aussi se méfier des valeurs de θ proches de π pour éviter les erreurs lors du calcul de $\tan(\theta/2)$. On utilisera alors :

$$\varphi = \frac{\theta.(\pi - \theta)}{4} + O\left((\pi - \theta)^3\right) \qquad \text{et} \qquad \omega = \frac{1-\varphi}{\theta^2}$$

7.3.6 Jacobien de la translation à droite de l'identité

La dérivation est complètement similaire à la translation à gauche, excepté que

$$\frac{\partial(q_2 * q_1)}{\partial q_2} = P_{q_1}^{\mathrm{T}} = \cos(\theta_1/2).I_4 + \frac{\sin(\theta_1/2)}{\theta_1}.\begin{bmatrix} 0 & -r_1^{\mathrm{T}} \\ r_1 & -S_{r_1} \end{bmatrix}$$

et on obtient :

Théorème 7.13 (Jacobien de la translation à droite de l'identité)
Soit r un vecteur rotation de norme $\|r\| = \theta$ et φ et ω les mêmes fonctions que pour la translation à gauche. Le jacobien de la translation à droite de l'identité est :

$$J_R(r) = \left.\frac{\partial(r \circ e)}{\partial e}\right|_{e=0} = \varphi.I_3 + \omega.r.r^T - \frac{S_r}{2} \tag{7.27}$$

7.3.7 Mesure invariante

Les déterminants des jacobiens de la translation à droite ou à gauche de l'identité sont

$$\det(J_L(r)) = \frac{\theta^2}{4.\sin(\theta/2)^2} \qquad \text{et} \qquad \det(J_R(r)) = \frac{-\theta^2}{4.\sin(\theta/2)^2}$$

Les mesures invariantes à droite et à gauche sont donc identiques, comme on pouvait s'y attendre puisque la métrique est bi-invariante.

Théorème 7.14 (Mesure bi-invariante)
Soit r un vecteur rotation de norme $\|r\| = \theta$. La mesure invariante (à droite comme à gauche) est

$$d\mathcal{SO}_3(r) = 4.\frac{\sin(\theta/2)^2}{\theta^2}.dr \tag{7.28}$$

7.3.8 Vérification des identités remarquables

Pour finir cette section sur le vecteur rotation et pour valider toutes ces formules, vérifions les identités remarquables du théorème (3.1). Pour commencer, notons que comme $S_{-r} = -S_r = S_r^{\mathrm{T}}$, on a :

$$J_L(r^{(-1)}) = J_L(-r) = J_L(r)^{\mathrm{T}} = J_R(r) = \varphi.I_3 + \omega.r.r^{\mathrm{T}} - \frac{S_r}{2}$$

Comme par définition $S_r.r = r \times r = 0$, on a $J_L(r^{(-1)}).r = (\varphi + \omega.\theta^2).r$, ce qui se simplifie, en sachant que $\omega = (1-\varphi)/\theta^2$, en

$$J_L(r^{(-1)}).r = r = -r^{(-1)}$$

De la même façon, on a

$$J_L(r).(-r) = -J_L(r).r = -r$$

L'identité (3.28) est donc vérifiée. L'identité (3.29) est trivialement vérifiée puisqu'elle s'exprime $(-I_3)^{\mathrm{T}}.(-r) = r$.

7.4 Transformations rigides en 3D et repères

Soit $\mathcal{B} = \{o, i, j, k\}$ la base orthonormée directe canonique de l'espace euclidien $\mathbb{R}^3$ (on a donc $o = (0,0,0)^{\mathrm{T}}$ et $[i,j,k] = I_3$). Considérons maintenant une autre base orthonormée directe $\mathcal{F} = \{t, i', j', k'\}$. Elle est formée d'un point de coordonnées t dans $\mathcal{B}$ et de trois vecteurs unitaires orthogonaux i', j' et k' qui forment un trièdre. $\mathcal{F}$ est donc un **repère**. Le mouvement rigide amenant de $\mathcal{B}$ à $\mathcal{F}$ est unique et s'écrit en coordonnées homogènes :

$$M = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad \text{avec} \quad R = \left[i', j', k'\right]$$

On peut donc écrire une transformation rigide (ou de manière équivalente un repère) comme le couple $\mathrm{f} = (R, t)$, avec $R \in \mathcal{SO}_3$ et $t \in \mathbb{R}^3$. Cette façon d'écrire une transformation rigide (ou *mouvement* pour d'alléger les écritures) met bien en évidence la structure de variété que l'on écrira $\mathcal{M}_3 = \mathcal{SO}_3 \ltimes \mathbb{R}^3$. En effet, $\mathcal{M}_3$ est alors une sous-variété de $\mathbb{R}^{3\times3} \times \mathbb{R}^3$ avec les contraintes déjà connues sur les matrices de rotation.

Pour déterminer les opérations de base sur les mouvements, et en particulier les règles de composition et d'inversion compatibles avec l'action des mouvements sur l'espace euclidien $\mathbb{R}^3$, observons (en passant au besoin par les matrices homogènes) que l'action de f sur un vecteur homogène est :

$$\mathrm{f} \star x = R.x + t$$

Pour exprimer la composition, il suffit de multiplier les matrices homogènes :

$$\mathrm{f}_1 \circ \mathrm{f}_2 = (R_1.R_2\ ,\ R_1.t_2 + t_1)$$

et l'identité étant $\mathrm{Id} = (I_3, 0)$, l'inversion est :

$$\mathrm{f}^{(-1)} = (R^{(-1)}\ ,\ -R^{(-1)}.t)$$

Nous avons les opérations de base sur le groupe, il nous faut maintenant une carte locale pour déterminer les géodésiques.

7.4.1 Carte principale

Pour obtenir une représentation minimale, il paraît raisonnable d'utiliser le vecteur rotation $r = \theta.n$ associé à R et le vecteur translation t. On peut alors représenter à la fois la transformation rigide de $\mathcal{B}$ à $\mathcal{F}$ et le repère $\mathcal{F}$ (tous deux exprimés dans $\mathcal{B}$) par le vecteur de dimension 6 :

$$f = \left| \begin{array}{c} r \\ t \end{array} \right.$$

Notons qu'on aurait également pu utiliser les angles d'Euler ou d'autres représentations minimales des rotations, mais cette représentation s'avérera la bonne puisqu'il s'agira en fait de la représentation exponentielle pour la métrique invariante *à gauche*. Pour simplifier les notations, nous écrirons comme pour les quaternions $f = (r, t) \in \mathcal{M}_3$ tout en considérant f comme un vecteur 6D.

Les opérations de composition et d'inversion se traduisent aisément dans cette carte en utilisant les opérations de la section précédente sur le vecteur rotation :

$$f_2 \circ f_1 = (r_2 \circ r_1\ ,\ r_2 \star t_1 + t_2) \qquad \text{et} \qquad f^{(-1)} = (r^{(-1)}\ ,\ r^{(-1)} \star (-t))$$

Notons que l'identité est bien le vecteur nul dans cette représentation.

7.4.1.1 Métrique invariante à gauche

Le jacobien de la translation à gauche s'exprime alors par

$$J_L(f) = \left.\frac{\partial (f \circ e)}{\partial e}\right|_{e=0} = \begin{bmatrix} \left.\frac{\partial (r \circ e_r)}{\partial e_r}\right|_{e_r=0} & 0 \\ 0 & R \end{bmatrix} = \begin{bmatrix} J_L(r) & 0 \\ 0 & R \end{bmatrix} \tag{7.29}$$

Choisissons une métrique Q à l'origine qui soit compatible avec la métrique euclidienne canonique sur $\mathbb{R}^3$ et avec la métrique canonique sur les rotations :

$$Q = \begin{bmatrix} I_3 & 0 \\ 0 & I_3 \end{bmatrix} \qquad \text{d'où} \qquad Q(f) = J_L(f)^{(-\mathrm{T})}.Q.J_L(f)^{(-1)} = \begin{bmatrix} Q(r) & 0 \\ 0 & I_3 \end{bmatrix}$$

Comme cette matrice est diagonale par bloc, les géodésiques partant de l'identité sont le produit direct des géodésiques sur les translations et les rotations : $f(s) = (s.r\ ,\ s.t)$.

Théorème 7.15 *La représentation $f = (r, t)$ munie du domaine $\mathcal{D} = \mathcal{B}_3(0, \pi) \times \mathbb{R}^3$ est la carte principale pour le groupe des transformations rigides $\mathcal{M}_3$ muni de la distance invariante à gauche canonique.*

Remarquons que la « norme » d'un mouvement est $N_L(\mathrm{f}) = \|f\| = \sqrt{\theta(R)^2 + \|t\|^2}$. Nous aurions pu ajouter un paramètre λ par exemple sur la métrique des rotations pour permettre de pondérer l'influence des radians (pour la rotation) sur les mètres (ou les inches...). On aurait alors obtenu :

$$N_\lambda(f)^2 = N_\lambda((r,t))^2 = \lambda^2.\|r\|^2 + \|t\|^2$$

On peut vérifier que cette « norme » définit bien une distance sur le groupe et vérifie les contraintes que nous avions développé à la section (3.4.3).

Preuve : (N_λ est une norme au sens de la section (3.4.3))
Cette norme est positive et nulle seulement pour $\|r\| = \|t\| = 0$, c'est-à-dire pour l'identité. Si $f = (r, t)$, la transformation inverse est $f^{(-1)} = (r^{(-1)}, r^{(-1)} \star (-t))$ et comme $\theta = \|r\|$ est une norme sur les rotations, on a :

$$N_\lambda(f^{(-1)})^2 = \lambda^2.\|r^{(-1)}\|^2 + \| - R^{\mathrm{T}}.t\|^2 = N_\lambda(f)^2$$

L'inégalité triangulaire se montre à partir de l'inégalité triangulaire sur les métriques des rotations et des translations. En effet, soient $f_1 = (r_1, t_1)$ et $f_2 = (r_2, t_2)$ deux mouvements :

$$f_1^{(-1)} \circ f_2 = \left(r_1^{(-1)} \circ r_2\ ;\ r_1^{(-1)} \star (t_2 - t_1)\right)$$

et donc $N_\lambda(f_1^{(-1)} \circ f_2)^2 = \lambda^2\ \|r_1^{(-1)} \circ r_2\|^2 + \|R_1^{\mathrm{T}}.(t_2 - t_1)\|^2$. L'inégalité triangulaire sur les rotations assure que $\theta(r_1^{(-1)} \circ r_2) \leq \theta_1 + \theta_2$ (où $\theta_i = \|r_i\|$) et nous avons sur les vecteurs : $\|t_2 - t_1\|^2 = \|t_1\|^2 + \|t_1\|^2 - 2\,\langle t_1 \mid t_2 \rangle$. Donc :

$$N_\lambda(f_1^{(-1)} \circ f_2)^2 \leq \theta_1^2 + \theta_2^2 + 2\theta_1\theta_2 + \|t_1\|^2 + \|t_2\|^2 + 2\|t_1\|\|t_2\|$$

Comme $(\theta_1\theta_2 + \|t_1\|\|t_2\|)^2 = (\theta_1^2 + \|t_1\|^2)(\theta_2^2 + \|t_2\|^2) - (\theta_1\|t_2\| - \theta_2\|t_1\|)^2$, on obtient

$$\begin{aligned} N_\lambda(f_1^{(-1)} \circ f_2)^2 &\leq (\theta_1^2 + \|t_1\|^2) + 2\sqrt{(\theta_1^2 + \|t_1\|^2)(\theta_2^2 + \|t_2\|^2)} + (\theta_2^2 + \|t_2\|^2) \\ &\leq \left(\sqrt{\theta_1^2 + \|t_1\|^2} + \sqrt{\theta_2^2 + \|t_2\|^2}\right)^2 \end{aligned}$$

Il suffit de prendre la racine carrée pour obtenir l'inégalité recherchée :

$$N_\lambda(f_1^{(-1)} \circ f_2) \leq N_\lambda(f_1) + N_\lambda(f_2)$$

■

7.4.1.2 Mesure invariante

Comme on a calculé le jacobien de la translation à gauche, on peut calculer celui de la translation à droite :

$$J_R(f) = \left.\frac{\partial(e \circ f)}{\partial e}\right|_{e=0} = \begin{bmatrix} \left.\frac{\partial(e_r \circ r)}{\partial e_r}\right|_{e_r=0} & 0 \\ \left.\frac{\partial(e_r \star t)}{\partial e_r}\right|_{e_r=0} & I_3 \end{bmatrix} = \begin{bmatrix} J_R(r) & 0 \\ -S_t & I_3 \end{bmatrix} \tag{7.30}$$

Nous ne résoudrons pas ici les équations des géodésiques pour la métrique invariante à droite, mais on peut vérifier qu'elles sont différentes. On peut également calculer les équations des courbes intégrales (les sous-groupes à un paramètre), et s'apercevoir que ces courbes ne correspondent pas non plus aux géodésiques pour la distance invariante à gauche.

Par contre, les mesures invariantes à droite et à gauche sont identiques : le groupe est unimodulaire. En effet, les déterminants des translations à gauche et à droite de l'identité sont :

$$\det(J_L(f)) = \det(J_L(r)).\det(R) \qquad \text{et} \qquad \det(J_R(f)) = \det(J_R(r)).\det(I_3)$$

et comme $\det(R) = 1$ et $\det(J_L(r)) = \det(J_R(r))$, on obtient :

$$d\mathcal{M}_3(f) = d_L\mathcal{M}_3(f) = d_R\mathcal{M}_3(r) = \frac{4.\sin^2(\theta/2)}{\theta^2}.dr.dt = \frac{4.\sin^2(\theta/2)}{\theta^2}.df \tag{7.31}$$

7.4.2 Opérations atomiques sur les mouvements

Maintenant que nous avons la carte principale, il ne nous reste plus qu'à déterminer les jacobiens de l'inversion et de la composition des mouvements, ainsi que la normalisation.

7.4.2.1 Inversion d'un mouvement

$$f = (r, t) \quad \longmapsto \quad f^{(-1)} = (r^{(-1)}\ ,\ r^{(-1)} \star (-t))$$

Comme l'inverse du vecteur rotation est $r^{(-1)} = -r$, le jacobien est

$$J_{Inv} = \begin{bmatrix} -I_3 & 0 \\ \frac{\partial(r^{(-1)} \star t)}{\partial r^{(-1)}} & -R^{\mathrm{T}} \end{bmatrix} \tag{7.32}$$

7.4.2.2 Composition des mouvements

$$f_2 = (r_2\ ,\ t_2)\ ,\ \ f_1 = (r_1\ ,\ t_1) \quad \longmapsto \quad f_2 \circ f_1 = (r_2 \circ r_1\ ;\ r_2 \star t_1 + t_2)$$

On notera J_1 le jacobien de $f = f_2 \circ f_1$ par rapport à r_1 et J_2 celui par rapport à r_2. Ils sont donnés par

$$J_1 = \frac{\partial(f_2 \circ f_1)}{\partial f_1} = \frac{\partial(r_2 \circ r_1\ ,\ r_2 \star t_1 + t_2)}{\partial(r_1, t_1)} = \begin{bmatrix} \frac{\partial(r_2 \circ r_1)}{\partial r_1} & 0 \\ 0 & R_2 \end{bmatrix} \tag{7.33}$$

$$J_2 = \frac{\partial(f_2 \circ f_1)}{\partial f_2} = \frac{\partial(r_2 \circ r_1\ ,\ r_2 \star t_1 + t_2)}{\partial(r_2, t_2)} = \begin{bmatrix} \frac{\partial(r_2 \circ r_1)}{\partial r_2} & 0 \\ \frac{\partial(r_2 \star t_1)}{\partial r_2} & I_3 \end{bmatrix} \tag{7.34}$$

7.4.2.3 Normalisation d'un mouvement (Domaine)

$$f = (r, t) \in \mathbb{R}^6 \quad \longmapsto \quad f' = (r', t') \in \mathcal{D}$$

Les seules contraintes de domaine sont sur le vecteur rotation : soit $r' = D(r)$ la normalisation du vecteur rotation et $J_D(r)$ son jacobien. La normalisation du mouvement f et son jacobien sont alors :

$$f' = D(f) = (D(r), t) \qquad \text{et} \qquad J_D(f) = \frac{\partial f'}{\partial f} = \begin{bmatrix} J_D(r) & 0 \\ 0 & I_3 \end{bmatrix}$$

7.4.3 Vérification des identités remarquables

Rappelons que le jacobien de la translation à gauche de l'identité est :

$$J_L(f) = \begin{bmatrix} J_L(r) & 0 \\ 0 & R \end{bmatrix}$$

Comme $J_L(r^{(-1)}).r = r$, il suffit de poser la multiplication pour voir que $J_L(f^{(-1)}).f = -f^{(-1)}$. De même, comme $J_L(r).(-r) = -r$, on obtient $J_L(f).f^{(-1)} = -f$. L'identité (3.28) est donc vérifiée.

Par contre, pour l'identité (3.29), nous devons observer que

$$J_{Inv}^{\mathrm{T}} = \begin{bmatrix} -I_3 & \frac{\partial\left(r^{(-1)} \star t\right)^{\mathrm{T}}}{\partial r^{(-1)}} \\ 0 & -R \end{bmatrix}$$

Or, en prenant la formule (7.21), on a $\frac{\partial\left(r^{(-1)} \star t\right)}{\partial r^{(-1)}} = -S_{(R.t)}.U_r$, soit : $\frac{\partial\left(r^{(-1)} \star t\right)^{\mathrm{T}}}{\partial r^{(-1)}} = U_{(-r)}^{\mathrm{T}}.S_{(R^{\mathrm{T}}.t)}$ d'où

$$\frac{\partial\left(r^{(-1)} \star t\right)^{\mathrm{T}}}{\partial r^{(-1)}}.R^{\mathrm{T}}.t = U_{(-r)}^{\mathrm{T}}.S_{(R^{\mathrm{T}}.t)}.(R^{\mathrm{T}}.t) = 0$$

car $S_y.y = y \times y = 0$. On a donc bien :

$$\frac{\partial f^{(-1)\,\mathrm{T}}}{\partial f}.f^{(-1)} = \left| \begin{array}{l} r + \frac{\partial\left(r^{(-1)} \star t\right)^{\mathrm{T}}}{\partial r}.R^{\mathrm{T}}.t \\ t \end{array} \right. = f$$

7.4.4 Opérations atomiques sur les repères

D'un point de vue purement différentiel, les repères sont identiques (difféomorphes) aux mouvements. On notera donc également $\mathcal{M}_3$ la variété des repères. On choisit bien évidemment l'origine des repères comme étant la base canonique : elle correspond dans notre représentation à l'identité des mouvements. Le groupe d'isotropie $\mathcal{H}$ étant réduit à l'identité, on a une parfaite identification des mouvements et des repères. On notera d'ailleurs $f = (r, x)$ un repère correspondant au mouvement $f = (r, x)$.

En particulier, la fonction de placement est l'identité : $f_f = f$ et son jacobien est donc $J(f_f) = I_6$, et il suffit de revenir à la définition de $f = (r, t)$ comme repère ou mouvement pour voir que l'action du mouvement g sur le repère f est simplement : $g \star f = g \circ f$. Le jacobien est évidemment également celui de la composition. Le domaine est strictement identique au cas des mouvements.

7.5 Repères semi-orientés et non-orientés

En imagerie médicale volumique, nous utiliserons beaucoup les **points extrémaux** définis dans (Thirion et Gourdon, 1995). Ce sont des points sur une iso-surface qui optimisent un critère de géométrie différentielle basé sur la courbure. Certains de ces points peuvent ainsi être vu comme la généralisation des points de coin. Étant définis sur une surface, ces points sont munis des deux directions principales de courbure (t_1, t_2) et de la normale à la surface, ce qui forme un repère car ces trois vecteurs sont orthonormés.

En fait, ce n'est pas tout à fait un repère, mais ce que l'on appellera un repère semi-orienté car les directions principales t_1 et t_2 ne sont que des *directions*, c'est-à-dire que l'on mesure indifféremment $\pm t_1$ et $\pm t_2$

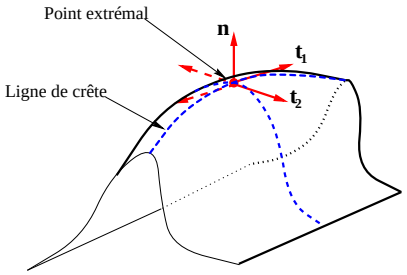


FIG. 7.1 – *Un point extrémal sur une iso-surface*

7.5.1 Trièdres semi-orientés

Le trièdre est formé des vecteurs $\{\pm t_1, \pm t_2, n\}$.On peut toujours choisir l'orientation du vecteur t_2 pour que ce trièdre soit orthonormé direct. Il nous reste alors les deux trièdres $\{\varepsilon.t_1, \varepsilon.t_2, n\}$ ($\varepsilon = \pm 1$) pour représenter la même primitive. Choisissons pour origine de notre variété le repère canonique (avec $t_1 = e_x$, $t_2 = e_y$ et $n = e_3$), alors la figure (7.2) montre que le groupe d'isotropie est $\mathcal{H} = \{I_3, \Pi_3\}$ où $\Pi_3 = \mathcal{R}(\pi, e_3)$ est la rotation d'angle π autour de $n = e_3$.

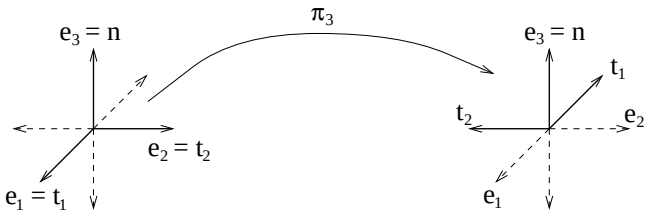


FIG. 7.2 – *Les deux trièdres représentant le trièdre semi-orienté « origine ».*

L'ensemble des trièdres semi-orientés est donc le quotient de l'espace des trièdres (c'est-à-dire des rotations) par $\mathcal{H}$: notre variété est $\mathcal{SO}_3/\mathcal{H}$. Comme le groupe d'isotropie $\mathcal{H}$ est discret et fini, il existe une distance invariante induite (voir section (3.4.4)) et on peut encore représenter une telle primitive par un vecteur rotation. Les géodésiques sont toujours les mêmes, seul le domaine de définition de la carte principale change.

7.5.1.1 Carte principale

Soit $s = \theta.n$ le vecteur rotation représentant l'un des deux trièdres semi-orientés possibles (avec $\theta = \|s\| \leq \pi$), le second étant $s' = s \circ \pi_3$, avec $\pi_3 = \pi.e_3$. Le vecteur qui appartient au domaine de définition de la carte principale est celui qui a la norme minimale. En passant par les quaternions, on a :

$$p = \left| \begin{array}{l} \cos(\theta/2) \\ \sin(\theta/2).n \end{array} \right. \qquad \text{et} \qquad q_3 = \left| \begin{array}{l} 0 \\ e_3 \end{array} \right.$$

soit :

$$p' = \pm \left| \begin{array}{l} \cos(\theta'/2) \\ \sin(\theta'/2).n \end{array} \right. = p * q_3 = \left| \begin{array}{l} \sin(\theta/2).\langle e_3 \mid n \rangle \\ \sin(\theta/2).n \times e_3 + \cos(\theta/2).e_3 \end{array} \right.$$

Comme les angles θ et θ' sont compris entre 0 et π, les sinus et cosinus sont positifs et on obtient : $\cos(\theta'/2) = \sin(\theta/2).|n_{[3]}|$, où $n_{[3]}$ est la composante de n selon e_3. Cosinus est une fonction décroissante sur le domaine qui nous intéresse et θ' est donc supérieur à θ si $\cos(\theta'/2) < \cos(\theta/2)$. On

obtient au final :

$$\theta \leq \theta' \quad \Longleftrightarrow \quad \tan\left(\frac{\theta}{2}\right) \leq \frac{1}{|n_{[3]}|} \tag{7.35}$$

Le domaine de définition est donc :

$$\mathcal{D} = \left\{ s = (s_{[1]}, s_{[2]}, s_{[3]})^{\mathrm{T}} \in \mathbb{R}^3 \quad / \quad \|s\| \leq \pi^2 \quad \text{et} \quad |s_{[3]}|.\tan\left(\frac{\|s\|}{2}\right) \leq \|s\| \right\}$$

7.5.1.2 Opérations atomiques sur les trièdres semi-orientés

La fonction de placement de $\mathcal{SO}_3/\mathcal{H}$ dans $\mathcal{SO}_3$ est facile à choisir : c'est l'identité. Avec ce choix, le jacobien de la translation de l'origine est $J_L(s)$, celui du vecteur rotation.

Domaine Pour normaliser le trièdre semi-orienté s, on commence par le normaliser comme si c'était un vecteur de rotation simple : $s' = D(s)$ avec le jacobien $J_D(s)$, puis on vérifie la contrainte supplémentaire :

$$\text{si} \qquad |s'_{[3]}|.\tan\left(\frac{\|s'\|}{2}\right) \leq \|s'\| \qquad \text{alors} \qquad s'' = s' \circ \pi_3 \quad \text{et} \quad J'' = \frac{\partial(s' \circ \pi_3)}{\partial s'}.J_D(s)$$

et on retourne s'' et J'', sinon on retourne s' avec le jacobien $J_D(s)$.

Action d'une rotation On note s notre trièdre semi-orienté, et $s' = r \circ s$ le résultat de la composition des vecteurs rotation, avec les jacobiens J_1 et J_2.

$$s' = r \circ s \qquad J_1 = \frac{\partial s'}{\partial r} = \frac{\partial(r \circ s)}{\partial r} \qquad J_2 = \frac{\partial s'}{\partial s} = \frac{\partial(r \circ s)}{\partial s}$$

Comme s' n'est pas forcément dans le domaine, il faut éventuellement le normaliser :

$$\text{si} \qquad |s'_{[3]}|.\tan\left(\frac{\|s'\|}{2}\right) \leq \|s'\| \qquad \text{alors} \qquad s'' = s' \circ \pi_3 \quad \text{et} \quad J' = \frac{\partial(s' \circ \pi_3)}{\partial s'}.$$

et on retourne $r \star s = s''$ avec les jacobiens $J''_1 = J'.J_1$ et $J''_2 = J'.J_2$, sinon on retourne $r \star s = s'$ avec les jacobiens J_1 et J_2.

7.5.2 Trièdres non-orientés

On peut ainsi continuer à réduire les caractéristiques de notre repère : supposons maintenant que l'on veuille utiliser les points extrémaux pour faire du recalage multi-modalité, par exemple entre une image IRM et une image scanner X. Comme ces images ne mesurent pas les mêmes caractéristiques des tissus, un certain nombre d'interfaces entre organes seront inversées. C'est en particulier le cas au niveau du bord de l'os. Cette fois-ci, même la normale du point extrémal n'est plus orientée : on peut mesurer n ou $-n$.

Le trièdre est cette fois-ci formé des vecteurs $\{\pm t_1, \pm t_2, \pm n\}$. On peut toujours choisir l'orientation des vecteurs pour qu'il soit orthonormé direct. Il nous reste alors quatre trièdres possibles pour représenter le même trièdre non-orienté, illustrés sur la figure (7.3). Choisissons pour origine de notre variété le repère canonique (avec $t_1 = e_x$, $t_2 = e_y$ et $n = e_3$), alors le groupe d'isotropie est $\mathcal{H} = \{I_3, \Pi_1, \Pi_2, \Pi_3\}$ où $\Pi_i = \mathcal{R}(\pi, e_i)$ est la rotation d'angle π autour du vecteur de base e_i.

Comme pour le cas des trièdres semi-orientés, notre variété est $\mathcal{SO}_3/\mathcal{H}$, et la carte principale est constituée du vecteur rotation mais avec un domaine plus restreint. La détermination de ce domaine et des opérations atomiques est calquée sur celle des trièdres semi-orientés.

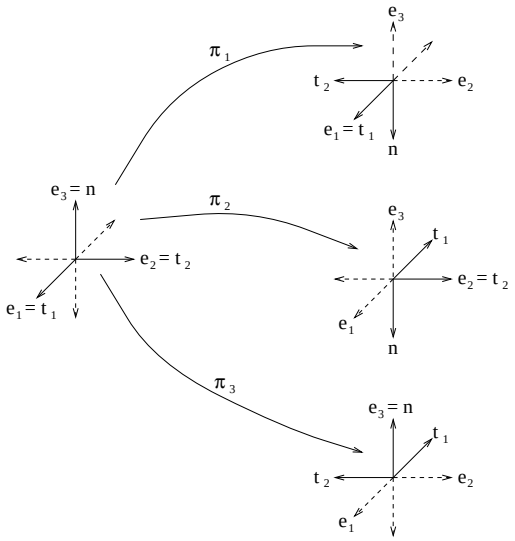


FIG. 7.3 – *Les quatre trièdres représentant le trièdre non-orienté « origine ».*

7.5.2.1 Carte principale

Soit $s = \theta.n$ le vecteur rotation représentant l'un des trièdres non-orientés possibles (avec $\theta = \|s\| \leq \pi$), les trois autres étant $s_i = s \circ \pi_i$, avec $\pi_i = \pi.e_i$. Le vecteur qui appartient au domaine de définition de la carte principale est celui qui a la norme minimale. En passant par les quaternions, on trouve cette fois-ci :

$$\cos(\theta_i/2) = \sin(\theta/2).|n_{[i]}|$$

Pour trouver le vecteur rotation de norme minimale, on peut déjà comparer les θ_i entre eux :

$$\theta_i \leq \theta_j \iff |n_{[i]}| \geq |n_{[j]}| \iff |s_{[i]}| \geq |s_{[j]}|$$

et parallèlement comparer θ et les θ_i

$$\theta \leq \theta_i \iff \tan\left(\frac{\theta}{2}\right) \leq \frac{1}{|n_{[i]}|}$$

Le domaine de définition est donc :

$$\mathcal{D} = \left\{ s = (s_{[1]}, s_{[2]}, s_{[3]})^{\mathrm{T}} \in \mathbb{R}^3 \quad / \quad \|s\| \leq \pi^2 \quad \text{et} \quad \forall\, i \in \{1,2,3\} \quad |s_{[i]}|.\tan\left(\frac{\|s\|}{2}\right) \leq \|s\| \right\}$$

7.5.2.2 Opérations atomiques sur les trièdres non-orientés

La fonction de placement de $\mathcal{SO}_3/\mathcal{H}$ dans $\mathcal{SO}_3$ est facile à choisir : c'est l'identité. Avec ce choix, le jacobien de la translation de l'origine est celui du vecteur rotation.

Domaine Pour normaliser le trièdre non-orienté s, on commence par le normaliser comme si c'était un vecteur de rotation simple : $s' = D(s)$ avec le jacobien $J_D(s)$, puis cherche l'indice k de la composante la plus grande (en valeur absolue) de ce vecteur s' :

$$k = \arg \max_{i \in \{1,2,3\}} |s'_{[i]}|$$

et on vérifie la contrainte supplémentaire :

$$\text{si} \quad |s'_{[k]}|.\tan\left(\frac{\|s'\|}{2}\right) \leq \|s'\| \quad \text{alors} \quad s'' = s' \circ \pi_k \quad \text{et} \quad J'' = \frac{\partial(s' \circ \pi_k)}{\partial s'}.J_D(s)$$

sinon on retourne $r \star s = s'$ avec le jacobien $J_D(s)$.

Action d'une rotation Le procédé est strictement calqué sur l'action d'une rotation sur un trièdre semi-orienté, mais avec le choix de l'indice k comme ci-dessus qui remplace l'indice fixe 3.

7.5.3 Repères semi et non-orientés

Pour construire maintenant des repères semi ou non-orientés, il suffit d'ajouter un point 3D qui représente la localisation de ce repère dans l'espace : $z = (s, x)$, exactement comme on l'a fait pour les repères, le trièdre semi ou non-orienté s et ses opérations atomiques remplaçant le vecteur rotation. Nous résumons ici les opérations atomiques obtenues.

- Fonction de placement : $f_z = z \qquad \frac{\partial f_z}{\partial z} = I_6$
- Translation de l'origine :
$$J(f_z) = \begin{bmatrix} J_L(s) & 0 \\ 0 & R(s) \end{bmatrix}$$
où $J_L(s)$ est le jacobien de la translation à gauche de l'identité pour le vecteur rotation et $R(s)$ la rotation associée à s (considéré comme vecteur rotation).
- Normalisation :
$$z' = D(z) = (D(s), x) \qquad J_D(z) = \begin{bmatrix} J_D(s) & 0 \\ 0 & I_3 \end{bmatrix}$$
où $D(s)$ est la normalisation des repères semi ou non-orientés et $J_D(s)$ son jacobien.
- Action d'un mouvement $f = (r, t)$:
$$f \star z = (r \star s\ ,\ R.x + t) \qquad \frac{\partial(f \star z)}{\partial z} = \begin{bmatrix} \frac{\partial(r \star s)}{\partial s} & 0 \\ 0 & R \end{bmatrix} \qquad \frac{\partial(f \star z)}{\partial f} = \begin{bmatrix} \frac{\partial(r \star s)}{\partial r} & 0 \\ \frac{\partial(r \star x)}{\partial r} & I_3 \end{bmatrix}$$
Il faut bien noter dans ces formules que $\frac{\partial(r \star x)}{\partial r}$ est le jacobien de l'action du vecteur rotation *sur le point* x tandis que $\frac{\partial(r \star s)}{\partial r}$ et $\frac{\partial(r \star s)}{\partial s}$ sont les jacobiens de l'action du vecteur rotation *sur le repère semi ou non-orienté* s.

7.6 Points

« Puisque le point n'a pas (théoriquement, du moins) d'existence, il est bizarre que l'idée que l'on se fait du point soit celle d'un rond. Mais ne pourrait-il pas y avoir des points carrés ou triangulaires ? »
Claude Cossette, Les Images démaquillées: approche scientifique de la communication par l'image.

Les points de $\mathbb{R}^3$ étant justement les primitives avec lesquelles on savait travailler sans notre théorie, il est intéressant de les inclure aussi dans notre implémentation et de vérifier qu'on obtient bien les résultats attendus.

7.6.1 Points de $\mathbb{R}^3$

Il est naturel de prendre comme origine $\mathrm{o} = (0,0,0)^{\mathrm{T}}$. Nous avons déjà vu que l'action d'une transformation $f = (r,t)$ est $f \star x = R.x + t$. Le groupe d'isotropie est obtenu en résolvant l'équation $R.0 + t = 0$, ce qui impose une translation nulle :

$$\mathcal{H} = \{(R,0) \ / \ R \in \mathcal{S}O_3\}$$

Les cosets sont donc les ensembles

$$\mathcal{F}_{\mathrm{x}} = \{(R,x) \ / \ R \in SO_3\}$$

Un représentant particulièrement agréable à manipuler est $f_x = (0,x)$, la translation de vecteur x. Cette fonction de placement conduit en particulier à des bruits homogènes additifs, ce qui est la modélisation usuelle sur les points. Le jacobien de la translation de l'origine est alors

$$J(f_x) = \frac{\partial(f_x^{(-1)} \star e)}{\partial e} = \frac{\partial(e-x)}{\partial e} = I_3$$

et la métrique invariante est donc bien la métrique canonique : $Q(x) = Id_3$. Il n'y a pas de problème de normalisation puisque le domaine de définition est $\mathbb{R}^3$ tout entier, et il ne reste donc plus qu'à déterminer les jacobiens de l'action d'une transformation rigide $f = (r,t)$:

$$\frac{\partial(f \star x)}{\partial f} = \frac{\partial(r \star x + t)}{\partial(r,t)} = \left[\frac{\partial(r \star x)}{\partial r} \ ; \ \frac{\partial t}{\partial t}\right] = \left[\frac{\partial(r \star x)}{\partial r} \ ; \ I_3\right]$$

$$\frac{\partial(f \star x)}{\partial x} = \frac{\partial(r \star x + t)}{\partial x} = R$$

7.7 Conclusion

Nous avons montré dans ce chapitre que l'on pouvait mener les calculs nécessités par la théorie des chapitres précédents sur un nombre important de primitives tridimensionnelles sous l'action du groupe des transformations rigides. Ayant implémenté les opérations atomiques dérivées dans ce chapitre, nous avons à notre disposition tous les algorithmes bas et moyen niveau développé précédemment.

Nous nous focaliserons dans la seconde partie de ce manuscrit sur des algorithmes haut niveau et leurs applications, utilisant pour cela les primitives définies dans ce chapitre.

Chapitre 8

Recalage et fusion de primitives : estimation et précision

« Errors using inadequate data are much less than those using no data at all. »
Charles Babbage

On s'intéresse dans ce chapitre à deux principaux problèmes d'estimation. Le premier est l'estimation d'un mouvement rigide à partir d'appariement de primitives et le calcul de l'incertitude sur cette estimation : c'est le **recalage**. Le second problème, déjà abordé dans la synthèse de la première partie (chap. 6), concerne l'obtention de la **moyenne** ou la **fusion** de primitives probabilistes. Nous nous attachons également à calculer l'incertitude sur l'estimation obtenue. Les solutions apportées à ces deux problèmes sont très similaires puisqu'elles passent toutes par une minimisation aux moindres carrés. Cependant, il n'y a guère que dans le cas des points où il existe une solution explicite, abondamment illustrée dans la littérature (Arun et al., 1987; Horn, 1987; Umeyama, 1991; Zhuang et Huang, 1994) et très peu de travaux s'intéressent à l'incertitude sur cette estimation (on pourra cependant noter (Kanatani, 1993) et (Csurka et al., 1995)).

Nous rappelons dans la section (8.1) les principales solutions explicites connues pour l'estimation de transformations aux moindres carrés à partir d'appariements de points et nous développons une méthode pour en estimer l'incertitude. Les solutions explicites ne se généralisant pas au cas de primitives quelconques, nous devons développer dans la section (8.2) des méthodes par descente de gradient ou filtrage de Kalman pour traiter le cas de primitives géométriques plus générales.

Par contre, les méthodes d'estimation de l'incertitude sur la transformation se généralisent assez bien (section 8.3), mais reposent sur la connaissance du bruit sur les données (les primitives). Nous étudions donc dans la section (8.4) l'estimation a posteriori du modèle de bruit (i.e. après fusion ou recalage), de manière à pouvoir réinjecter cette connaissance dans les algorithmes de fusion et de recalage pour en déterminer l'incertitude : c'est la base de l'algorithme d'estimation pratique présenté à la section (8.5).

8.1 Recalage à partir d'appariements de points

Pour introduire le problème du recalage, nous supposerons dans cette section que les objets sont représentés par des ensembles de N points appariés $\{x_i\}$ et $\{y_i\}$ dans $\mathbb{R}^3$, et que la transformation entre ces objets est rigide. S'il n'y avait aucune erreur d'appariement ni de mesure, on aurait donc $y_i = R.x_i + t$ pour tout i, et on pourrait déterminer les paramètres du mouvement avec seulement trois couples de points appariés en position générique : on a alors 9 équations scalaires pour 6 inconnues (3 pour la rotation et 3 pour la translation) et 3 invariants (les distances entre les points).

Malheureusement, on n'a jamais des données exactes, et l'on doit donc prendre en compte les erreurs de mesures. A cause des 3 invariants, il est très peu probable que le système évoqué ci-dessus ait une solution car il est sur-contraint. La solution usuelle est alors de minimiser l'erreur aux moindres carrés. La résolution de ce problème est bien connue, soit par les quaternions (Horn, 1987; Ayache, 1991; Faugeras, 1993; Horaud et Monga, 1993), soit par la décomposition en valeurs singulières (SVD) (Arun et al., 1987; Umeyama, 1991), et sera rappelée à la section (8.1.1). Les mêmes techniques nous permettrons également de faire une courte incursion dans le cas non-rigide pour calculer la similitude et la transformation affine aux moindres carrés (section 8.1.2).

Le second problème est que la transformation que l'on calcule à partir des données bruitées n'est évidemment pas la transformation exacte entre les primitives exactes. Il est donc important, voire vital dans certaines applications médicales (planification de chirurgie par exemple), d'estimer l'incertitude que l'on a sur notre transformation. Nous verrons une méthode qui permet de le faire pour la transformation aux moindres carrés et une méthode basée sur le filtrage de Kalman qui permet en plus de prendre en compte l'incertitude sur les points dans notre estimation.

8.1.1 Calcul de la transformation rigide aux moindres carrés

On cherche les paramètres de la rotation R et la translation t minimisant le critère :

$$C(R,t) = \sum_i \|y_i - R.x_i - t\|^2 \tag{8.1}$$

8.1.1.1 Calcul de la translation

La translation optimale est caractérisée par une dérivée nulle du critère :

$$\frac{\partial C}{\partial t} = -2.\sum_i (y_i - R.x_i - t)^{\mathrm{T}}. = 0 \qquad \Longleftrightarrow \qquad \sum_i y_i - R.\left(\sum_i x_i\right) = N.t$$

En notant $\bar{x} = \frac{1}{N}.\sum_i x_i$ et $\bar{y} = \frac{1}{N}.\sum_i y_i$ les barycentres des deux ensembles de points, on obtient donc :

$$\hat{t} = \bar{y} - R.\bar{x} \tag{8.2}$$

et en exprimant les points en repère barycentrique : $x'_i = x_i - \bar{x}$ et $y'_i = y_i - \bar{y}$, le critère à minimiser se réécrit :

$$C'(R) = \sum_i \|y'_i - R.x'_i\|^2$$

Notons que cette caractérisation de la translation optimale ne suppose pas que R soit une rotation, mais que ce soit simplement un opérateur linéaire. L'équation (8.2) est donc encore valide si l'on recherche une similitude ou une transformation affine au lieu d'un mouvement. De même, cette caractérisation est valable dans n'importe quelle dimension, et pas seulement en dimension 3.

Dans la suite de cette section, nous supposerons que les points x_i et y_i sont exprimés en repère barycentrique, ce qui permet de s'affranchir de la translation.

8.1.1.2 Rotation : méthode des quaternions

Soit q un quaternion rotation (donc unitaire) et $\{x_i\}$ et $\{y_i\}$ les deux ensembles de vecteurs appariés, identifiés à des quaternions purs. Le critère aux moindres carrés se réécrit alors (voir la section (7.2.2)) :

$$C(q) = \sum_i \|y_i - q * x_i * \overline{q}\|^2 = \sum_i \|y_i * q - q * x_i\|^2 . \|\overline{q}\|^2 = \sum_i \|y_i * q - q * x_i\|^2$$

puisque $\|q\| = 1$. En utilisant les quaternions matriciels, on a $y_i * q = Q_{y_i}.q$ et $-q * x_i = -P_{x_i}^{\mathrm{T}}.q = P_{x_i}.q$ car x_i est un vecteur. Ceci permet donc de simplifier le critère en

$$C(q) = q^{\mathrm{T}}.A.q \qquad \text{avec} \qquad A = \left(\sum_i (Q_{y_i} + P_{x_i})^{\mathrm{T}}.(Q_{y_i} + P_{x_i}) \right)$$

Notons que comme x_i et y_i sont des quaternions purs, on a : $P_{x_i}^{\mathrm{T}} = -P_{x_i}$ et $Q_{y_i}^{\mathrm{T}} = -Q_{y_i}$. La matrice A peut donc être calculée plus efficacement par $A = -\sum_i (Q_{y_i} + P_{x_i})^2$. Pour minimiser ce critère sous la contrainte $\|q\|^2 = 1$, on écrit le lagrangien

$$\Lambda(q) = C(q) - \lambda.(\|q\|^2 - 1) = q^{\mathrm{T}}.(A - \lambda.I_4).q + \lambda$$

celui-ci doit être stationnaire à l'optimum :

$$\frac{\partial \Lambda(q)}{\partial q} = 0 \qquad \Longleftrightarrow \qquad (A - \lambda I_4).q = 0$$

Avec la contrainte $\|q\| = 1$, les optimums sont donc les vecteurs propres unitaires de la matrice A, et le minimum absolu est, parmi les 4 vecteurs propres, celui qui est associé à la plus petite des valeurs propres (puisque C est une somme de carrées, les valeurs propres sont positives).

Théorème 8.1 (Méthode des quaternions pour les moindres carrés 3D)
Le quaternion rotation minimisant le critère aux moindres carrées

$$C(q) = \sum_i \|y_i - q * x_i * \overline{q}\|^2$$

est donné par le vecteur propre unitaire $\hat{q}$ associé à la plus petite valeur propre λ de la matrice :

$$A = -\left(\sum_i (Q_{y_i} + P_{x_i})^2 \right) \tag{8.3}$$

La valeur du critère au minimum est alors $\hat{C} = \lambda$.

8.1.1.3 Rotation : méthode SVD

Cette méthode de résolution est inspirée de (Umeyama, 1991). On suppose bien sûr que les points x_i et y_i est exprimés en repère barycentrique, mais on se place ici en dimension quelconque n, la dimension 3 n'amenant pas de simplification particulière. Observons tout d'abord que $\langle x \mid y \rangle = x^{\mathrm{T}}.y = \mathrm{Tr}(y.x^{\mathrm{T}})$. On peut donc simplifier notre critère en :

$$\begin{aligned} C(R) &= \textstyle\sum_i \|y_i\|^2 + \sum_i \|x_i\|^2 - 2.\sum_i \langle y_i \mid R.x_i \rangle \\ &= \textstyle\sum_i \|y_i\|^2 + \sum_i \|x_i\|^2 - 2.\mathrm{Tr}\left(\sum_i R.x_i.y_i^{\mathrm{T}}\right) \end{aligned}$$

et le minimum du critère C est obtenu pour la matrice de rotation $\hat{R}$ maximisant le critère $G(R) = \mathrm{Tr}(R.K^{\mathrm{T}})$, où K est la matrice de corrélation $K = \sum_i y_i.x_i^{\mathrm{T}}$. Pour résoudre ce nouveau critère, écrivons le lagrangien :

$$\Lambda = \mathrm{Tr}(R.K^{\mathrm{T}}) - \mathrm{Tr}\,(L.(R.R^{\mathrm{T}} - Id)) - 2.g.(\det(R) - 1)$$

où L est une matrice symétrique de multiplicateurs de Lagrange pour prendre en compte la contrainte $R.R^{\mathrm{T}} = Id$ et g un multiplicateur scalaire pour la contrainte $\det(R) = +1$. Un optimum est caractérisé par un lagrangien stationnaire : d'après l'appendice B (équations (B.2) et (B.10)), on a :

$$\frac{\partial(\mathrm{Tr}(R.K^{\mathrm{T}}))}{\partial R} = K \qquad \frac{\partial(\mathrm{Tr}(L.R.R^{\mathrm{T}}))}{\partial R} = 2.L.R \qquad \text{et} \qquad \frac{\partial \det(R)}{\partial R} = R^* = \det(R).R^{(-\mathrm{T})} = R$$

et comme R est une matrice de rotation, la dérivée du lagrangien devient :

$$\frac{\partial \Lambda}{\partial R} = 2.K - 2.R.L - 2.g.R = 0 \tag{8.4}$$

Il nous reste maintenant à déduire de cette équation et des contraintes les valeurs de R et des multiplicateurs L et g. En introduisant la matrice $L' = L + g.\,Id$, l'équation (8.4) se réduit à $R.L' = K$, et nous avons déjà éliminé un multiplicateur.

Soit maintenant $K = U.D.V^{\mathrm{T}}$ une décomposition en valeurs singulières de K. On rappelle que U et V sont des matrices orthogonales (qui peuvent être impropres) et D est la matrice diagonale des valeurs singulières (positives). Comme L est une matrice symétrique, on a :

$$L'^2 = (L'.R^{\mathrm{T}})(R.L') = K^{\mathrm{T}}.K = V.D^2.V^{\mathrm{T}}$$

Les matrices L'^2 et L' commutent évidemment, et peuvent donc être diagonalisées dans la même base : la matrice L' est donc $L' = V.D.S.V^{\mathrm{T}}$ où $S = \mathrm{DIAG}(s_1, \ldots, s_n)$ avec $s_i = \pm 1$. Comme la matrice S est égale à son inverse et commute avec D, on trouve en reportant dans $R.L' = K$:

$$R.V.D = U.S.D$$

La matrice de rotation $R = U.S.V^{\mathrm{T}}$ est toujours solution de cette équation, mais elle n'est pas forcément la seule. Elle l'est si les valeurs singulières sont non nulles (K est alors de rang maximal), et (Umeyama, 1991) montre que c'est encore le cas si l'on a une seule valeur singulière nulle ($\mathrm{rang}(K) = n - 1$). La dernière contrainte, celle du déterminant, se traduit alors par : $\det(S) = \det(U).\det(V)$. En reportant$R = U.S.V^{\mathrm{T}}$ dans la valeur du critère, on trouve que la valeur de celui-ci à l'optimum (qui dépend de S) est :

$$\hat{G} = \mathrm{Tr}(R.K^{\mathrm{T}}) = \mathrm{Tr}(U.S.D.U^{\mathrm{T}}) = \mathrm{Tr}(D.S) = \sum_{i=1}^{n} d_i.s_i$$

En supposant que les valeurs singulières soient triées par ordre décroissant, le maximum du critère est donc obtenu pour $s_1 = \ldots = s_{n-1} = +1$ et $s_n = \det(U).\det(V)$.

Théorème 8.2 (Méthode SVD pour les moindres carrés n-D)
Soit $U.D.V^T = K$ une décomposition singulière de la matrice K, dont les valeurs singulières (positives) sont triées par ordre décroissant. Alors le maximum du critère $G(R) = Tr(R.K^T)$ sur les rotations (propres) est atteint pour

$$\hat{R} = U.S.V^T \qquad avec \qquad S = DIAG(1, \ldots 1, \det(U).\det(V)) \tag{8.5}$$

avec la valeur $\hat{G} = Tr(D.S)$. *Le maximum est unique si le rang de* K *est* $n-1$ *ou* n.

Cette rotation est également la solution du critère aux moindres carrés $C = \|y_i - R.x_i\|^2$ *si* K *est la matrice de corrélation* $K = \sum_i y_i.x_i^T$, *et le minimum est atteint avec la valeur* $\hat{C} = \sum_i \|y_i\|^2 + \sum_i \|x_i\|^2 - 2.Tr(D.S)$.

Il est évident, puisque l'on minimise le même critère, que la solution par les quaternions et celle par la SVD sont identiques (si ces solutions sont uniques). Une étude récente (Lorusso et al., 1995) compare d'ailleurs ces deux algorithme d'un point de vue numérique, plus la décomposition polaire (Horn et al., 1988) et la méthode des quaternions duaux (Walker et Shao, 1991). Rappelons que toutes ces méthodes déterminent la solution du même problème : la transformation rigide aux moindres carrés entre points appariés. Cette étude conclue que la différence de précision est insignifiante (du niveau de la précision numérique de la machine), les temps de calculs sont comparables (15% maximum de différence, pouvant être occasionnés par un défaut d'optimisation dans l'implémentation) et que la seule différence (à peine) sensible est la sensibilité des algorithmes lorsque les points utilisés s'approchent d'une configuration dégénérée.

8.1.2 Similitude et transformation affine aux moindres carrés

Le même type de critère, les moindres carrés, peut servir à déterminer le recalage par d'autres transformations, comme les similitudes ou les transformations affines :

$$C(s,R,t) = \sum_i \|y_i - s.R.x_i - t\|^2 \qquad \text{ou} \qquad C(A,t) = \sum_i \|y_i - A.x_i - t\|^2$$

Nous avons déjà vu que la translation est déterminée par les barycentres et la partie linéaire optimale (rappelons que $\bar{x}$ et $\bar{y}$ sont les barycentres des deux ensembles de points) :

$$\hat{t} = \bar{y} - \hat{A}.\bar{x}$$

On peut donc considérer que les points sont en repère barycentrique et oublier la translation.

8.1.2.1 Similitude aux moindres carrés

Notons $\sigma_x^2 = \sum_i \|x_i\|$, $\sigma_y^2 = \sum_i \|y_i\|$ et bien sûr $K = \sum_i y_i.x_i^{\mathrm{T}}$. Le critère est alors :

$$C = \sum_i \|y_i - s.R.x_i - t\|^2 = s^2.\sigma_x^2 - 2.s.\mathrm{Tr}(R.K^{\mathrm{T}}) + \sigma_y^2$$

Le minimum sur la rotation R est donc obtenu pour la même rotation $\hat{R}$ que dans le cas rigide, et le minimum sur l'échelle s pour

$$\frac{\partial C}{\partial s} = 2.s.\sigma_x^2 - 2.\mathrm{Tr}(R.K^{\mathrm{T}}) = 0$$

soit pour

$$\hat{s} = \frac{\mathrm{Tr}(\hat{R}.K^{\mathrm{T}})}{\sigma_x^2} \tag{8.6}$$

Remarquons que, dans le cas 3D, on aurait pu obtenir aussi la solution par la méthode des quaternions : si le quaternion unitaire q représente la rotation R, le quaternion $p = \sqrt{s}.q$ représente

la similitude $(s, R) : p*x*\overline{p} = s.(q*x*\overline{q}) \equiv s.R.x$. On peut vérifier que la rotation $\hat{q}$ reste inchangée par rapport au cas rigide (ce que l'on a déjà observé avec la méthode SVD), et que l'on obtient

$$\hat{s} = \frac{\sigma_x^2 + \sigma_y^2 - \hat{C}_R}{2.\sigma_x^2}$$

où $\hat{C}_R$ est la valeur du critère au minimum *pour les moindres carrées rigides.*

8.1.2.2 Moindres carrés affines

Le critère est :

$$C(A) = \sum_i \|y_i - A.x_i\|^2 = \sigma_y^2 - 2.\sum_i y_i^{\mathrm{T}}.A.x_i + \sum_i x_i.A^{\mathrm{T}}.A.x_i$$

est les optimums sont caractérisés par une dérivée nulle. Grâce aux formules de l'appendice B, on a :

$$\frac{\partial(y^{\mathrm{T}}.A.x)}{\partial A} = y.x^{\mathrm{T}} \qquad \text{et} \qquad \frac{\partial(x^{\mathrm{T}}.A^{\mathrm{T}}.A.x)}{\partial A} = 2.A.x.x^{\mathrm{T}}$$

En notant $\Sigma_{xx} = \sum_i x_i.x_i^{\mathrm{T}}$, $\Sigma_{yy} = \sum_i y_i.y_i^{\mathrm{T}}$ les matrices de dispersion et $\Sigma_{yx} = K = \sum_i y_i.x_i^{\mathrm{T}}$ la matrice de corrélation habituelle, un optimum est donc caractérisé par

$$\frac{\partial C}{\partial A} = \sum_i 2.A.x_i.x_i^{\mathrm{T}} - 2.\sum_i y_i.x_i^{\mathrm{T}} = 0 \qquad \Longleftrightarrow \qquad A.\Sigma_{xx} = \Sigma_{yx}$$

Si Σ_{xx} est inversible, le minimum est unique et est obtenu pour

$$\hat{A} = \Sigma_{yx}.\Sigma_{xx}^{(-1)} \tag{8.7}$$

Dans le cas contraire, le minimum n'est pas unique, et l'un des minimums est obtenu avec la pseudo-inverse Σ_{xx}^{+} au lieu de de l'inverse $\Sigma_{xx}^{(-1)}$.

8.1.2.3 Discussion sur l'estimation des similitudes et transformation affines

Contrairement au cas des transformations rigides, la métrique sur les points n'est pas invariante par les similitudes ou les transformations affines. Un certain nombre de propriétés d'invariance ou de stabilité de notre estimation de la transformation ne sont donc plus valides (voir par exemple la section (3.4.1)). En particulier, la transformation $\hat{\mathrm{f}}_{xy}$ estimé entre les x_i et les y_i n'est plus l'inverse de la transformation $\hat{\mathrm{f}}_{yx}$ estimée entre les y_i et les x_i.

En effet, si l'on prend le cas affine, on a en général $\Sigma_{yx}.\Sigma_{xx}^{(-1)} \neq \Sigma_{yy}.\Sigma_{xy}^{(-1)}$. Ceci est dû au fait que l'on estime pas $\hat{\mathrm{f}}_{xy}$ et $\hat{\mathrm{f}}_{yx}$ dans le même repère de l'espace euclidien. Plus précisément, on estime $\hat{\mathrm{f}}_{xy}$ en supposant que l'espace des y est muni de la métrique canonique, alors que l'on estime $\hat{\mathrm{f}}_{yx}$ en supposant que la métrique canonique est sur l'espace des x.

La prise en compte de cette modification de la métrique (et du fait que les géodésiques sont cependant globalement conservées) pourrait conduire à l'extension de notre théorie de l'incertitude (première partie) aux groupes de type affine et non plus seulement rigide.

8.1.3 Estimation de l'incertitude sur la transformation

Jusqu'à présent, nous n'avons pas supposé de modèle de bruit sur nos points et nous avons simplement considéré un estimateur de la transformation rigide. L'incertitude sur cette estimation dépend cependant de l'incertitude sur les données : on notera $\Sigma_{\boldsymbol{x_i x_i}}$ et $\Sigma_{\boldsymbol{y_i y_i}}$ les matrices de covariance représentant l'incertitude sur nos points. On peut donc voir nos données comme des ensembles de points aléatoires appariés $\boldsymbol{x_i} \sim (x_i, \Sigma_{\boldsymbol{x_i x_i}})$ et $\boldsymbol{y_i} \sim (y_i, \Sigma_{\boldsymbol{y_i y_i}})$.

Pour quantifier l'incertitude sur une transformation, et en l'occurrence ici un mouvement rigide, la matrice de covariance définie dans la première partie de ce manuscrit semble tout à fait adaptée. On utilise donc la carte principale $\vec{\mathrm{f}} = (r, t)$ pour les mouvements.

8.1.3.1 Transformation rigide estimée aux moindres carrés

On rangeant toutes nos données (les coordonnées des points) dans un grand vecteur χ dont la covariance est diagonale par blocs (en supposant que tous les points soient indépendants) : $\boldsymbol{\chi} \sim (\chi, \Sigma_{\boldsymbol{\chi\chi}})$ où

$$\begin{aligned}\chi &= (x_1^{\mathrm{T}}, \dots x_N^{\mathrm{T}}, y_1^{\mathrm{T}}, \dots y_N^{\mathrm{T}})^{\mathrm{T}} \\ \Sigma_{\boldsymbol{\chi\chi}} &= \mathrm{DIAG}(\Sigma_{\boldsymbol{x_1 x_1}}, \dots \Sigma_{\boldsymbol{x_N x_N}}, \Sigma_{\boldsymbol{y_1 y_1}}, \dots \Sigma_{\boldsymbol{y_N y_N}})\end{aligned}$$

On peut se ramener au formalisme développé à la section (2.14) : le mouvement estimé est

$$\hat{\vec{\mathrm{f}}} = \arg\min_{\vec{\mathrm{f}}} C(\vec{\mathrm{f}}, \chi) \qquad \text{avec} \qquad C(\vec{\mathrm{f}}, \chi) = \frac{1}{2}.\sum_i z_i^{\mathrm{T}}.z_i \qquad \text{où} \qquad z_i = y_i - \vec{\mathrm{f}} \star x_i$$

Nous avons ajouté ici un facteur 1/2 qui ne change rien à l'optimisation ni au résultat mais qui simplifie les calculs. La fonction implicite caractérisant un optimum est donc

$$\Phi(\chi, \vec{\mathrm{f}}) = \frac{\partial C}{\partial \vec{\mathrm{f}}}^{\mathrm{T}} = \sum_i \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}}.z_i = 0$$

Pour calculer les dérivées secondes, on néglige les termes en $z.\ddot{z}$ par rapport aux termes en $\dot{z}^2$ (voir par exemple (Gill et al., 1981, section 4.7)) :

$$\begin{aligned}H = \frac{\partial \Phi}{\partial \vec{\mathrm{f}}} &\simeq \textstyle\sum_i \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}} . \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right) \\ \frac{\partial \Phi}{\partial \chi} &\simeq \textstyle\sum_i \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}} . \left(\frac{\partial z_i}{\partial \chi}\right)\end{aligned}$$

et l'équation (2.14) nous donne la matrice de covariance sur $\vec{\mathrm{f}}$:

$$\begin{aligned}\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} &= H^{(-1)}.\left(\frac{\partial \Phi}{\partial \chi}\right).\Sigma_{\boldsymbol{\chi\chi}}.\left(\frac{\partial \Phi}{\partial \chi}\right)^{\mathrm{T}}.H^{(-1)} \\ &= H^{(-1)}.\left(\sum_i \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}} . \left(\frac{\partial z_i}{\partial \chi}\right).\Sigma_{\boldsymbol{\chi\chi}}.\left(\frac{\partial z_i}{\partial \chi}\right)^{\mathrm{T}} . \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)\right).H^{(-1)}\end{aligned}$$

Le résultat est donc :

$$\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = H^{(-1)}.\left(\sum_i \left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}}.\Sigma_{\boldsymbol{z_i z_i}}.\left(\frac{\partial z_i}{\partial \vec{\mathrm{f}}}\right)\right).H^{(-1)} \tag{8.8}$$

Dans le cas d'un bruit **isotrope**, **identique et indépendant** sur tous les points d'une image ($\Sigma_{\boldsymbol{x_i x_i}} = \sigma_x^2.\, Id$ et $\Sigma_{\boldsymbol{y_i y_i}} = \sigma_y^2.\, Id$), le vecteur d'erreur $\boldsymbol{z_i} = \boldsymbol{y_i} - \vec{\mathrm{f}} \star \boldsymbol{x_i}$ est encore isotrope :

$\Sigma_{\boldsymbol{z_i z_i}} = (\sigma_x^2 + \sigma_y^2).Id = \sigma_z^2.Id$. Notons que c'est à cette condition que les moindres carrés fournissent une estimation optimale. Le terme central de la dernière équation (la somme) se simplifie donc pour donner $\sigma_z^2.H$, ce qui se simplifie avec $H^{(-1)}$:

$$\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = \sigma_z^2.H^{(-1)} \qquad \text{avec} \qquad H = \sum_i \left(\frac{\partial(\vec{\mathrm{f}} \star x_i)}{\partial \vec{\mathrm{f}}}\right)^{\mathrm{T}} . \left(\frac{\partial(\vec{\mathrm{f}} \star x_i)}{\partial \vec{\mathrm{f}}}\right) \tag{8.9}$$

8.1.3.2 Filtrage de Kalman

Si l'on suppose maintenant que l'on a une information d'incertitude sur chaque point grâce à sa matrice de covariance, on peut toujours utiliser les moindres carrés, mais cet estimateur n'est plus optimal. Si l'un des appariement est par exemple dix fois plus précis que les autres, ou plus précis dans une direction qu'une autre, on ne peut pas utiliser quantitativement cette information.

L'idée pour remédier à cela est de pondérer chaque terme dans les moindres carrés en fonction de son information, et donc d'utiliser la matrice de covariance comme métrique : on ne minimise plus alors la distance classique mais la distance de Mahalanobis. Le filtrage de Kalman est un algorithme incrémental pour réaliser ce type de minimisation (nous renvoyons le lecteur à l'appendice C pour les explications et les équations du filtre de Kalman) et qui actualise à chaque étape la matrice de covariance sur l'estimation (qui sera pour nous un mouvement).

Dérivons tout d'abord l'équation de mesure. Comme nous n'utilisons que des paires de points appariés, l'équation de mesure est identique pour tout i on peut oublier pour l'instant les indices. Si les points x et y étaient mesurés exactement, on aurait par hypothèse $y - \vec{\mathrm{f}} \star x = 0$. La mesure de ces points est en fait corrompue par un bruit que l'on suppose additif et indépendant : on ne mesure que $\boldsymbol{x} = x + \boldsymbol{\delta x}$ et $\boldsymbol{y} = y + \boldsymbol{\delta y}$. On suppose que ces bruits sont centrés et de covariance connue : $\boldsymbol{\delta x} \sim (0, \Sigma_{\boldsymbol{xx}})$ et $\boldsymbol{\delta y} \sim (0, \Sigma_{\boldsymbol{yy}})$. En reportant dans l'équation $y - \vec{\mathrm{f}} \star x = 0$, on obtient :

$$\boldsymbol{y} - \boldsymbol{\delta y} - R.\boldsymbol{x} - t + R.\boldsymbol{\delta x} = 0$$

et on isole l'erreur pour obtenir l'équation de mesure (ou vecteur d'erreur) :

$$\boldsymbol{z} = \boldsymbol{y} - \vec{\mathrm{f}} \star \boldsymbol{x} = \boldsymbol{\delta y} - R.\boldsymbol{\delta x} \sim (0, \Sigma_{\boldsymbol{zz}}) \qquad \text{avec} \qquad \Sigma_{\boldsymbol{zz}} = \Sigma_{\boldsymbol{yy}} + R.\Sigma_{\boldsymbol{xx}}.R^{\mathrm{T}}$$

On cherche donc à minimiser le critère :

$$C(\vec{\mathrm{f}}) = \sum_i \mu^2(\boldsymbol{z_i}, 0) = \sum_i \hat{z}_i^{\mathrm{T}}.\Sigma^{(-1)}_{\boldsymbol{z_i z_i}}.\hat{z}_i$$

où $\hat{z}_i$ est cette fois-ci la valeur calculée à partir des mesures.

Pour satisfaire au formalisme du filtre de Kalman (étendu), il faut en fait rajouter dans ce critère la distance de Mahalanobis par rapport à une estimation initiale de l'état, et calculer la matrice M qui exprime localement l'influence du vecteur d'état sur le vecteur d'erreur (c'est une approximation au premier ordre). Dans notre cas, on a :

$$M = \frac{\partial z}{\partial \vec{\mathrm{f}}} = -\frac{\partial \vec{\mathrm{f}} \star x}{\partial \vec{\mathrm{f}}}$$

Ce jacobien fait partie des opération atomiques que nous avons développé sur les points au chapitre 7 (section 7.6.1).

On peut donc résumer l'algorithme d'estimation du mouvement par filtrage de Kalman comme ceci :

- Initialiser l'état $\mathbf{f_0}$ avec l'identité ou l'estimation aux moindres carrés et une matrice de covariance suffisamment grande (entre 100 et 500 fois la covariance estimée pour les moindres carrés par exemple) pour minimiser l'influence de cet état initial sur la transformation finale et surtout sur l'estimation de son incertitude.

  On pourra bien sûr utiliser une estimée initiale si on en possède une.
- Pour chaque couple de points appariés $(\boldsymbol{x_i}, \boldsymbol{y_i})$:
  - Calculer $\boldsymbol{z_i} = \boldsymbol{y}_i - \vec{\mathrm{f}}_{(i-1)} \star \boldsymbol{x}_i$, et la matrice $M_i = \dfrac{\partial z_i}{\partial \vec{\mathrm{f}}} = -\dfrac{\partial(\vec{\mathrm{f}} \star x_i)}{\partial \vec{\mathrm{f}}}$ estimée en $\vec{\mathrm{f}}_{(i-1)}$,
  - Mettre à jour l'état de $\mathbf{f_{(i-1)}} \sim (\vec{\mathrm{f}}_{(i-1)}, \Lambda_{(i-1)})$ à $\mathbf{f_i} \sim (\vec{\mathrm{f}}_i, \Lambda_i)$ en utilisant les équations du filtre (C.1).
  - Normaliser la transformation $\mathbf{f_i}$: comme le filtre est additif, il peut nous faire sortir du domaine de la carte principale.

Nous avons donc obtenu deux méthodes pour estimer la transformation rigide et son incertitude entre deux ensembles de points appariés : on utilisera les moindres carrés classiques si on ne connaît pas le modèle de bruit sur les points (cela revient en fait à supposer un bruit isotrope identique et indépendant sur chacun des deux ensembles) et le filtre de Kalman si on a la covariance sur les points, ou si l'on veut supposer un modèle de bruit non isotrope. Nous discuterons des avantages de l'un ou l'autre des algorithmes à la section (8.2.3).

8.2 Recalage à partir d'appariements de primitives

Puisque l'on a défini dans la première partie de ce manuscrit la distance entre deux primitives déterministes et la distance de Mahalanobis entre deux primitives probabilistes, il est possible de généraliser les deux méthodes précédentes au cas d'appariement de primitives homogènes ayant une distance invariante. Toutefois, la complexité de la recherche d'une solution explicite pour le recalage des points n'engage pas à faire de même pour des primitives plus complexes, d'autant que ces solutions explicites sont intrinsèquement spécifiques à chaque type de primitive et de transformation. On se contentera donc d'algorithmes génériques itératifs.

8.2.1 Moindres carrés simples

Soient $\{\mathrm{x_i}\}$ et $\{\mathrm{y_i}\}$ deux ensembles de primitives appariées. L'erreur de superposition pour une transformation f est

$$C(\mathrm{f}) = \frac{1}{2}.\sum_i \mathrm{dist}(\mathrm{y_i}, \mathrm{f} \star \mathrm{x_i})^2 = \frac{1}{2}.\sum_i \mathrm{dist}\left((\mathrm{f}_{\mathrm{y_i}}^{(-1)} \circ \mathrm{f}) \star \mathrm{x_i}\ ,\ \mathrm{o}\right)^2 = \frac{1}{2}.\sum_i \|\vec{\mathrm{z}}_i\|^2$$

où $\vec{\mathrm{z}}_i$ est le vecteur d'erreur dans la carte principale :

$$\vec{\mathrm{z}}_i = (\vec{\mathrm{f}}_{\vec{\mathrm{y}}_i}^{(-1)} \circ \vec{\mathrm{f}}) \star \vec{\mathrm{x}}_i$$

L'idée de base est donc de faire une descente de gradient pour obtenir le minimum. Il faut donc dériver le vecteur d'erreur $\vec{\mathrm{z}} = (\vec{\mathrm{f}}_{\vec{\mathrm{y}}}^{(-1)} \circ \vec{\mathrm{f}}) \star \vec{\mathrm{x}}$ par rapport à la transformation $\vec{\mathrm{f}}$. Comme on retrouvera également ce vecteur d'erreur pour la minimisation de la distance de Mahalanobis, qui nécessitera de plus le calcul de la covariance $\Sigma_{\mathbf{zz}}$, il est utile d'ajouter le « vecteur d'erreur » à la liste des

opérations de base sur les primitives géométriques. Pour cela, il suffit de détailler les jacobiens de cette opération et de les exprimer en fonction des opérations atomiques sur les transformations et les primitives considérées.

8.2.1.1 Le vecteur d'erreur pour le recalage $\vec{z} = (\vec{f}_{\vec{y}}^{(-1)} \circ \vec{f}) \star \vec{x}$

Observons tout d'abord que l'on peut écrire ce vecteur sous la forme $\vec{z} = \vec{f}_{\vec{y}}^{(-1)} \star (\vec{f} \star \vec{x})$, forme qui est plus adaptée à la dérivation. Calculons en premier :

$$\overrightarrow{x_f} = \vec{f} \star \vec{x} \qquad \text{avec les jacobiens} \qquad J_1 = \frac{\partial(\vec{f} \star \vec{x})}{\partial \vec{x}} \qquad \text{et} \qquad J_2 = \frac{\partial(\vec{f} \star \vec{x})}{\partial \vec{f}}$$

puis $\vec{f}_{\vec{y}}^{(-1)}$ et $J_3 = \dfrac{\partial \vec{f}_{\vec{y}}^{(-1)}}{\partial \vec{f}_{\vec{y}}} . \dfrac{\partial \vec{f}_{\vec{y}}}{\partial \vec{y}}$ et enfin

$$\vec{z} = \vec{f}_{\vec{y}}^{(-1)} \star \overrightarrow{x_f} \qquad \text{avec les jacobiens} \qquad J_4 = \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \overrightarrow{x_f})}{\partial \overrightarrow{x_f}} \qquad \text{et} \qquad J_5 = \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \overrightarrow{x_f})}{\partial \vec{f}_{\vec{y}}^{(-1)}}$$

A partir de ces jacobiens, on peut aisément calculer

$$J_{\vec{x}} = \frac{\partial \vec{z}}{\partial \vec{x}} = J_4 . J_1 \qquad J_{\vec{y}} = \frac{\partial \vec{z}}{\partial \vec{y}} = J_5 . J_3 \qquad J_{\vec{f}} = \frac{\partial \vec{z}}{\partial \vec{f}} = J_4 . J_2$$

Pour les moindres carrés simples, on se contentera de retourner $\vec{z}$ et $J_{\vec{f}}$, et on implémente l'opération suivante sur les primitives probabilistes pour la minimisation de la distance de Mahalanobis par filtrage de Kalman ou descente de gradient :

- **Vecteur d'erreur pour le recalage**

$$\left(\mathbf{x} \sim (\vec{x}\,,\, \Sigma_{\mathbf{xx}}) \quad , \quad \mathbf{y} \sim (\vec{y}\,,\, \Sigma_{\mathbf{yy}}) \quad , \quad \vec{f}\right) \qquad \longmapsto \qquad \left(\boldsymbol{z} \sim (\vec{z}\,,\, \Sigma_{\mathbf{zz}}) \quad , \quad J_{\vec{f}}\right)$$

  où la covariance est $\Sigma_{\mathbf{zz}} = J_{\vec{x}} . \Sigma_{\mathbf{xx}} . J_{\vec{x}}^{\mathrm{T}} + J_{\vec{y}} . \Sigma_{\mathbf{yy}} . J_{\vec{y}}^{\mathrm{T}}$

8.2.1.2 Descente de gradient

Supposons que l'on ait une estimation courante $\vec{f}_t$ au temps t de la transformation recherchée. On calcule alors pour chaque paire de primitives appariées le vecteur d'erreur $\vec{z}_i$ et son jacobien $J_{\vec{f}_{(t,i)}}$ par rapport à la transformation courante. On a alors

$$\frac{\partial C}{\partial \vec{f}_t} = \sum_i \vec{z}_i^{\mathrm{T}} . J_{\vec{f}_{(t,i)}} \qquad \text{soit} \qquad \Phi_t = \left(\frac{\partial C}{\partial \vec{f}_t}\right)^{\mathrm{T}} = \sum_i J_{\vec{f}_{(t,i)}}^{\mathrm{T}} . \vec{z}_i$$

Pour calculer la matrice hessienne, on néglige comme d'habitude les termes en $\ddot{f} . f$ par rapport aux termes en $\dot{f}^2$:

$$H_t = \frac{\partial \Phi_t}{\partial \vec{f}_t} = \sum_i J_{\vec{f}_{(t,i)}}^{\mathrm{T}} . J_{\vec{f}_{(t,i)}}$$

On avance donc d'un temps unité suivant la géodésique partant de $\vec{f}_t$ avec le vecteur tangent $\overrightarrow{\delta f}_t \in T_{\vec{f}}\mathcal{G}$ suivant :

$$\overrightarrow{\delta f}_t = -H_t^{(-1)} . \Phi_t = -\left(\sum_i J_{\vec{f}_{(t,i)}}^{\mathrm{T}} . J_{\vec{f}_{(t,i)}}\right)^{(-1)} . \left(\sum_i J_{\vec{f}_{(t,i)}}^{\mathrm{T}} . \vec{z}_i\right) \tag{8.10}$$

Comme ce vecteur $\overrightarrow{\delta \mathrm{f}}_t$ est exprimé dans la carte exponentielle en $\vec{\mathrm{f}}_t$, l'estimation au temps $t+1$ est donnée par l'équation (3.24) :

$$\vec{\mathrm{f}}_{t+1} = \exp_{\vec{\mathrm{f}}_t}(\overrightarrow{\delta \mathrm{f}}_t) = \vec{\mathrm{f}}_t \circ \left(J_L(\vec{\mathrm{f}}_t)^{(-1)}.\overrightarrow{\delta \mathrm{f}}_t \right) \tag{8.11}$$

Il reste à spécifier un point de départ (une estimation initiale de la transformation) et un critère d'arrêt pour notre descente de gradient : en l'absence d'information supplémentaire, on pourra toujours partir de l'identité, et s'arrêter lorsque la norme de la transformation d'ajustement est inférieure à un seuil fixé ou que le nombre d'itération devient trop grand. En pratique, nous avons observé que cet algorithme converge toujours en environ 10 itérations pour un seuil $\varepsilon = 10^{-10}$ sur $\|\overrightarrow{\delta \mathrm{f}}\|_{\vec{\mathrm{f}}} = \|J_L(\vec{\mathrm{f}})^{(-1)}.\overrightarrow{\delta \mathrm{f}}\|$, soit quasiment à la précision numérique de la machine. quelque-soit le nombre d'appariements.

8.2.1.3 Estimation de l'incertitude au minimum

Puisque nous avons minimisé réellement la norme de vecteurs d'erreur (grâce à la carte exponentielle), les développements effectués sur les points sont toujours valides. En supposant une covariance $\Sigma_{\mathbf{z_i z_i}}$ sur le vecteur d'erreur $\vec{z}_i$, on a encore d'après (2.14) :

$$\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = H^{(-1)}.\left(\sum_i \frac{\partial \Phi}{\partial \vec{z}_i} \right).\Sigma_{\boldsymbol{\chi}\boldsymbol{\chi}}.\left(\frac{\partial \Phi}{\partial \vec{z}_i} \right)^{\mathrm{T}}.H^{(-1)} = H^{(-1)}.\left(\sum_i J_{\vec{\mathrm{f}}_i}^{\mathrm{T}}.\Sigma_{\mathbf{z_i z_i}}.J_{\vec{\mathrm{f}}_i} \right).H^{(-1)}$$

où les jacobiens et la matrice hessienne sont bien sûr estimés au minimum. Comme dans le cas des points, cette minimisation aux moindres carrés n'est optimale que si $\Sigma_{\mathbf{z_i z_i}} = \sigma_z^2.\,Id$, et l'incertitude sur la transformation se simplifie alors en

$$\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = \sigma_z^2.H^{(-1)} \qquad \text{avec} \qquad H = \sum_i J_{\vec{\mathrm{f}}_i}^{\mathrm{T}}.J_{\vec{\mathrm{f}}_i} \tag{8.12}$$

Notons que cette hypothèse est beaucoup plus forte que la simple isotropie et qu'elle demande de bien choisir la métrique invariante : par exemple sur les repères, la variance σ_θ^2 sur le vecteur rotation doit être du même ordre de grandeur que la variance σ_x^2 sur la position. Le choix des unités est donc des plus importants.

8.2.2 Minimisation de la distance de Mahalanobis

Dérivons tout d'abord l'équation de mesure. Comme nous n'utilisons que des appariements de primitives d'un seul type, l'équation de mesure est identique pour tout i on peut oublier pour l'instant les indices. Si les primitives x et y étaient exactes, on aurait $\mathrm{y} = \mathrm{f} \star \mathrm{x}$, soit l'équation de mesure (vectorielle) :

$$(\vec{\mathrm{f}}_{\vec{\mathrm{y}}}^{(-1)} \circ \vec{\mathrm{f}}) \star \vec{\mathrm{x}} = 0$$

En fait, on n'a accès qu'aux valeurs bruitées $\mathbf{x} \sim (\vec{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ et $\mathbf{y} \sim (\vec{y}, \Sigma_{\mathbf{yy}})$ que l'on supposera centrées et de covariance connue. En remplaçant les valeurs exactes par leur mesure, on obtient le vecteur d'erreur probabiliste :

$$\vec{\mathbf{z}}_{\boldsymbol{i}} = (\vec{\mathbf{f}}_{\vec{\mathbf{y}}}^{\,(-1)} \circ \vec{\mathrm{f}}) \star \vec{\mathbf{x}} \tag{8.13}$$

On cherche donc à minimiser le critère :

$$C(\mathrm{f}) = \sum_i \mu^2(\mathbf{z}_{\boldsymbol{i}}, 0) = \sum_i \vec{z}_i^{\mathrm{T}}.\Sigma_{\mathbf{z_i z_i}}^{(-1)}.\vec{z}_i$$

8.2.2.1 Filtrage de Kalman

Pour le filtrage de Kalman, il faut cependant rajouter à ce critère la distance de Mahalanobis par rapport à une estimation initiale de l'état. Il n'y a alors guère de différence avec l'algorithme pour l'estimation à partir de points :

- Initialiser l'état $\mathbf{f_0}$ avec l'identité ou l'estimation aux moindre carrés et une matrice de covariance suffisamment grande (50 fois la covariance estimée pour les moindres carrés par exemple) pour minimiser l'influence de cet état initial sur la transformation finale et surtout sur l'estimation de son incertitude.

  On pourra bien sûr utiliser une estimée initiale si on en possède une.
- Pour chaque couple de primitives appariés $(\mathbf{x_i}, \mathbf{y_i})$:
  - Calculer $\mathbf{z_i}$ et la matrice $M = J_{\vec{\mathrm{f}}}$ estimée en $\vec{\mathrm{f}}_{(i-1)}$,
  - Mettre à jour l'état de $\mathbf{f_{(i-1)}} \sim (\vec{\mathrm{f}}_{(i-1)}, \Lambda_{(i-1)})$ à $\mathbf{f_i} \sim (\vec{\mathrm{f}}_i, \Lambda_i)$ en utilisant les équations du filtre (C.1),
  - Normaliser la transformation $\mathbf{f_i}$.

La dernière étape, la normalisation, pourrait certainement être supprimée en reprenant les équations du filtre et en exprimant que $\overrightarrow{\delta \mathrm{f}}$ n'est pas une erreur additive mais un vecteur de $T_{\vec{\mathrm{f}}}\mathcal{G}$ (comme ci-dessus ou ci-dessous). Par contre, il faudrait également voir ce que cela implique sur la mise à jour de la matrice de covariance. Comme nous avons une implémentation très optimisée du filtre de Kalman classique, il est presque sûr que les temps de calcul en seraient augmentés.

8.2.2.2 Descente de gradient

La méthodologie est celle des moindres carrés simples, mais sur le critère des distances de Mahalanobis :

$$C(\mathrm{f}) = \frac{1}{2}.\sum_i \mu^2(\mathbf{z_i}, 0) = \frac{1}{2}.\sum_i \vec{\mathrm{z}}_i^{\mathrm{T}}.\Sigma^{(-1)}_{\mathbf{z_i}\mathbf{z_i}}.\vec{\mathrm{z}}_i$$

La dérivée et la matrice hessienne de ce critère sont donc :

$$\Phi_t = \left(\frac{\partial C}{\partial \vec{\mathrm{f}}_t}\right)^{\mathrm{T}} = \sum_i J^{\mathrm{T}}_{\vec{\mathrm{f}}_{(t,i)}}.\Sigma^{(-1)}_{\mathbf{z_i}\mathbf{z_i}}.\vec{\mathrm{z}}_i \qquad \text{et} \qquad H_t = \frac{\partial \Phi_t}{\partial \vec{\mathrm{f}}_t} = \sum_i J^{\mathrm{T}}_{\vec{\mathrm{f}}_{(t,i)}}.\Sigma^{(-1)}_{\mathbf{z_i}\mathbf{z_i}}.J_{\vec{\mathrm{f}}_{(t,i)}}$$

Le vecteur d'ajustement est alors $\overrightarrow{\delta \mathrm{f}}_t = -H_t^{(-1)}.\Phi_t$ et l'équation d'évolution reste :

$$\vec{\mathrm{f}}_{t+1} = \exp_{\vec{\mathrm{f}}_t}(\overrightarrow{\delta \mathrm{f}}_t) = \vec{\mathrm{f}}_t \circ \left(J_L(\vec{\mathrm{f}}_t)^{(-1)}.\overrightarrow{\delta \mathrm{f}}_t\right) \tag{8.14}$$

Comme pour les moindres carrés simples, il faut un point de départ et un critère d'arrêt. En pratique, nous avons observé une convergence à 10^{-10} en environ 15 itérations si l'on part de l'identité, et en 5 à 10 itération si l'on démarre de l'estimation aux moindres carrés simples.

Il reste à déterminer l'incertitude sur la transformation à l'optimum : cela se fait encore une fois exactement comme pour les moindres carrés simples, mais sans aucune hypothèse sur la covariances des vecteurs d'erreur, et on obtient :

$$\Sigma_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = H^{(-1)} \qquad \text{avec} \qquad H = \sum_i J^{\mathrm{T}}_{\vec{\mathrm{f}}}.\Sigma^{(-1)}_{\mathbf{z_i}\mathbf{z_i}}.J_{\vec{\mathrm{f}}} \tag{8.15}$$

8.2.3 Comparaison des algorithmes

Nous noterons en abrégé LSQ, KAL et MAHA les trois méthodes de recalage développées ci-dessus. Les résultats de ces algorithmes dépendent de nombreux paramètres et il importe de ne sélectionner que les plus importants. Nous avons choisi de regarder la précision relative de ces algorithmes et les temps de calculs en fonction du nombre de primitives, en gardant un modèle de bruit fixé. La précision absolue de chaque méthode est théoriquement prédite par la matrice de covariance de chaque estimation, et nous en vérifierons la validité au chapitre 9. Nous considérons ici que le modèle de bruit sur les données est connu.

Pour réaliser ces statistiques, nous choisissons pour chaque estimation un nombre N de primitives réparties aléatoirement dans une « image » 512x256x128, une transformation rigide aléatoire (avec une translation limitée à 512) que nous appliquons à ces primitives pour obtenir le second ensemble de primitives, et nous bruitons ensuite toutes les primitives de manière indépendante avec un bruit homogène de covariance fixée à $\Sigma = \mathrm{DIAG}(0.0024, 0.0030, 0.038; 0.22, 0.31, 0.084)$ pour les primitives de type repère et $\Sigma_{\mathbf{ee}} = \mathrm{DIAG}(0.2, 0.3, 0.09)$ pour les points. Cette matrice de covariance n'est pas innocente, c'est le modèle du bruit que nous obtiendrons au prochain chapitre (section 9.3.3) en analysant les point extrémaux dans certaines images IRM.

8.2.3.1 Précision relative des méthodes

Nous nous attachons ici à quantifier la précision relative des trois méthodes entre elles. Les expériences ci-dessous ont été réalisées avec des repères semi-orientés, simulant ainsi des points extrémaux pour être le plus proche possible des conditions d'application de nos algorithmes en imagerie médicale.

Sur le même ensemble d'appariements synthétique, on calcule la transformation par LSQ, KAL et MAHA et, pour chacune de ces trois estimations, on calcule la transformation résiduelle par rapport à la transformation exacte. Pour simplifier, on ne conserve que l'angle θ de la rotation résiduelle et la norme d de la translation. Comme ces valeurs d'erreur changent beaucoup avec l'ensemble d'appariement considéré, et que de plus on ne veut étudier que la précision *relative* des trois méthodes, nous avons choisi de prendre l'erreur de MAHA comme référence (la plupart du temps c'est la plus précise) : on calcule donc les rapports $\varepsilon_{\theta_{\mathrm{LSQ}}} = \theta_{\mathrm{LSQ}}/\theta_{\mathrm{MAHA}}$ et $\varepsilon_{d_{\mathrm{LSQ}}} = d_{\mathrm{LSQ}}/d_{\mathrm{MAHA}}$ et de manière similaire pour KAL. Ces rapport indiquent donc si la méthode LSQ (resp. KAL) est plus précise que MAHA ($\varepsilon_{\mathrm{LSQ}} < 1$) ou moins précise ($\varepsilon_{\mathrm{LSQ}} > 1$).

Pour avoir des statistiques valables, nous avons effectué 150 recalages pour un nombre de primitive fixé, et nous présentons dans la figure (8.1) la moyenne et l'écart-type[1] des précisions relatives de LSQ et KAL par rapport à MAHA.

Dans ces expériences, le filtrage de Kalman (KAL) est initialisé avec le résultat de LSQ, en multipliant par 50 la covariance sur l'estimation pour ne pas trop biaiser l'estimation finale de l'incertitude. La précision relative est alors très similaire à celle de MAHA. La solution est d'ailleurs très proche (excepté une légère sous-évaluation de l'incertitude sur le recalage). Notons par contre que KAL diverge souvent si l'état initial est trop éloigné de la solution ou si l'incertitude sur cet état initial est trop élevée.

Par contre, LSQ est en moyenne 1.2 fois moins précis sur la rotation et 1.4 fois mois précis sur la translation. La solution donné par LSQ est par ailleurs quasiment identique à celle fournie par la solution explicite aux moindres carrés sur les points par la méthode des quaternions (on note QUAT

1. Comme la précision relative est multiplicative (2 fois plus précis correspond à $\varepsilon = 0.5$), nous avons calculé la moyenne et l'écart-type de $\log(\varepsilon)$, mais nous présentons le résultat sous forme de rapport pour une compréhension plus aisée.

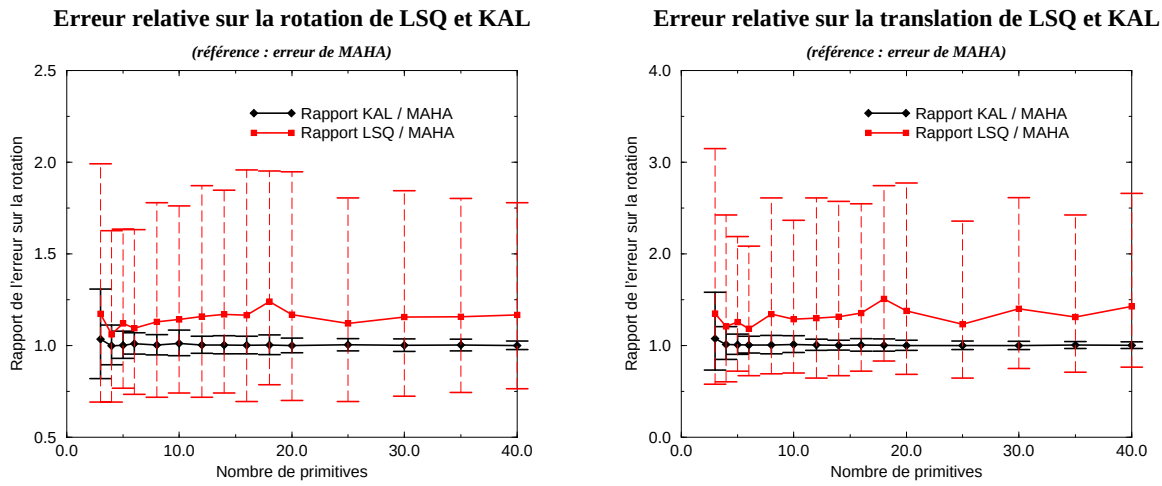


FIG. 8.1 – *Précision relative sur la rotation et la translation des méthodes LSQ et KAL par rapport à MAHA, en fonction du nombre de primitive.*

cet algorithme). Ceci est dû au fait que l'on utilise la métrique canonique sur les mouvements (i.e. la rotation en radians et la translation en voxels), métrique qui donne une influence beaucoup plus importante aux points qu'aux trièdres.

Pour finir, notons que le gain de précision obtenu en utilisant l'information d'incertitude sur les repères dépend pour une large part de cette incertitude initiale. Avec un modèle de bruit où le trièdre a un écart-type de 1 degré (soit environ 0.017 rad) et le point un écart-type de 1 voxel, on peut obtenir une estimation 2 fois plus précise avec MAHA ou KAL (initialisé correctement) qu'avec LSQ ou QUAT.

8.2.3.2 Temps de calculs

Les calculs sont effectués sur une DEC alpha 3000 à 166 Mhz, et les temps donnés sont des moyennes sur 100 à 500 estimations. Pour obtenir un temps de calcul fiable, nous avons utilisé le profiler GNU *gprof* qui permet d'obtenir le temps passé dans une procédure en incluant le temps passé dans tous les descendants (les fonctions appelées). Notre référence temporelle est le temps de calcul des moindres carrés classiques (méthode QUAT) : environ 20 ms pour 100 appariements de points.

La première expérience concerne l'influence du nombre de points sur le temps de calcul du recalage. Les temps donnés ci-dessous sont des moyennes sur les quatre types de primitives confondues (repères, repères semi et non-orienté, points). Les algorithmes KAL et MAHA sont initialisés avec le résultat de LSQ : nous donnons le décompte du temps passé dans l'algorithme proprement dit et le temps total avec l'initialisation.

Nombre d'appariements

	10	50	100	500
LSQ	87 ms	355 ms	640 ms	2970 ms
KAL	19 ms / 106 ms	102 ms / 457 ms	195 ms / 835 ms	957 ms / 3927 ms
MAHA	105 ms / 192 ms	397 ms / 752 ms	740 ms / 1380 ms	3510 ms / 6480 ms

La relation entre le nombre d'appariements et le temps est approximativement linéaire, et on en conclu que, pour recaler 100 primitives, il faut compter environ 650 ms par primitive pour LSQ, 650+200=850 ms pour le filtrage de Kalman et 650+750=1400 ms avec MAHA. En comparaison des 20 ms de la solution explicite des moindres carrés sur les points, il est clair qu'on y perd, bien que les temps de calculs restent raisonnables pour des applications qui ne sont pas « temps réel ».

Le deuxième paramètre intéressant à isoler est l'influence du type de primitive. Nous avons porté dans le tableau ci-dessous le temps de calcul moyen pour le recalage de 100 primitives.

	Repères	Repères semi-orientés	Repères non-orientés	Points
QUAT				20 ms
LSQ	694 ms	785 ms	820 ms	275 ms
KAL	216 ms / 910 ms	245 ms / 1030 ms	250 ms / 1070 ms	87 ms / 362 ms
MAHA	792 ms / 1486 ms	880 ms / 1665 ms	870 ms / 1690 ms	368 ms / 643 ms

Comme toutes nos primitives possèdent actuellement un point (l'origine pour les repères), on peut oublier un instant ce qu'il y a autour du point et initialiser les algorithmes KAL et MAHA par QUAT au lieu de LSQ : les temps de convergence restent similaires, et l'initialisation ne compte plus que pour 20 ms.

8.2.3.3 Discussion

Les trois algorithmes KAL, MAHA et LSQ fonctionnant directement sur les primitives ont leurs avantages et leurs inconvénients : les moindres carrés simples sont en général efficaces et relativement rapides, mais sont sensibles au choix de la métrique, en particulier pour l'estimation de l'incertitude sur la transformation. Par contre, ils fournissent une bonne estimation de la transformation, même en l'absence l'information sur l'incertitude des primitives appariées.

Les méthodes KAL et MAHA supposent que l'on connaisse (ou que l'on ait estimé) la covariance sur les primitives. La descente de gradient sur la distance de Mahalanobis est le plus lent des trois algorithmes (quoique l'on puisse l'accélérer notamment en l'initialisant avec les moindres carrées simples), mais semble donner une estimation fiable de la transformation et de son incertitude dans tous les cas. Par contre, le filtrage de Kalman est très rapide mais extrêmement sensible à l'estimation initiale : démarrer avec l'identité est une covariance élevée conduit souvent à une estimation fantaisiste de la transformation. On pourrait penser à itérer le filtre, pour avoir un meilleur point de départ, mais on perd son avantage principal : l'incrémentalité. De plus, on risque d'estimer une incertitude erronée puisque les mesures ne sont plus indépendantes et que la donnée d'une covariance plus faible sur l'état initial (qui stabilise le filtre) diminue celle de l'état final, mais dans une proportion difficilement calculable.

En fait, le filtrage de Kalman est efficace pour faire de la fusion ou de la mise à jour : si l'on a déjà une estimation de la transformation et de son incertitude, et que l'on ne peut pas conserver les appariements (par exemple pour des raisons de place mémoire), on peut toujours mettre à jour incrémentalement cette transformation grâce au filtre lorsque d'autres appariements se présentent. Au lieu d'estimer directement avec le filtrage de Kalman depuis une transformation initiale incohérente, on pourra conserver les premiers appariements (par exemples les 10 premiers), calculer la transformation initiale avec MAHA, puis utiliser le filtre normalement.

8.2.3.4 Conclusion : utilisation des algorithmes

Il reste cependant que ces trois algorithmes ont des temps de calculs très élevés par rapport à « simplification » de nos primitives en points et l'utilisation de QUAT, pour un apport de précision finalement relativement faible en général. Il semble donc approprié d'utiliser QUAT, au moins en initialisation de tous les algorithmes dès que l'on peut le faire (à partir de 3 primitives). LSQ est alors disqualifié puisqu'il donne à très peu de choses près le même résultat (avec notre métrique et notre modèle de bruit).

On pourra s'arrêter ici si l'on recherche plus la rapidité que l'estimation. Pour améliorer quand même le résultat, on peut utiliser une passe de KAL (en ne multipliant la covariance de QUAT que par 10 ou 20 pour garantir la convergence), puis MAHA pour fignoler l'estimation et recalculer une valeur non biaisée de l'incertitude (typiquement 3 itérations suffisent pour converger). On obtient alors les temps suivant pour recaler 100 repères : 20 ms pour QUAT, 220 ms pour KAL, et 620 ms pour MAHA, soit un total de 850 ms. On peut descendre à 440 ms environ en n'autorisant qu'une seule itération dans MAHA, juste pour recalculer l'incertitude.

Notons pour finir que l'on peut estimer une transformation unique avec un seul appariement de repères. Par contre, il y a deux (resp. 4) solutions avec les repères semi (resp. non) orientés. Cette multiplicité disparaît avec deux appariements de repères semi ou non-orientés en position générique, mais l'algorithme LSQ converge parfois dans ce cas vers un minimum local à cause du faible poids du trièdre vis-à vis des points. On peut en général détecter cette convergence parasite par la lenteur de la convergence (typiquement plus de 30 itérations au lieu d'une quinzaine), mais il convient de toute façon de tester la descente de gradient depuis les différents points de départ possibles pour être sûr d'obtenir le minimum global. L'algorithme MAHA semble être capable de passer outre ces minimums locaux : nous n'avons pas observé de convergence parasite.

8.3 Fusion de primitives ou de transformations

Nous avons développé dans la partie théorique (section 4.3.2) un algorithme par descente de gradient pour estimer la moyenne d'un ensemble de mesures $\{\mathrm{x}_i\}$ supposées être des réalisations de la même primitive aléatoire $\mathbf{x}$. Le problème que l'on se pose ici est de calculer l'incertitude sur cette estimation, comme on l'a fait à la section précédente pour la transformation, et de fusionner plusieurs estimations munies de leur incertitude.

Pour cela, nous pouvons utiliser les mêmes algorithmes que pour estimer une transformation, mais les critères sont plus simples et le problème mieux posé.

8.3.1 Moyenne : moindres carrés simples

Rappelons brièvement l'algorithme de la section (4.3.2) : étant donné un ensemble de N primitives $\{\mathrm{x}_i\}$ (censées être des réalisations de la même primitive aléatoire), la moyenne de Karcher ou de Fréchet est l'élément minimisant le critère des moindres carrés :

$$C(y) = \frac{1}{2}.\sum_i \mathrm{dist}(\mathrm{y},\mathrm{x}_i)^2 = \frac{1}{2}.\sum_i \|\vec{\mathrm{f}}_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}}_i\|^2 = \frac{1}{2}.\sum_i \|\vec{\mathrm{z}}_i\|^2$$

où $\vec{\mathrm{z}}_i = \vec{\mathrm{f}}_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}}_i$ est le vecteur d'erreur pour la moyenne. La fonction implicite Φ définissant le minimum s'écrit (équation (4.20)) :

$$\Phi = \left(\frac{\partial C}{\partial \vec{\mathrm{y}}}\right)^{\mathrm{T}} = -\sum_i J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(-\mathrm{T})}.(\vec{\mathrm{f}}_{\vec{\mathrm{y}}}^{(-1)} \star \vec{\mathrm{x}}_i) = -J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(-\mathrm{T})} \sum_i \vec{\mathrm{z}}_i$$

et ses dérivées sont la matrice hessienne (équation 4.22) et la dérivée croisée :

$$H = N.Q(\vec{\mathrm{y}}) = N.J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(\text{-T})}.J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(\text{-1})} \qquad \text{et} \qquad \frac{\partial \Phi}{\partial \vec{\mathrm{z}}_i} = -J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(\text{-T})}$$

Le vecteur d'ajustement est alors

$$\overrightarrow{\delta \mathrm{y}} = -H^{(\text{-1})}.\Phi = \frac{1}{N}.J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}}) \sum_i \vec{\mathrm{z}}_i$$

et l'équation d'évolution se simplifie en

$$\vec{\mathrm{y}}_{t+1} = \exp_{\vec{\mathrm{y}}_t}(\overrightarrow{\delta \mathrm{y}_t}) = \vec{\mathrm{f}}_{\vec{\mathrm{y}}_t} \star \exp\left(J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}_t})^{(\text{-1})}.\overrightarrow{\delta \mathrm{x}_t}\right) = \vec{\mathrm{f}}_{\vec{\mathrm{y}}_t} \star \left(\frac{1}{N}\sum_i \left(\vec{\mathrm{f}}^{(\text{-1})}_{\vec{\mathrm{y}}_t} \star \vec{\mathrm{x}}_i\right)\right) \tag{8.16}$$

On prend comme point de départ l'une de nos mesures et on s'arrête comme d'habitude lorsque la norme du vecteur d'ajustement devient inférieure à un seuil ε. En pratique, nous avons observé que 5 à 10 itérations suffisent pour une convergence à 10^{-10} avec les repères (éventuellement semi ou non-orientés). Avec les points, on converge bien sûr en une seule itération (à environ 10^{-14}).

Incertitude de l'estimation Le minimum $\hat{\mathrm{y}}$ est défini par le fonction implicite $\Phi = 0$: on utilise encore une fois l'équation (2.14) pour obtenir

$$\Sigma_{\hat{\mathbf{y}}\hat{\mathbf{y}}} = H^{(\text{-1})}.\left(\sum_i \frac{\partial \Phi}{\partial \vec{\mathrm{z}}_i}.\Sigma_{\mathbf{z_i z_i}}.\frac{\partial \Phi}{\partial \vec{\mathrm{z}}_i}^{\mathrm{T}}\right).H^{(\text{-1})} = \frac{1}{N^2}.\sum_i J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}}).\Sigma_{\mathbf{z_i z_i}}.J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{\mathrm{T}}$$

Les mesures x_i étant supposées être des observations de la même primitive aléatoire $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$, il suffit d'approximer la moyenne $\bar{\mathrm{x}}$ par son estimation $\hat{\mathrm{y}}$ pour voir que tous les vecteurs d'erreur suivent la même loi : $\vec{z}_i \sim (0, \Sigma_{\mathbf{zz}})$, avec $\Sigma_{\mathbf{zz}} = J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(\text{-1})}.\Sigma_{\mathbf{xx}}.J(\vec{\mathrm{f}}_{\vec{\mathrm{y}}})^{(\text{-T})}$. On obtient donc au final la covariance très simple sur notre estimée :

$$\Sigma_{\hat{\mathbf{y}}\hat{\mathbf{y}}} = \frac{1}{N}.\Sigma_{\mathbf{xx}} \tag{8.17}$$

Notons au passage qu'on peut estimer $\Sigma_{\mathbf{zz}}$ par :

$$\Sigma_{\mathbf{zz}} = \frac{1}{N-1}.\sum_i \vec{\mathrm{z}}_i.\vec{\mathrm{z}}_i^{\mathrm{T}}$$

8.3.2 Fusion : minimisation de la distance de Mahalanobis

Dérivons tout d'abord l'équation de mesure : on suppose que nos mesures x_i sont toutes des réalisation de primitives aléatoires de même moyenne : $\mathbf{x_i} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{x_i x_i}})$. Les vecteurs d'erreur

$$\vec{\mathbf{z}}_{\boldsymbol{i}} = \vec{\mathrm{f}}_{\vec{\mathrm{y}}} \star \vec{\mathbf{x}}_{\boldsymbol{i}}$$

sont donc de moyenne nulle si notre estimation y correspond bien à la moyenne $\bar{\mathrm{x}}$. Pour prendre en compte l'information contenue dans chacune des matrices de covariance des mesures, on cherche donc à minimiser la distance de Mahalanobis de chaque mesure avec la moyenne :

$$C(\mathrm{y}) = \sum_i \mu^2(\mathbf{z_i}, 0) = \sum_i \vec{\mathrm{z}}_i^{\mathrm{T}}.\Sigma^{(\text{-1})}_{\mathbf{z_i z_i}}.\vec{\mathrm{z}}_i$$

Pour le filtrage de Kalman et la descente de gradient, nous aurons besoin de la dérivée du vecteur d'erreur par rapport à l'état courant $\vec{y}$ et de la matrice de covariance de $\vec{\mathbf{z}}_{\boldsymbol{i}}$. Malheureusement, l'introduction des matrices de covariance ne permet pas les simplifications utilisées pour les moindres carrés simples, et on a seulement :

$$J_{\vec{y}_i} = \frac{\partial \vec{z}_i}{\partial \vec{y}} = \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \vec{x}_i)}{\partial \vec{y}} = \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \vec{x}_i)}{\partial \vec{f}_{\vec{y}}^{(-1)}} . \frac{\partial(\vec{f}_{\vec{y}}^{(-1)})}{\partial \vec{f}_{\vec{y}}} . \frac{\partial \vec{f}_{\vec{y}}}{\partial \vec{y}}$$

$$\Sigma_{\mathbf{z_i z_i}} = \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \vec{x}_i)}{\partial \vec{x}_i} . \Sigma_{\mathbf{x_i x_i}} . \frac{\partial(\vec{f}_{\vec{y}}^{(-1)} \star \vec{x}_i)}{\partial \vec{x}_i}^{\mathrm{T}}$$

8.3.2.1 Filtrage de Kalman

Nous n'avons plus cette fois-ci de problème pour initialiser le filtre, on démarre avec pour état la première mesure : $\mathbf{y_1} = \mathbf{x_1}$. Ensuite, pour chaque autre mesure $\mathbf{x_i}$:

- Calculer $\mathbf{z_i}$ et la matrice $M = J_{\vec{y}_i}$ estimée en $\vec{y}_{(i-1)}$,
- Mettre à jour l'état de $\mathbf{y}_{(i-1)} \sim (\vec{y}_{(i-1)}, \Lambda_{(i-1)})$ à $\mathbf{y_i} \sim (\vec{f}_i, \Lambda_i)$ en utilisant les équations du filtre (C.1),
- Normaliser la primitive probabiliste $\mathbf{y_i}$.

8.3.2.2 Descente de gradient

Si l'on écrit le critère $C(\mathrm{y}) = \sum_i \vec{z}_i^{\mathrm{T}} . \Sigma_{\mathbf{z_i z_i}}^{(-1)} . \vec{z}_i$, la dérivée et la matrice hessienne de ce critère pour l'estimation $\vec{y}_t$ sont donc :

$$\Phi_t = \left(\frac{\partial C}{\partial \vec{f}_t}\right)^{\mathrm{T}} = \sum_i \frac{\partial \vec{z}_i}{\partial \vec{y}_t}^{\mathrm{T}} . \Sigma_{\mathbf{z_i z_i}}^{(-1)} . \vec{z}_i \qquad \text{et} \qquad H_t = \frac{\partial \Phi_t}{\partial \vec{f}_t} = \sum_i \frac{\partial \vec{z}_i}{\partial \vec{y}_t}^{\mathrm{T}} . \Sigma_{\mathbf{z_i z_i}}^{(-1)} . \frac{\partial \vec{z}_i}{\partial \vec{y}_t}$$

Le vecteur d'ajustement est alors $\overrightarrow{\delta \mathrm{y}_t} = -H_t^{(-1)} . \Phi_t$ et l'équation d'évolution reste :

$$\vec{y}_{t+1} = \exp_{\vec{y}_t}(\overrightarrow{\delta \mathrm{y}_t}) = \vec{f}_{\vec{y}_t} \circ \left(J(\vec{f}_{\vec{y}_t})^{(-1)} . \overrightarrow{\delta \mathrm{y}_t} \right) \tag{8.18}$$

Comme pour les autres algorithmes, on prend comme point de départ l'une de nos mesures, et pour le critère d'arrêt un seuil sur la norme du vecteur d'ajustement (en pratique, nous avons observé une convergence à 10^{-10} en une quinzaine d'itérations). Pour déterminer l'incertitude sur la transformation à l'optimum, on a exactement le même type d'équations que pour le recalage : $\Sigma_{\hat{\mathbf{y}}\hat{\mathbf{y}}} = H^{(-1)}$, où H est la matrice hessienne à l'optimum (par exemple la dernière calculée).

8.3.3 Comparaison des algorithmes

On note comme pour le recalage en LSQ, KAL et MAHA les trois méthodes pour la fusion développées ci-dessus.

8.3.3.1 Précision des résultats

Nous avons réalisé le même type d'expériences que pour le recalage, mais il n'y a pas ici de méthode qui donne un résultat meilleur en général si toutes les primitives ont la même incertitude : les précisions relatives se tiennent toutes à moins de 5% près. Par contre, on peut observer une précision un peu plus importante pour MAHA et KAL dans le cas de la fusion de deux ou trois primitives ayant des bruits très différents.

8.3.3.2 Temps de calculs

Les expériences sont réalisées dans les mêmes conditions que pour le recalage, mais notre référence temporelle est ici le temps de calcul du barycentre de 100 points (sans détermination de l'incertitude sur l'estimée) : 0.2 ms.

La première expérience concerne l'influence du nombre de points sur le temps de calcul de la primitive moyenne. Les temps donnés ci-dessous sont des moyennes sur les quatre types de primitives confondues (repères, repères semi et non-orienté, points). L'algorithme MAHA est initialisé avec le résultat de LSQ; nous donnons le décompte du temps passé dans l'algorithme proprement dit et le temps total avec l'initialisation.

	Nombre de primitives			
	10	50	100	500
LSQ	6 ms	30 ms	60 ms	285 ms
KAL	10 ms	53 ms	109 ms	540 ms
MAHA	10 ms / 16 ms	42 ms / 72 ms	83 ms / 143 ms	537 ms / 822 ms

Il est clair sur ces chiffres que la relation est linéaire, et on en conclut qu'il faut compter environ 60 ms pour fusionner 100 primitives avec LSQ, 110 ms par le filtrage de Kalman et 160 ms par la descente de gradient sur la distance de Mahalanobis. Notons que ce dernier temps peut être multiplié par un facteur 5 à 10 si l'initialisation est simplement la première primitive.

Le deuxième paramètre intéressant à isoler est l'influence du type de primitive. Nous avons porté dans le tableau ci-dessous le temps de calcul de la moyenne de 100 primitives.

	Repères	Repères semi-orientés	Repères non-orientés	Points
LSQ	68 ms	75 ms	78 ms	11 ms
KAL	122 ms	132 ms	133 ms	51 ms
MAHA	94 ms / 162 ms	100 ms / 175 ms	90 ms /168 ms	30 ms / 41 ms

Comme on pouvait s'y attendre, les estimations faisant intervenir des repères sont plus longues que les estimations (linéaires) sur les points, mais restent relativement raisonnables.

8.3.3.3 Conclusion

Les trois algorithmes donnent de très bons résultats pour le calcul de la moyenne. Dans le cas de la fusion (matrices de covariances différentes sur les divers mesures), on utilisera plutôt MAHA ou KAL pour prendre en compte le maximum d'information. Du point de vue de la complexité, les moindres carrés simples sont les plus rapides, suivis par le filtrage de Kalman et MAHA est le plus lent mais la différence entre les trois n'est pas critique.

En résumé, on utilisera les moindres carrés simples pour faire une moyenne (sans information de covariance), la descente de gradient sur la distance de Mahalanobis ou le filtrage de Kalman si l'on veut un résultat plus précis en fusion.

Notons que, contrairement au cas du recalage, on ne peut pas ici simplifier nos primitives pour en faire une moyenne approximative plus rapidement. Par contre, le problème est bien posé et les temps de calculs restent faibles.

8.4 Estimation du modèle de bruit sur les primitives

Dans les deux problèmes d'estimation précédents, le recalage et la fusion, nous avons supposé que le modèle de bruit sur les primitives était connu. Même si l'on peut souvent en avoir une idée a priori, il est intéressant de le calculer a posteriori, après recalage ou fusion, pour pouvoir vérifier si nos hypothèses sont valides ou éventuellement le réinjecter dans les algorithmes pour obtenir des estimations plus précises. Par ailleurs, si ce bruit de mesure provient de la succession de divers traitements, comme par exemple en imagerie l'acquisition, la numérisation, les pré-traitements puis l'extraction des primitives, ce calcul a posteriori est souvent le seul moyen de connaître ou de modéliser le bruit sur nos primitives. Nous serons ainsi en mesure de modéliser le bruit sur les points extrémaux (dans certaines images IRM) à la section (9.3.3) grâce à de telles mesures a posteriori.

8.4.1 Estimation du bruit après fusion

Observons tout d'abord le cas des points : on a N mesures $\{x_i\}$ provenant de la même distribution $\boldsymbol{x} \sim (\overline{x}, \Sigma_{\boldsymbol{xx}})$. La covariance est normalement :

$$\Sigma_{\boldsymbol{xx}} = \int_{\mathbb{R}^n} (y - \overline{x}).(y - \overline{x})^{\mathrm{T}}.p_{\boldsymbol{x}}(y).dy \simeq \frac{1}{N} \sum_i (x_i - \overline{x}).(x_i - \overline{x})^{\mathrm{T}}$$

En fait, on ne connaît pas exactement la moyenne $\overline{x}$ mais on en a obtenu une estimation $\hat{x}$ par fusion (en l'occurrence ici avec le barycentre). L'idée est donc d'utiliser $\hat{x}$ à la place de $\overline{x}$ dans l'équation ci-dessus, mais comme notre estimation minimise l'erreur (aux moindres carrés), elle est légèrement plus proche des données que ne le serait la vraie moyenne. On peut corriger ce biais en utilisant la formule bien connue (Bard, 1974; Press, 1972) :

$$\hat{\Sigma}_{\boldsymbol{xx}} = \frac{1}{N-1}. \sum_i (x_i - \hat{x}).(x_i - \hat{x})^{\mathrm{T}}$$

Le cas des primitives est tout à fait similaire : nous avons défini la covariance empirique comme (équation 4.28) :

$$\Sigma_{\mathbf{xx}} = \mathbf{E}\left[\overrightarrow{\bar{\mathrm{x}}\mathbf{X}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{X}}^{\mathrm{T}} \right] \simeq J(\vec{\mathrm{f}}_{\overline{\mathrm{x}}}). \left(\frac{1}{N} \sum_i (\vec{\mathrm{f}}_{\overline{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i).(\vec{\mathrm{f}}_{\overline{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i)^{\mathrm{T}} \right) .J(\vec{\mathrm{f}}_{\overline{\mathrm{x}}})^{\mathrm{T}}$$

Il ne reste donc qu'à corriger le biais que l'on introduit en utilisant l'estimation $\hat{\mathrm{x}}$ au lieu de la moyenne exacte $\overline{\mathrm{x}}$:

$$\hat{\Sigma}_{\mathbf{xx}} = \frac{1}{N-1} .J(\vec{\mathrm{f}}_{\hat{\mathrm{x}}}). \left(\sum_i (\vec{\mathrm{f}}_{\hat{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i).(\vec{\mathrm{f}}_{\hat{\mathrm{x}}}^{(-1)} \star \vec{\mathrm{x}}_i)^{\mathrm{T}} \right) .J(\vec{\mathrm{f}}_{\hat{\mathrm{x}}})^{\mathrm{T}} \tag{8.19}$$

8.4.2 Estimation du bruit après recalage

Ce type d'estimation est l'un des plus important car il est utilisable avec des données réelles sans avoir à concevoir une expérience spécifique. Nous étudions ici le cas des points et des repères, avant de discuter des possibilités de généralisation au cas de primitives génériques.

8.4.2.1 Estimation du modèle de bruit sur les points

Bruit additif anisotrope On considère donc deux ensembles de points appariés $\{x_i\}$ et $\{y_i\}$, images l'un de l'autre par une transformation : $y_i = \mathrm{f} \star x_i$. On ne mesure en fait que les valeurs bruitées $\boldsymbol{x_i} = x_i + \boldsymbol{\delta x_i}$ et $\boldsymbol{y_i} = y_i + \boldsymbol{\delta y_i}$, où les bruits additifs sont supposés être indépendants et de covariance $\Sigma_{\boldsymbol{xx}}$ et $\Sigma_{\boldsymbol{yy}}$. En supposant que l'on connaisse la transformation exacte, le vecteur d'erreur est

$$\boldsymbol{z_i} = \boldsymbol{y_i} - \mathrm{f} \star \boldsymbol{x_i} = \boldsymbol{\delta y_i} - R.\boldsymbol{\delta x_i}$$

La covariance sur $\boldsymbol{z_i}$ est donc $\Sigma_{\boldsymbol{zz}} = \Sigma_{\boldsymbol{yy}} + R.\Sigma_{\boldsymbol{xx}}.R^{\mathrm{T}}$, et peut être estimée par :

$$\hat{\Sigma}_{\boldsymbol{zz}} = \frac{1}{N}\sum_i \hat{z}_i.\hat{z}_i^{\mathrm{T}} \tag{8.20}$$

Si l'un des ensembles de points est exact, on peut facilement estimer la covariance sur l'autre :

$$\Sigma_{\boldsymbol{xx}} = 0 \implies \hat{\Sigma}_{\boldsymbol{yy}} = \hat{\Sigma}_{\boldsymbol{zz}} \qquad \text{et} \qquad \Sigma_{\boldsymbol{yy}} = 0 \implies \hat{\Sigma}_{\boldsymbol{xx}} = R^{\mathrm{T}}.\hat{\Sigma}_{\boldsymbol{zz}}.R$$

Le problème est plus difficile si l'on suppose que les deux ensembles de points proviennent d'acquisitions similaires et qu'ils ont donc le même bruit de covariance Σ : on doit résoudre l'équation $\hat{\Sigma}_{\boldsymbol{zz}} = \Sigma + R.\Sigma.R^{\mathrm{T}}$. Une méthode est présentée dans (Koch, 1988) pour résoudre ce type d'équation. Dans notre cas, on peut vérifier avec un logiciel de calcul symbolique (par exemple Maple) que la solution est unique, à moins que la rotation R ait un angle $\theta = \pi/2$. Cependant, dans un grand nombre de cas en imagerie médicale, le patient a grossièrement la même position dans la machine lors des différentes acquisitions et la rotation est faible : on peut alors approximer la covariance sur les points par :

$$\hat{\Sigma}_{\boldsymbol{xx}} = \hat{\Sigma}_{\boldsymbol{yy}} = \frac{\hat{\Sigma}_{\boldsymbol{zz}}}{2} = \frac{1}{2.N}.\sum_i z_i.z_i^{\mathrm{T}}$$

Nous avons jusqu'à présent supposé que l'on connaissait la transformation exacte f. En fait, après le recalage, nous n'en avons qu'une estimée $\hat{\mathrm{f}}$ aux moindres carrés, et l'utilisation de cette estimée dans les formules ci-dessus introduit un biais : l'estimation de l'erreur est légèrement plus faible qu'elle ne devrait être. En suivant (Bard, 1974), nous devons donc remplacer le nombre de mesures N par $N - l/m$, où m est la dimension des équations de mesure (vectorielles) et l le nombre de paramètres que l'on a estimés. Dans notre cas, on a estimé un mouvement rigide 3D ($l = 6$) à partir de points ($m = 3$).

On obtient donc au final une estimation du bruit additif anisotrope sur les points après recalage avec :

$$\Sigma \simeq \hat{\Sigma}_{\boldsymbol{xx}} = \hat{\Sigma}_{\boldsymbol{yy}} = \frac{\hat{\Sigma}_{\boldsymbol{zz}}}{2} = \frac{1}{2.(N-2)}.\sum_i \hat{z}_i.\hat{z}_i^{\mathrm{T}} \tag{8.21}$$

Bruit isotrope On a dans ce cas $\Sigma_{\boldsymbol{xx}} = \sigma_x^2.I_d$ et $\Sigma_{\boldsymbol{yy}} = \sigma_y^2.I_d$. La covariance sur le vecteur d'erreur est donc $\Sigma_{\boldsymbol{xx}} = \sigma_z^2.I_d$, avec $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$. La matrice de covariance sur le vecteur d'erreur peut toujours être estimée avec l'équation (8.20), et on obtient en prenant la trace :

$$\mathrm{Tr}(\hat{\Sigma}_{\boldsymbol{zz}}) = 2.\hat{\sigma}_z^2.d = \frac{1}{N}\sum \|\hat{z}_i\|^2$$

où d est la dimension de l'espace : $d = 3$ dans notre cas. On peut remarquer que $\hat{C} = \sum_i \|z_i\|^2$ est la valeur du critère (8.1) au minimum. On peut donc récrire notre estimation en incluant la correction

du biais :

$$\sigma_x^2 + \sigma_y^2 \simeq \hat{\sigma}_z^2 = \frac{\hat{C}}{2.d.(N-2)}$$

Il n'y a plus de problème cette fois-ci pour estimer les variances en supposant que l'un ou l'autre des ensembles soit exact, ni en supposant qu'ils aient la même variance : $\sigma_x^2 = \sigma_y^2 \simeq \hat{\sigma}_z^2/2$.

8.4.2.2 Estimation du modèle de bruit sur les repères

Bruit homogène ou isotrope Ce cas est particulier puisque les repères sont identifiables à des transformations rigides, et cela amène des simplifications dans les calculs. Pour bien marquer que nous utilisons cette particularité, nous notons dans cette section $\{\mathrm{f}_{m_i}\}$ et $\{\mathrm{f}_{s_i}\}$ les deux ensembles de repères appariés (m pour modèle et s pour scène). Pour alléger les notations, nous oublions également les flèches sur ces transformations, bien que tout ceci ne soit valide que dans la carte principale[2]. En fait, nous n'avons accès qu'à leur mesures bruitées $\mathbf{f_{m_i}} = \mathrm{f}_{m_i} \circ \mathbf{e_{m_i}}$ et $\mathbf{f_{s_i}} = \mathrm{f}_{s_i} \circ \mathbf{e_{s_i}}$. En supposant que l'on connaisse exactement la transformation f du modèle vers la scène, le vecteur d'erreur est :

$$\mathbf{e_i} = \mathbf{f_{s_i}}^{(-1)} \circ \mathrm{f} \circ \mathbf{f_{m_i}} = \mathbf{e_{s_i}}^{(-1)} \circ \mathbf{e_{m_i}}$$

Avec l'hypothèse d'un bruit homogène identique pour les deux ensembles de repères, on a $\mathbf{e_{s_i}} \sim (0, \Sigma)$ et de même pour $\mathbf{e_{m_i}}$. Comme le jacobien de l'inversion est $-I_6$ à l'identité, le vecteur d'erreur du recalage a pour loi $\mathbf{e_i} = (0, 2.\Sigma)$.

Comme dans le cas des points, un estimateur de la covariance Σ est donné par :

$$\hat{\Sigma} = \frac{1}{2.(N-1)} . \sum_i \hat{\mathrm{e}}_i . \hat{\mathrm{e}}_i^{\mathrm{T}}$$

car on utilise l'estimation $\hat{\mathrm{f}}$ de la transformation fournie par le recalage au lieu de la transformation exacte et il faut remplacer N par $N-1$ pour corriger le biais (la dimension des équations de mesure est cette fois-ci $m = l = 6$).

Un modèle de bruit simplifié Dans certain cas, par exemple avec un petit nombre d'appariements, l'estimation de la matrice de covariance ci-dessus est très instable. Nous utilisons alors un modèle de bruit simplifié, inspiré du bruit isotrope sur les points : on impose une covariance diagonale « isotrope » par bloc, c'est-à-dire avec une variance σ_θ^2 commune à chacune des coordonnées du vecteur rotation et σ_d^2 commune aux coordonnées de la translation. En séparant le vecteur d'erreur en une composante rotation et une composante translation : $\mathrm{e}_i^{\mathrm{T}} = (\mathrm{e}_{\theta_i}^{\mathrm{T}}, \mathrm{e}_{d_i}^{\mathrm{T}})$, on peut estimer ces variances par

$$\hat{\sigma}_\theta^2 = \frac{\sum \|\hat{\mathrm{e}}_{\theta_i}\|^2}{6.(N-1)} \quad \text{et} \quad \hat{\sigma}_d^2 = \frac{\sum \|\hat{\mathrm{e}}_{d_i}\|^2}{6.(N-1)}$$

8.4.3 Discussion sur l'estimation du bruit

Les méthode pour estimer le bruit que nous venons de présenter reposent à chaque fois sur des hypothèses bien précises : bruit additif pour les points, identification avec les transformations pour les repères. Ces hypothèses impliquent à chaque fois qu'une covariance homogène sur les données induise une covariance *identique* sur le vecteur d'erreur. Il n'est pas clair que cela soit vrai pour tous les types de primitives ou même pour les points avec une fonction de placement différente (bruit non

2. Rappelons que cette représentation est constituée du vecteur rotation et du vecteur translation.

additif). Une étude supplémentaire serait à mener pour voir si l'on peut quand même concevoir des méthodes pour estimer un bruit homogène sur les données à partir des valeurs du vecteur d'erreur.

Dans le cas d'un bruit isotrope, il faut de plus contraindre la matrice de covariance estimée $\hat{\Sigma}$ à être invariante par l'action du groupe d'isotropie $\mathcal{H}$. Là encore, une étude plus approfondie serait nécessaire pour savoir s'il faut ajouter une opération de ce type dans la liste de nos opérations atomiques (i.e. dépendantes du type de primitive).

Par ailleurs, nous avons vu avec le bruit additif général sur les points que ce problème d'estimation peut être mal posé. Supposons par exemple la covariance suivante sur des points en deux dimension : $\Sigma = \mathrm{DIAG}(\sigma_1^2\ ,\ \sigma_2^2)$, identique sur les x et les y, une rotation de 90 degrés donne en effet une covariance isotrope sur z : $\Sigma_{zz} = \mathrm{DIAG}(\sigma^2\ ,\ \sigma^2)$ avec $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Il y a donc plusieurs solutions au problème inverse, et il faut régulariser pour en choisir une. Une bonne solution pour cela consiste à prendre la covariance Σ (solution du problème d'estimation) qui minimise l'information, donc qui maximise $\log(\det(\Sigma))$. Dans l'exemple ci-dessus, cela revient à maximiser $\sigma_1^2.\sigma_2^2$ sous la contrainte $\sigma^2 = \sigma_1^2 + \sigma_2^2$, ce qui donne la solution (unique) : $\sigma_1^2 = \sigma_2^2 = \sigma^2/2$. Ce problème de régularisation devrait être intégré dans une théorie plus générale de l'estimation du modèle de bruit sur les primitives.

En pratique, pour estimer un modèle de bruit sur des repères semi ou non-orientés, nous considérons que l'orientation des repères appariés est fixée par le recalage en fixant au hasard celle de l'un des repères de chaque couple et en choisissant celle qui minimise la distance après recalage pour l'autre. Nous pouvons ainsi utiliser les techniques d'estimation du bruit développées pour les repères.

8.5 Un algorithme pratique et générique pour le recalage

On récupère en général, à la sortie d'un algorithme de mise en correspondance, deux ensembles de primitives appariées, sans avoir forcément une estimation de l'incertitude sur ces mesures. De plus, certains appariements peuvent être aberrants vis à vis du mouvement principal : il ne faut donc pas les prendre en compte dans le recalage.

Pour pouvoir prédire l'incertitude sur le recalage, nous utilisons donc la méthodologie suivante : une première estimation aux moindres carrés permet d'estimer grossièrement un bruit homogène ou isotrope sur les primitives, que l'on peut utiliser pour trier les appariements par ordre de vraisemblance et éliminer les plus aberrants. On recalcule alors le recalage sur les mesures fiables munies de leur incertitude en minimisant la distance de Mahalanobis, et les algorithmes que nous avons développés dans la section (8.2) nous permettent également d'estimer la précision de ce résultat. Ce processus itératif peut être continué jusqu'à la convergence.

8.5.1 Rejet des mesures aberrantes

Le problème que l'on se pose ici est de déterminer si un appariement est compatible ou aberrant vis a vis d'une certaine transformation. En clair, la question est de décider si $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ en ne connaissant que les valeurs mesurées $\hat{\mathrm{y}}$ et $\hat{\mathrm{x}}$. Un problème lié est de classer les appariements par ordre de vraisemblance, par exemple pour stabiliser le filtre de Kalman lors de l'estimation du recalage.

Nous avons déjà vu la solution à ces problèmes avec la **distance de Mahalanobis** $\mu^2(\mathbf{y}, \mathrm{f}\star\mathbf{x})$ et le **test du** $\boldsymbol{\chi}^2$ (section 5.3). L'interprétation de la distance de Mahalanobis comme une distribution du χ^2 suppose bien sûr que l'on se place dans le cadre de l'approximation gaussienne pour une covariance Σ faible (section 5.2.5) : on suppose donc que les covariances sur les primitives aléatoires

$\mathbf{y}$ et $\mathbf{x}$ sont faibles, et grâce au principe du minimum d'information, on peut supposer que la distribution est gaussienne sur la variété (voir la section 5.2), ce que l'on peut approcher efficacement par une gaussienne dans $\mathbb{R}^n$ (n étant la dimension de nos primitives).

Notons que la distance de Mahalanobis $\mu^2(\mathbf{y}, \mathrm{f} \star \mathbf{x})$ que l'on calcule ici est égale à la distance de Mahalanobis $\mu^2(\mathbf{z}, \mathrm{o})$ du vecteur d'erreur pour le recalage $\mathbf{z} = \mathbf{f_y}^{(-1)} \star (\mathrm{f} \star \mathbf{x})$ par rapport à l'origine. Si la transformation f n'est pas connue exactement mais estimée par $\hat{\mathrm{f}}$ lors d'une recalage, on aura tendance à sous-estimer légèrement la distance de Mahalanobis entre nos deux primitives. Il est donc important de ne pas utiliser dans notre distance de Mahalanobis la transformation probabiliste $\hat{\mathbf{f}}$ munie de l'incertitude que l'on a pu estimer, ce qui tendrait à minimiser encore plus la distance résultante, mais simplement la transformation déterministe $\hat{\mathrm{f}}$ à la place de f.

Le test du χ^2 vise alors à vérifier que la valeur observée $\hat{\mathrm{z}}$ du vecteur d'erreur est bien compatible avec sa covariance théorique. Rappelons que la distance de Mahalanobis est, avec ces notations et dans la carte principale :

$$\mu^2 = \hat{\mathrm{z}}^{\mathrm{T}}.\Sigma_{\mathbf{zz}}^{(-1)}.\hat{\mathrm{z}}$$

Le test statistique consiste à accepter l'hypothèse $\mathbf{y} = \mathrm{f} \star \mathbf{x}$ si $\mu^2 \leq \varepsilon$ et à la rejeter sinon. Le seuil ε est choisi de telle manière que l'on accepte l'hypothèse avec une probabilité α si elle est vraie. Nous présentons dans la table (8.1) quelques exemples de ces seuils ε en fonction de la probabilité α pour des tests du χ^2 en dimension 3 (pour les points) et 6 (pour les divers repères). Par exemple, si la distance de Mahalanobis entre deux points est $\mu^2 > 11.34$, la probabilité que $\boldsymbol{y}$ soit identique à $\mathrm{f} \star \boldsymbol{x}$ est inférieure à 1%. Nous pouvons donc rejeter cet appariement.

α	Dim	50%	90 %	95 %	99%
ε	3	2.37	6.25	7.81	11.34
ε	6	5.35	10.65	12.59	16.81

TAB. 8.1 – *Table de la distribution du χ^2 à 3 et 6 degrés de liberté.*

8.5.2 Un processus itératif d'estimation globale

Nous pouvons maintenant intégrer toutes les étapes d'estimation décrites précédemment dans un seul algorithme itératif. On suppose donc que l'on a au départ deux ensembles de primitives appariées, sans information d'incertitude. La connaissance de cette information supprime bien évidemment les étapes d'estimation sur modèle de bruit sur les primitives et permet de commencer directement le processus itératif sans initialisation.

Pour calculer une première estimation de l'incertitude sur les primitives, il nous faut un premier recalage qui ne peut se faire qu'avec les moindres carrés. L'initialisation est donc la suivante :

1. estimer un recalage grossier avec l'algorithme LSQ de la section (8.2). En pratique, les moindres carrés sur les points seulement (QUAT) sont souvent suffisants et beaucoup plus rapides. On peut également concevoir une estimation robuste (voir section 8.5.3),
2. estimer le modèle de bruit sur les primitives (section 8.4),
3. écarter les appariements aberrants de la liste par un test du χ^2,
4. faire une estimation grossière de l'incertitude sur la transformation pour obtenir $(\hat{\mathrm{f}}_0, \Sigma_{\mathbf{f_0 f_0}})$.

Si une transformation initiale est fournie, on ne réalise pas la première étape. La seconde étape, l'estimation du modèle de bruit, n'est nécessaire que si le type de données est nouveau. A partir du moment où l'on a une estimation fiable de ce bruit, il n'est plus nécessaire de le recalculer à chaque fois.

Pour réaliser un compromis acceptable entre le temps de calcul et la précision, le processus itératif générique est alors le suivant:

1. trier les appariements par distance de Mahalanobis croissante,
2. faire une passe de filtrage de Kalman (algorithme KAL) sur les appariements acceptés pour obtenir $(\hat{f}_{i+1}, \Sigma_{\mathbf{f_{i+1}f_{i+1}}})$, en utilisant comme transformation initiale $(\hat{f}_i, 20 * \Sigma_{\mathbf{f_i f_i}})$.

(3.) éventuellement réestimer le modèle de bruit sur les primitives,

4. écarter les appariements aberrants de la liste par un test du χ^2, en vérifiant également les anciens appariements aberrant (qui pourrait redevenir acceptables).

Ce processus est répété jusqu'à la convergence (i.e. une distance entre deux transformation successives inférieure à 10^{-12} par exemple), ou un nombre maximal d'itération (typiquement 5 à 10 sont suffisantes).

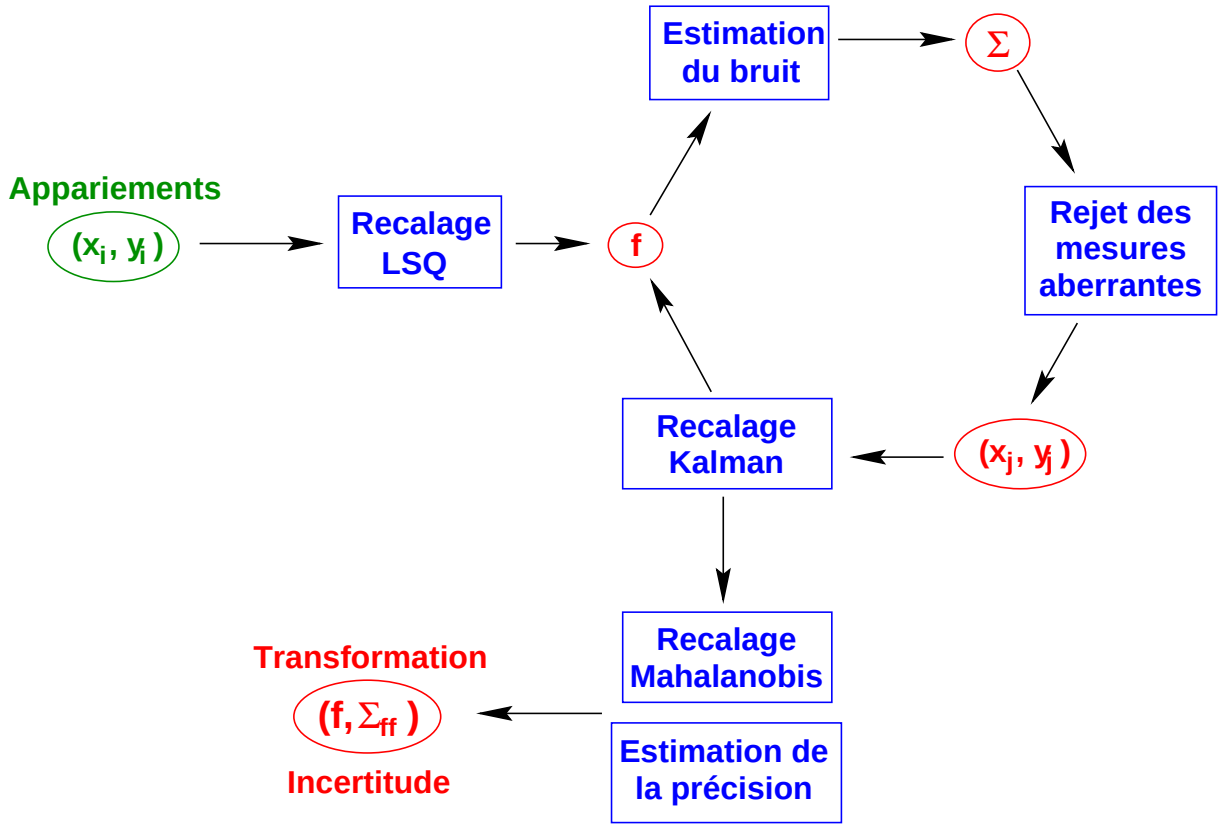


FIG. 8.2 – *Algorithme modulaire de recalage*

Le facteur 20 utilisé à chaque passe du filtre de Kalman pour multiplier la covariance de l'estimation précédente, ainsi que le tri des appariements, permettent de stabiliser le filtre de Kalman et d'assurer la convergence, mais induisent une sous-estimation de l'incertitude sur la transformation de plus en plus importante au cours des itérations. Pour obtenir à la fois une transformation précise et une bonne estimation de son incertitude, nous finissons donc le processus par :

– une descente de gradient sur la distance de Mahalanobis (MAHA) en n'utilisant que les appariements acceptés et l'estimation de l'incertitude sur le mouvement au minimum.

Le point faible de cette encapsulation d'algorithmes est l'estimation du modèle de bruit sur les primitives, qui est relativement sensible et sujet à l'erreur, en particulier avec un faible nombre

d'appariement. Cependant, c'est souvent la seule façon d'avoir une idée du bruit sur les données lorsque l'on utilise de nouvelles données. En pratique, on peut fusionner plusieurs estimations du bruit lors d'une phase d'apprentissage pour être plus précis, et éventuellement continuer à mettre à jour cette estimation plus robuste au cours de recalages du même type de données. Cependant, les résultats obtenus sur les données réelles de la section (9.3.3) semblent être relativement stables.

8.5.3 Variations sur cet algorithme et robustesse

Les divers repères sont des primitives qui contiennent un point. On peut donc sacrifier la précision à la vitesse en simplifiant nos repères en points. On peut également choisir d'estimer un bruit homogène sur les repères ou un bruit simplifié, ou bien un bruit additif isotrope ou non sur les points. Le choix du modèle de bruit est plutôt guidé par le nombre d'observations disponibles : nous présenterons une étude de ces paramètres à la section (9.2).

En terme de robustesse, l'algorithme que nous proposons ici appartient à la classe des M-estimateurs redescendants, c'est-à-dire la classe des algorithmes robustes basés sur une itération de moindres carrés (voir par exemple (Huber, 1981; Meer et al., 1991) pour un état de l'art des statistiques robustes). Cependant, le point de chute (« breakdown point ») est peu élevé. Cela veut dire que l'on ne peut pas supporter un pourcentage élevé de mesures aberrantes sans influer de manière significative sur le résultat, en l'occurrence la justesse de la transformation estimée. On peut rendre cet algorithme plus robuste jusqu'à un point de chute de 0.5 (i.e. supportant 50% de mesures aberrantes) en initialisant, non plus avec un moindre carrés (LSQ), mais avec une moindre médiane des carrés (LMS pour « Least Median of Squares »). Ceci est particulièrement facile et peu cher en complexité pour les repères : chaque appariement détermine une transformation unique, avec laquelle on peut classer les autres appariements par distance (ou distance de Mahalanobis). On attribue à l'appariement servant de base un score correspondant à la distance médiane après classement, et on choisit comme transformation celle dont l'appariement de base minimise le score de distance médiane. Si nous avons N appariements, la complexité théorique est de $O(N^2 \log N)$, mais on peut la réduire considérablement grâce à un échantillonnage de Monté-Carlo (Meer et al., 1991).

Il est difficile de mettre un tel algorithme dans un cadre complètement générique, puisque pour les repères semi-orienté, nous avons deux transformations possibles par appariement et quatre pour les repères non-orientés. La complexité de cet algorithme augmente beaucoup pour les points (théoriquement $O(N^4 \log N)$) puisqu'il nous faut cette fois-ci trois appariements pour déterminer (aux moindres carrés) une transformation unique.

Notons qu'il est impératif de conserver l'étape d'ajustement itératif de la transformation pour atteindre l'objectif d'efficacité de l'estimation, c'est-à-dire pour réduire au maximum l'incertitude sur l'estimation.

Enfin, il serait souhaitable, pour obtenir un algorithme entièrement robuste, d'avoir également des techniques d'estimation robustes de la matrice de covariance du bruit sur les primitives.

8.6 Conclusions sur le recalage et sa précision

Nous avons étudié dans ce chapitre les principaux algorithmes de recalage sur les points et montré que l'on pouvait les généraliser de différentes manières à des primitives génériques, en estimant de plus l'incertitude sur le résultat. La fusion de primitives est assimilable à une version simplifiée de ce problème et peut se résoudre par des algorithmes similaires. Nous avons également montré

que ces algorithmes génériques, même s'il engendrent un sur-coût de calcul non négligeable, restent abordables avec une implémentation un peu optimisée.

Par contre, l'estimation du modèle de bruit sur les primitives après recalage reste un problème ouvert dans le cas général, que nous n'avons résolu ici que par des techniques ad hoc spécifiques à chaque type de primitive. La conception d'un cadre général pour ce problème semble cependant possible.

Enfin, nous avons montré dans la dernière section un exemple d'organisation de ces briques algorithmiques pour fabriquer un algorithme dédié à un problème général relativement difficile : l'estimation simultanée du recalage, de son incertitude du modèle de bruit sur les primitives et des appariements aberrants. La solution que nous proposons ne requiert qu'un seul paramètre : le seuil ε pour le test du χ^2, mais nécessite cependant un nombre suffisant d'appariements pour pouvoir réaliser des statistiques fiables.

Il reste toutefois quelques résultats importants à vérifier en ce qui concerne la précision que nous prédisons sur le recalage : c'est le problème de la **validation** qui fera l'objet du prochain chapitre.

Chapitre 9

Validation du recalage d'images médicales

« I could prove God statistically. »
George Gallup

Nous avons obtenu dans le chapitre précédant divers algorithmes pour calculer l'incertitude sur le recalage. La question que l'on se pose maintenant est de savoir si cette estimation est réaliste. De manière plus générale, le problème de la validation du recalage d'images médicales est crucial, voire vital, par exemple si l'on utilise ce recalage pour guider une opération chirurgicale. Nous proposons dans la première section une méthode statistique pour réaliser cette validation et nous étudions dans la section (9.2) l'influence de divers paramètres sur la justesse de l'incertitude estimée avec des données synthétiques. Enfin, les sections (9.3) (9.4) sont consacrées à des données réelles en imagerie médicales.

9.1 Validation des méthodes de recalage

La question traditionnelle de la validation est : « quelle est la distance entre le résultat de notre technique et une référence que l'on considère comme la vérité terrain ? ». En d'autres termes, il nous faut trouver un moyen de mesurer l'incertitude sur la transformation estimée. Nous analysons dans la section suivante divers manières de prédire ou de mesurer l'incertitude.

Si l'on utilise maintenant des statistiques d'un ordre supérieur dans le recalage, on obtient en même temps que l'estimation du mouvement une évaluation de l'incertitude sur ce résultat. La question est alors de vérifier si les divers hypothèses statistiques utilisées dans la technique de recalage sont vérifiées et si l'évaluation de l'incertitude est réaliste. C'est le sujet de la section (9.1.2).

9.1.1 Prédiction de l'incertitude

La méthode idéale pour prédire l'erreur finale sur le recalage consisterait à modéliser analytiquement le processus entier de recalage, de la position physique de l'objet à l'estimation finale du

mouvement. Pour les cas réels, cela implique de modéliser les déformations de l'objet (rien n'est réellement rigide, surtout pour des « objets » anatomiques), évaluer les distorsions dues à l'acquisition et à la reconstruction de l'image, évaluer les erreurs de l'extraction des primitives, avant de considérer les erreurs dues au recalage. Ceci est généralement impossible à faire en pratique, quand tout cela dépend en plus de la forme de l'objet et du réglage de l'appareil d'acquisition.

L'algorithme que nous avons présenté à la section (8.5) s'inscrit dans ce cadre, mais ne modélise l'erreur que dans les traitements haut niveau sur les primitives, en cherchant à déterminer à posteriori un modèle de bruit sur celles-ci regroupant toutes les sources d'erreur des traitements bas niveau. La qualité de l'estimation finale de l'incertitude dépend évidemment de l'adéquation du modèle de bruit supposé aux données (voir aussi la section (9.3.3)).

Nous proposons ici une méthode statistique qui cherche à *mesurer* l'incertitude sur le recalage entièrement a posteriori, qui peut donc s'appliquer à n'importe quel algorithme de recalage, mais qui nécessite un grand nombre de recalages sur des données très homogènes.

9.1.1.1 Estimation *a posteriori* des erreurs sur le recalage

Considérons un algorithme de recalage comme une « boite noire » qui prend en entrée une liste d'appariements de primitives et qui donne une estimation du mouvement en sortie. En supposant l'indépendance des paires de primitives appariées, on peut séparer la liste des appariements en plusieurs sous-listes pour obtenir plusieurs estimées du même mouvement inconnu. Par exemple, on peut séparer les appariements en deux listes de taille approximativement égales et déterminer deux estimations indépendantes de f :

$$\hat{\mathrm{f}}_1 = \mathrm{f} \circ \hat{\mathrm{e}}_1 \qquad \text{et} \qquad \hat{\mathrm{f}}_2 = \mathrm{f} \circ \hat{\mathrm{e}}_2$$

Comme on ne connaît pas le mouvement réel f, on étudie leur « différence » : le vecteur d'erreur

$$\hat{\vec{z}} = \hat{\vec{\mathrm{f}}}_2^{(-1)} \circ \hat{\vec{\mathrm{f}}}_1 = \hat{\vec{\mathrm{e}}}_2^{(-1)} \circ \hat{\vec{\mathrm{e}}}_1$$

qui ne dépend plus du mouvement exacte f et qui est centré autour de l'identité (le vecteur nul).

Si les appariements sont sous-échantillonés aléatoirement en deux listes de m paires, pour conserver l'indépendance et une distribution similaire dans l'espace, alors les erreurs d'estimation $\hat{\mathrm{e}}_1$ et $\hat{\mathrm{e}}_2$ suivent la même loi et ont en particulier la même covariance $\Sigma(m)$. En supposant que l'extraction des primitive et l'algorithme de recalage ne soient pas biaisés, ces erreurs sont de plus centrées : $\hat{\mathrm{e}}_1$ et $\hat{\mathrm{e}}_2$ sont donc des réalisation des primitives aléatoires $\mathbf{e}_1 \sim \mathbf{e}_2 \sim (\mathrm{Id}, \Sigma(m))$. On peut alors calculer que $\mathbf{e_2}^{(-1)} \sim (\mathrm{Id}, \Sigma(m))$, ce qui donne au final :

$$\mathbf{z} \sim (\mathrm{Id}, 2.\Sigma(m))$$

En répétant cette expérience avec n ensembles d'appariements différents (pour avoir des mesures indépendantes) mais homogènes (provenant du même type de données acquises dans les mêmes conditions), on peut estimer l'incertitude (à l'origine) sur l'estimation du mouvement f à partir de m appariements par la covariance empirique observée sur les n mesures $\hat{\vec{\mathrm{z}}}_i$ du vecteur d'erreur $\mathbf{z}$:

$$\Sigma(m) = \frac{1}{2} \int_{\mathcal{M}} \vec{\mathrm{z}}.\vec{\mathrm{z}}^{\mathrm{T}}.p_{\boldsymbol{z}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) \simeq \frac{1}{2.n} \sum_{i=1}^{n} \hat{\vec{\mathrm{z}}}_i.\hat{\vec{\mathrm{z}}}_i^{\mathrm{T}}$$

9.1.2 Validation de l'incertitude

Dans l'optique où nous avons une estimation de l'incertitude sur le recalage, la question de la validation se transforme de « a quelle distance notre estimation est-elle de la vérité ? » en « quelle est la validité de notre estimation de l'incertitude ? ». Le point difficile dans ce problème reste le même pour les deux questions : qu'est-ce que la vérité et comment la mesurer ? Nous pensons qu'aucune technique ne peut fournir la vérité exacte, mais que si une technique A a une incertitude sur le résultat nettement plus faible qu'une autre technique B, alors A peut servir de référence pour valider B. Cela signifie que toutes les techniques de validation sont en fin de compte statistiques.

Ainsi, dans les données synthétiques que nous utilisons à la section (9.2), la transformation est connue à la précision numérique de la machine, ce qui nous autorise à négliger son incertitude face à celle du recalage. Par contre, il n'est pas évident que le modèle de bruit que l'on simule sur les primitives soit soit représentatif de la vérité : on peut donc simplement valider nos estimations dans le cadre des hypothèses synthétiques. Une technique intermédiaire entre cette validation synthétique et les données réelle utilise un « fantôme », c'est-à-dire un objet synthétique dont on connaît plus ou moins la forme et la position lors de l'acquisition. Le problème dans ces expériences est de pouvoir estimer l'incertitude que l'on a sur la position dite exacte.

Le même problème se pose lorsque l'on cherche à valider le recalage de données réelles basé sur l'image par rapport à à des marqueurs externes ajoutés sur « l'objet » (ou le patient). L'avantage de ces marqueurs, qui sont la plupart du temps des cadres stéréotaxiques, est qu'ils produisent dans l'image des primitives très visible que l'on peut physiquement discriminer et identifier facilement. On peut donc obtenir une mise en correspondance quasiment exacte à tous les coups. Par contre, en ce qui concerne le recalage, il ne faut pas oublier que la mesure de ces marqueurs est soumise à l'erreur de la même façon que les autres primitives et le mouvement calculé sur le cadre est donc entaché d'une certaine incertitude qu'il serait intéressant de connaître. Il semble que cette incertitude soit inférieure à celle des techniques de recalage basées sur l'image pour du recalage multi-modalité (voir par exemple l'étude réalisée dans (West et al., 1996)), bien que certains auteurs, par exemple (Van den Elsen, 1993), mettent ce jugement en doute dans le cas d'images médicales de haute résolution. Pour le recalage mono-modalité et mono-patient, il semble assuré aujourd'hui que les techniques basées sur l'image soient plus précises que les techniques utilisant des marqueurs externes. Il est alors illusoire de considérer ces dernières techniques comme la référence pour valider les premières.

On peut donc conclure qu'il n'existe pas de référence absolue, mais uniquement des références relatives par rapport auxquelles on peut valider d'autres techniques ayant une incertitude plus grande. Dès que l'on sort du cadre des expérimentations synthétiques, nous avons donc besoin d'une technique de validation permettant de vérifier *sans référence extérieure* si l'incertitude est correctement prédite ou estimée.

9.1.2.1 Validation *a posteriori* de l'incertitude sur le recalage

L'algorithme modulaire de recalage proposé à la section (8.5) produit en sortie une estimation probabiliste du mouvement : $\mathbf{f} \sim (\hat{\mathrm{f}}\ ,\ \Sigma_{\mathbf{ff}})$. Le but de cette section est de vérifier *a posteriori* que la primitive aléatoire $\mathbf{f}$ a bien pour moyenne le mouvement exact f avec la covariance prédite. Nous considérons tout d'abord le cas des données synthétiques où le mouvement exact (ou presque exact) est connu. Comme chaque expérience de recalage est basée sur un mouvement exact différent et produit une matrice de covariance différente, nous avons à « normaliser » nos résultats pour pouvoir comparer différentes réalisations de la même distribution.

Pour normaliser par rapport au mouvement exact, on calcule pour chaque recalage le vecteur

d'erreur $\hat{\vec{z}} = \vec{\mathrm{f}}^{(-1)} \circ \hat{\vec{\mathrm{f}}}$ (pour simplifier les notation, nous avons ici oublié l'indice relatif à l'expérience de validation). La distribution théorique de la transformation aléatoire dont provient cette mesure est, puisque f est exact : $\mathbf{z} \sim \left(\mathrm{Id}\,,\ J_L(\vec{\mathrm{f}}).\Sigma_{\mathbf{ff}}.J(\vec{\mathrm{f}})^{\mathrm{T}} \right)$, avec $J(\vec{\mathrm{f}}) = \partial\vec{\mathrm{z}}/\partial\vec{\mathrm{f}}$.

Un nouveau changement de variable est nécessaire pour normaliser cette distribution vis a vis de la covariance : avec l'hypothèse gaussienne, la distance de Mahalanobis μ^2 d'une réalisation $\hat{\mathrm{z}}$ du vecteur d'erreur $\mathbf{z}$ avec l'identité devrait distribuée comme un χ^2 à 6 degrés de libertés, la dimension du groupe des transformations rigides (on se situe ici dans le cadre de l'hypothèse Σ faible de la section (5.3) :

$$\mu^2 = \hat{\vec{z}}^{\mathrm{T}}.\Sigma_{\mathbf{zz}}^{(-1)}.\hat{\vec{z}} \quad \sim \quad \chi^2_6$$

On peut maintenant répéter cette expérience avec m paires d'images différentes pour obtenir m valeurs indépendantes μ_i^2 et vérifier qu'elles ont vraiment une distribution χ^2_6. Le test de Kolmogorov-Smirnov (Press et al., 1991) est très bien adapté pour faire cela (on l'appellera dorénavant le test K-S), mais comme il ne donne qu'une réponse binaire, nous vérifions également que la moyenne soit proche de 6 (ce qui est vérifié même en dehors de l'hypothèse Σ faible, simplement avec l'hypothèse d'une distribution gaussienne sur la variété d'après l'équation (5.10)) et que la variance soit proche de 12. On appelle **indice de validation** la valeur moyenne estimée de μ_i^2 :

$$I = \bar{\mu}^2 = \frac{1}{m}.\sum \mu_i^2$$

dont la variance est calculée par :

$$\sigma_I^2 = \frac{1}{m-1}.\sum (\mu_i^2 - \bar{\mu}^2)^2$$

Cet indice peut s'interpréter comme une indication de combien notre méthode d'estimation sous-estime ($I > 6$) ou surestime ($I < 6$) l'erreur sur le mouvement. C'est en quelque sorte une erreur relative sur l'incertitude.

On peut facilement généraliser cette méthode pour valider notre algorithme de recalage sur des données réelles : on sépare aléatoirement notre liste d'appariements en deux ensembles (approximativement de même taille), puis on calcule deux estimées indépendantes $\mathbf{f}_1 \sim (\hat{\mathrm{f}}_1, \Sigma_{\mathbf{11}})$ et $\mathbf{f}_2 \sim (\hat{\mathrm{f}}_2, \Sigma_{\mathbf{22}})$. On utilise alors la distance de Mahalanobis $\mu^2(\mathbf{f}_1, \mathbf{f}_2)$ exactement comme ci-dessus et on peut fusionner les deux transformations (section 8.3) pour récupérer une seule transformation plus précise que l'on peut utiliser maintenant dans d'autres algorithmes.

9.1.3 Une mesure simplifiée de l'incertitude sur le recalage

La matrice de covariance sur le mouvement rigide est une information souvent trop riche pour l'utilisateur final du système de traitement d'images médicales. Cette matrice doit bien sûr être conservée si l'on veut utiliser le recalage obtenu dans d'autres algorithmes, mais elle doit être simplifié pour pouvoir donner une idée simple et intuitive de la qualité du recalage à l'utilisateur final.

A partir de $\mathbf{f} \sim (\hat{\mathrm{f}}, \Sigma_{\mathbf{ff}})$, on peut calculer pour chaque point de l'image ou de l'objet de départ sa position finale ainsi que son incertitude *dûe au recalage* (on ne considère pas ici l'incertitude dûe à l'extraction de la primitive elle-même) : $\boldsymbol{y} = \mathbf{f} \star x$. Pour simplifier encore l'information, on calcule l'écart-type (ou RMS pour root mean square) en ce point : si $y = \mathrm{f} \star x$ est la position exacte du point (exact) x, on le trouvera en fait en $\hat{y} = \hat{\mathrm{f}} \star x$ à cause de l'erreur sur la transformation. La distance moyenne du point mesuré avec le point exact est donc :

$$\sigma(\hat{y}) = \sqrt{E\left((\hat{y}-y)^{\mathrm{T}}.(\hat{y}-y)\right)} = \sqrt{\mathrm{Tr}(\Sigma_{\boldsymbol{yy}})}$$

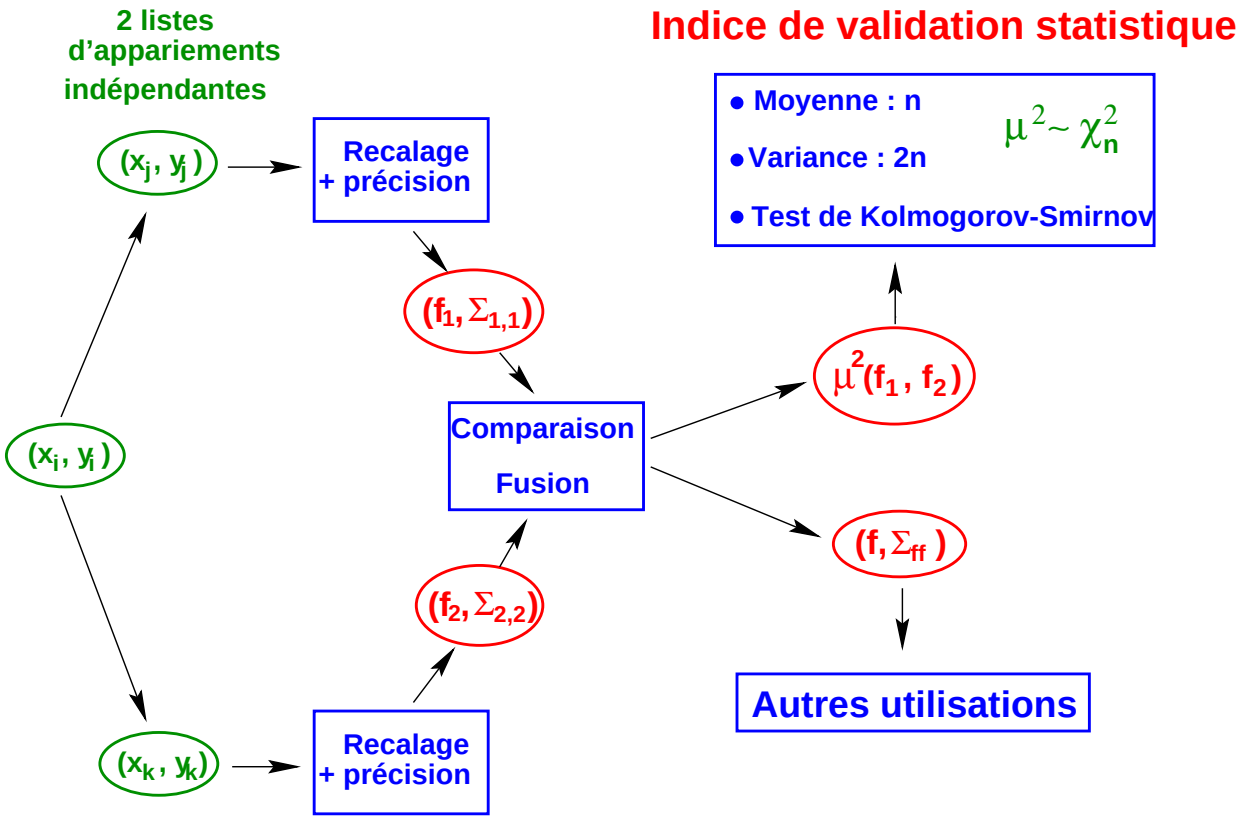


FIG. 9.1 – *Validation du recalage sur des données réelles.*

Dans le cas de l'imagerie médicale, nous proposons de caractériser la précision du recalage par deux mesures simples : **la précision moyenne sur la frontière** (la moyenne de l'écart-type sur les 8 coins de l'image), qui donne une idée de l'erreur maximale due au recalage dans l'image, et **la précision moyenne sur l'objet** (l'écart-type moyen sur les points ou les primitives recalées). Notons que, dans la précision moyenne sur l'objet, ce n'est pas la distance moyenne entre les points après recalage que nous mesurons, mais l'incertitude prévue sur ces points dûe à l'erreur sur la transformation. Nous n'utilisons donc que le positionnement de ces primitives pour définir la zone d'intérêt sur laquelle nous voulons connaître l'influence de l'erreur du recalage. On pourrait également utiliser un autre ensemble de points dans la zone d'intérêt, dédié à cette vérification et n'ayant pas servi au recalage. C'est ce qui se fait en général pour pouvoir *mesurer* (et non pas prédire) l'erreur de recalage (West et al., 1996) (l'erreur moyenne porte alors le nom de « Target Registration Error »). Cela implique toutefois que ces points soient très précisément localisés et que l'erreur de mesure de ces points soit très inférieure à l'erreur de recalage.

9.2 Expériences sur des données synthétiques

Pour ces expériences, nous avons fabriqué des listes de primitives (points, repères semi, non ou complètement orientés) avec une erreur « gaussienne » sur la position et l'orientation et un placement aléatoire uniforme dans une image de 256 x 256 x 256 voxels d'un millimètre (la section (6.4.2) décrit la façon de réaliser ces tirages aléatoires). Le mouvement exacte est choisi aléatoirement, mais en restreignant la translation à un tiers de la taille de l'image pour conserver une superposition entre les

images. La matrice de covariance utilisée pour simuler le bruit de mesure est donnée par le résultat de l'analyse sur des images réelles, et est présentée à la section (9.3.3). Nous avons utilisé un χ^2 de 16 pour les primitives de type repère et de 8 pour les points, ce qui correspond à peu près à un intervalle de confiance de 97%.

Pour être encore plus proche des données réelles, nous avons conduit une série d'expériences où les primitives sont extraites d'une vraie image (et non plus réparties uniformément dans l'image) que l'on bouge globalement et dont on perturbe les mesures avec une erreur « gaussienne ». Ces expériences ont donné des résultats très similaires à ceux que nous décrivons dans la suite de cette section, indiquant ainsi que l'hypothèse de distribution uniforme des primitives dans l'image est réaliste pour les mesures de précision. Notons cependant que lorsque nous travaillons avec des points extrémaux dans une image IRM ou scanner, nous obtenons environ 3000 primitives dont 500 sont appariées très précisément. L'hypothèse de distribution uniforme ne serait sans doute plus applicable pour un faible nombre de primitives (par exemple de 3 à 10) provenant de marqueurs externes.

L'objectif de cette section est de valider l'algorithme de recalage modulaire présenté au chapitre précédent et nous devons pour cela étudier les limitations des différents modules en fonction des combinaison de « paramètres » utilisés.

9.2.1 Validation de l'incertitude sur le recalage

Nous observé au chapitre précédent que le point faible théorique de notre algorithme de recalage est l'estimation du bruit de mesure sur les primitives. Pour valider les autres modules de l'algorithme et en particulier l'estimation de l'incertitude sur le recalage (en connaissant le bruit de mesure des primitives), nous avons conduit une série d'expériences en utilisant dans l'algorithme de recalage le modèle de bruit exact sur les primitives. Dans le cas réel, cela correspondrait à l'utilisation courante de l'algorithme de recalage pour une application déterminée, avec des images acquises de manière similaire : on peut alors supposer que le bruit de mesure des primitives a été estimé suffisamment correctement lors d'une phase d'apprentissage ou de mise au point et qu'il ne reste qu'à calculer le recalage et son incertitude. Cette phase d'apprentissage sera détaillée à la section (9.2.2.2).

9.2.2 Recalage avec une covariance exacte sur les primitives

Pour valider l'enchaînement d'algorithmes (LSQ - KAL - MAHA) permettant d'estimer la transformation et son incertitude, nous avons mené une série d'expériences en supposant la covariance sur les primitives connue de manière exacte. Nous présentons dans la figure (9.2) l'indice de validation (synthétique) $I = \overline{\mu}$ et les résultats du test de K-S pour les repères semi-orientés et les points. Ces statistiques sont réalisées avec 500 recalages synthétiques pour chaque mesure. Les résultats pour les repères et les repères non-orientés sont très similaires. Nous avons tracé la moyenne et l'écart-type de l'indice de validation, dont les valeurs théoriques pour une distribution χ^2_6 (6 et $\sqrt{12} = 3.46$) sont représentés par les lignes en pointillés. Le résultat du test de Kolmogorov-Smirnov est également tracé et le test accepte la distribution empirique comme un χ^2_6 dans tous les cas avec une confiance supérieure à 1%. Quelque soit le nombre et le type d'appariements utilisés, notre algorithme produit donc une estimation remarquablement précise de l'incertitude sur le recalage.

9.2.2.1 Introduction de mesures aberrantes

Pour tester la robustesse de notre algorithme, nous avons introduit (aléatoirement) dans les appariements synthétiques 5% de mesures aberrantes. Les résultats présentés dans la figure (9.3)

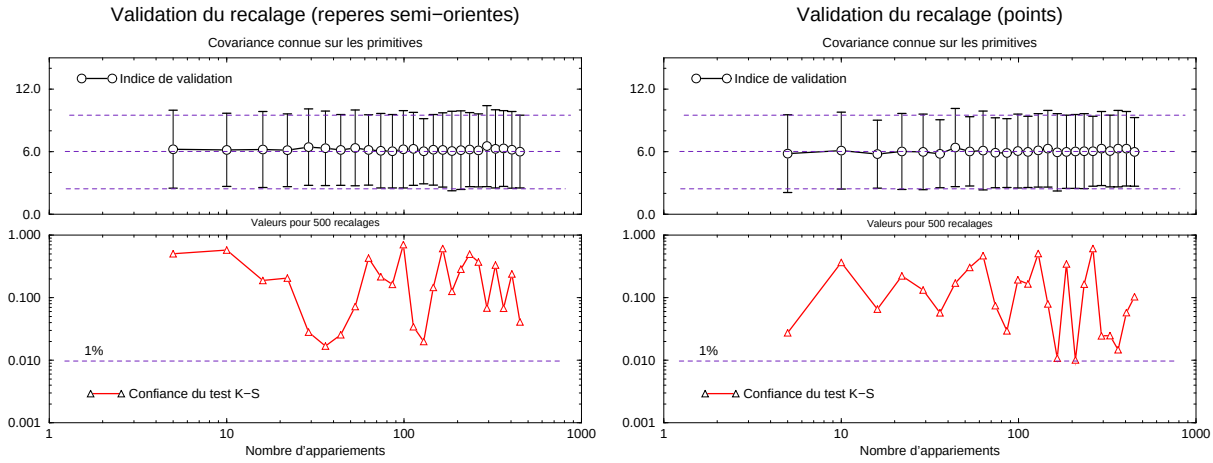


FIG. 9.2 – *Validation du recalage avec une une covariance connue sur les primitives. Haut : moyenne et variance de l'indice de validation synthétique en fonction du nombre de primitives appariés. Bas : valeur de confiance du test K-S. Celui-ci valide le recalage dans tous les cas avec une confiance supérieure à 1%.*

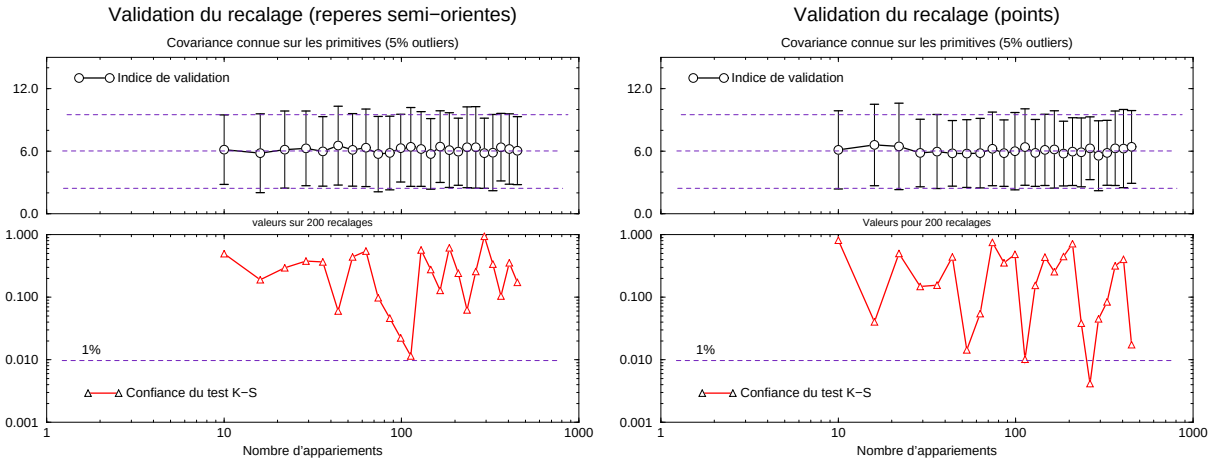


FIG. 9.3 – *Validation du recalage avec une une covariance connue sur les primitives et 5% de mesures aberrantes. Haut : moyenne et variance de l'indice de validation synthétique en fonction du nombre de primitives appariés. Bas : valeur de confiance du test K-S. Celui-ci valide le recalage dans tous les cas avec une confiance supérieure à 1%.*

sont ici réalisés avec 200 recalages synthétiques : l'algorithme est encore validé.

9.2.2.2 Apprentissage du modèle de bruit sur les primitives

Dans une application réelle, les données proviennent la plupart du temps du même type d'appareillage et l'on cherche toujours à mesurer le même type de paramètres. Il est alors utile d'avoir une estimation fiable du modèle de bruit sur les primitives et de ne pas chercher à le réestimer à chaque fois. Cependant, il nous faut quand même ajuster ce modèle de bruit pour qu'il soit le plus précis possible. Nous proposons pour cela de mettre à jours l'estimation du bruit au cours des premières expériences (phase d'apprentissage), jusqu'à ce que l'estimation soit suffisamment fiable (il faut pour cela un minimum de 1000 à 10000 appariements d'après ce que l'on a pu observer).

On peut ainsi utiliser l'estimation courante du modèle de bruit à chaque expérience et vérifier de plus que ce modèle décrit bien le bruit sur les primitives dans le recalage en cours : en calculant les distances de Mahalanobis entre primitives appariées (munies de la covariance du modèle de bruit courant), on peut vérifier grâce au test de Kolmogorov-Smirnov que cette distribution est compatible avec un χ^2 à d degrés de libertés (où d est la dimension des primitives). Si la distribution est compatible (confiance du test K-S supérieure à 0.01), on peut mettre à jour notre estimation du modèle de bruit et prédire une incertitude fiable. Si la distribution n'est pas compatible, c'est que les données ne proviennent vraisemblablement pas du même appareillage d'acquisition ou que les paramètres ont changés. Il convient alors de relancer une nouvelle phase d'apprentissage du modèle de bruit avec ce nouveau type de données.

9.2.3 Modèle de bruit anisotrope ou simplifié

Maintenant que notre algorithme de recalage et d'estimation de la transformation est validé lorsque l'on connaît le bruit sur les données, il nous reste à évaluer son comportement lorsque l'on inclut l'évaluation de ce bruit dans l'algorithme, ce qui correspond à la phase d'apprentissage évoquée ci-dessus. Nous avons présenté à la section (8.4) plusieurs modèles de bruits et en particulier un modèle isotrope et anisotrope (homogène) sur les points, ainsi que leurs équivalents pour les primitives de type repère, que nous appellerons modèle de bruit complet (homogène et isotrope sur les repères : la matrice de covariance n'est pas contrainte) et simplifié (la matrice de covariance est contrainte à être diagonale avec une variance unique pour la rotation et de même pour la translation). La question que l'on se pose ici est de savoir quel est le « meilleur » modèle de bruit à utiliser, à la fois en terme de précision et de validation de l'estimation d'incertitude sur le recalage.

Il paraît évident que le paramètre le plus influent pour le choix du modèle de bruit est le nombre d'appariements. En effet, l'estimation des 21 paramètres de la matrice de covariance pour le bruit complet sur les repères (resp. 6 pour le bruit anisotrope sur les points) est moins stable que l'estimation des 2 paramètres du modèle simplifié (resp. 1 pour le bruit isotrope sur les points). Nous aurons donc intérêt à choisir le modèle simplifié pour un faible nombre d'appariements et le modèle plus complet lorsqu'il y en a suffisamment.

Pour déterminer ce seuil sur le nombre d'appariements et le domaine de validation de chacun des modèles de bruit, nous présentons dans les figures (9.5) et (9.4) l'indice de validation et les précisions moyennes obtenus sur le recalage, en indiquant également les estimations non validées par le test de Kolmogorov-Smirnov.

Comme les modèles de bruits complets (ou anisotropes) apportent un gain de précision de l'ordre de 25 % pour les repères et de 15 à 40 % pour les points, il convient d'utiliser ces modèles de bruits aussi souvent que possible. Par contre, il ne sont pas validés avec moins de 30 à 40 appariements pour les repères et 15 à 20 pour les points. A partir de maintenant, nous utiliserons donc le modèle

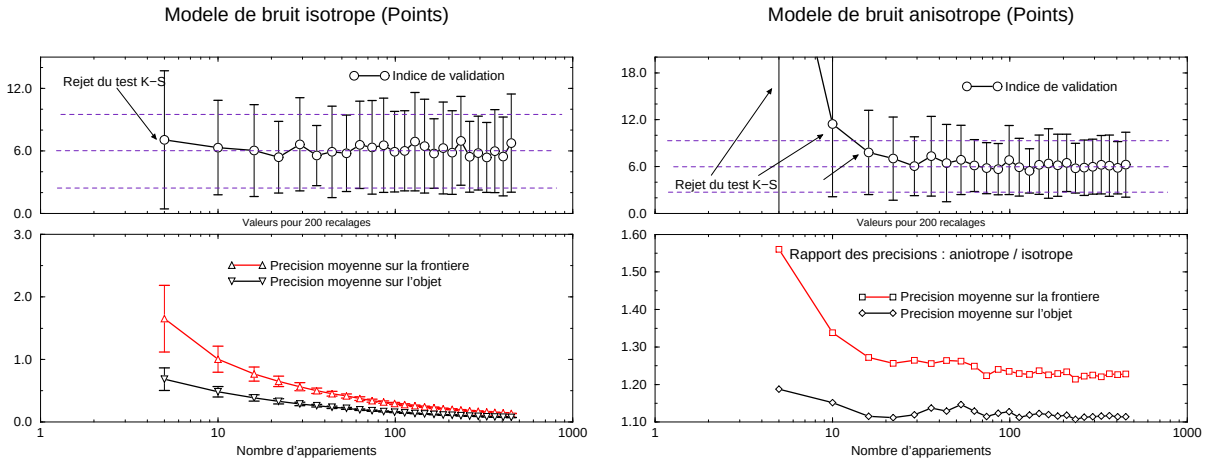


FIG. 9.4 – *Indice de validation en fonction du nombre d'appariements pour le modèle de bruit anisotrope (en haut à droite) et isotrope (en haut à gauche) sur les points. Les précision moyennes sont présentées pour le cas isotrope (en bas à gauche), la figure en bas à droite présentant le rapport des précision anisotropes / isotropes, c'est-à-dire le gain obtenu sur la précision en utilisant le modèle de bruit anisotrope.*

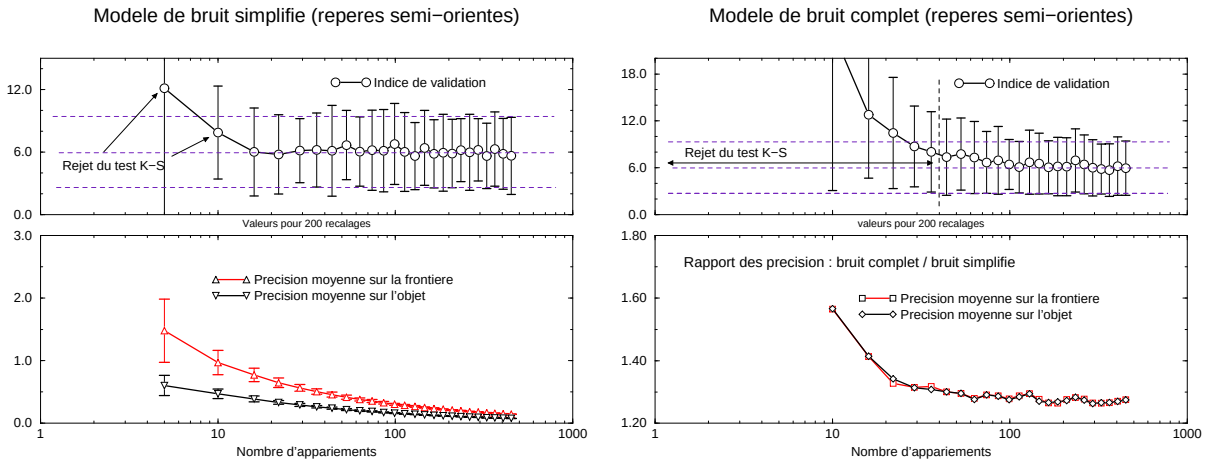


FIG. 9.5 – *Indice de validation en fonction du nombre d'appariements pour le modèle de bruit complet(en haut à droite) et simplifié (en haut à gauche) sur les repères semi-orientés. Les précision moyennes sont présentées pour le cas isotrope (en bas à gauche), la figure en bas à droite présentant le rapport des précision, c'est-à-dire le gain obtenu sur la précision en utilisant le modèle de bruit complet.*

de bruit simplifié s'il y a moins de 40 repères (resp. 20 points) et le modèle de bruit complet au delà.

9.2.4 Comparaison de la précision du recalage basé sur les points et les repères

Enfin, pour finir ces expériences synthétiques, il est intéressant de connaître le gain apporté sur la précision du recalage par l'utilisation de primitives de type repères au lieu de simples points. Comme nous ajoutons un trièdre à un point pour former un repère, on s'attend à être au moins aussi précis avec ceux-ci qu'avec les points, mais, en contrepartie, l'extraction de ce trièdre dans les images repose souvent sur des dérivées d'ordre supérieur, ce qui signifie que ce trièdre est en général plus bruité que la position du point. On ne s'attend donc pas à gain énorme sur la précision du recalage, sauf si la localisation des points devient significativement plus bruitée que celle des trièdres.

Nous avons déjà effectué une comparaison similaire à la section (8.2.3.1) pour comparer la précision relative des méthodes de recalage KAL, MAHA et LSQ, dont le résultat est (à cause de la métrique utilisée) très proche de QUAT et donc de la précision obtenue à partir d'appariements de points. De cette expérimentation, on peut conclure que le nombre d'appariements n'est pas un paramètre majeur dans cette comparaison. Par contre, le rapport entre l'incertitude sur le trièdre et celle sur le point (l'origine du trièdre) détermine le gain de précision.

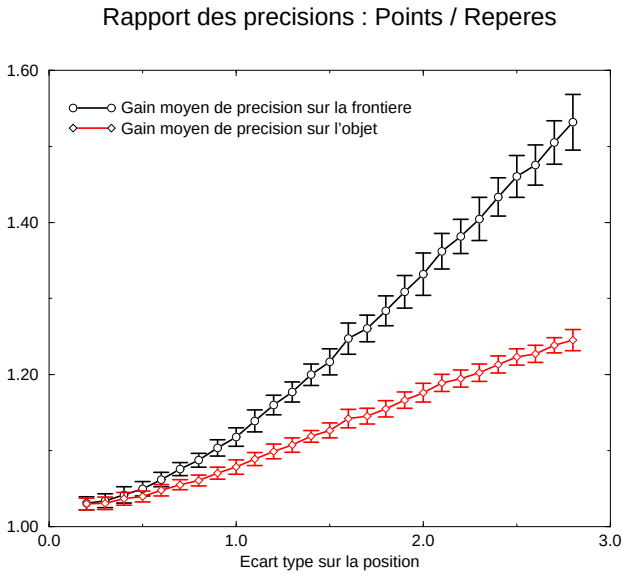


FIG. 9.6 – *Rapport de la précision prédite sur la frontière et sur l'objet en fonction de l'écart-type sur le position de la primitive (le point est l'origine du repère). L'écart-type sur la variance du repère est fixée. Les valeurs présentées sont la moyenne sur 200 recalages.*

Nous avons donc choisi pour chaque recalage entre 150 et 250 appariements de primitives (repères semi-orientés) bruitées avec un écart-type fixé de $\sigma_\theta = 0.02\,rad$ sur l'orientation et un écart-type variant de $\sigma_d = 0.2\,mm$ à $2.8\,mm$ sur la mesure de la position du repère (i.e. le point). Comme la précision prédite varie avec la configuration des appariements, nous avons effectué à chaque fois un recalage basé sur les repères et un recalage basé uniquement sur les points avec les même données et nous présentons dans la figure (9.6) le rapport entre les précisions moyennes à la frontières et

sur l'objet, c'est-à-dire le rapport des écart-types prévus *dûs au recalage* sur les points considérés de l'image. La valeurs présentés sont les moyennes et écart-types de ces rapports sur 50 recalages.

L'utilisation des trièdres en plus des points, pour former des repères, peut donc amener une augmentation sensible de la précision, en particulier avec un faible nombre de primitives ou lorsque la position devient très bruitée par rapport à l'orientation. Nous avons pu également observer un autre effet qui peut amener une augmentation nette de la précision pour les repères : c'est l'adéquation du modèle de bruit. En effet, un bruit « compositif » sur les repères (où le bruit sur la position est exprimé dans le repère local) est immanquablement interprété comme un bruit isotrope si l'on ne considère que les points car les repères sont orientés dans toutes les directions. On ne peut donc pas profiter de cette anisotropie pour améliorer le recalage et sa précision en n'utilisant que les points et le recalage en utilisant les repères peut être jusqu'à deux fois plus précis !

9.3 Recalage mono-patient d'images IRM 3D du cerveau

Les images utilisées dans cette section pour réaliser les expériences ont été fournies par le Dr R. Kikinis du *Brigham and Woman's Hospital* (Boston, USA) et font partie d'une étude extensive de l'évolution de la sclérose en plaque. Chaque patient subit une IRM (Imagerie par résonance magnétique) plusieurs fois par an (typiquement 24 acquisitions 3D différentes) et le but est de recaler très précisément en 3D toutes ces images pour segmenter et mesurer finement l'évolution des lésions de la sclérose en plaque. Ces lésion apparaissent comme des taches blanches dans la figure (9.7).

Les images sont des IRM T1 (premier écho), et ont $256 \times 256 \times 54$ voxels de taille 1 x 1 x 3 mm. (Thirion, 1994) a déjà présenté un algorithme pour recaler automatiquement de telles images mais le but est ici de déterminer la précision du recalage. En effet, il est visuellement impossible de déterminer « à l'oeil » un défaut de recalage occasionnant une erreur de superposition des images de l'ordre de la taille du voxel. Pour donner une idée du jugement visuel, nous présentons dans la figure (9.7) quatre coupes se correspondant après recalage dans quatre images différentes (ces coupes sont rééchantillonnées par interpolation tri-linéaire après transformation). Si ces coupes semblent se correspondre, il est difficile de juger de la qualité de cette correspondance. Nous présentons pour cela la différences de chacune de ces images avec la première dans la figure (9.8). Le gris représente une absence de différence, tandis que le blanc et le noir signifient des différences d'intensité marqués. Il apparaît que le cerveau est très bien recalé, mais qu'il y a eu de petits mouvements au niveau de la peau et entre le cerveau et la boite crânienne. On peut également visualiser très efficacement l'évolution des lésions : l'une d'elle, en bas à gauche croît au cours du temps, tandis qu'une autre, en haut à droite régresse. Cependant, la précision du recalage influe beaucoup sur les images de différences et on peut obtenir des mesures d'évolution des lésions aberrantes si le recalage est par trop imprécis. Il est donc très important de quantifier l'incertitude du recalage.

9.3.1 Points et repères dans les images médicales 3D

L'algorithme de mise en correspondance repose sur l'extraction de points particuliers dans les images 3D, définis à partir d'un critère de géométrie différentielle (voir figure 9.9). Dans notre cas, ce sont les **points extrémaux** de (Thirion et Gourdon, 1993), qui sont les points sur la surface considérée pour lesquels les deux courbures principales sont extrémales. Avec ces points, nous obtenons non seulement des invariants (les courbures principales), mais aussi les deux directions principales qui, associées à la normale a la surface, forment un trièdre semi-orienté, comme nous l'avons expliqué à la section (7.5) (voir en particulier la figure 7.1).

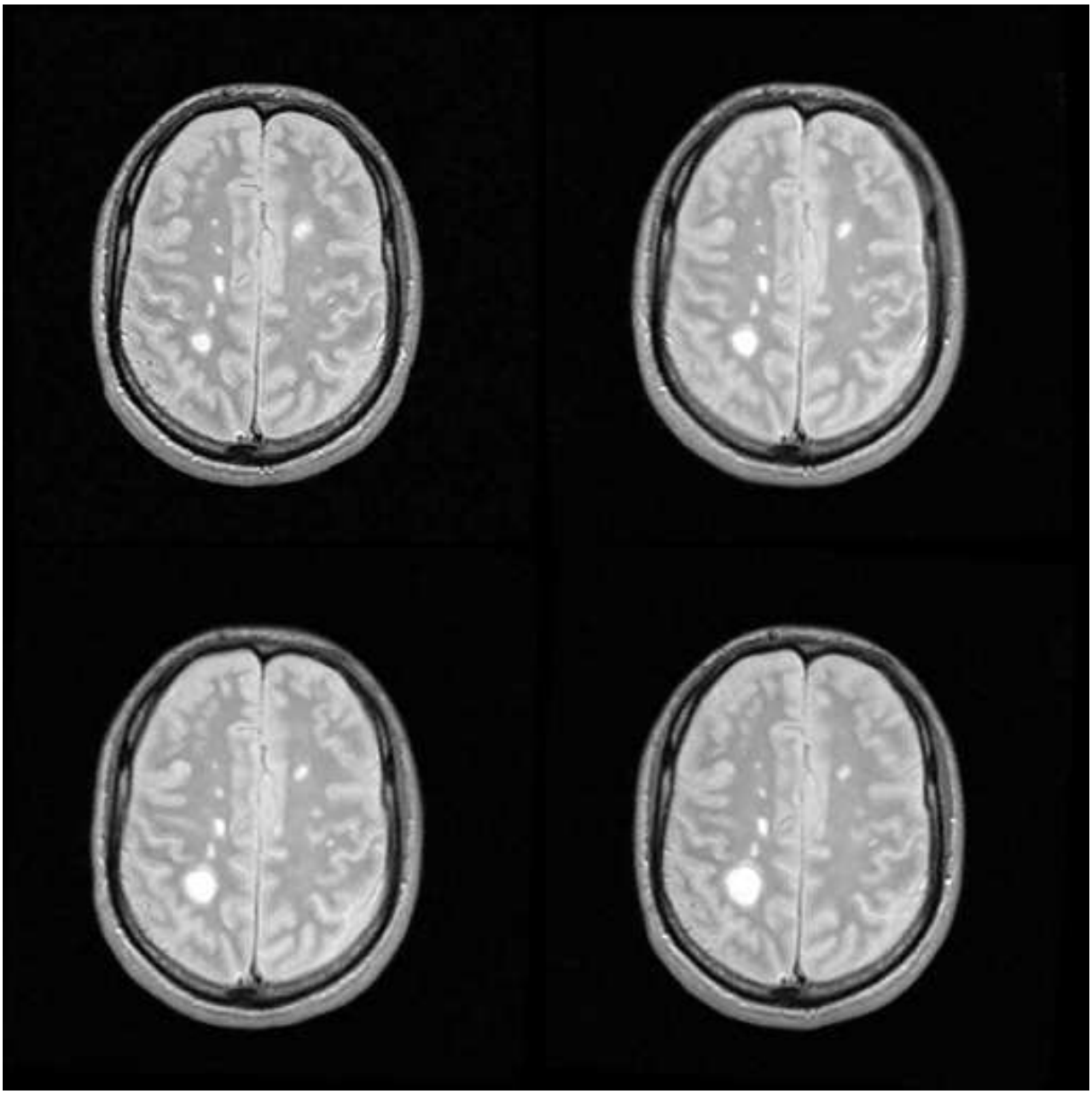


FIG. 9.7 – *La même coupe de trois images IRM 3D de l'étude de la sclérose en plaque après recalage. Il y a deux semaines entre chaque acquisition. Noter l'évolution de deux lésions (taches blanches) : l'une croît dans l'hémisphère antérieur gauche et l'autre diminue dans l'hémisphère postérieur droit.*

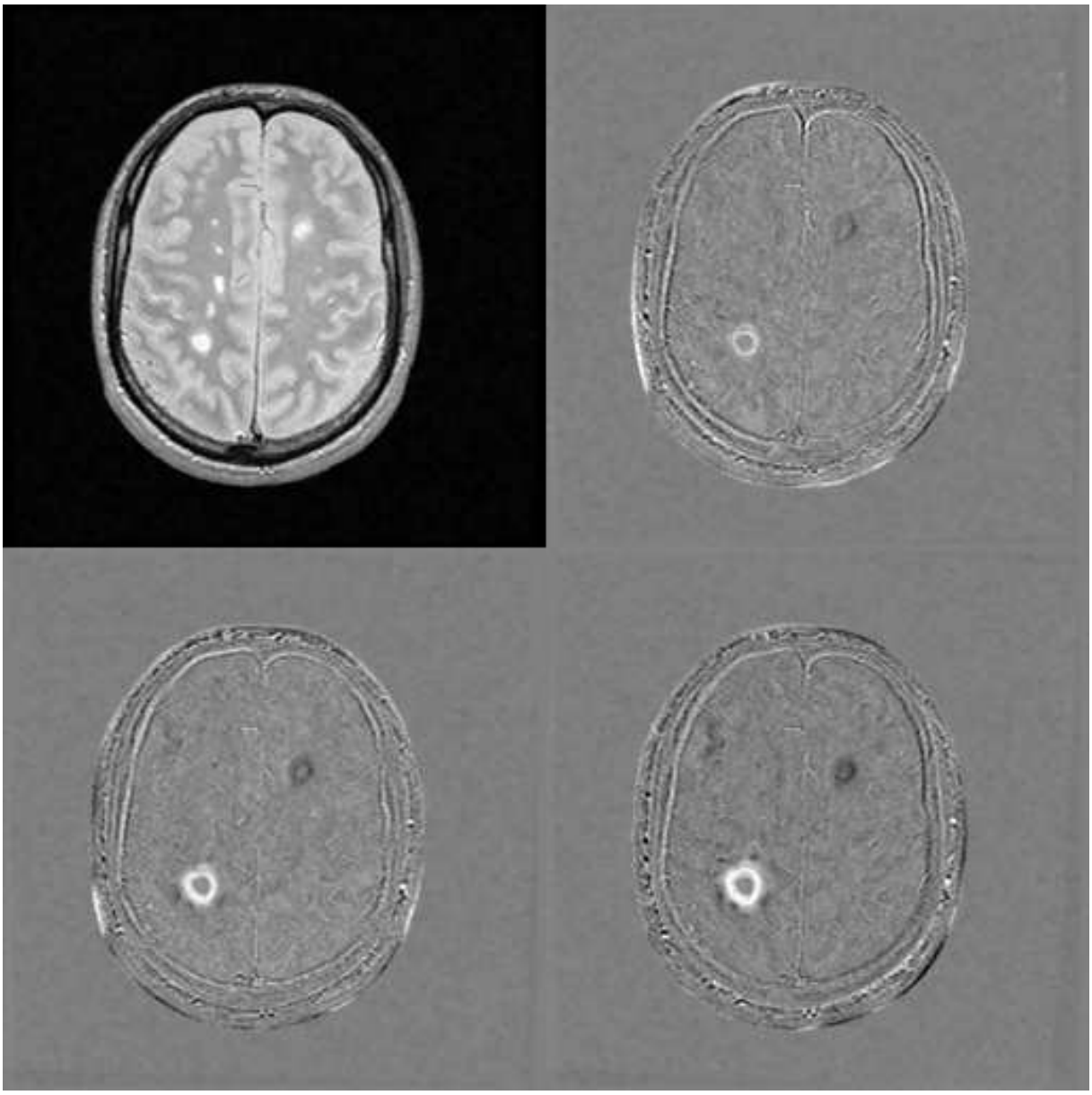


FIG. 9.8 – *Différence des images après recalage, par rapport à la première. L'intensité est multipliée par 5 et recentrée de telle sorte que l'absence de différence corresponde au gris. Les lésions croissantes apparaissent comme des disques blancs et les lésions régressives comme un un disque noir.*

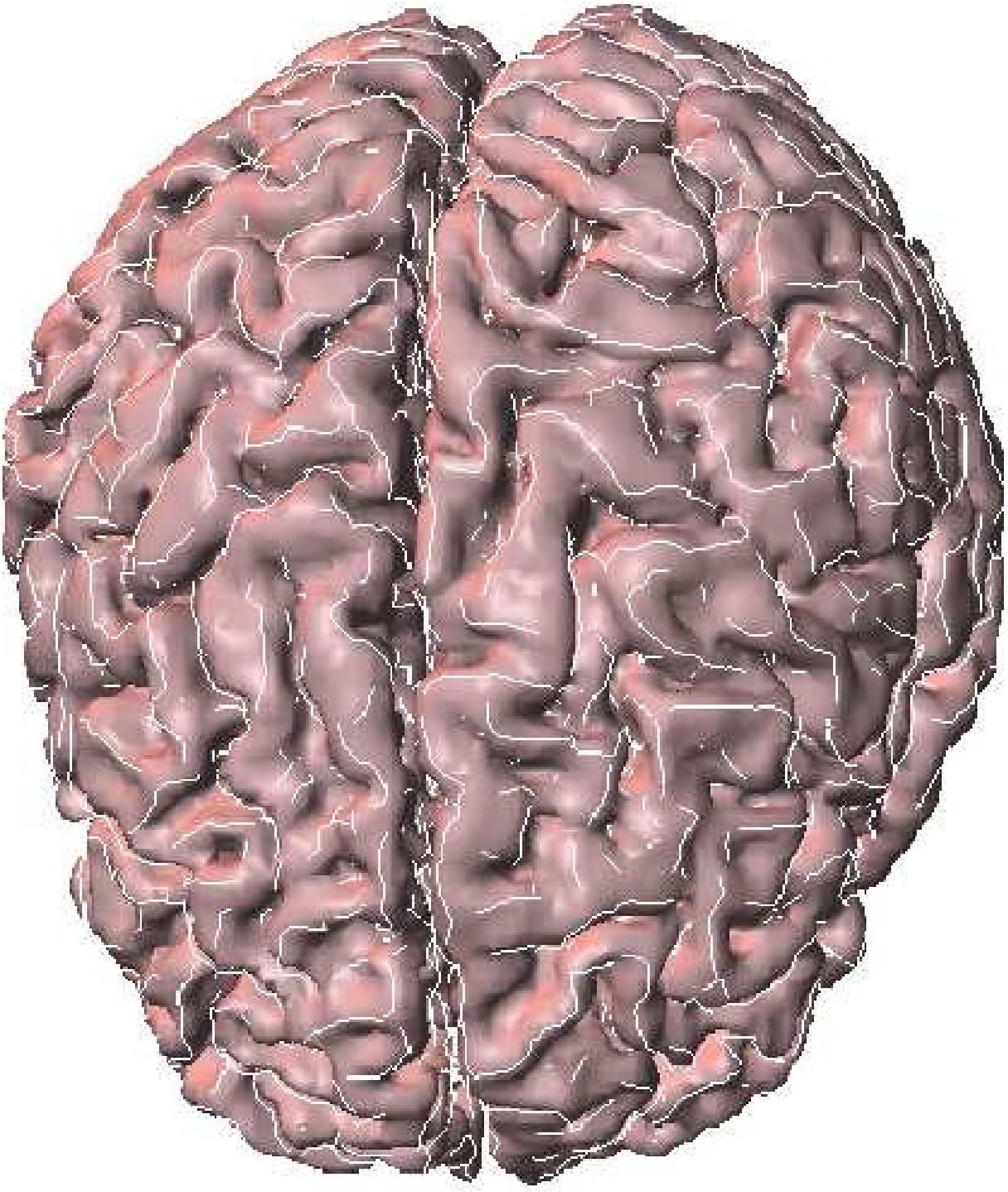


FIG. 9.9 – *Lignes de courbure extrémale sur les circonvolutions du cerveau. Les points extrémaux qui sont utilisées pour la mise en correspondance et le recalage sont des points spécifiques sur ces lignes.*

Dans une image IRM, on extrait typiquement 2000 à 3000 points extrémaux dans les 3.5 millions de voxels. L'algorithme de mise en correspondance détermine environ 600 appariements entre deux images d'un même patient, avec un écart-type résiduel (RMS) d'environ 1 mm. La « densité » de points extrémaux dans l'image est donc de 0.1 % dans l'image et la probabilité de faux appariements est très faible (voir section 10.3).

9.3.2 Résultats

Avec 24 images d'un même patient, nous aurions pu générer 576 couples d'images, dont seulement 23 sont indépendants. Pour avoir suffisamment de mesures de validation sans que le corrélation soit trop forte et que le temps de calcul ne soit prohibitif (l'extraction des points extrémaux prend environ 4mn CPU sur une DEC alpha workstation), nous avons utilisé 60 couples d'images tirés aléatoirement. Pour chaque recalage, on sépare (aléatoirement) la liste des appariements fournis par l'algorithme de mise en correspondance en deux listes de taille approximativement égales, à partir desquelles on calcule deux estimations de la même transformation, munies de leur incertitude. La distance de Mahalanobis entre ces deux transformation nous donne l'indice de validation réel μ_i^2. On fusionne alors les deux transformation pour obtenir l'estimation finale du recalage et de son incertitude, selon le schéma décrit à la figure (9.1).

Les estimations sont réalisées ici en modélisant les point extrémaux par des repères semi-orientés. Nous avons trouvé une **précision moyenne** sur la frontière de $\sigma_{fr} = 0.119\,mm$ et une **précision moyenne** sur l'objet de $\sigma_{obj} = 0.061\,mm$. L'indice de validation est de $I = \bar{\mu} = 6.33$ avec une variance de $\sigma_I^2 = 16.4$ (rappelons que les valeurs théoriques sont de 6 et 12). Le test K-S valide pleinement ces résultats avec une importance de 0.59. Les estimations sont réalisées ici en modélisant les point extrémaux par des repères semi-orientés et montrent une augmentation de précision de l'ordre de 20 % par rapport au recalage n'utilisant que les points, ce qui est en accord avec nos expérimentations sur les données synthétiques.

9.3.3 Analyse du modèle de bruit estimé sur les repères

Ces expériences ont également montré une différence intéressant entre les modèles de bruits estimés sur les points et sur les repères (semi-orientés). En effet, le modèle de bruit estimé sur les points est quasiment isotrope avec un écart-type de $\sigma = 0.5$, tandis que le modèle de bruit estimé sur les repères est présenté dans la figure (9.10). Rappelons qu'avec le modèle de bruit homogène ou compositif, l'erreur $\mathbf{f} = \mathrm{f} \circ \mathbf{e}$ sur le repère est en fait exprimée dans le repère local $\{x, t_1, t_2, n\}$, où t_1 et t_2 sont les directions principales de la surface au point extrémal x et n la normale.

$$\Sigma_{\mathbf{ee}} = \begin{array}{c} \\ \boldsymbol{e_r} \\ \\ \\ \boldsymbol{e_t} \\ \\ \end{array} \overset{\begin{array}{ccc|ccc} \multicolumn{3}{c}{\boldsymbol{e_r}^\top} & \multicolumn{3}{c}{\boldsymbol{e_t}^\top} \end{array}}{\left[\begin{array}{ccc|ccc} 0.0024 & 0.0000 & -0.0000 & 0.0002 & -0.0011 & 0.0000 \\ 0.0000 & 0.0030 & -0.0000 & 0.0001 & -0.0000 & 0.0000 \\ -0.0000 & -0.0000 & 0.0373 & -0.0006 & -0.0003 & 0.0001 \\ \hline 0.0002 & 0.0001 & -0.0006 & 0.2276 & 0.0011 & 0.0008 \\ -0.0011 & -0.0000 & -0.0003 & 0.0011 & 0.3157 & -0.0015 \\ 0.0000 & 0.0000 & 0.0001 & 0.0008 & -0.0015 & 0.0838 \end{array}\right]}$$

FIG. 9.10 – *Matrice de covariance du modèle de bruit estimé sur les points extrémaux considérés comme des repères semi-orientés.*

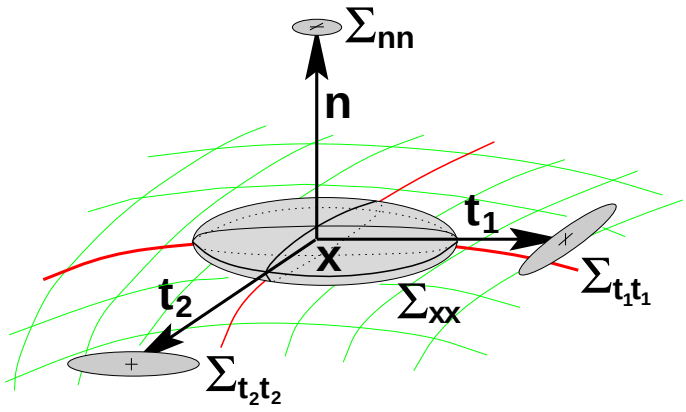


FIG. 9.11 – *Interprétation graphique du modèle de bruit estimé sur les points extrémaux. L'incertitude sur l'origine (le point x) est environ 2 fois plus grande dans le plan tangent que selon la normale et l'incertitude de la normale est isotrope, tandis que les directions principales t_1 et t_2 sont 4 fois plus incertaines dans le plan tangent.*

Cette covariance est grossièrement diagonale, avec des écart-types de $\sigma_{t_1} = 0.05$, $\sigma_{t_2} = 0.055$ et $\sigma_n = 0.2$ pour le vecteur rotation (en radians) et $\sigma_{x_{t_1}} = 0.5$, $\sigma_{x_{t_2}} = 0.55$ et $\sigma_{x_n} = 0.25$ pour la position (en mm), ce qui donne un écart-type moyen sur la position de $\sigma = 0.46$, comparable avec les valeurs mesurées par le modèle de bruit additif sur les points seulement.

En ce qui concerne le trièdre, on peut voir que l'erreur de rotation autour de la normale est environ 4 fois plus grande que l'erreur de rotation autour des direction principales, ce qui indique que la normale est l'axe le plus stable, les directions principales pouvant être beaucoup plus corrompues tout en restant dans le plan tangent. Pour la position, on observe que la coordonnée selon la normale (orthogonalement à l'iso-surface) est environ 2 fois plus stable que les deux autres coordonnées positionnant le point sur la surface. Ceci est tout à fait en accord avec ce que l'on attendait puisque l'extraction de la surface et de la normale ne font intervenir que des dérivées d'ordre 0 et 1 de l'image, alors que celle des directions principales demande des dérivées d'ordre 2 et 3, qui sont forcément plus instables.

L'utilisation d'un modèle de bruit adapté sur les repères nous a donc permis d'exhiber une information supplémentaire totalement invisible avec un modèle de bruit additif sur les points : l'absence d'orientation sur ceux-ci conduit effectivement à calculer un modèle de bruit qui est l'intégrale sur toutes les orientations possibles, ce qui conduit à un bruit quasiment isotrope que l'on observe effectivement avec un recalage où l'on n'utilise que les points. Cet effet constitue donc une justification *a posteriori* de notre modèle de bruit homogène (ou compositif) sur les repères et montre la validité de notre approche théorique rigoureuse dans le traitement de l'erreur sur les primitives géométriques.

9.3.4 Analyse de l'incertitude prédite sur la transformation

Nous avons également calculé l'incertitude moyenne prédite sur la transformation lors des 60 recalages : celle-ci est grossièrement diagonale et isotrope pour le vecteur rotation d'erreur, avec un écart-type de $\sigma_{r_i} = .00039$ sur chaque composante. Ceci signifie que l'erreur sur la rotation est de l'ordre de 0.00067 rad, soit 0.04 degrés ! L'erreur est ici exprimée dans le repère d'origine de l'image transformée lors du recalage. Les transformation mises en jeu dans ce type de recalage étant faibles (le patient a toujours une position approximativement identique dans l'IRM), elle correspond aussi à peu de choses près à l'erreur dans le repère de l'autre image. En ce qui concerne la translation,

nous pouvons observer un écart-type de $\sigma_x \simeq \sigma_y \simeq .055\, mm$ dans le plan des coupes et un écart-type $\sigma_z = .065\, mm$ dans l'axe d'empilement des coupes. Cette valeur un peu plus élevé reflète l'anisotropie de cette direction dans laquelle l'échantillonnage a un pas 3 fois plus grand.

9.3.5 Discussion

L'algorithme de mise en correspondance utilisé avant le recalage cherche à estimer la mise en correspondance impliquant le maximum de de points extrémaux communs et qui soit compatible avec le mouvement rigide d'une unique structure, le cerveau dans le cas présent. Cependant, avec ce niveau de précision à la fois sur l'extraction des points extrémaux et du recalage, cette hypothèse d'un unique mouvement ne tient plus et l'on peut distinguer plusieurs structures ayant des mouvements très proches. Le crâne, par exemple, peut bouger (quasiment rigidement) par rapport au cerveau, et la peau est sujette à des déformations importantes qui sont très visibles avec la séquence animée des 24 images d'une année recalées. Les correspondances de points extrémaux relatives à ces structures ont été ici rejetées comme aberrantes, mais on pourrait concevoir un algorithme de mise en correspondance permettant de gérer des mouvements multiples que l'on pourrait discriminer grâce aux outils développés dans ce manuscrit.

Même pour le cerveau, il y a des déformations locales, comme par exemple celles qui sont dûes aux lésions de la sclérose en plaques, qui sont significativement plus grandes que la précision moyenne sur la frontière ou sur l'objet que nous avons calculé. Cependant, comme nous sommes intéressés par le mouvement moyen du cerveau, l'obtention d'une telle précision sur le recalage conserve tout son sens et nous permet en fait de visualiser pour la première fois l'effet dynamique des lésions sur les tissus environnants.

9.4 Analyse du mouvement relatif des os du bassin

Le but de cette expérience est de mesurer quantitativement les mouvements relatifs des os du bassin lors de certains mouvements. Il s'agit en l'occurrence du bassin d'une danseuse et de mouvements spécifiques des jambes amenant en configuration dite « bassin ouvert » et « bassin fermé ». Une acquisition IRM (pondération T1) a été réalisée dans ces deux configurations à la Fondation Lenval (Nice). Une coupe de l'une de ces images est présentée dans la figure (9.12). Ces images ne sont pas d'une grande résolution puisque qu'elles sont constituées de 256 x 256 x 28 voxels de taille 1.33 x 1.33 x 5.6 mm, soit une anisotropie d'un facteur supérieur à 4 dans la direction orthogonale aux coupes.

Le bassin est principalement constitué, au niveau osseux, du sacrum et des os iliaques droit et gauche (figure 9.13), formant une structure quasi rigide, les fémurs étant libres de tourner dans la cavité prévue à cet effet dans les os iliaques. Les mouvements mis en jeu entre les os iliaques et le sacrum étant très faibles et parasités par le mouvement global du bassin entre les deux configurations, il est impossible de les mesurer directement sur l'image (contrairement au mouvement des fémurs). De plus, une telle mesure serait immanquablement trop imprécise pour décider s'il y a effectivement un mouvement. Il est donc indispensable de mesurer à la fois le mouvement relatif de ces os et son incertitude pour pouvoir décider statistiquement (grâce au test du χ^2) s'il y a un mouvement et de quelle ampleur.

9.4.1 Segmentation manuelle

Pour pouvoir mesurer les mouvements relatifs des os, il nous faut estimer plusieurs mouvements rigides entre les deux images. Malheureusement, l'algorithme de mise en correspondance précédem-

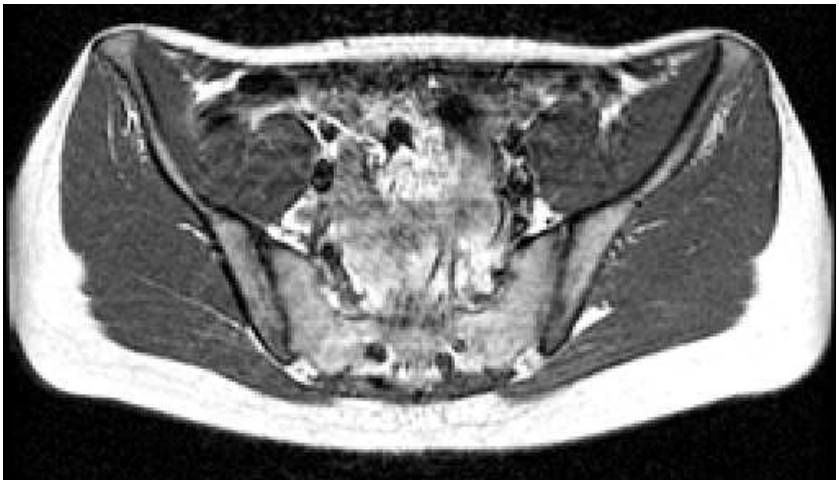


FIG. 9.12 – *Une coupe axiale d'une image IRM en pondération T1 du bassin. On peut distinguer au centre en bas le sacrum et les os iliaques gauche et droit sur les côtés.*

ment utilisé ne permet que d'estimer les correspondances de primitives relatives à un seul objet rigide. De plus, les techniques d'extractions des points extrémaux reposent sur une l'hypothèse que l'objet recherché dans l'image est délimité par une surface d'iso-intensité, ce qui est faux dans le cas des os pour une image IRM. En effet, ceux-ci apparaissent en signal intermédiaire (la moelle) avec simplement une bordure de signal plus faible (la corticale : voir la figure 9.12). Cette bordure étant très fine, elle disparaît par endroit à cause du bruit et de l'effet de volume partiel, ce qui la rend en particulier peu marquée au contact sacrum iliaques. Elle est également asymétrique (plus importante sur le côté gauche que sur le côté droit de l'os) à cause de l'effet de « déplacement chimique » (particularité de certaines acquisitions IRM). L'intensité de cette bordure sombre variant considérablement, il est donc impossible de déterminer une valeur adéquate pour délimiter les os par une surface d'iso-intensité.

Pour effectuer quand même des mesures, nous avons fait segmenter manuellement chacun des os par un spécialiste des images IRM (M le Pr. Dourthe) qui a délimité sur chaque coupe le contour des 5 os nous intéressant, produisant ainsi 5 images binaires. Nous avons ensuite rempli ces contours à l'aide d'opérations morphologiques et lissé ces images binaires par filtrage gaussien avant d'extraire la surface de chacun des os comme une surface d'iso-intensité. Nous avons visualisé et étiqueté le résultat de cette segmentation de l'une des images dans la figure (9.13).

9.4.2 Recalages

Dans les images lissés des os, nous avons extrait les lignes de crêtes et les points extrémaux, ce qui nous a permis de recaler chacun des os de la première configuration avec l'os correspondant de la seconde. Ces recalages sont toutefois très imprécis puisqu'ils ne font intervenir que 30 appariements pour les iliaques et 35 pour le sacrum. De plus, le bruit estimé sur les primitives est considérablement plus élevé que dans le cas du cerveau : l'algorithme estime une écart-type d'environ 25 degrés sur l'angle de la rotation résiduelle du trièdre et de 1 mm sur la distance résiduelle en position. Ceci conduit à des incertitudes élevés sur les paramètres de la transformation : un écart-type de 3.5 degrés sur la rotation et de 1.5 mm sur la translation en moyenne.

Nous présentons dans la figure (9.14) la superposition des bassins obtenue en utilisant le recalage du sacrum : on considère ainsi que le sacrum est fixé et on peut visualiser le mouvement des iliaques.

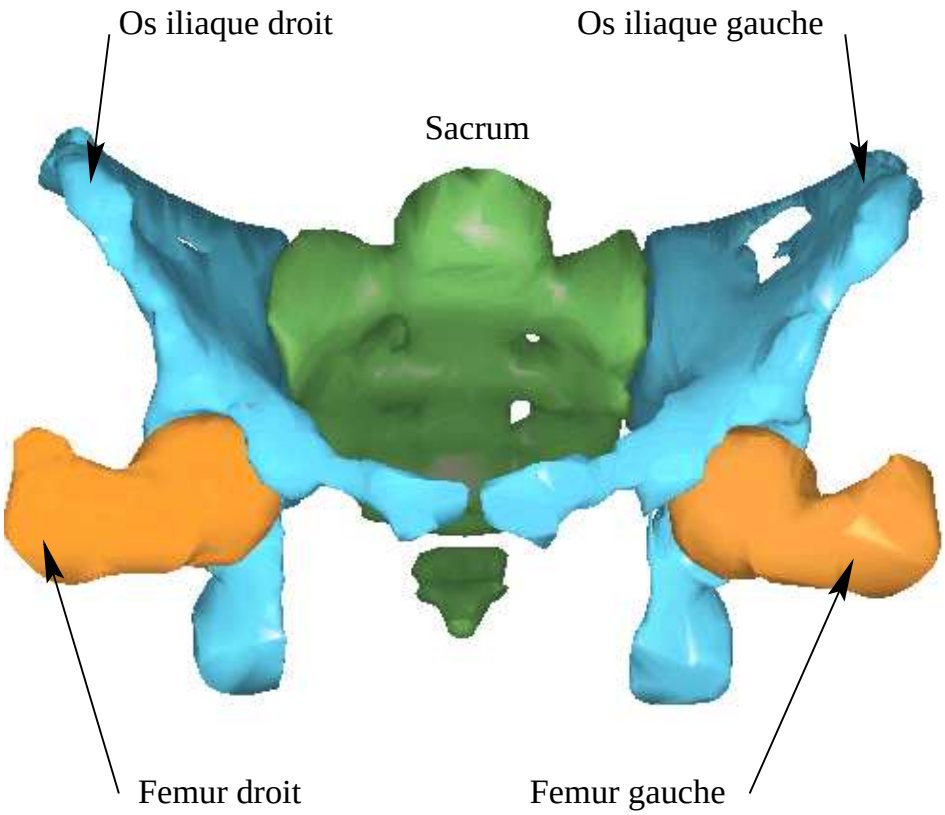


FIG. 9.13 – *Résultat de la segmentation manuelle de l'une des images IRM. On dispose de la surface de chaque os, que l'on peut étiqueter comme sur cette vue.*

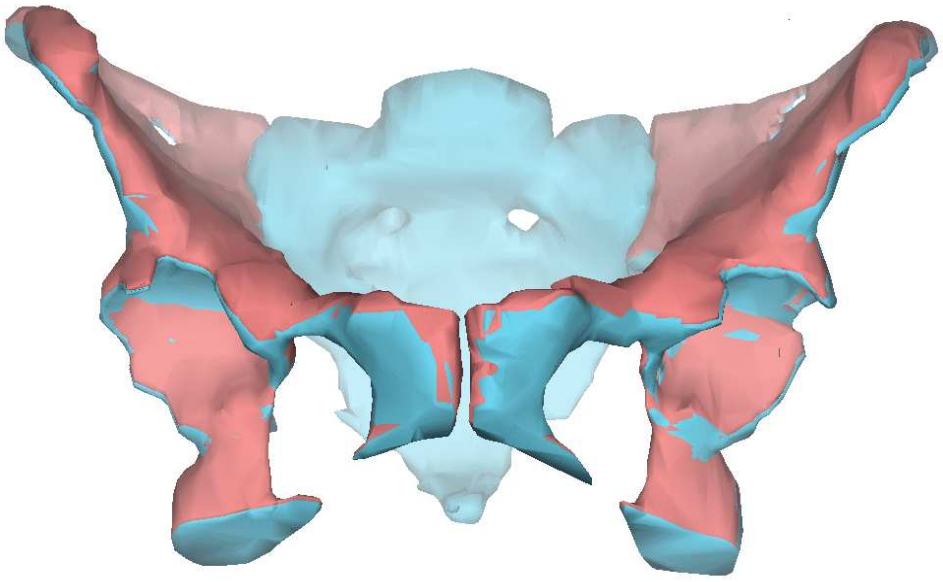


FIG. 9.14 – *Superposition des bassins en utilisant le recalage du sacrum : il semble que les iliaques en configuration « bassin ouvert » (en bleu) se soient légèrement écartés par rapport à la configuration « bassin fermé » (en rouge).*

Il semble effectivement sur cette figure que les iliaques en configuration « bassin ouvert » (en bleu) se soient légèrement écartés par rapport à la configuration « bassin fermé » (en rouge). En effet, lorsque l'on compare les transformation rigides obtenues pour les iliaques droits et gauches (f_d et f_g), on obtient une transformation résiduelle $e = f_d^{(-1)} \circ f_g$ composée d'une rotation d'environ 1.2 degrés autour de l'axe vertical et une translation de 3 à 4 mm qui tend à éloigner les iliaques l'un de l'autre et à les ramener en arrière du sacrum.

Cependant, lorsque l'on calcule l'incertitude sur toutes ces transformation, les variances s'ajoutent et on se rend compte que les transformations résiduelles entre les iliaques et le sacrum comme la transformation résiduelle entre les iliaques ne sont pas significatives : les distances de Mahalanobis sont respectivement de 16 et 18, ce qui est approximativement la limite d'un test du χ^2_6 à 99 %. On ne peut donc pas conclure qu'il y a eu effectivement un mouvement relatif et pourtant il est difficile de conclure que les transformation résiduelles ne sont dûes qu'aux erreurs de mesures à cause de la symétrie du mouvement relatif des iliaques.

9.4.3 Discussion

L'apport des statistiques sur l'incertitude du recalage dans cette expérience est indéniable puisqu'il nous permet d'éviter de conclure abusivement qu'il y a un mouvement relatif entre les os du bassin. Cependant, la précision est insuffisante pour pour que l'on puisse conclure dans l'autre sens que le bassin est strictement rigide.

Divers effet se conjuguent en fait pour donner au final cette incertitude très élevé sur nos transformations : la faible résolution de l'image et surtout sa forte anisotropie verticale (d'un facteur supérieur à 4) favorisent la délocalisation des amers géométriques. La segmentation manuelle ajoute sans soute un bruit non négligeable car les contours ont été tracés indépendemment dans chaque coupe avec une largeur de 1 a 2 pixels. De plus, il nous a fallu lisser abondamment la surface pour « effacer » les marches d'escalier dues à la forte épaisseur des coupes. Ceci a évidement fortement diminué le nombre de points de forte courbure sur ces surfaces, donc le nombre de primitives que nous avons pu utiliser pour les recalages ainsi que pour les statistiques de bruits. Avec 30 à 35 appariements, nous considérons ces dernières comme très incertaines.

Pour résoudre ces problèmes, il faudrait des images au moins deux fois moins anisotropes, mais aussi une méthode automatique pour extraire des primitives fiables sur les structures de type osseuses dans les images IRM. L'idée serait pour cela de remplacer le modélisation « iso-surface » de la frontière de l'objet par une modélisation « crête et vallées » (au sens géographique) dans le paysage d'intensité de l'image. Enfin, pour finir, un algorithme de reconnaissance de sous-structures, comme celui que nous présentons pour les protéines au chapitre 11, pourrait tenter de déterminer les ensembles d'appariements compatibles et ainsi déterminer automatiquement et indépendemment le mouvement de chacun des os. Cela nécessiterait cependant d'avoir suffisamment de primitives très bien localisée sur chacun des os et qu'il y ait un mouvement relatif entre ceux-ci. Dans le cas contraire, on trouvera une seule structure rigide rassemblant les trois os.

9.5 Conclusion

Nous avons montré dans cette section comment on pouvait valider l'estimation de l'incertitude sur le recalage à la fois sur des donnés synthétiques et des données réelles, pourvu que le nombre d'appariements soit suffisant. Il apparaît que le point le plus sensible dans l'algorithme modulaire de recalage présenté au chapitre précédent est l'estimation du bruit sur les primitives. Si cet élément est connu, notre recalage est validé par le test de Kolmogorov-Smirnov sur les données synthétiques

quels que soient les paramètres et le primitives. L'estimation de ce modèle de bruit lors du processus de recalage doit par contre se faire avec beaucoup d'attention et il est préférable de la réaliser lors d'une phase d'apprentissage. Nous avons proposé pour cela une méthode relativement générique.

Nous avons également validé cet algorithme de recalage sur des données réelles du cerveau et montré qu'une précision 10 fois inférieure à la taille du voxel peut être atteinte dans ce cas. Le modèle de bruit (homogène) estimé sur les points extrémaux a pu être interprété et correspond à ce que l'on attendait. Il exhibe de plus une « anisotropie » qui n'est pas décelable par un bruit additif sur les points et apporte ainsi un gain de précision de l'ordre de 20 % sur le recalage par rapport à l'utilisation des points sans les trièdres. Cela justifie a posteriori notre analyse rigoureuse de l'incertitude sur les primitives géométriques (en tous cas pour les primitives de type repère) et démontre son adéquation aux images médicales.

Enfin, pour finir, nous avons présenté une application où la vérification des incertitudes sur les transformations évite de conclure trop rapidement qu'il y a un mouvement relatif entre divers structures (en l'occurrence les os iliaques et le sacrum dans le bassin) et met en avant la sensibilité de la précision du recalage vis à vis des traitements d'image bas-niveau fournissant les primitives. Avec ce type d'outil, il devient d'ailleurs possible de comparer statistiquement a posteriori divers procédés d'extraction de primitives et de caractériser le domaine d'application où il est préférable de les utiliser. La dernière section a également stigmatisé le lien très fort qui unit les algorithmes de recalages et les algorithmes de mise en correspondance : il convient souvent de coupler les deux pour créer un algorithme de reconnaissance efficace. Nous nous intéressons à ce problème dans le chapitre suivant.

Chapitre 10

Mise en correspondance, reconnaissance et robustesse

« Vision: the art of seeing things invisible. »
Jonathan Swift

Le problème de la reconnaissance est sans doute l'un des plus étudiés en vision par ordinateur. Sauf exception, nous ne rentrons donc pas dans les détails mais nous essayons plutôt de donner une approche globale (et donc forcément incomplète) des méthodes utilisées. De plus, dans le cadre de ce manuscrit, nous ne nous intéressons qu'à la mise en correspondance de primitives de même type (typiquement 2D-2D ou 3D-3D), et nous excluons donc les nombreuses techniques de mise en correspondance 2D-3D développées en vision par ordinateur. Le support de la synthèse rapide présentée dans la première section provient principalement de (Ayache, 1989; Grimson, 1990; Brown, 1992; Zhang, 1993). D'un point de vue plus historique, le lecteur pourra également consulter (Besl et Jain, 1985; Chin et Dyer, 1986).

Nous nous intéressons ensuite à la gestion de l'erreur dans ces algorithmes, afin d'assurer que si un objet est présent, il sera bien reconnu. Nous autorisons pour cela une certaine tolérance sur les mesures, mais la présence même de cette zone d'erreur sur les primitives augmente sérieusement la probabilité de trouver par hasard un nombre fixé de primitives dans une configuration suffisamment proche dans les deux images pour que la reconnaissance soit acceptée. Nous nous intéressons dans la section (10.3) à caractériser la fréquence de ces faux positifs d'un point de vue qualitatif. Au passage, cela nous permet de confirmer que les primitives complexes (comme les repères) sont bien plus sélectives que les points seuls. Leur utilisation est donc souhaitable dans les algorithmes de mise en correspondance mais certains problèmes restent à résoudre pour généraliser ces algorithmes au cadre générique des primitives. Nous avons identifié quatre points clés, qui sont d'ordre théorique comme les invariants ou la classification (clustering), ou bien d'ordre algorithmique comme le plus proche voisin ou l'indexation incertaine.

10.1 Algorithmes de mise en correspondance

Selon le modèle de structuration de l'information présenté en introduction (section 1.1), les traitements bas niveau sur les images produisent des ensembles (non structurés) de primitives que nous appelons **scènes**. Nous avons vu au chapitre 8 diverses méthodes pour recaler des objets si l'on connaît les correspondances entre les primitives de deux scènes. Le problème que l'on se pose ici est de déterminer ces correspondances. En fait, ces deux problèmes sont la plupart du temps couplés dans la réalité pour former un problème de **reconnaissance** impliquant **la mise en correspondance** (l'appariement) des primitives constituant un objet et la localisation de celui-ci, c'est-à-dire le calcul de la transformation entre les deux instances de l'objet (**recalage**).

On dispose donc d'un ensemble de m primitives $\mathcal{X} = \{\mathrm{x}_1, \dots \mathrm{x}_m\} \in \mathcal{M}^m$ qui constitue la scène modélisant la première image et d'un ensemble de n primitives $\mathcal{Y} = \{\mathrm{y}_1, \dots \mathrm{y}_n\} \in \mathcal{M}^n$ modélisant la seconde. L'une des scènes peut également être le modèle structuré d'un objet, mais il suffit d'en oublier la structure pour traiter la mise en correspondance comme précédemment. Pour différencier les deux scènes nous utiliserons dans ce chapitre le vocabulaire hérité de la « reconnaissance à base de modèle » en vision par ordinateur, c'est-à-dire que nous appelons ici **modèle** l'ensemble des primitives $\mathcal{X}$ comme si celles-ci provenaient d'un modèle conservé dans une bibliothèque d'**objets**, et **scène** l'ensemble des $\mathcal{Y}$ extraites d'une image dans laquelle on cherche à reconnaître les objets précédemment modélisés.

Si l'on avait deux **acquisitions parfaites** d'un même objet unique (avec également des traitements bas niveau parfaits), le modèle et la scène auraient le même nombre de primitives, simplement transformées par une certaine transformation $\mathrm{f} \in \mathcal{G}$ et indicées dans un ordre différent. Nous n'aurions donc qu'une permutation des indices à trouver, et les primitives des deux scènes se correspondraient alors exactement à une transformation près. Dans la pratique, les conditions sont rarement aussi idéales : l'objet que l'on cherche à retrouver n'est pas forcément seul et il peut donc y avoir de nombreuses primitives supplémentaires dans les deux images qui ne doivent pas être appariés mais que l'on ne peut pas supprimer à priori (clutter). Par ailleurs, même si l'objet est connu parfaitement (grâce à un modèle CAO par exemple), son observation dans une image peut être suffisamment bruitée pour que certaines primitives soient inobservables. C'est ce que l'on appelle l'**occultation**, par référence à la vison par ordinateur, même si le phénomène conduisant à ce résultat est totalement différent dans les images volumiques. Enfin, pour finir, la mesure des primitives est évidemment bruitée et la superposition des primitives appariées après recalage ne sera jamais parfaite.

En général, on considère que l'on a **reconnu** un objet s'il existe une transformation $\mathrm{f} \in \mathcal{G}$ qui superpose (à une certaine erreur près) un nombre suffisant de primitives entre les deux images. On cherche donc à maximiser à la fois le nombre d'appariements entre le modèle et la scène et la qualité de ces appariements, c'est-à-dire l'adéquation de la superposition des primitives appariées par la transformation trouvée.

On appelle **fonction d'appariement** l'application π qui fait correspondre à tout indice i d'une primitives x_i du modèle $\mathcal{X}$ soit l'indice $j = \pi(i)$ d'une primitive y_j de la scène $\mathcal{Y}$, soit la primitive nulle $*$, auquel cas la primitive n'a pas de correspondant dans la scène (elle est occultée ou l'objet est absent). La fonction d'appariement π peut donc être considérée comme une application de l'ensemble $\mathcal{X}$ dans $\mathcal{Y} \cup \{*\}$ ou comme une application de $\{1 \dots m\}$ dans $\{1 \dots n+1\}$. L'ensemble Π de ces applications possède donc $(n+1)^m$ éléments. L'idée de base de la mise en correspondance est de maximiser le nombre d'appariements, mais ce problème seul est sous-contraint : on peut en effet apparier les m primitives du modèle à un seule primitive de la scène et obtenir ainsi un maximum d'appariements avec une solution aberrante. Il faut donc imposer des contraintes supplémentaires,

comme par exemple l'unicité ou la symétrie des appariements, mais ce n'est pas suffisant et il faut en général optimiser un critère qui rend compte de la vraisemblance de chacun des appariements après recalage : notons α cette fonction de vraisemblance. On prend en général $\alpha(\mathrm{z}, \mathrm{z}') = 1$ si $\mathrm{dist}(\mathrm{z}, \mathrm{z}') \leq \varepsilon$ et 0 sinon. La fonction d'appariement est alors définie comme celle qui maximise le **score de matching** :

$$\hat{\pi} = \arg\max_{\pi \in \Pi} \left(\sum_i \alpha \left(\mathrm{f} \star \mathrm{x}_i \, , \, \mathrm{y}_{\pi(i)} \right) \right) \tag{10.1}$$

Parallèlement, on veut optimiser la qualité des appariements et trouver la **transformation** qui envoie le modèle dans la scène, par exemple aux moindres carrés, avec la convention que $\mathrm{dist}(\mathrm{x}, *) = 0$. On cherche à minimiser l'erreur moyenne entre la transformée d'une primitive du modèle et son correspondant. La transformation retenue minimise alors :

$$\hat{\mathrm{f}} = \arg\min_{\mathrm{f} \in \mathcal{G}} \left(\sum_i \mathrm{dist} \left(\mathrm{f} \star \mathrm{x}_i \, , \, \mathrm{y}_{\pi(i)} \right) \right) \tag{10.2}$$

On voit donc que ce problème couple la recherche dans l'espace des correspondances (trouver $\pi \in \Pi$) avec la recherche dans l'espace des transformations (trouver $\mathrm{f} \in \mathcal{G}$). Chacun de ces deux sous-problèmes indépendamment est (relativement) simple, mais leur couplage rend la tâche beaucoup plus ardue et nécessite la définition d'un compromis. Nous rappelons rapidement dans les paragraphes suivants les principales techniques utilisées en traitement d'image et en vision par ordinateur pour résoudre ce problème.

10.1.1 Arbres d'interprétation

Examinons les équations posées : la recherche de la transformation f dépend de la solution π trouvée. Par contre, on peut rechercher π dans l'espace Π des correspondances et s'occuper à posteriori de f qui fournira le validation de l'interprétation. Cette idée est à la base de la recherche arborescente proposée par (Grimson, 1990). L'algorithme de base consiste donc à tester à chaque nœud de l'arbre l'appariement d'une primitive de $\mathcal{X}$ non encore utilisée avec chacune des primitives de $\mathcal{Y} \cup \{*\}$, puis lorsqu'on arrive à une feuille, on recherche s'il existe une transformation de $\mathcal{G}$ qui convient.

Diverses méthodes classiques de l'intelligence artificielle permettent d'élaguer la recherche dans cet arbre exponentiel ($(n+1)^m$ feuilles), en particulier l'introduction de contraintes géométriques sous forme d'**invariants unaires**. Supposons qu'on travaille avec des segments comme primitives : la longueur du segment vu dans la scène ne peut être qu'inférieure (à l'erreur près) à celle du segment modèle, puisque les seules modifications possibles sont les erreurs de mesure et l'occultation : c'est une contrainte unaire sur l'appariement. Supposons maintenant l'appariement de deux paires de segments vérifiant les contraintes unaires ; les angles entre les deux droites supports dans la scène et dans le modèle doivent être quasiment identiques : c'est une **contrainte d'invariance binaire**. On peut ainsi développer pour chaque type de primitives un ensemble de contraintes géométriques unaires, binaires et d'ordres supérieurs qui permettent de conserver lors de la descente dans l'arbre une consistance locale de l'interprétation, la recherche de la transformation fournissant en fin de compte et s'il y a lieu la preuve de la consistance globale. L'introduction de l'erreur s'effectue très simplement en propageant l'erreur de mesure sur les primitives dans le calcul des contraintes. Une étude complète de ces techniques est développée dans (Grimson, 1990). La complexité reste cependant $O((n+1)^m)$ dans le cas le pire.

10.1.2 Plus proche voisin itéré (ICP)

L'algorithme du plus proche voisin itéré, ou ICP pour « Iterative Clostest Point », a été introduit par (Besl et McKay, 1992) et (Zhang, 1994), et utilisé intensivement en imagerie médicale par (Feldmar, 1995). Cela consiste en l'optimisation alternée des appariements et de la transformation. A partir d'une transformation initiale f_o, on réalise les deux étapes suivantes :

- **Mise en correspondance** : PPV (plus proche voisin)
  On apparie chaque primitive x_i du modèle transformé avec la primitive y_j la plus proche dans la scène :
  $$\pi_t(j) = \arg\min_j \ \mathrm{dist}\,(\mathrm{f}_t \star \mathrm{x}_i, \mathrm{y}_j) \qquad \text{soit} \qquad \mathrm{y}_{\pi_t(j)} = \mathrm{PPV}(\mathrm{f}_t \star \mathrm{x}_i)$$
- **Recalage**
  La transformation est généralement calculée aux moindres carrés, surtout si l'on travaille avec des points : les solutions explicites de la section (8.1.1) ont l'avantage d'être rapides, ce qui n'est pas négligeable dans un algorithme itératif comme celui-ci. Si l'on possède une information d'incertitude, on peut l'utiliser dans les étapes terminales pour affiner la solution.

Le problème de la distance comme critère d'appariement est que la solution n'est généralement pas symétrique : on peut avoir $\mathrm{y}_j = \mathrm{PPV}(\mathrm{x}_i)$ et $\mathrm{x}_k = \mathrm{PPV}(\mathrm{y}_j)$ avec $k \neq i$. Pour résoudre ce problème, nous avons proposé dans (Pennec et Ayache, 1998) (voir aussi la figure 11.6) une version symétrique du plus proche voisin dans laquelle on apparie x_i à $\mathrm{y}_j = \mathrm{PPV}(\mathrm{x}_i)$ si et seulement si : $\mathrm{PPV}\big(\mathrm{PPV}(\mathrm{x}_i)\big) = \mathrm{x}_i$

Sous certaines conditions, on peut prouver la convergence de l'algorithme vers un minimum local (Feldmar, 1995) : il s'agit essentiellement de considérer les deux étapes comme des minimisations alternées *du même critère* selon deux ensembles de paramètres distincts, en l'occurrence la fonction d'appariement et la transformation. (Cohen, 1996) montre alors la convergence du processus.

Un problème générique pour ce type d'algorithme itératif est le critère de terminaison. Puisque nous ne nous intéressons ici qu'à des primitives identifiées et physiquement séparées dans l'espace (par opposition à des points répartis « continûment » sur une courbe ou une surface), l'ensemble des appariements possibles est discret et chaque fonction d'appariement donne lieu à une transformation unique. Lorsque l'algorithme a convergé, non seulement les appariements restent les mêmes lors de l'itération suivante, mais la transformation calculée est strictement identique. Il suffit donc d'arrêter l'algorithme lorsque les transformations sont identiques d'une itération sur l'autre ou lorsqu'on a atteint un nombre maximal d'itérations.

Le principal problème de cet algorithme est cependant sa sensibilité aux conditions initiales, particulièrement lorsque la partie commune au modèle et à la scène est faible par rapport au nombre total de primitives présentes (comme pour les protéines au chapitre 11). Il est alors impératif de démarrer avec une transformation ou un ensemble d'appariements très proche de la solution espérée. C'est pour cette raison que cet algorithme est souvent utilisé comme un algorithme de *vérification* à la fin du processus de reconnaissance, pour affiner les appariements et la transformation obtenus par l'une des autres techniques présentées dans cette section.

10.1.3 Transformée de Hough

La transformée de Hough a été introduite en 1962 par Paul Hough dans pour détecter des objets géométriques simples comme des droites à partir de points en accumulant des évidences dans leur espace paramétrique. Une bonne synthèse de cette optique est réalisée dans (Leavers, 1993). La méthode a été généralisée pour la mise en correspondance, par exemple dans (Stockman, 1987), de la manière suivante : soit k le nombre minimum d'appariements nécessaires pour calculer une

transformation unique (ou au moins un nombre fini) entre le modèle et la scène. Pour chacun de ces k-uplet d'appariements, on calcule cette transformation (si elle existe) et on accumule dans l'espace des transformations les évidences pour cerner la transformation qui permet d'apparier le maximum de primitives.

L'accumulation des évidences peut se faire par une discrétisation de l'espace des transformations et un vote des transformations calculées ci-dessus, une passe de balayage de l'accumulateur permettant à la fin de l'algorithme de détecter les meilleurs hypothèses de transformation. On peut également utiliser un algorithme de clustering dans l'espace des transformations sans discrétisation comme nous le ferons au chapitre 11.

Cet algorithme a donné lieu à beaucoup d'études, en particulier pour traiter les erreurs de mesure : celles-ci sont reprises dans (Grimson, 1990). La complexité est ici en $O(m^k.n^k)$, mais l'algorithme nécessite le stockage en mémoire de l'espace des transformations et son parcours en fin d'algorithme pour trouver le maximum. Si les transformations rigides en 2D n'occupent qu'un espace de dimension 3, on passe à une dimension 6 pour le 3D et à 9 pour les transformations affines 3D, ce qui peut donner lieu à des temps de calcul rédhibitoires si l'on ne gère pas efficacement cette recherche.

10.1.4 Alignement ou prédiction-vérification

La technique est un mélange de la transformée de Hough et de l'arbre d'interprétation : on commence par supposer un k-uplet d'appariements $(\mathrm{x}_i, \mathrm{y}_j)$ qui permet de calculer la transformation superposant ces primitives, puis on vérifie ces hypothèses en calculant les transformées des primitives du modèle et en recherchant dans une certaine zone autour de ces primitives prévues des primitives de la scène pouvant convenir. Le score de qualité de cette hypothèse est simplement le nombre d'appariements trouvés, ou peut être un critère plus élaboré. Cette technique a été principalement développée par (Ayache et Faugeras, 1986; Huttenlocher et Ullman, 1987; Huttenlocher et Ullman, 1990).

En théorie, on doit exécuter ce schéma pour tous les k-uplets d'appariements possibles et conserver les hypothèses de score maximal. Prenons l'exemple des points 3-D : il faut trois appariements pour spécifier une transformation rigide ($k = 3$). Il y a $C_m^3 = \binom{m}{3} = \frac{m!}{3!.(m-3)!}$ façons de choisir 3 points parmi les m du modèles, $C_n^3 = \binom{n}{3}$ possibilités pour la scène et 3! manières d'apparier les deux triplets. Nous avons donc $\binom{m}{3}.\binom{n}{3}.3! = O(m^3.n^3)$ alignements ou prédictions à vérifier. En pratique, on s'arrête dès qu'on estime avoir reconnu et placé le ou les objets communs aux deux images. De plus, dans le cas des points 3D, on peut utiliser les trois invariants du triplet (les distances inter-points) pour indexer le triplet dans une table de hachage, ce qui permettra de retrouver en temps presque constant les triplets compatibles dans l'image. La complexité est alors réduite à $O(m^3 + n^3)$ pour l'étape de prédiction.

L'étape de vérification est particulièrement importante puisqu'elle doit rejeter les mauvaises hypothèses mais conserver les bonnes. Il s'agit en général d'une (ou plusieurs) étapes de plus proche voisin itératif : on transforme les primitives du modèle et on les apparie aux primitives les plus proches dans la scène (avec cependant un seuil sur la distance). On peut alors utiliser ces nouveau appariements pour améliorer la précision de la transformation, et éventuellement itérer pour s'assurer que les appariements et la transformation sont stables. Ces itérations de raffinement sont importantes si l'on veut arrêter l'algorithme dès qu'il a trouvé *une* bonne solution. Si l'on teste toutes les prédictions, on peut se contenter de raffiner la ou les solutions les meilleures.

10.1.5 Hachage géométrique

Cette technique a été introduite par (Lamdan et Wolfson, 1988; Wolfson, 1990) et repose sur l'idée de pré-compiler les informations des modèles de la base d'objets dans une table de hachage avec une représentation invariante (vis-à-vis du groupe de transformation G considéré), redondante et basée sur des primitives locales pour permettre de les reconnaître malgré les occultations. Lors de la reconnaissance, nous n'avons alors qu'à calculer la représentation correspondante de la scène puis à accumuler les évidences pour l'appariement d'une partie de la scène avec un objet. Un exemple d'utilisation de cette technique est présenté à la section (11.3.1).

Plus précisément, le hachage géométrique s'intéresse au cas d'un sous-groupe des transformations affines agissant sur des points k-D : on peut alors définir une base $\mathcal{B}$ d'un modèle choisissant au plus $k+1$ points de celui-ci et exprimer les autres points dans cette base locale. Par exemple, deux points suffisent à définir une base orthonormée en dimension deux (en fait un point et une direction suffisent), ou trois points non colinéaires en dimension trois, mais il faut trois points exactement en dimension deux pour une base affine et quatre en dimension trois. Les coordonnées sont alors invariantes : si f est une transformation affine et x un point, les coordonnées de $\mathrm{f} \star x$ dans $\mathcal{B}' = \mathrm{f} \star \mathcal{B}$ et celles de x dans $\mathcal{B}$ sont identiques.

L'idée est d'indexer dans une table de hachage, lors d'une étape de pré-traitement, chaque « base » possible (i.e. chaque k-uplet du modèle) par les coordonnées des autres points du modèle dans cette base. On pourra consulter les figures (10.5) et (10.6) pour des exemple de tables de hachage. Lors de la reconnaissance, on choisit une base image et et on exprime les coordonnées des autres points image dans cette base. Ces coordonnées étant invariantes, on retrouve et on vote, au moyen de la table de hachage, pour les bases modèles possédant des points dans la même configuration (voir par exemple les figures 11.4 et 11.5). Si la base choisie dans la scène fait partie d'un objet, le nombre de votes obtenus pour la base modèle correspondante sera le nombre de points de l'objet présents dans la scène (à k près). La complexité est donc en $O(n^{k+1})$ dans le pire des cas pour la reconnaissance, si le temps d'accès à un bucket est constant. Ceci est obtenu en considérant n votes pour chacune des n^k bases possibles dans la scène. La vérification n'est pas prise en compte puisque dans l'algorithme normal, les meilleurs résultats sont directement retenus.

L'utilisation d'une table de hachage fournit un algorithme permettant une reconnaissance sous-linéaire en nombre de modèles lorsqu'on considère une bibliothèque d'objets grâce à l'indexation de tous ces objets dans une seule table de hachage (on ajoute alors dans les informations conservées lors de l'indexation quel est le modèle concerné).

10.1.6 Indexation d'invariants géométrique

La généralisation du hachage géométrique à des appariements de primitives géométriques autres que des points n'est pas complètement évidente, contrairement à la prédiction-vérification ou à la transformée de Hough. En effet, la notion de base et de coordonnées des autres points dans cette base est liée à la structure d'espace vectoriel. On peut toutefois concevoir un algorithme similaire mais faisant intervenir les invariants de plusieurs primitives à la place des coordonnées, l'accumulation des évidences se faisant toujours dans l'espace des appariements.

Prenons l'exemple des points 3-D : l'équivalent du hachage géométrique présentée ci-dessus serait d'utiliser dans la table de hachage les invariants caractéristiques de la forme constituée des quatre points ordonnés (x_i, x_j, x_k, x_l) au lieu des coordonnées de x_l dans une base construite à partir de (x_i, x_j, x_k). Au niveau de l'accumulation des évidences, il paraît alors naturel d'incrémenter les appariements possibles individuellement au lieu de le faire pour la base : si (y'_i, y'_j, y'_k, y'_l) un quadruplet de points compatible dans la scène, on votera alors pour chaque appariement (x_i, y'_i)

du quadruplet. Cette modification faite, il n'est plus réellement utile de considérer des quadruplets de points : on peut tout à fait réduire le nombre de points à trois, voire simplement à deux, ce qui permet de réduire considérablement la complexité : on passe pour la reconnaissance de $O(n^4)$ à $O(n^3)$ voire $O(n^2)$.

Il faut cependant faire attention que, dans le cas où l'on ne considère que les invariants d'un couple de points, la probabilité de faux positif devient vite très importante et les résultats ont une grande chance d'être aberrants (voir section 10.3). Si l'on considère un triplet de points, on retrouve un algorithme très proche de la transformée de Hough, sauf que l'accumulation n'a plus lieu dans l'espace des transformations mais dans celui des appariements. Il serait sans doute intéressant de combiner les deux approches, par exemple en regroupant les transformations compatibles associées à chacun des votes pour un appariement. Nous verrons au chapitre 11 une autre imbrication possible de ces deux algorithmes, qui fonctionne sur les repères et utilise les invariants binaires entre ceux-ci.

Ce type d'algorithme a été utilisé par (Fischer et al., 1992a; Fischer et al., 1992b) pour descendre la complexité du recalage basé sur les points de $O(n^4)$ à $O(n^3)$, l'accumulation se faisant ici sur l'appariement d'une pseudo-base constituée des deux premiers points de chaque triplet. Toutefois, le passage à d'autre types de primitives pose un problème de taille : celui du calcul des invariants. Nous détaillerons ce point à la section (10.4.3). Un autre exemple d'algorithme de ce type utilisant des primitives complexes (points munis de descripteurs de forme locaux) a été développé par (Califano et Mohan, 1991).

10.1.7 Isomorphisme de graphes

Cette technique considère les primitives d'un modèle comme les nœuds d'un graphe, étiquetés par les invariants unaires, et dont les arêtes sont munies d'un attribut qui caractérise la relation entre les deux primitives constituant les extrémités. Dans notre cas, on peut relier toutes les primitives d'un modèle entres elles et caractériser cette relation par les invariants binaires correspondants.

Au problème de la mise en correspondance de primitives se substitue alors la question suivante : quelle est la correspondance entre le graphe du modèle et le graphe de la scène ? Le problème principal de ce type de techniques est que l'occultation et la dispersion des primitives communes dans une scène plus dense, ainsi que le bruit, amènent à résoudre un problème beaucoup plus complexe : celui de *matching inexact de graphes*, que l'on sait NP-complet. Divers auteurs ont toutefois présenté des solutions approchés (Shapiro et Haralick, 1981; Eshera et Fu, 1986) contourné la complexité en réutilisant la géométrie (Davies, 1991). On peut également ramener la technique « Local Feature Focus » de (Bolles et Cain, 1982) dans ce formalisme.

L'avantage de cette technique est toutefois de pouvoir modéliser une information plus importante que la simple géométrie des primitives, par exemple en ne reliant dans le graphe que les primitives qui sont effectivement reliées par un « edge » ou, dans le cas de l'imagerie 3D, une ligne de crête. On pourrait ajouter également le type de ligne ou de surface reliant deux primitives dans l'étiquetage de l'arête ou toute autre information que l'on voudrait conserver dans la modélisation d'un objet et qui permet de simplifier grandement le problème de la reconnaissance.

En fait, il apparaît que cette technique est adaptée à une structuration haut-niveau de l'information, et donc à la reconnaissance des objets déjà modélisés ou d'objets articulés. Elle est par contre relativement inadaptée à la comparaison de scènes non structurées comme pour la reconnaissance de sous-structures dans les protéines que nous présentons au chapitre 11.

10.2 Gestion de l'erreur

Nous avons décrit jusqu'à présent les algorithmes théoriques qui correspondent au cas de mesures quasiment exactes. Dans la pratique, nous savons que toutes nos mesures sont sujettes à l'erreur (section 2.1). La présence de cette erreur nous oblige à reconsidérer les algorithmes de mise en correspondance, même dans le cas simple des points. Pour le hachage géométrique, par exemple, les erreurs de mesure dans la scène se répercutent sur le calcul des invariants et continuer à voter ponctuellement pour l'invariant trouvé (à la discrétisation près de la table de hachage) conduirait à manquer un grand nombre d'appariements; on obtiendrait alors des **faux négatifs**, c'est-à-dire qu'on ne reconnaîtrait pas un objet qui est présent. La même chose peut se produire avec les techniques d'alignement si la zone de recherche des correspondants n'est pas adaptée à l'erreur sur la localisation des primitives, augmentée de l'incertitude due à l'erreur sur l'estimation de la transformation.

10.2.1 Zones d'erreur et compatibilité

Afin de remédier à ces défauts, on supposera connue une estimation de l'erreur de mesure sur les primitives, soit sous forme d'une borne sur les valeurs, soit sous forme d'une loi de probabilité. L'approche de Grimson et Huttenlocher (Grimson et al., 1991; Grimson et Huttenlocher, 1990a) ou Lamdan et Wolfson (Lamdan et Wolfson, 1991) consiste à propager la borne sur l'erreur de mesure des points dans la méthode de calcul des invariants afin d'obtenir une borne sur la zone de vote pour le hachage géométrique ou la zone de recherche pour l'alignement assurant de ne manquer aucun appariement possible.

Contrairement à l'idée communément admise au sujet ces méthodes, on ne fait pas l'hypothèse que la distribution soit uniforme sur la zone d'erreur, mais seulement que la distribution soit à support compact inclus dans la zone d'erreur. Pour être sûr de ne pas rater d'appariement, on cherche à calculer une propagation **conservative** (i.e. une majoration) de la zone d'erreur lors des divers manipulations sur les primitives. Le caractère conservatif de la zone d'erreur permet ainsi d'assurer la **correction** de l'algorithme. En contrepartie, l'utilisation récursive de majorations finit par produire des zones d'erreur énormes qui ne sont plus comparables aux données. Pour illustrer ce point, nous avons calculé dans (Pennec, 1993a) la propagation des bornes sur l'erreur pour l'alignement et le hachage géométrique avec les points 3D.

La figure (10.1) (tirée de (Pennec, 1993a), page 11, section (2.4.2) : étude statistique de la précision des bornes) montre les valeurs mesurés pour le maximum et la moyenne d'un type d'erreur en fonction de la borne ε prédite sur celle-ci. Nous avons pu établir en calculant les droites de régression que

$$\text{Erreur}_{moy} = \frac{\varepsilon}{13.2} \qquad \text{et} \qquad \text{Erreur}_{max} < \frac{\varepsilon}{2}$$

La borne conservative prédite pour cet exemple relativement simple est donc déjà deux fois plus grande que la borne statistiquement observée. De plus, ces bornes sont relativement difficiles à calculer et chaque opération sur les points nécessite un calcul de borne particulier.

Une seconde approche est de considérer l'erreur sous forme probabiliste et de propager ceci dans les algorithmes, en conservant tout au long l'aspect probabiliste, ce qui permet en particulier de mêler l'aspect qualitatif des appariements dans le critère de mise en correspondance qui ne comprend d'habitude que l'aspect quantitatif. Différentes méthodes statistiques sont ainsi étudiées dans (Wells, 1993), mais les principaux travaux sur le hachage géométrique probabiliste à base de points sont dues à Rigoutsos et Hummel (Rigoutsos et Hummel, 1991b; Rigoutsos et Hummel, 1991a; Rigoutsos, 1992), une extension étant présentée par Tsai dans (Tsai, 1993) pour le cas de droites.

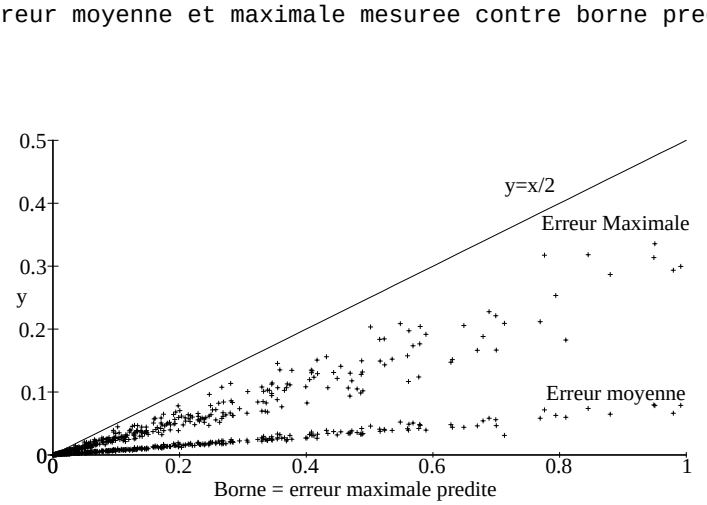


FIG. 10.1 – *Erreur maximale et moyenne observée en fonction de l'erreur maximale prédite. La borne d'erreur calculée ici n'est représentative que de l'incertitude sur l'orientation du trièdre calculé comme base pour le hachage géométrique. Les 500 valeurs moyennes et maximales sont mesurés pour 1000 perturbations d'une base constituée de 3 points.*

Dans le cadre probabiliste que nous avons développé dans ce manuscrit, il paraît plus adapté d'utiliser les modèles probabilistes, en modélisant les erreurs sur les primitives par une moyenne et une covariance : $\mathbf{x} \sim (\overline{\mathrm{x}}\ ,\ \Sigma_{\mathbf{xx}})$, d'autant que nous savons parfaitement propager cette incertitude dans les opérations élémentaires. La plupart des méthodes de reconnaissance probabilistes évoquées ci-dessus supposent une distribution gaussienne des données pour calculer la combinaison des probabilités dans le score de matching. Cela nous convient tout à fait puisque nous savons que c'est la densité qui minimise l'information en connaissant la moyenne et la covariance. Toutefois, le support de la gaussienne recouvre tout l'espace des primitives et on peut donc théoriquement apparier n'importe quelle primitive avec n'importe quelle autre avec une probabilité la plupart du temps très faible, mais non nulle. Cet effet est catastrophique pour la complexité théorique des algorithmes et on utilise en général une borne sur l'erreur admissible au niveau de la distance de Mahalanobis (un test du χ^2 dans l'hypothèse gaussienne). On peut rapprocher cette modélisation de la gaussienne contaminée décrite à la section (2.2.5.5). La zone d'appariement admissible autour de $\mathbf{x}$ est donc définie par :

$$\mathcal{Z}_\nu(\mathbf{x}) = \left\{ \mathrm{z} \in \mathcal{M} \ / \ \mu^2(\mathbf{x}, \mathrm{z}) = \overrightarrow{\overline{\mathrm{x}}\mathrm{z}}^{\mathrm{T}} . \Sigma_{\mathbf{xx}}^{(-1)} . \overrightarrow{\overline{\mathrm{x}}\mathrm{z}} \leq \nu^2 \right\}$$

où ν^2 est un seuil (en général global) qui peut être interprété *dans le cas gaussien* comme un χ^2. En comparaison, le domaine admissible utilisé dans la précédente modélisation était :

$$\mathcal{Z}_\varepsilon(\mathbf{x}) = \left\{ \mathrm{z} \in \mathcal{M} \ / \ \mathrm{dist}(\overline{\mathrm{x}}, \mathrm{z})^2 = \overrightarrow{\overline{\mathrm{x}}\mathrm{z}}^{\mathrm{T}} . \overrightarrow{\overline{\mathrm{x}}\mathrm{z}} \leq \varepsilon^2 \right\}$$

où la borne ε est ici métrique, et donc plus difficile à choisir qu'un seuil adimensionnel. Le parallèle entre ces deux zones d'erreur fait apparaître notre « modélisation probabiliste tronquée » comme une extension du modèle d'erreur bornée, où la matrice d'information $\Sigma_{\mathbf{xx}}^{(-1)}$ est utilisée en tant que métrique locale. Par ailleurs, (Faugeras, 1993, chap. 5, p.152) montre que le mécanisme de propagation des covariances par les jacobiens peut s'interpréter de façon déterministe comme la propagation au premier ordre de ces zones d'erreurs : si $\boldsymbol{x}$ est un vecteur aléatoire, f une fonction vectorielle

(continue, différentiable...) et $\mathcal{Z}_\nu(\mathrm{x})$ la zone d'erreur définie ci-dessus, alors la zone d'erreur $\mathcal{Z}_\nu(f(\boldsymbol{x}))$ est une approximation au premier ordre de la zone d'erreur transformée $f(\mathcal{Z}_\nu(\mathrm{x}))$.

En utilisant la modélisation probabiliste $\mathbf{x} \sim (\overline{\mathrm{x}}\,,\, \Sigma_{\mathbf{xx}})$ des primitives et un paramètre adimensionnel global ν, nous définissons donc des zones d'erreur pour la mise en correspondance qui sont cohérentes avec notre modélisation probabiliste et qui peuvent s'interpréter de manière déterministe ou statistique. Par contre, la propagation n'est en général plus conservative et n'est qu'une approximation au 1^{er} ordre (sauf dans le cas de l'action d'une transformation fixe puisque la propagation des covariances est exacte). Ceci permet toutefois d'éviter les majorations récursives des bornes dans leur propagation et de conserver des zones d'erreur proches de la réalité.

10.2.2 Influence sur les algorithmes

Les modifications que nous proposons dans cette section supposent que l'on connaisse *a priori* l'incertitude sur les primitives. Il est évidemment souhaitable de vérifier celle-ci *a posteriori* en utilisant les techniques décrites à la section (8.4). Pour une phase d'apprentissage, on peut imaginer de recommencer la reconnaissance avec la nouvelle estimation du bruit sur les primitives, mais il est nécessaire de connaître quelques exemples pour vérifier que le résultat de ces estimations récursives est correct, stable et robuste.

Arbres d'interprétation La prise en compte de l'erreur dans cet algorithme est relativement simple : il suffit de propager l'incertitude dans le calcul des invariants et d'utiliser un test sur la distance de Mahalanobis entre les invariants pour vérifier la satisfaction des contraintes.

Transformée de Hough On calcul la transformation $\mathbf{f} \sim (\hat{\mathrm{f}}\,,\, \Sigma_{\mathbf{ff}})$ permettant de superposer un k-uplet de primitives du modèle avec un k-uplet de primitives de la scène grâce aux techniques de recalage présentées à au chapitre 8. On vérifie ensuite que ces k appariements sont corrects grâce aux distances de Mahalanobis après recalage entre les primitives appariées, mais comme la transformation est justement calculée en minimisant cette distance, ces valeurs sont biaisées et en l'occurrence sous-estimées (voir la section 8.5.1). Il convient donc d'utiliser la valeur $\hat{\mathrm{f}}$ de la transformation comme une valeur déterministe dans les distances de Mahalanobis et de vérifier que $\mu^2(\hat{\mathrm{f}} \star \mathbf{x}_i, \mathbf{y}_i) \leq \nu^2$ pour chacun des k appariements de base. Puisque k appariements sont nécessaires pour calculer la transformation, celle-ci n'est pas fiable si l'un des tests échoue, et on rejette alors la transformation. Si les k tests sont concluants, la transformation est possible et on l'ajoute dans l'accumulateur *avec son incertitude*.

Pour cette accumulation, il existe deux techniques majeures. La plus répandue est sans doute le vote pour les transformations compatibles dans une discrétisation de cet espace, ce qui est à l'origine du terme « accumulateur », et le parcours de cette discrétisation en fin d'algorithme pour déterminer la ou les transformations ayant les scores les plus élevés. Avec l'introduction de l'erreur, il faudrait voter pour tous les buckets de l'accumulateur intersectant la zone de compatibilité $\mathcal{Z}_\nu(\mathbf{f})$, ce qui pose certains problèmes que nous détaillerons dans la section (10.4.4). Il reste également le problème de la complexité du parcours de cet accumulateur en fin d'algorithme. La seconde solution est simplement d'ajouter la transformation probabiliste $\mathbf{f}$ que nous avons trouvé à la liste des transformations possibles et d'utiliser un algorithme de **clustering** (section 10.4.2) en fin d'algorithme pour extraire et fusionner les ensembles de transformation compatibles. C'est la solution que nous avons adopté comme seconde phase de l'algorithme présenté au chapitre 11.

Alignement Le calcul de la transformation entre k primitives et la vérification des k appariements de la base ainsi formée est rigoureusement identique au cas précédent. Par contre, pour la vérification, l'incertitude de la transformation doit être prise en compte. On recherche donc pour chaque primitive $\mathbf{x_i}$ du modèle la primitive $\mathbf{y_j} = \mathrm{PPV}(\mathbf{f} \star \mathbf{x_i})$ la plus proche (au sens de la distance canonique ou de la distance de Mahalanobis et éventuellement avec la contrainte de symétrie du PPV), et on accepte l'appariement si le test suivant est réussi :

$$\mu^2(\mathbf{f} \star \mathbf{x_i}, \mathbf{y_j}) < \nu^2$$

Si le nombre d'appariements est suffisant, il convient alors de faire une vérification de la cohérence globale et un affinage de la transformation. On peut en effet sérieusement réduire l'incertitude cette son estimation en utilisant tous les appariements.

ICP et vérification de la cohérence globale Comme nous venons de calculer les appariements, l'étape suivante est de recalculer aux moindres χ^2 la transformation prenant en compte tous ces appariements. Pour affiner l'estimation, on peut éliminer les outliers en vérifiant la distance de Mahalanobis sur les appariements, mais cette fois-ci sans prendre en compte l'incertitude de la transformation (puisque celle-ci est calculée aux moindres χ^2 sur les appariements que l'on vérifie) :

$$\mu^2(\hat{\mathbf{f}} \star \mathbf{x_i}, \mathbf{y_j}) < \nu^2$$

et on itère ainsi jusqu'à la convergence. Pour faire de cette vérification un vrai ICP, il suffit de recalculer le plus proche voisin de chaque primitive du modèle et ne conserver que les appariements qui vérifient le test ci-dessus au lieu de n'exclure que les outliers.

Hachage et indexation géométrique Pour l'indexation géométrique, il suffit de calculer l'incertitude sur les invariants comme pour les arbres d'interprétation, mais il faut ensuite les indexer *avec leur zone d'erreur*, ce qui pose les mêmes problèmes que pour indexer les transformations dans la transformée de Hough : voir (10.4.4).

Le hachage géométrique est plus complexe, même si l'on ne considère que le cas des points 3D : il faut tout d'abord fixer la manière de calculer la base à partir de 3 points pour qu'on puisse également calculer l'incertitude de la transformation entre cette base et la base canonique. Il n'est pas clair de savoir quelle est la meilleure façon de le faire, même si une solution aux moindres carrés (entre ces 3 points et 3 points fixes dans la base canonique) semble adaptée : on estime directement dans ce cas la transformation en question et son incertitude. Le changement de base correspond alors à l'action de la transformation et on peut déterminer l'incertitude sur les coordonnées invariantes. Il ne reste plus qu'à indexer ces coordonnées dans la table de hachage *avec leur zone d'erreur* : on retrouve encore une fois le problème de l'indexation incertaine.

10.2.3 Faux positifs

En gérant correctement l'erreur sur les données, nous avons garanti l'exactitude des algorithmes (pas de faux négatifs). En contrepartie, nous avons une probabilité beaucoup plus de **faux positifs** (ou reconnaissances fantômes). En effet, l'utilisation d'une zone de compatibilité au lieu d'un point unique autorise l'appariement de points situés dans ces zones par hasard, et lorsque la probabilités de tels appariements est suffisamment élevée, les faux appariements se conjuguent (conspiration) pour donner un score de mise en correspondance suffisamment important pour que l'on estime avoir reconnu un objet. Notons que si l'on parle de reconnaissance fantôme, c'est que l'on peut décider si une mise en correspondance est correcte ou non.

On peut distinguer deux sources principales de faux positifs. La première est due à la recherche d'une consistance locale insuffisamment contrainte, par exemple lorsqu'on utilise uniquement les invariants unaires et binaires pour prédire des appariements globaux, ce qui est le cas de quasiment tous les algorithmes présentés. Une étape de vérification de la consistance globale des appariements est nécessaire en fin d'algorithme. Cette vérification permet d'affiner la transformation et d'éliminer les appariements aberrants (outliers). L'étape de vérification pouvant être coûteuse en temps et il convient de ne pas la multiplier inutilement et donc de limiter si possible le nombre de faux positifs. Nous verrons dans la section (10.3) comment quantifier la **robustesse** des algorithmes en calculant le nombre moyen de faux positifs pour certains algorithmes, ce qui permettra également d'estimer leur complexité moyenne.

L'autre source principale de faux positifs est plus sournoise et provient de la simplification effectuée lors de la modélisation de l'image (ou des données) par des primitives : un certain nombre d'informations pertinentes peuvent ainsi être oubliées dans le processus d'appariement, ce qui conduit à des reconnaissance correctes au niveau des primitives (i.e. globalement consistantes), mais incorrectes au niveau des données. Ce problème est inhérent au principe de structuration ascendant de l'information, mais on peut en minimiser l'influence en modélisant les données par des primitives plus adaptés et plus sélectives. Nous verrons ainsi à la section (10.3.2) que l'ajout d'un trièdre sur les points pour former un repère permet d'éliminer plus de 80% des faux appariements, même avec une erreur très large sur le trièdre, et nous avons vérifié dans une application réelle sur les protéines (chap. 11) que l'emploi des repères comme primitives au lieu des points permettait d'éliminer certains appariements biologiquement non significatifs mais géométriquement cohérents.

10.3 Étude de la robustesse : faux positifs

Nous avons vu jusqu'à présent le calcul des zones d'erreur (ou de compatibilité) des primitives géométriques et les modifications que cela apporte aux algorithmes de reconnaissance. La contrepartie de la prise en compte de l'erreur est la possibilité de conspirations, c'est-à-dire de reconnaissances fantômes. Le but de ce chapitre est d'étudier d'un point de vue probabiliste la fréquence de ce phénomène.

En effet, il est intéressant de connaître qualitativement le comportement des algorithmes de reconnaissance en fonction du bruit de mesure et du rapport signal sur bruit des deux images (nombre de primitives communes recherchées par rapport au nombre de primitives présentes). On peut ainsi décider avant même de l'implémenter si une méthode est efficace ou non. Un problème relié mais plus global est d'estimer, indépendamment de la méthode utilisée, la « complexité théorique de la mise en correspondance » ou plutôt le seuil de détection : quelle est le nombre moyen d'appariement de τ primitives si le modèle et la scène sont aléatoires, et à partir de quelle rapport signal sur bruit peut-on espérer reconnaître quelque-chose ? On peut ainsi se rendre compte que dans le cas des images médicales 3D que nous avons traité, la probabilité de faux positif est extrêmement faible et valide pleinement la mise en correspondance.

10.3.1 Modélisation du problème

La robustesse des algorithmes de reconnaissance a été étudiée par (Grimson et Huttenlocher, 1990a; Grimson et Huttenlocher, 1990b; Grimson et Huttenlocher, 1991; Lamdan et Wolfson, 1991; Grimson et al., 1994) pour les problèmes de reconnaissance à partir d'une bibliothèque de modèles exacts basés sur des points et des droites. Pour simplifier l'analyse des faux positifs, nous nous plaçons également dans le cadre d'un modèle exact et d'une scène bruitée, mais contenant des

primitives. Notre apport principal dans cette section est le développement de la méthode de calcul par intégration sur les transformations admissibles entre le modèle et la scène.

On suppose donc un modèle exact de m primitives et une scène bruitée de n primitives, dans lesquels τ primitives sont dans la même configuration (à l'erreur près). Les $(m-\tau)$ et $(n-\tau)$ autres primitives sont supposées être réparties aléatoirement dans les images modèle $\mathcal{I}_m$ et scène $\mathcal{I}_s$. En fait, si cette phrase est claire lorsque l'on parle de points, elle l'est beaucoup moins pour des primitives. Par « image », nous entendons l'ensemble des primitives mesurables dans une image réelle ou dans notre ensemble de données. Dans une image volumique, par exemple, la position d'un repère est contrainte à être dans le volume effectif de l'image $\mathcal{U} \subset \mathbb{R}^3$, mais le trièdre n'est à priori pas contraint. La zone image pour nos repères est donc : $\mathcal{I} = \mathcal{SO}_3 \times \mathcal{U} \subset \mathcal{M}$. La répartition « aléatoire » des primitives dans une « image » $\mathcal{I}$ est bien évidemment la distribution uniforme (i.e. invariante) sur cet ensemble.

Pour calculer la probabilité d'un faux positif, il nous faut également faire des hypothèses sur l'algorithme de mise en correspondance : on utilise ici le score de matching le plus simple : le nombre d'appariements, et on accepte une mise en correspondance (i.e. un ensemble d'appariements) si le score après vérification est supérieur à un certain seuil. Les algorithmes de reconnaissance modifiés étant corrects, on est assuré de ne pas rater d'appariement et on peut prendre τ comme score.

10.3.2 Sélectivité : probabilité d'un faux appariement

Supposons que nous ayons une hypothèse f de transformation entre le modèle et la scène. Pour vérifier cette hypothèse, on transforme le modèle et on recherche les primitives compatibles dans la scène : la primitive y est appariée avec $\mathbf{x}$ si $\mu^2(\mathrm{f} \star \mathrm{x}, \mathbf{y}) \leq \nu^2$, c'est-à-dire si $\mathrm{f} \star \mathrm{x} \in \mathcal{Z}_\nu(\mathrm{y})$.

On appelle **sélectivité** la probabilité d'accepter cet appariement dans l'hypothèse où l'une des primitives est en fait un outlier dont la position est aléatoire uniforme dans l'image. Par symétrie et pour simplifier les calculs, nous supposons ici que c'est la primitive x qui est uniforme. La probabilité d'un faux appariement s'écrit alors sous forme conditionnelle :

$$P(\mathrm{f} \star \mathrm{x} \leftrightarrow \mathbf{y}) = P\left((\mathrm{f} \star \mathrm{x}) \in \mathcal{Z}_\nu(\mathbf{y}) | \mathrm{x} \in \mathcal{I}_m\right)$$

et comme la loi uniforme est la mesure invariante normalisée par le volume, on obtient en notant $\mathcal{V}(\mathcal{X})$ le volume de l'ensemble $\mathcal{X}$:

$$P(\mathrm{f} \star \mathrm{x} \leftrightarrow \mathbf{y}) = \frac{\int_{(\mathrm{f} \star \mathcal{I}_m) \cap \mathcal{Z}_\nu(\mathbf{y})} d\mathcal{M}(\mathrm{x})}{\int_{\mathcal{I}_m} d\mathcal{M}(\mathrm{x})} = \frac{\mathcal{V}\left((\mathrm{f} \star \mathcal{I}_m) \cap \mathcal{Z}_\nu(\mathbf{y})\right)}{\mathcal{V}(\mathcal{I}_m)}$$

Si le volume $\mathcal{V}(\mathcal{Z}_\nu(\mathbf{y}))$ de la zone d'erreur est suffisamment faible par rapport au volume $\mathcal{V}(\mathcal{I}_m)$ de l'image, on peut considérer que l'image transformée $\mathrm{f}(\mathcal{I}_m)$ contient entièrement la zone d'erreur ou ne l'intersecte pas du tout. Ceci nous permet d'approcher la probabilité précédente par :

$$P(\mathrm{f} \star \mathrm{x} \leftrightarrow \mathbf{y}) = \varepsilon \frac{\mathcal{V}\left(\mathcal{Z}_\nu(\mathbf{y})\right)}{\mathcal{V}(\mathcal{I}_m)} \qquad \text{avec} \qquad \begin{cases} \varepsilon = 1 & \text{si} \quad \mathrm{f}^{(-1)} \star \overline{\mathrm{y}} \in \mathcal{I}_m \\ \varepsilon = 0 & \text{autrement} \end{cases}$$

Nous avons vu à la section (5.1) qu'un modèle de bruit homogène est une hypothèse raisonnable pour modéliser le bruit de mesure des primitives. Le volume de la zone d'erreur est dans ce cas invariant : il ne dépend que de la covariance à l'origine $\Sigma = J(\mathrm{f_y})^{(-1)}.\Sigma_{\mathbf{yy}}.J(\mathrm{f_y})^{(-\mathrm{T})}$ et pas de l'emplacement $\overline{\mathrm{y}}$ autour duquel il est centré :

$$\mathcal{V}\left(\mathcal{Z}_\nu(\mathbf{y})\right) = \mathcal{V}_0 = \int_{\vec{\mathrm{x}}^{\mathrm{T}}.\Sigma.\vec{\mathrm{x}} \leq \nu^2} d\mathcal{M}(\vec{\mathrm{x}})$$

Le cas des repères Prenons comme exemple les repères : on considère au chapitre 11 que deux repères sont appariés si la distance entre leurs points est inférieure à un seuil d_0 et si la rotation d'ajustement de leurs trièdres a un angle inférieure à un seuil θ_0. En fait, cette zone d'erreur correspond au modèle de bruit simplifié sur les repères de la section (8.4). Le volume de cette zone d'erreur est donc invariant et peut être calculé à l'origine : en utilisant la mesure invariante (section 7.4.1.2) sur le repère $\vec{\mathrm{x}} = (r, t)$, on a :

$$\mathcal{V}_0 = \int_{\theta<\theta_0} \int_{\|t\|<d_0} d\mathcal{M}(r,t) = \left(\int_{\|r\|<\theta_0} \frac{\sin^2(\|r\|/2)}{\|r\|^2}.dr \right) . \left(\int_{\|t\|<d_0} dt \right)$$

$$\mathcal{V}_0 = \left(\int_{\theta<\theta_0} \sin^2(\theta/2).d\theta \right) . \left(\int_{n\in\mathcal{S}_2} dn \right) . \left(\int_{\|t\|<d_0} dt \right) = 2.\pi.(\theta_0 - \sin\theta_0). \left(\frac{4\pi}{3} d_0^3 \right)$$

En supposant une image cubique de taille l (256 par exemple), cela donne un volume euclidien de $V_I = l^3$ pour les points et, les trièdres n'étant pas contraints, un volume de $2\pi^2$ pour $\mathcal{SO}_3$. On obtient finalement la sélectivité suivante :

$$P(\mathrm{f} \star \mathrm{x} \leftrightarrow \mathbf{y}) = \varepsilon.\eta$$

$$\eta = \left(\frac{\theta_0 - \sin\theta_0}{\pi} \right) \frac{4}{3} \left(\frac{d_0}{l} \right)^3$$

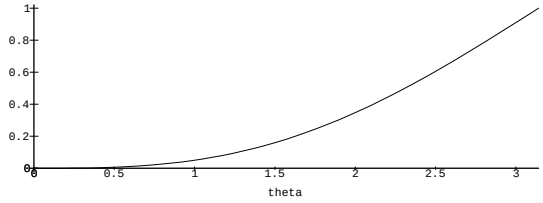


FIG. 10.2 – *Probabilité d'un faux appariement et sélectivité η pour un repère avec une borne θ sur l'angle de la rotation d'ajustement et d_0 sur la position. A droite : sélectivité du trièdre seul (terme de gauche dans l'équation).*

Nous avons isolé ici dans le premier terme la probabilité d'un faux appariement dû au trièdre seul, qui reflète le gain de sélectivité obtenu en utilisant les repères au lieu des points seuls. Cette fonction est tracée dans la figure (10.2) et montre un gain très intéressant : même pour une borne de $\theta_0 = \pi/2 = 90\,\mathrm{deg}$ sur le trièdre, plus de 80% des faux appariements sont rejetés. Pour une borne plus réaliste de $\theta_0 = \pi/10 \simeq 20\,\mathrm{deg}$, la probabilité d'un faux appariement du trièdre chute à 0.0016 : il faudrait diviser l'incertitude sur la position par 10 pour obtenir une telle sélectivité en utilisant seulement les points !

Nous verrons dans la section suivante qu'il est parfois utile de supposer une image sphérique de diamètre d (avec $d = \sqrt{3}.l$ par exemple). Le volume de l'image est dans ce cas $\mathcal{V} = 4.\pi.(d/2)^3/3$ et la sélectivité devient :

$$\eta = \left(\frac{\theta_0 - \sin\theta_0}{\pi} \right) \left(\frac{2.d_0}{d} \right)^3$$

10.3.3 Nombre moyen d'hypothèses (Hough et alignement)

Pour ces deux algorithmes, on commence par trouver les appariements possibles de k primitives (k étant le nombre minimal de primitives à apparier pour obtenir une transformation unique). Une telle hypothèse est correcte si les k primitives appariées sont de vrais appariements : il suffit en effet que l'un au moins des appariements soit faux pour que la transformation trouvée soit incorrecte.

On choisit donc k primitives ordonnées du modèles : il y a $A_k^\tau = \binom{\tau}{k}.k!$ arrangements possibles en prenant k primitives correctes et A_k^m arrangements possibles au total. La probabilité de choisir

une « base » incorrecte dans le modèle est donc (approximation pour $k \ll \tau$) :

$$p_{(m,k)} = 1 - \frac{A_k^\tau}{A_k^m} = 1 - \frac{\tau!.(m-k)!}{m!.(\tau-k!)} \simeq 1 - \left(\frac{\tau}{m}\right)^k$$

Proportion de « bases compatibles » Supposons maintenant qu'une base soit choisie dans le modèle. On lui trouve une « base » correspondante dans la scène s'il existe une transformation qui apparie ses k primitives avec k primitives de la scène. Avec une transformation fixée f et en supposant que toutes les primitives (même celles qui sont correctes) soient réparties a peu près uniformément dans la scène, la probabilité qu'aucune des n primitives de la scène ne tombe dans la zone compatible d'une des primitives de la base transformée est : $(1 - \varepsilon.\eta)^n$. La probabilité qu'on trouve au moins un appariement pour chacune des k primitives de la base est donc (toujours avec une transformation f fixée) :

$$p_k(\mathrm{f}) = \prod_{i=1}^{k}(1 - (1 - \varepsilon_i.\eta)^n) \qquad \text{avec} \qquad \varepsilon_i = \begin{cases} 1 & \text{si} \quad \mathrm{f} \star \mathrm{x}_i \in \mathcal{I}_s \\ 0 & \text{sinon} \end{cases}$$

Il ne nous reste plus qu'à intégrer sur toutes les transformations possibles pour trouver le nombre moyen de bases compatibles dans la scène. Pour cela, on peut noter que l'expression ci-dessus est constante et non nulle si tous les ε_i sont égaux à un, c'est-à-dire uniquement quand la « base modèle » est transformé dans la scène par f. On peut donc écrire cette probabilité comme :

$$P_k = \int_{\mathrm{f}\in\mathcal{G}} p_k(\mathrm{f}).d\mathcal{G} = (1 - (1 - \eta)^n)^k . \int_{\mathrm{f}\in\mathcal{G}} \prod_{i=1}^{k} \varepsilon_i(\mathrm{f}).d\mathcal{G}$$

On peut majorer l'intégrale α du second terme en demandant simplement que les images s'intersectent :

$$\prod_{i=1}^{k} \varepsilon_i(\mathrm{f}) \simeq 1 \qquad \text{si} \qquad \mathrm{f} \star \mathcal{I}_m \cap \mathcal{I}_s \neq \varnothing$$

ce qui permet de l'estimer grossièrement pour une image 3D en faisant l'observation suivante (pour les transformations rigides) : soit d le diamètre de l'image que l'on suppose sphérique (par exemple $d = \sqrt{3}.l$ pour une image l^3). En mettant l'origine au centre, on peut faire n'importe quelle rotation, suivie d'une translation de norme inférieure au diamètre de l'image (et non pas au rayon). On peut donc estimer notre intégrale par :

$$\alpha \simeq 2.\pi^2 . \int_{\|t\|<d} dt = \frac{8}{3}.\pi^3.d^3 = (2.\pi.l)^3$$

Une majoration du volume de transformation possible entre deux images 3D de taille d est $\alpha \simeq (2.\pi.d)^3$. La probabilité qu'une base de k primitive dans le modèle ait une base correspondante dans l'image est :

$$P_k = \alpha.(1 - (1 - \eta)^n)^k$$

Proportion d'hypothèses correctes Maintenant, on sait que si une base image est correcte, on trouve obligatoirement la base associée dans la scène, mais on peut également trouver d'autres bases qui correspondent et qui sont des faux positifs qu'il faudra quand même compter. Pour chacune des A_k^m bases possibles, on a en moyenne P_k fausses hypothèses, soit un total de $P_k.A_k^m$ en moyenne, chiffre auquel il faut rajouter les A_k^τ hypothèses correctes.

Dans le cas de la transformée de Hough, c'est la densité des transformations correctes parmi les transformation correspondantes à toutes les hypothèses qui nous intéresse :

$$P_{hyp} = \frac{A_k^\tau}{A_k^\tau + P_k.A_k^m} = \left(1 + \alpha.(1-(1-\eta)^n)^k.\frac{\tau!.(m-k)!}{m!.(\tau-k!)}\right)^{(-1)} \simeq 1 - \alpha.\left(1-(1-\eta)^n.\frac{\tau}{m}\right)^k$$

10.3.4 Probabilité d'acceptation d'une vérification

On suppose ici que l'on a une hypothèse de transformation que l'on veut vérifier. Un modèle ayant m points dont k sont utilisés pour former la base, il reste $m-k$ points à mettre en correspondance. Dans le cas de l'alignement il y aura donc $m-k$ zones de recherche dans la scène La probabilité qu'un point de la scène ne tombe dans aucune de ces zones est $(1-\bar{\eta})^{m-k}$, donc la probabilité qu'il tombe dans l'une au moins de ces zones et qu'il soit apparié est donc

$$p = 1 - (1-\bar{\eta})^{m-k}$$

Nous avons utilisé ici la sélectivité $\bar{\eta}$ pour bien spécifier que celle-ci n'est pas forcément celle que l'on a déterminé précédemment. En effet, dans le cas de la vérification d'une hypothèse pour l'alignement, la transformation est déterminée avec k appariements et il faut prendre en compte son incertitude lors de l'appariement des autres primitives. S'il est facile de calculer en ligne cette sélectivité moyenne, il est très difficile d'en faire une estimation théorique puisque celle-ci dépend de la configuration de la « base ». Par contre, lors de la vérification finale, l'incertitude de la transformation n'est pas prise en compte (puisqu'elle est calculée à partir des appariements que l'on cherche à vérifier), et la sélectivité $\bar{\eta}$ est bien celle que l'on a calculé précédemment.

Comme il y a $n-k$ primitives dans la scène, la probabilité que j primitives sur ces $n-k$ soient acceptées comme appariements est la binomiale :

$$B_{(n-k,p)}(j) = \binom{n-k}{j} p^j (1-p)^{n-k-j}$$

La probabilité d'acceptation d'une vérification est donc la la somme des probabilités d'avoir plus de τ appariements :

$$q = \sum_{j \geq \tau} B_{(n-k,p)}(j) = 1 - \sum_{j=0}^{\tau} \binom{n-k}{j} p^j (1-p)^{n-k-j} = I_p(\tau+1, n-\tau)$$

où $I_p(a,b)$ est la fonction Bêta incomplète normalisée.

Dans l'hypothèse $p \ll 1$, c'est-à-dire $\bar{\eta}m \ll 1$, et $n \gg k$, on peut approcher la binomiale $B_{(n-k,p)}$ par la loi de Poisson de paramètre $\lambda = (n-k).p$:

$$q \simeq 1 - e^{-\lambda} \sum_{j=0}^{\tau} \frac{\lambda^j}{j!}$$

Proportion de faux positifs On sait que l'on a A_k^m bases possibles dans le modèle et en moyenne P_k fausses hypothèses pour chacune de ces bases, chaque hypothèse étant acceptée avec une probabilité q. On accepte donc en moyenne $A_k^m.P_k.q$ mises en correspondances incorrectes, en sachant que les A_k^τ bases modèle correctes donne aussi lieu à des mises en correspondances acceptées (qui sont bien sûr redondantes). La proportion de faux positifs est donc :

$$P_{align} = \frac{A_k^\tau}{A_k^\tau + A_k^m.P_k.q} \simeq 1 - \frac{A_k^m.P_k.q}{A_k^\tau} \simeq\simeq 1 - \alpha.q.\left(1-(1-\eta)^n.\frac{\tau}{m}\right)^k$$

La probabilité q peut donc être directement interprétée comme le taux de filtrage des hypothèses grâce à la vérification.

10.3.4.1 Probabilité de faux positif *a posteriori*

Dans les deux section précédentes, nous avons cherché à caractériser le comportement de deux algorithmes, en propageant, étape par étape, la probabilité des appariements incorrect. On peut également se poser la question des faux positifs *a posteriori*, une fois que l'on a mis en correspondance τ primitives du modèle et de la scène. Nous formalisons le problème ainsi : quelle est la probabilité de trouver τ appariements si le modèle et la scène sont en fait constitués de m et n primitives réparties uniformément ?

En fait, les sections précédentes nous ont permis de développer les outils nécessaires pour répondre à cette question. En effet, si l'on a trouvé k appariements, c'est qu'il existe une transformation f qui nous permet de les valider. Il nous suffit donc de calculer la probabilité de trouver au moins k appariements pour une transformation fixée, puis d'intégrer cette probabilité sur les transformation possibles.

Pour une transformation f fixée, nous avons (au plus) m zones de compatibilité dans la scène après transformation. En simplifiant tout de suite les expressions, la probabilité qu'un point de la scène tombe dans une de ces zones est $p = \eta.m$, et la probabilité qu'il en tombe j sur les n de la scène est la binomiale $B_{(n,p)}(j)$. La probabilité de trouver au moins τ appariements est donc :

$$q = \sum_{j \geq \tau} B_{(n,p)}(j) = I_p(\tau + 1, n) \simeq 1 - e^{-\lambda} \sum_{j=0}^{\tau} \frac{\lambda^j}{j!}$$

avec $\lambda = n.p = n.m.\eta$. Pour obtenir le nombre moyen de faux positifs (la probabilité non normalisé), il suffit maintenant d'intégrer cette expression sur toutes les transformation admissibles (i.e. superposant un morceau des deux images) ce qui revient à multiplier par α.

Le nombre moyen d'appariements de τ primitives entre un modèle et une scène aléatoires (uniformes) de m et n primitives est donc avec notre modèle d'image sphérique de diamètre d (et à la louche) :

$$\Phi = q.\alpha = (2.\pi.l)^3.I_{\eta.m}(\tau + 1, n).\frac{8}{3}.\pi^3.d^3$$

Cette expression n'est évidemment significative que lorsque $\eta.m.n \ll 1$ et, vu l'aspect très qualitatif de notre analyse, on ne pourra y croire que lorsqu'elle est extrêmement faible. Par contre, on peut observer que l'intégration sur les transformations admissibles n'influence ce nombre moyen que par l'intermédiaire du facteur α et la qualité de son estimation est donc relativement peu importante par rapport à celle de la sélectivité η qui influence très directement la probabilité d'acceptation q. Pour s'en convaincre, on pourra voir sur la figure (10.3) la différence des résultats entre points et repères, en sachant qu'une variation de α ne produit qu'une faible translation verticale en échelle logarithmique.

10.3.4.2 Exemple d'application

Prenons le cas des données IRM du cerveau de la section (9.3). Les images ont une taille 256 x 256 x 165 mm et on y extrait typiquement 2500 points extrémaux dont 500 sont finalement appariés.

Sélectivité des repères En combinant l'erreur sur les repères des deux images (modèle et scène), on obtient selon la section (9.3.3) une matrice de covariance grossièrement diagonale :

$$\Sigma = \mathrm{DIAG}(0.005\ ,\ 0.006\ ,\ 0.065\ ;\ 0.45\ ,\ 0.60\ ,\ 0.165)$$

Sur une zone aussi petite, on peut approcher la mesure invariante sur les rotation par la mesure de Lebesgue :

$$\frac{\sin^2(\theta/2)}{\theta^2}.d\theta \simeq (1+O(\theta^2)).\frac{d\theta}{4}$$

et donc approximer le volume de la zone compatible par le volume de l'ellipsoïde d'incertitude euclidien défini par la matrice Σ et un χ^2 de $\nu^2 = 16$:

$$\mathcal{V}_0 = \int_{x^{\mathrm{T}}.\Sigma^{(-1)}.x \leq \nu^2} \frac{dx}{4}$$

Pour calculer cette intégrale, on fait le changement de variable $x = \nu.\Lambda.y$, où Λ est une racine carrée de Σ, et on obtient :

$$\mathcal{V}_0 = \frac{1}{4}.\int_{y^{\mathrm{T}}.y<1} \nu^6.|\det(\Lambda)|.dy = \nu^6.\sqrt{\det(\Sigma)}.\frac{\pi^3}{6}.\frac{1}{4} \simeq 16^3.(2.94\ 10^{-4}).\frac{\pi^3}{24} \simeq 1.56$$

puisque l'intégrale n'est en fait que le volume de la sphère $\mathcal{S}_6$. Par ailleurs, le volume de l'image est de $256.256.165 = 1.08\ 10^7$ pour la position du trièdre (ou pour un point) et pour le trièdre de

$$\int_{\|r\|\leq\pi} \frac{\sin^2(\|r\|/2)}{\|r\|^2}.dr = 2.\pi^2$$

En fait, les repères ne sont que semi-orientés, et comme deux repères représentent le même repère semi-orienté, il faut diviser ce volume par 2. On obtient donc un volume image total pour les repères (semi-orientés) de :

$$\mathcal{V}(\mathcal{I}) = 256.256.165.\pi^2 = 1.067\ 10^8$$

ce qui donne une sélectivité de $\eta = \frac{\mathcal{V}_0}{\mathcal{V}(\mathcal{I})} = 1.46\ 10^{-8}$

Sélectivité des points Il est intéressant de comparer avec la sélectivité des points : la covariance obtenue sur le vecteur d'erreur est isotrope d'écart type $\sigma = 0.70 = \sqrt{2}.0.5$. On utilise ici un χ^2 de $\nu^2 = 6$. Le volume de la zone d'erreur est cette fois-ci $\mathcal{V}_0 = \nu^3.\sigma^3.\frac{4}{3}.\pi \simeq 21.7$, ce que l'on compare au volume simple de l'image : $\mathcal{V}(\mathcal{I}) = 1.08\ 10^7$. On obtient une sélectivité de $\eta_{pt} = 2.01\ 10^{-6}$. On gagne donc ici deux ordres de grandeur en utilisant les repères au lieu des points !

Probabilité de faux positif Pour obtenir une majoration du nombre moyen de faux positifs, on prend comme diamètre de l'image $d = \sqrt{3}.l$ avec $l = 256$, ce qui majore le volume des transformation admissibles et donne un facteur $\alpha = (2.\pi.l)^3 = 4.16\ 10^9$. Le nombre de primitives est $n = m = 2500$ et la probabilité d'un appariement de τ primitives pour une transformation fixée est donc $q = I_{\eta.n}(\tau+1, n)$.

Nous avons tracé dans la figure (10.3) le nombre moyen de faux positifs $\Phi = \alpha.q$ en fonction du nombre τ de primitives appariés, pour les points et les repères. Dans les deux cas, la probabilité d'obtenir un faux positif avec 500 appariements (nombre moyen de primitives appariées dans ces images) est rigoureusement nulle (à la précision numérique de la machine). A titre d'indication, nous avons tracé sur la figure la ligne de probabilité $\Phi = 10^{-10}$, à partir de laquelle on peut

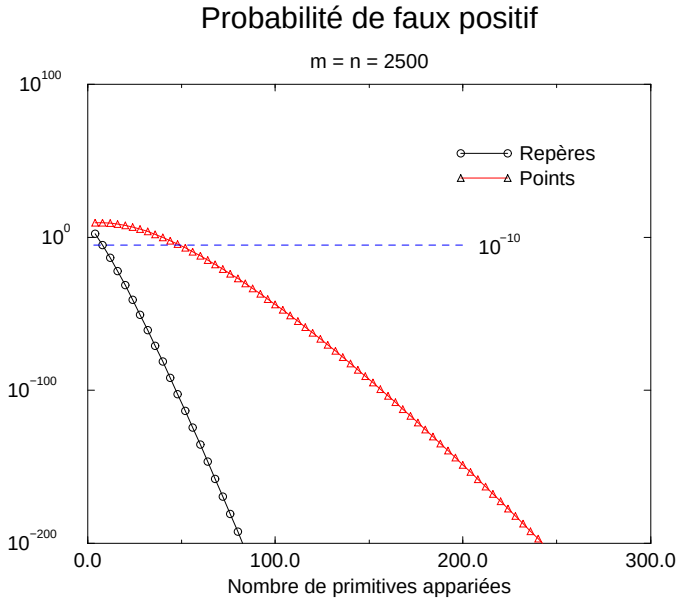


FIG. 10.3 – *Probabilité de faux positif pour les points et les repères semi-orientés. La sélectivité des repères est de* $\eta_{rep} = 1.5\ 10^{-8}$ *tandis que celle des points est de* $\eta_{pt} = 2\ 10^{-6}$. *Les images sont de taille 256 x 256 x 165 et contiennent 2500 primitives. La barre des* $\Phi = 10^{-10}$ *est passée pour 10 repères et 55 points.*

raisonnablement penser qu'on n'observera plus aucun faux positif. Étant donné que nous avons largement surestimé le volume α des transformations admissibles, cette borne est d'autant plus crédible. La limite est passée pour 10 appariements de repères et 56 appariements de points. En fait en peut grossièrement estimer qu'il faut 5 fois plus d'appariements de points que d'appariements de repères pour obtenir la même probabilité de faux positif ! L'intérêt d'utiliser des primitives complexes est donc particulièrement clair.

10.3.5 Discussion

En ce qui concerne les estimations de faux positifs pour l'alignement et la transformée de Hough, on peut continuer les statistiques pour calculer par exemple le nombre moyen d'hypothèses à vérifier avant d'en trouver une bonne, ou la complexité moyenne des algorithmes, mais tous ces calculs reposent sur un fonctionnement précis des algorithmes de reconnaissance. De plus, pour pouvoir mener cette analyse d'un point de vue théorique, nous avons été obligés de faire des hypothèses relativement restrictives (modèle exact, distribution uniforme...) qui ne sont souvent pas vérifiés. Les divers analyses présentées ici doivent donc être considérés comme simplement indicatives d'un comportement et non pas comme des estimations quantitatives.

Par contre, dans le cadre d'algorithmes génériques sur les primitives géométriques, il serait intéressant de calculer la sélectivité des appariements en ligne (lors de l'algorithme) et de la propager dans les différentes opérations utilisées comme nous l'avons fait pour l'incertitude. On pourrai ainsi vérifier en tout point de l'algorithme l'état des hypothèses et abandonner celles qui sont trop incertaines sur des bases quantitatives. Cela est à relier directement avec les critères de matching probabilistes développés dans (Rigoutsos et Hummel, 1991b; Rigoutsos et Hummel, 1991a; Rigoutsos, 1992), intégrant directement dans le critère d'appariement la probabilité de faux positifs.

10.4 Quelques problèmes clés

Dans la première section, nous avons répertorié un certain nombre d'algorithmes de mise en correspondance généralement utilisés avec des points et nous les avons formalisé grossièrement en terme de primitives. Nous avons ensuite analysé les modification nécessaires pour ces algorithmes continuent à fonctionner **correctement** (i.e. sans faux négatifs) en présence d'erreur. Nous nous intéressons dans cette section à quelques problèmes clés, sous-tendus par les algorithmes présentés, et qui constituent les points difficiles pour l'utilisation de ces techniques de mise en correspondance avec des primitives générales ou simplement à cause de l'introduction de l'erreur.

En ce qui concerne l'algorithme ICP, nous avons déjà développé le formalisme nécessaire au calcul de la transformation, mais il manque une implémentation efficace pour l'algorithme du **plus proche voisin dans une variété**, munie bien sûr d'une métrique riemannienne non euclidienne. Un problème relativement proche qui se pose pour la transformée de Hough est celui du **clustering** de primitives ou de transformations (un terme équivalent en français serait *groupement* ou *regroupement*). Tous les autres algorithmes utilisent plus ou moins des **invariants** binaires, ternaires ou d'ordre encore supérieur. Le problème est non seulement de les calculer, mais aussi de savoir les comparer correctement et efficacement. Nous proposons une approche basée sur les **espaces de forme** qui devrait permettre de replacer ces invariants dans un cadre de géométrie riemannienne similaire à celui que nous avons développé pour les primitives homogènes. Le dernier point que nous abordons dans cette section concerne les techniques d'accélération de la recherche de primitives ou d'invariants compatibles, c'est-à-dire principalement les techniques de **hachage**.

10.4.1 PPV dans une variété

Le problème du plus proche voisin est ancien et a été beaucoup étudié d'un point de vue géométrie algorithmique. La notion mathématique correspondante est celle de **diagramme de Voronoï**. On pourra consulter par exemple (Aurenhammer, 1991; Okabe et al., 1992; Boissonnat et Yvinec, 1995). Cependant, il est souvent plus facile d'implanter des techniques moins rigoureuses mais plus efficaces comme les « k-D trees » (Preparata et Shamos, 1985) qui subdivisent itérativement l'espace suivant chacune des dimensions et en fonction des points donnés. Cette technique repose sur l'équivalence de la norme euclidienne L_2 avec la norme L_∞, séparable selon les axes, ce qui permet d'obtenir un encadrement de la norme euclidienne en fonction de la différence maximale sur les coordonnées. La généralisation de cet algorithme à une variété riemannienne semble difficile car il n'existe plus de système de coordonnées globales qui nous permette de définir une norme similaire à L_∞ et d'alterner la recherche suivant les axes. Cependant, les géodésiques passant par un point étant des droites passant par l'origine dans la carte exponentielle en ce point, il n'est pas impossible que l'on puisse mettre au point une technique de subdivision de la variété basée sur cette propriété.

La notion de diagramme de Voronoï est par contre généralisable à une variété riemannienne, mais très peu de travaux s'y sont intéressés. On peut toutefois citer (Boissonnat et Yvinec, 1995, chap.18) pour une variété hyperbolique, mais sa détermination repose sur le fait que cette variété est de courbure négative partout et qu'il existe un difféomorphisme qui permet de se ramener à $\mathbb{R}^n$ (en l'occurrence une projection). Ce type de traitement ne peut évidemment pas s'appliquer à des variétés de courbure positives comme les sphères ou les espaces projectifs (ce qui inclut les rotations 3D), mais d'autres techniques sont possibles. Watson a ainsi développé une méthode pour calculer le diagramme de Voronoï sur les sphères $\mathcal{S}_n$ munies de leur métrique canonique (voir (Watson, 1988) et figure 10.4). Il est possible que cette méthode soit généralisable à nos variétés homogènes (munies de la métrique riemannienne invariante) en utilisant une fois de plus les propriétés des cartes exponentielles. Il reste toutefois le problème de l'efficacité de ces algorithmes.

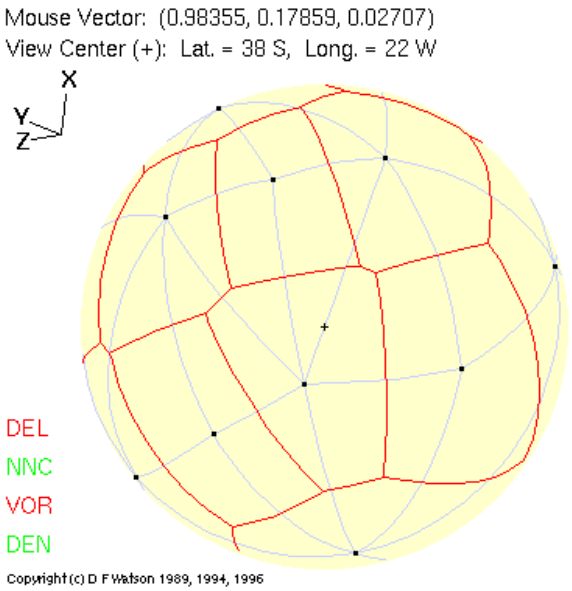


FIG. 10.4 – *Diagramme de Voronoï et triangulation de Delaunay duale sur la sphère $\mathcal{S}_2$ générés interactivement et en temps réel par l'applet java ModeMap de Dave Watson accessible à l'URL : http://www.iinet.com.au/∼watson/modemap.html.*

Enfin, pour conclure ce tour d'horizon très rapide du problème, on pourra noter que toutes ces méthodes ne peuvent pas calculer le plus proche voisin au sens de la distance de Mahalanobis. En effet, elles reposent sur un pré-traitement des données en connaissant la métrique, alors que celle-ci n'est fixée que lorsqu'on pose la question pour la distance de Mahalanobis puisque cette distance fait intervenir la covariance de la primitive testée. Il reste donc un travail important à réaliser dans ce domaine pour pouvoir inclure proprement un algorithme du plus proche voisin dans notre méthodologie générique sur les primitives.

10.4.2 Clustering de primitives géométriques probabilistes

Clustering peut se traduire en français par *regroupement* ou *classification* suivant le sens principal sur lequel on veut insister. Dans notre cas, le principal problème de clustering est de regrouper les transformation compatibles entre elles pour la transformée de Hough. On pourra consulter (Olson, 1994) pour une analyse récente des différentes techniques utilisée dans cette optique et (Jolion et al., 1991) pour une approche robuste, sans oublier les ouvrages classiques (Duda et Heart, 1973; Jain et Dubes, 1988).

Comme on n'utilise pas ici l'action des transformations sur des primitives, le problème générique est celui du clustering de primitives dans une variété. Avec notre formalisation, le problème est relativement simplifié par rapport au cadre classique puisque nous possédons en plus des mesures une information d'incertitude associée : la matrice de covariance. Par contre, Grimson a montré dans (Grimson et Huttenlocher, 1990b; Grimson et Huttenlocher, 1991) que l'introduction de l'erreur rendait ce type d'algorithme particulièrement sensible aux faux positifs. De plus, nous voudrions pouvoir rechercher plusieurs objets simultanément, c'est-à-dire effectuer un clustering sans connaître le nombre de classes recherchées.

Il semble possible de généraliser un certain nombre de techniques de clustering sur les points à

notre problème en remplaçant la distance entre points par la distance riemannienne ou de Mahalanobis entre primitives. Par contre, les techniques d'échantillonage de l'espace se heurtent à des problèmes importants dûs au fait que nous ne sommes plus dans un espace vectoriel (voir cependant à ce sujet la section 10.4.4).

Nous avons choisi dans le chapitre 11 d'utiliser une technique relativement peu efficace, peu rigoureuse mais simple à implanter : on choisit comme hypothèse de départ la transformation la plus informative de la liste (en utilisant l'opération statistique « Information » de la section 6.3.4) et on fusionne itérativement la transformation compatible la plus proche de l'état courant (au sens de la distance de Mahalanobis), en enlevant à chaque fois la transformation fusionnée de la liste. On recommence ensuite l'opération pour estimer une nouvelle classe, jusqu'à ce que la liste originelle soit vide. Il ne reste plus alors qu'à filtrer les clusters obtenus pour ne conserver que les quelques meilleurs.

Une autre solution, sans doute plus rigoureuse, consisterait à faire un moindre χ^2 selon l'algorithme développé à la section (8.3.2) en partant du même point de départ, mais en n'utilisant à chaque étape de la descente de gradient, que les transformations compatibles avec l'estimée courante (celle-ci étant considérée comme déterministe puisque estimée à partir des transformations compatibles). Lorsque l'algorithme a convergé, on enlève les transformations compatibles de la liste et on cherche à estimer une nouvelle classe par le même principe. Cet algorithme reste toutefois à tester, particulièrement en ce qui concerne la convergence et la stabilité. De manière plus générale, un travail important reste à effectuer pour adapter les algorithmes existants à notre formalisme, tester leur efficacité, leur stabilité et leur précision.

10.4.3 Invariants n-aires : espace des formes

Nous avons parlé précédemment d'invariants binaires, ternaires, etc, et plus spécialement des « invariants caractéristiques de la forme constituée des k points ordonnés », en restant volontairement flou sur cette notion. On imagine assez aisément l'utilisation de la distance comme invariant de deux points pour les transformations rigides et nous avons déjà parlé des trois distances inter-points pour les invariants d'un triplet de points. Il est par contre plus dur d'imaginer ce que sont les invariants rigides d'un quadruplet de points, sans parler des invariants projectifs des droites ou d'autres primitives. Beaucoup de travaux se sont focalisés sur la recherche de tels invariants et l'on pourra consulter utilement à ce sujet (Mundy et Zisserman, 1992; Weinshall, 1993; Rothe et al., 1994).

Nous présentons rapidement ici une autre approche basée sur la théorie des formes et principalement développée par Kendall et Le dans (Kendall, 1989; Le et Kendall, 1993; Le, 1995). On pourra également se reporter au chapitre 12 pour un exemple d'utilisation de cette théorie sur les points. L'idée est de caractériser l'espace des configurations de k primitives. Par configuration, nous entendons ce qui est invariant sous l'action d'un groupe de transformation fixé : **la forme**. Un k-uplet de primitives est ainsi un élément de $\mathcal{M}^k$ et pour obtenir la forme, on identifie tous les k-uplets superposables par une transformation $f \in \mathcal{G}$: l'espace $\mathcal{I}_k$ des formes à k primitives est donc l'espace quotient $\mathcal{I}_k = \mathcal{M}^k/\mathcal{G}$, noté $\Sigma(\mathcal{M}, \mathcal{G}, k)$ par Kendall. On peut reformuler ceci de la manière suivante : un k-uplet de primitives $X \in \mathcal{M}^k$ peut se « factoriser » en une forme $i \in \mathcal{I}_k$ et la « position » $f \in \mathcal{F} \subset \mathcal{G}$ de cette forme dans l'espace.

Supposons que nous ayons obtenu une caractérisation de l'espace des k-formes et une carte locale. Pour pouvoir comparer des formes, nous devons déterminer une métrique riemannienne, puis les géodésiques et enfin les carte exponentielles. Nous pensons qu'il est possible de déterminer les conditions d'existence d'une métrique riemannienne sur $\mathcal{I}_k$ compatible avec les métriques sur $\mathcal{M}$ et

sur $\mathcal{G}$ de façon similaire à ce que nous avons fait à la section (3.5.4) pour la métrique invariante sur une variété homogène. La détermination des géodésiques se fait alors de la même façon que pour n'importe quelle variété, mais contrairement au cas des variétés homogènes, nous n'avons pas de groupe de transformation pour identifier la carte exponentielle en un point fixé (origine) à la carte exponentielle en un autre point, ce qui fait que, d'un point de vue informatique, nous devrions gérer toutes les cartes exponentielles...

Si les notions de primitives aléatoires, de moyenne et de covariance semblent se transportent relativement aisément sur ces variétés d'invariants, le problème de la gestion informatique d'un « invariant probabiliste » reste donc à résoudre. Les opérations élémentaires sur les invariants probabilistes sont par contre plus restreintes : nous avons essentiellement identifié les trois opérations suivantes :

- Traduction d'un k-uplet en k-forme et position : $\mathcal{X} \in \mathcal{M}^k \leftrightarrow (\mathrm{i}, \mathrm{f}) \in \mathcal{I}_k \times \mathcal{G}$.
  Les différentielles doivent également être calculées pour pouvoir implanter aussi cette opération au niveau des primitives et invariants probabilistes.
- Distance entre deux k-formes : $\mathrm{dist}(\mathrm{i}_1, \mathrm{i}_2)$.
  Cette opération de base est a priori implantable ne utilisant la distance euclidienne dans la carte exponentielle en l'un des deux points, mais encore faut-il avoir l'expression de cette carte en tout point.
- Distance de Mahalanobis entre deux k-formes probabilistes : $\mu(\mathbf{i}_1, \mathbf{i}_2)$.

On peut également rajouter à cette liste des opérations du type « modélisation statistique » et « mesure du bruit » (voir section 6.3.4), mais il faut tout d'abord en définir le sens. Pour conclure, nous pensons que l'étude des invariants n-aires par les espaces de formes devrait permettre une étude théorique rigoureuse de ces objets géométriques, mais leur gestion informatique nécessitera sans doute des aménagements par rapports à la méthodologie que nous avons développé jusqu'ici pour gérer les primitives et les transformations aléatoires.

10.4.4 Indexation incertaine

Dans tous les algorithmes utilisant des invariants se pose le même problème : retrouver efficacement parmi un ensemble d'invariants, éventuellement muni de leur incertitude, ceux qui sont compatibles avec un invariant donné. La solution la plus simple, mais aussi la moins efficace, consiste à comparer individuellement l'invariant en question avec tous les autres, ce qui a une complexité en $O(N)$ si N est le nombre d'invariants dans la « base de donnée ».

Il y a principalement deux type de techniques pour diminuer la complexité : on peut échantillonner récursivement l'espace de recherche en fonction des données (techniques de type k-D trees), ce qui donne généralement un complexité de recherche de $O(\log N)$, ou bien échantillonner l'espace de recherche de manière fixe, mais avec une technique d'indexation de cette échantillonage qui permette de retrouver en temps quasi-constant les données relatives à un emplacement : ce sont les méthodes de hachage, auxquelles nous nous intéressons dans cette section.

Supposons pour l'instant qu'il n'y ait pas de bruit sur les données. Il nous faut déterminer un échantillonage de l'espace et une fonction de hachage qui transforme les coordonnées d'un « bucket » (une case de l'échantillonnage) en un code, qui est l'index d'un tableau de listes regroupant les données de même code. La fonction de hachage doit être choisie de façon à disperser au maximum les codes produits afin d'obtenir des listes de longueur faible (idéalement 1 pour chaque code).

L'introduction de l'erreur pose un certain nombre de problèmes dans ce formalisme puisque nous devons indexer ou retrouver nos invariants avec une zone d'erreur : celle-ci intersecte en général plusieurs cases de l'échantillonnage. Il nous faut tout d'abord déterminer quels sont ces cases.

Ensuite, ces différentes cases créent des entrés supplémentaires dans la table de hachage. Nous n'aurons donc plus N entrées comme précédemment, mais de l'ordre de $6.N$ entrées si toutes les zones d'erreurs sont comme dans la figure (10.5). Il est important que quantifier le nombre moyen de cases indexées pour calculer le nombre moyen d'objets par index de la table (la longueur moyenne de la liste). La complexité réelle du hachage à la reconnaissance est proportionnelle à ces deux facteurs. La fonction de hachage peut en effet associer le même index à deux cases de coordonnées arbitrairement différentes, et une étape de vérification de compatibilité entre l'invariant en question et ceux de même index est nécessaire.

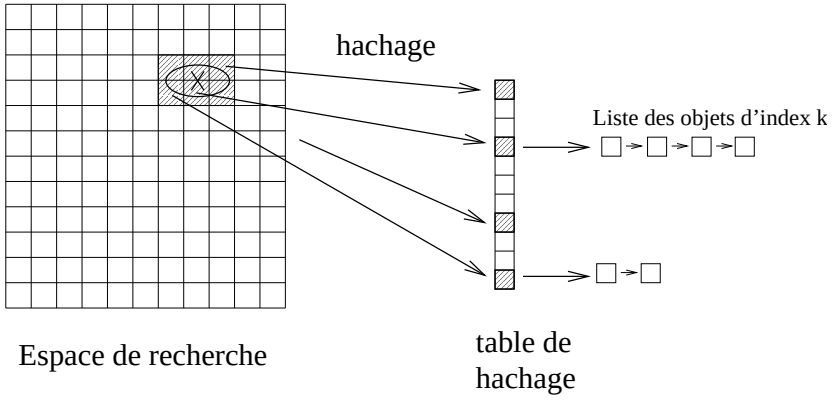


FIG. 10.5 – *Hachage de données multi-dimensionnelles en présence d'erreur*

Dans (Pennec, 1993b), nous avons analysé les différents types de hachage à base de points utilisés par les algorithmes de reconnaissance. Prenons un exemple simple : l'indexation d'un point de $\mathbb{R}^d$ avec une zone d'erreur bornée en norme par ε avec une discrétisation cartésienne de l'espace de pas isotrope l. On montre alors que le nombre moyen de cases indexées (i.e. recoupant la zone compatible) est de l'ordre de $\bar{n} = (1 + 2.\varepsilon/l)^d$. Sur des critères de probabilité de faux positifs, nous avions conclu qu'un pas de hachage $l \simeq \varepsilon$ était un bon compromis, ce qui donne déjà en dimension 3 un nombre moyen de cases indexées de 9. Pour une dimension fixée, ce n'est qu'un facteur multiplicatif dans la complexité, mais ce facteur est exponentiel vis à vis de la dimension. Ainsi, si l'on regarde le hachage géométrique avec des points 3D (rigides), nous avons trois invariants pour la base (les distances entre les trois points de celles-ci) et les trois coordonnées du quatrième point dans cette base. Le nombre moyen de cases indexées est cette fois-ci de 729 !

Par ailleurs, même en supposant la même borne d'erreur sur tous les points, la propagation de celle-ci dans le calcul des invariants donne une borne qui varie en fonction de la position des invariants, comme on peut le voir dans la figure (10.6). Dès lors, on peut se demander si une discrétisation adaptative de l'espace ne serait pas plus appropriée.

Enfin, pour finir, nous n'avons vu jusqu'ici que les problèmes liés aux invariants simples des points. Si l'on considère par exemple un couple de repères, l'invariant binaire associé est lui aussi un repère (c'est la transformation de l'un des repères à l'autre, exprimée dans l'un des repères). L'espace des invariants est donc cette fois-ci $\mathcal{SO}_3 \times \mathbb{R}^3$ et non plus $\mathbb{R}^d$. La première question est comment échantillonner de manière régulière un espace compact bouclant sur lui-même comme la sphère ou $\mathcal{SO}_3$. Ce problème est relié au calcul des cases intersectées par la zone d'erreur, qui doit être très efficace pour que le hachage conserve un intérêt. On pourrait imaginer un échantillonnage « cartésien » dans la carte principale, mais si la rotation est proche du bord de la carte, la zone d'erreur peut être séparée en deux morceaux, c'est-à-dire sortir d'un côté de la carte pour y rentrer à l'opposé. De plus, si la zone d'erreur est un vrai ellipsoïde pour un vecteur rotation centré dans

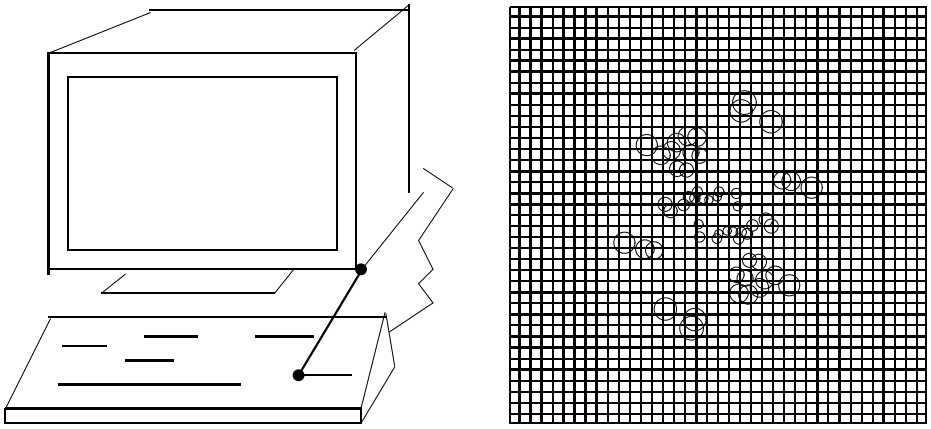


FIG. 10.6 – *A droite, une image 512*512. Les deux points servant de base sont indiqués en gras avec la taille de leur zone d'erreur (borne de 5 pixels). A gauche, la tranche de la table de hachage correspondante à la longueur mesurée de la base. On peut voir la sensibilité des invariants au bruit de mesure de la base en comparant la taille des zones d'erreur dans l'espace images (à gauche) et dans l'espace des coordonnées invariantes (à droite).*

la carte, c'est de moins en moins vrai lorsqu'on s'approche du bord.

L'adaptation des techniques de hachage à des primitives ou de invariants généraux en présence d'erreur n'est donc pas un problème facile, et le gain d'efficacité pour lequel ces techniques étaient utilisées à l'origine n'est pas forcément conservé. La recherche efficace des invariants compatibles est pourtant un problème crucial à résoudre pour conserver une complexité faible à nos algorithmes de reconnaissance sur les primitives.

10.5 Discussion

Nous avons formalisé dans ce chapitre divers algorithmes de reconnaissance en terme de primitives géométriques et montré comment on pouvait les modifier pour qu'ils prennent en compte l'erreur de façon correcte. La contrepartie est l'apparition de faux positifs et nous avons montré avec une étude très qualitative que l'emploi de primitives plus complexes mais plus sélectives était pleinement justifié pour les algorithmes de reconnaissance. Ce type d'étude serait d'ailleurs à affiner pour pouvoir calculer et propager la probabilité de faux positif dans les divers opérations des algorithmes de mise en correspondance. Par contre, un certain nombre de points délicats sont encore à résoudre pour pouvoir utiliser couramment ces algorithmes dans le cadre générique que nous avons développé dans ce manuscrit.

Nous avons identifié en particulier quatre points clés. Le premier est lié à la structure des invariants et nous l'avons relié à la théorie des espaces de forme. Nous pensons qu'il serait ainsi possible, non seulement de caractériser les variété mises en jeu, mais aussi leur structure métrique induite. Le second concerne la classification dans une variété, et plus particulièrement en présence d'une information d'incertitude sur chacune des mesures. Nous avons développé une solution à ce problème mais elle n'est pas entièrement satisfaisante. Les deux derniers points clés sont liés à l'efficacité algorithmique des problèmes de recherche. Il s'agit dans un cas de recherche le point le plus proche, au sens de la métrique riemannienne ou de la distance de Mahalanobis, et dans l'autre cas de recherche les primitives (probabilistes) compatibles avec une primitive donnée, au sens du χ^2. Nous avons pu observer que les techniques de hachage posaient des problèmes de complexité

très important, d'une part à cause de la gestion de l'erreur et, d'autre part, à cause de la topologie non plane des variétés concernées. Ces problèmes deviennent cruciaux avec l'augmentation de la dimension de la variété. Par contre, nous pensons que certaines techniques de subdivision adaptative de l'espace pourraient être généralisables relativement facilement à nos variétés homogènes, en utilisant les propriétés des cartes exponentielles.

Enfin, pour finir, nous avons montré avec une étude très qualitative de la probabilité de faux positifs que l'emploi de primitives plus complexes mais plus sélectives était pleinement justifié pour les algorithmes de reconnaissance. Ce type d'étude serait d'ailleurs à affiner pour pouvoir calculer et propager la probabilité de faux positif dans les divers opérations des algorithmes de mise en correspondance.

Chapitre 11

Problèmes de reconnaissance en biologie moléculaire

Ce chapitre a été publié dans *Shape and Pattern Matching in Computational Biology* en 1994 (Pennec et Ayache, 1994; Pennec et Ayache, 1998) et a été l'une des motivations majeures du développement de la théorie présentée dans ce manuscrit. La section (11.4.1.1) est une expérimentation récente utilisant les résultats du chapitre 8 et qui confirme le gain de sélectivité obtenu grâce à la gestion rigoureuse de l'incertitude sur les repères.

Résumé

La plupart des actions biologiques des proteines dépendent de parties précises de leur structure 3D que l'on nomme *motifs*. Pour decouvrir automatiquement les motifs 3D se correspondant entre proteines, nous proposons un nouvel algorithme de reconnaissance et recalage de sous-structures 3D basé sur des techniques de geometric hashing. Le point clé de la méthode est l'introduction d'une *base euclidienne 3D* attachée à chaque acide aminé, ce qui permet de calculer 6 invariants par couple d'acides aminés, et par là même de réduire drastiquement la complexité, typiquement en $O(n^2)$.

Une analyse précise de la propagation des erreurs dans l'étape de prétraitement et l'introduction du filtre de Kalman étendu assurent l'efficacité et la robustesse lors de la reconnaissance.

Nos résultats expérimentaux confirment la validité de l'approche et la stabilité remarquable des 6 invariants utilisés pour le matching. Nous croyons que ce nouvel algorithme permettra dans un futur proche la comparaison systématique de très grandes structures grâce à la réduction de la complexité qu'il autorise.

11.1 Introduction

11.1.1 Background

Most biological actions of proteins, such as catalysis or regulation of the genetic message (transcription, maturation, etc.) depend on some typical parts of their three-dimensional structure, called

3D *structural* or *binding motifs*.

Proteins with similar 3D motifs often show similar biological properties, and it is therefore highly desirable to search for similar 3D motifs between proteins (Branden et Tooze, 1991). Since proteins are composed of thousands of atoms, this search requires efficient and fully automated methods. Applications in biology and medicine are immense, ranging from drug design and disease prediction to protein design and engineering, without forgetting research on the understanding of action mechanisms.

The protein structure is typically modeled at three different scales:

- the primary structure is the linear succession of amino acids,
- the secondary structure is the *local* organization of the chain into structural motifs,
- and the tertiary structure is the *complete* description of the positions of atoms in space, and can reveal the presence of 3D non linear motifs.

Most of existent techniques for comparing proteins deal with the comparison of the primary structure only, using character string comparison algorithms, and especially dynamic programming based ones (Myers, 1991; Lacroix et Codani, 1991). These approaches can not easily find most of the structural or binding motifs, whose nature is intrinsically 3D. For instance, polypeptide chains that form a specific structural motif frequently show no or very low sequence homology, even when they are associated with the same specific function (Branden et Tooze, 1991, page 99). Hence, only the comparison of the tertiary structures of the chains can reveal 3D correspondences.

Moreover, the availability through international data banks of an everyday increasing number of 3D structures allows for the systematic search of motifs present in large sets proteins, provided that efficient and fast 3D substructure matching algorithms do exist.

In this spirit, Fischer and his colleagues (Fischer et al., 1992a; Fischer et al., 1992b) have exploited the geometric hashing paradigm previously introduced in computer vision by Lamdan and Wolfson (Lamdan et Wolfson, 1988; Wolfson, 1990). They proposed substructure matching methods based on preprocessing and recognition algorithms of complexity $O(n^3)$, where n is the number of atoms of interest (either in the motif or in the protein). A key point of their approach is the possibility to refer to 2 rigid invariants (the "distance coordinates") of any atom of the protein with respect to two other atoms picked arbitrarily as forming a geometric "basis". The results reported in their publications were encouraging, and motivated our work.

11.1.2 An $O(n^2)$ 3D substructure matching algorithm

Following their pioneering work, our main idea was to reduce the size of a "basis" from two to a single atom. To achieve this goal, we introduce a *3D reference frame* attached to each amino acid. Doing this, we can now choose a single amino acid as a basis, and compute 6 rigid invariants (the parameters of translation and rotation) attached to any other amino acid. This allows to drastically reduce the complexity of both the preprocessing and recognition stages of geometric hashing, typically from $O(n^3)$ to $O(n^2)$.

A thorough analysis of the propagation of the errors in geometric hashing (Pennec, 1993a) and the introduction of extended Kalman filtering for the clustering of found transformations (Ayache, 1991; Guéziec et Ayache, 1992) guided our implementation to insure efficiency and robustness of the approach.

Our experimental results confirm the validity of the approach, and the remarkable stability of the 6 rigid invariants used for matching. We believe that this new algorithm, because of the reduction of the algorithmic complexity it implies, will allow the systematic comparison of very large structures in the near future.

Our paper is organized as follows: first we detail the reference frame attached to each amino-acid, and then describe the new geometric hashing algorithm we propose for matching. Third we report our experimental study, and finally we present a few potential extensions of our work.

11.2 Protein structure modeling

Proteins are composed of possibly several chains of amino acids linked each others by peptide bonds. Three groups of atoms in each amino acid constitute the backbone of the chain : the central atom C_α to which are attached on each side an NH group and a carbonyl group $C' = O$ (see figure 11.1). The residue R, also bound to the C_α, characterizes the nature of the amino acid but does not take part to the backbone of the chain.

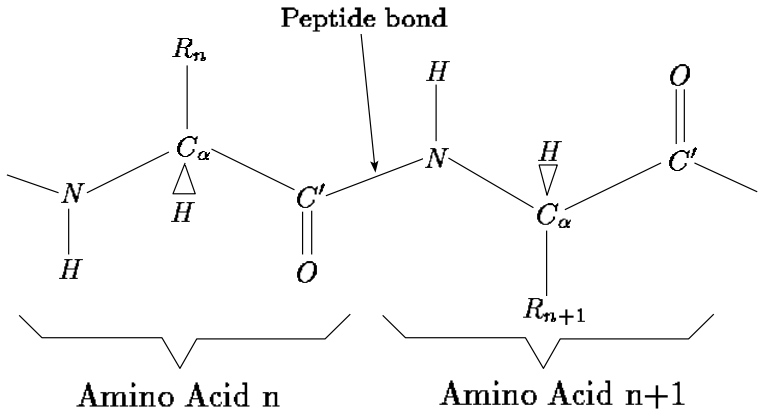


FIG. 11.1 – *Structure of a Protein chain : a peptide bond links amino acids between each others in a linear way.*

Topologically, the backbone of the chain is linear, but its geometry is very complex. Even if rotations are allowed around the bonds $C_\alpha - C'$ and $C_\alpha - N$, and hence the geometry of the chain is weakly constrained, the geometry of the atoms attached to the C_α is perfectly determined. In particular, the three atoms N, C_α, C' form a known triangle from which we can define a basis (see figure 11.2). We shall see later the function of this basis. It is sufficient to say for the moment that

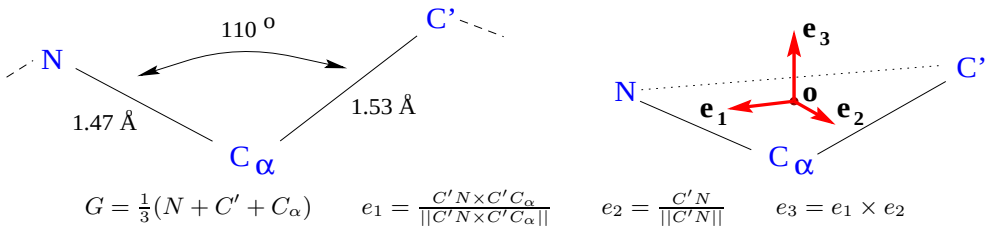


FIG. 11.2 – *Geometry of an amino acid around the C_α and definition of a basis.*

this basis uniquely defines the position and orientation of the amino acid in space. We will hence model an amino acid by a couple (point, trihedron) with possibly a label (the type of amino acid), and a protein by the set of these couples.

The structure comparison problem is thus stated as follows : given two such sets, find all rigid transformations that match a minimum number of amino acids of the two structures. The problem

can be extended to the comparison of a target molecule with a data-base of proteins. We will see that the geometric hashing paradigm is especially well suited for such a research.

11.3 Matching proteins

The problem we are confronted with is very close to recognition problems in volume image analysis, especially in the medical field. In this case, one has to process points extracted from surfaces with their associated Frenet trihedron (Thirion, 1993; Ayache, 1993; Guéziec et Ayache, 1993). So in both cases, the model adopted to reduce the data is a set of couples (point, trihedron). Classical techniques rely on model-based approach of object recognition. See for instance (Grimson, 1990) for a survey. Given a data-base of modeled objects (called models), the aim is to recognize in a scene what objects are present, and how they are placed. The simplest problem where the data-base is reduced to only one object is called simply matching or sometimes registration.

11.3.1 The geometric hashing algorithm

The geometric hashing algorithm was introduced by Lamdan and Wolfson for model based recognition in computer vision (Lamdan et Wolfson, 1988; Wolfson, 1990). The basic idea is to store in a data-base at preprocessing time a redundant representation of models, based on local features to allow for occlusion, and invariant by rigid transformation. Doing so, the representation of the scene computed at recognition time will present some similarities with that of some objects of the data-base. Accumulating these evidences will allow the recognition and registration of objects present in the scene and in the data-base. This approach can also be related to shape autocorelation operators (Califano et Mohan, 1991).

– *Invariant description*
In our case, local features are frames. However, in the absence of labels, any model frame can be matched with any scene frame. To obtain an invariant description, we thus have to consider binary constraints between frames. Indeed, a couple of frames has 6 invariants given by the parameters of the rigid transformation from the last frame to the first (figure 11.3). For ease of work, we use in fact the 3 coordinates of the translation (expressed in frame i), and the 3 angular values between the corresponding axis of the two frames. These values will constitute the 6D invariant vector of a frame couple. In order to deal with occlusion, the representation

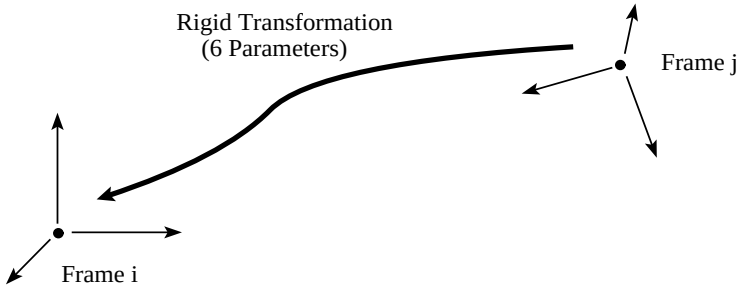


FIG. 11.3 – *If frame i is considered as the basis of the Euclidean space (reference frame), the 6 parameters of the rigid transformation mapping frame j on the origin (frame i) are invariants.*

of one frame has to be redundant: each frame will then be associated with any other frame of the object to compute the set of 6D invariant vectors characterizing this reference frame (see

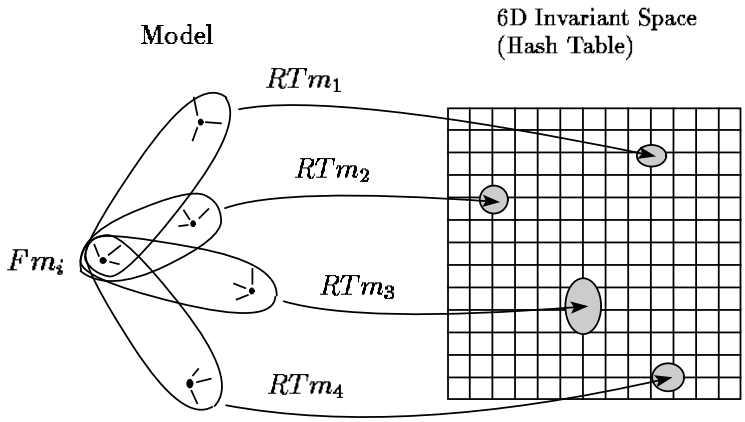


FIG. 11.4 – *Preprocessing: the 6D invariant vector associated with every couple of frames is computed with its error zone (see Error handling section) and used as an index for the couple in the hash table. For the sake of simplicity, we only deal here with the representation of model frame i (Fm_i).*

for example figure 11.4). The global representation is then the set of every frame couple of the model, each one being an entry for the hash table, with the 6D invariant vector as index.

– *Preprocessing*
In order to optimize the access to the representation for recognition time, the geometric hashing algorithm uses a hash table for storing models. Indeed, given one object, just compute the 6D invariant vector associated with each possible couple (reference frame, other model frame), and set it as an index in a 6D hash table for the couple. Each model is processed independently, but stored in the same hash table. The complexity of the step is $O(M.m^2)$, where M is the number of models and m the mean number of amino acids per model. Typical values for m range from 15 for motifs to a few hundreds for big proteins. The complexity in space for the hash table is the same since it only depends on the number of entries. This step is performed without any knowledge of the scene to be matched and hence can be done once for all.

– *Recognition*
Choose a reference frame; for each different scene frame, compute the 6D invariant vector and retrieve the compatible model couples (reference frame, other model frame) in quasi constant time thanks to the hash table. During the process, maintain a list of the model reference frames found, and for each one accumulate the number of compatible couples. This will be the score for the matching of these model reference frames with the considered scene reference frame.

The process is repeated with each scene frame taken as the reference frame. The output is the list of model and scene matching reference frames with their associated score (see figure 11.5). We only keep the matches with a score above a threshold. This parameter is either static or dynamically adapted during the algorithm. It is also possible the keep a fixed number of matches (usually the best ones).

The simple matching of a model and scene reference frames is sufficient for computing afterward a rigid transformation, but every compatible couple can brings up some additional information (the matching of secondary frames) that can be used to refine it at a small cost during this process, using for example an extended Kalman filter.

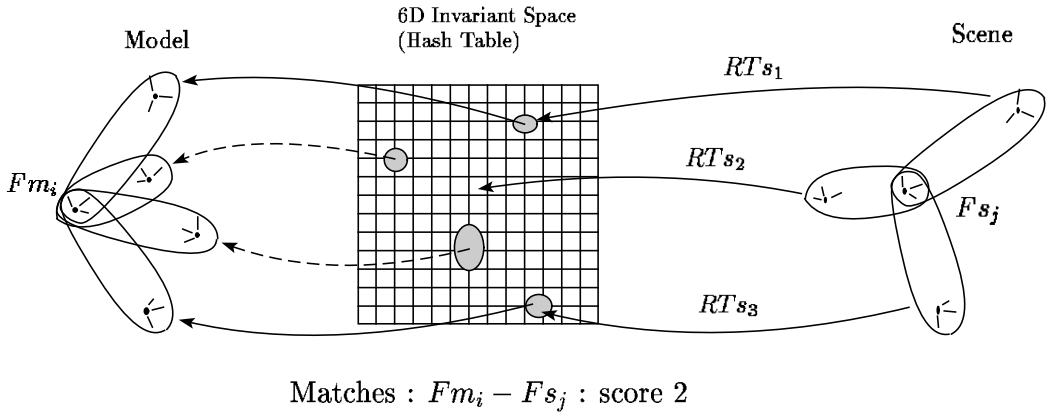


FIG. 11.5 – *Recognition: for each scene frame couple, compute the 6D invariant vector and retrieve through the hash table every compatible model frame couple. For each such couple, tally a vote for the matching of the reference frames (here Fm_i - Fs_j).*

Considering that the access to the hash table is done in constant time, which is true for well distributed tables, the complexity of the whole stage is $O(n^2)$ where n is the number of frames (amino acids) in the scene.

– *Error handling*

Due to the resolution of the X-ray determination of protein structure, conformational deformations and even structural differences between molecules that induce different constraints on the motif, one has to deal with errors in atom positions. A previous study on the propagation of errors within geometric hashing shows that, considering a bounded error ε on the position of every atom, one can determine bounds on the error for the invariant vector used to hash (Pennec, 1993a). We do only provide here the practical bounds we used and redirect the interested reader to the original paper for further developments. Let r be the radius of the inscribed circle to the triangle used to form the reference frame, and d be the distance between the two amino acids within the considered couple. The error in the coordinates of the amino acid in the reference frame (first part of the 6D vector) is bounded by $\varepsilon_{gh} = \varepsilon\left(1 + \frac{d}{3r}\right)$. A strong supplementary constraint is given by the bound $\varepsilon_d = 2\,\varepsilon$ on the distance d. Concerning the 3 angular values (second part of the 6D vector), a theoretical error bound on each angle is $\varepsilon_\theta = \frac{2\,\varepsilon}{l} + O(\varepsilon^2)$ where l is the length of the considered triangle edge, but half this bound is a good practical value. As numerical values, we compute that $r = 0.38$ Å and use an average value $l = 1.7$ Å for triangle edges. Each bucket of the hash table intersecting the volume error of a 6D invariant vector is defined as an entry for the couple. Since the hash table can be coarsely discretized, we verify at recognition time that the couples found have compatible invariant vectors.

As long as the hash table is sparsely distributed, this bounded error model for geometric hashing is sufficient. When it comes to very big proteins or big data bases, a probabilistic scheme (Rigoutsos, 1992; Rigoutsos et Hummel, 1993) would be more adapted.

11.3.2 Clustering and extension

From now on, we use a probabilistic scheme. Hence, each amino acid frame is given an associated covariance matrix, which is propagated through the computations. The original covariance matrix

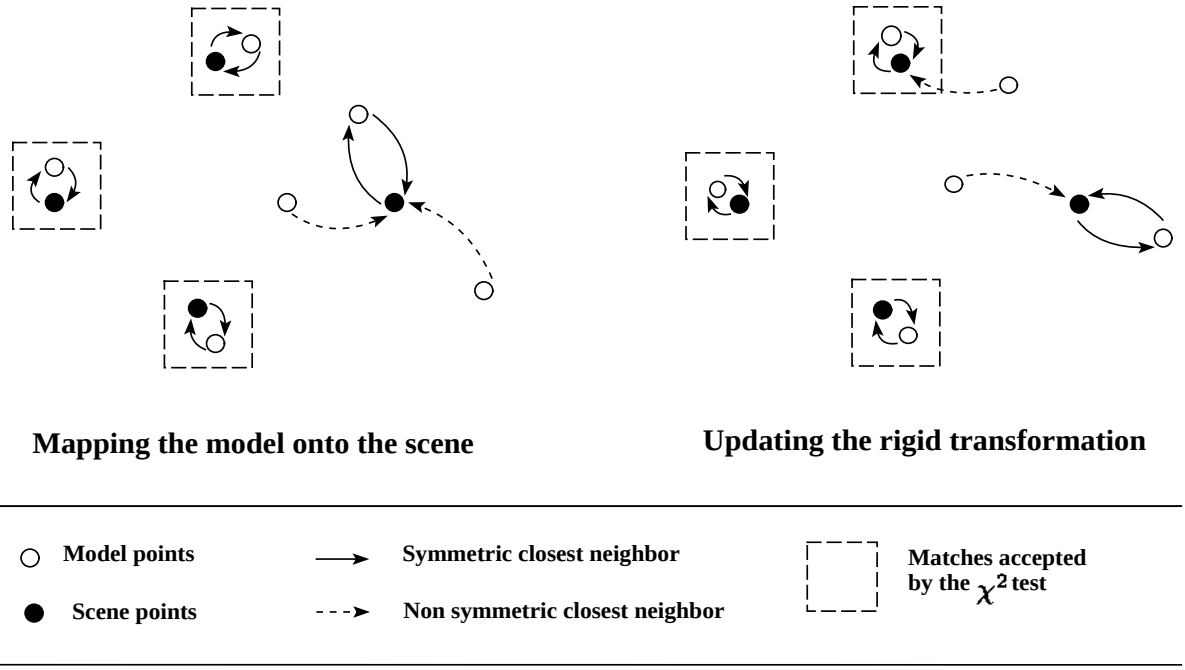


FIG. 11.6 – *Left: the model is mapped onto the scene and three matches are satisfying the symmetric closest point and χ^2 test constraints. Right: these three matches are used to recompute the rigid transformation and update the position of the model. Since the same matches are satisfying the constraints, the algorithm can stop.*

is diagonal with a value σ_d^2 on the translation vector and σ_θ^2 for error on the rotation vector. In our experiments we generally use $\sigma_d = 0.5$Å and $\sigma_\theta = 0.2$ rad.

- *Clustering*
  We have to aggregate the redundant information obtained from the geometric hashing stage. In order not to mix up the different possible models, we split the list of matching reference frames into one list for each model. For each such list, we consider the first match as a cluster, and examine each different match as follows:
  - Decide with a χ^2 test on Mahalanobis distance between the two transformations if they are compatible.
  - If so, merge them together using the extended Kalman filter : the cluster transformation is updated with the new 3 pairs of atoms matched together.

  Repeat the process until each match be incorporated into a cluster. The complexity is $O(k_m.k_c)$ with a number k_m of matches and k_c of clusters on output, but k_c is quasi constant in practice, a few tens at the very most.
- *Extension*
  Clusters must now be checked and their matching list extended. This is done using an alignment test: using the rigid transformation previously determined, the model is mapped onto the scene and the possible matches are verified. For the sake of simplicity, only the position of the C_α is now considered in each amino acid. Each C_α of the model is examined as follows (see figure 11.6).

- Map the model C_α onto the scene and search for the closest C_α of the scene. In order to keep the algorithm symmetrical between the model and the scene, map back the scene C_α to the model and verify that the original model C_α is its closest neighbor. If not, reject the model C_α.
- Compute the Mahalanobis distance between the transformed model C_α and the scene one, and decide using a χ^2 test if this match is valid. If not, reject the model C_α.
- Update the rigid transformation of the cluster with this new match using the extended Kalman filter.

This process is repeated until convergence (stability of matches) or a maximum number of iterations (usually about 10). Seeking the closest neighbor is performed using k-D trees (Preparata et Shamos, 1985). The complexity of constructing a k-D tree is $O(n.\log n)$ with $O(n)$ storage. The search for a closest neighbor is sub-linear, and almost constant in practice. Hence, the whole stage has a complexity $O(n.\log n + n.k)$.

11.3.3 Algorithm analysis

- *Simplifications*: The following heuristics can help speeding up the geometric hashing step:
  - Only keep couples with inter-distances under a threshold (typically around 20 Å). The underlying justification is that amino acids of the motif are usually close in space.
  - Label the reference frames (or even every amino acid) with their residue name. Indeed, since we only seek in this step some initial matches, we do not need to obtain every match and it can be sufficient to consider amino acids with the same residue. This is however not used in our experiments.
- *Complexity*: Let m be the mean number of amino acids in the models, M the number of models, n the number of amino acids in the scene, and k the mean number of clusters found by model. The theoretical complexity is $O(M.m^2)$ for the preprocessing stage (comprising the k-D trees preprocessing of models). The whole recognition stage is in $O(n^2 + M.k.n)$. In fact, as far as the number of models M is low compared to n, the real complexity is dominated by the geometric hashing stage: $O(n^2)$.
- *Parameters*: There is a small number of parameters that need to be adjusted in the algorithm, such as the bound ε on atom coordinates errors and the variances σ_d and σ_θ on the associated frame, the threshold on the score for geometric hashing, and the thresholds χ_c (clustering) and χ_v (verification) used for χ^2 tests. We evaluate the values of these parameters in a learning step: knowing two matched motifs, we register them and compute the variances and the bound ε. Using these informations, we compute in a second step the minimal thresholds. In order to keep some control over the algorithm, we choose to parameterize the variances by the bound ε in a linear way and keep the χ values constant. Hence, in most cases, the only parameters we have to play with are ε and the minimal score for geometric hashing.

11.4 Experimental results

For all our experiments, we use the atoms coordinates of proteins provided by Brookhaven National Laboratory's Protein Data Bank (Bernstein et al., 1977; Abola et al., 1987). Visualization is done using the RasMol program of R. Sayle (Sayle et Bissel, 1992). For rigid transformations, we provide the translation in angströms and the rotation vector in radian. Experiments were done with

and without the distance constraint of section (11.3.3) without any difference. The labeling scheme was not used, and hence amino acids are not discriminated in the process.

11.4.1 Detection of a structural motif: the Helix-Turn-Helix motif

Structural motifs can be defined as the super-secondary structure. They are the simple combination of a few secondary structure elements. Some of them are associated with particular functions or are simple parts of larger structural and functional assemblies. Therefore, the Helix-Turn-Helix motif is responsible for the binding of DNA within many procaryotic proteins. Some of them bind tightly to the DNA at a promoter of a gene, preventing RNA polymerase from fixing and hence blocking the initiation of the transcription. They are repressors. Conversely, activators bind next to the promoter and help polymerase to bind.

We choose to compare the tryptophan repressor for E. Coli (PDB code 2WRP (Lawson et al., 1988)) and phage 434 CRO (PDB code 2CRO (Mondragon et al., 1989)), whose Helix-Turn-Helix sequence are known to be (Brennan et Matthews, 1989; Harrison et Aggarwal, 1990) :

Protein	*Position*	*Sequence*		
$2CRO$	$15-37$	MT	$QTELATKAGV$	$KQQSIQLIEAG$
$2WRP$	$66-88$	MS	$QRELKNELGA$	$GIATITRGSNS$

In this experiment, we choose the whole protein 2CRO as the model and look for common substructures with 2WRP: the two corresponding sequences are correctly matched, and 4 other non linear matches are found (table 11.1).

```
Cluster 1 : score 27
Rotation Vector : -2.53546   0.291995  0.345864
Translation     : -17.9746  -1.90902  -0.211789

   ---------------       15 MET - 66 MET       24 ALA - 75 LEU       33 LEU - 84 ARG
     2CRO - 2WRP         16 THR - 67 SER       25 GLY - 76 GLY       34 ILE - 85 GLY
   ---------------       17 GLN - 68 GLN       26 VAL - 77 ALA       35 GLU - 86 SER
    9 ARG - 63 ARG       18 THR - 69 ARG       27 LYS - 78 GLY       36 ALA - 87 ASN
      ...       ...      19 GLU - 70 GLU       28 GLN - 79 ILE       37 GLY - 88 SER
   11 ILE - 64 GLY       20 LEU - 71 LEU       29 GLN - 80 ALA          ...       ...
      ...       ...      21 ALA - 72 LYS       30 SER - 81 THR       43 ARG - 50 ALA
   15 MET - 66 MET       22 THR - 73 ASN       31 ILE - 82 ILE          ...       ...
   16 THR - 67 SER       23 LYS - 74 GLU       32 GLN - 83 THR       45 LEU - 53 THR
                                                                     ---------------

Cluster 2 : score 21
[ .............. ]
```

TAB. 11.1 – *Detected matches between the proteins 2CRO and 2WRP. The output of the algorithm is compacted for a better understanding.*

We do not assess any biological significance to these supplementary matches. They are only the result of a geometric matching. On the other hand, the fact that other matches do not score more than 21 shows that the HTH motif is the main common structural motif between the two proteins.

11.4.1.1 Improvements of the algorithm

The last step (of the algorithm (the extension) is an alignment test realized as an improved iterative closest point (Besl et McKay, 1992) : the matching step looks for the symmetric closest point and the registration step discards outliers using a χ^2 test and recomputes the transformation. These two steps were previously realized only with points : we keep the matching part on points (we

would like to be able to do it directly on frames), but we use the outlier rejection and registration scheme of section (8.5) on frames using a fixed noise model that determines the type of common substructures the algorithm extracts : a small noise only keep a small but very accurate set of matches, whereas a large noise authorizes more uncertain matches, and thus favours bigger and more global substructures matches.

Looking for a 3D binding motif, we used here a quite small model of isotropic noise ($\sigma_\theta = 0.1\,\text{rad} = 5\,\text{deg}$ and $\sigma_d = 0.35$ Å). The algorithm ends up with only the correct matches from (15 MET - 66 MET) to (36 ALA - 87 ASN), discarding the four supplementary matches because they have different orientation. The last match (37 GLY - 88 SER) is also eliminated, and is indeed very arguable considering the distance after registration and especially the difference in orientation. The typical object precision due to the registration (on the 22 matched amino-acids) is given to $\sigma_{obj} = 0.29$ Å. We show in figure (11.7) the two proteins and in figure (11.8) the registration found. The only two other common substructures found score now 13 and 8 matches and correspond to alpha helices, which are very stable secondary structure elements.

In order to test the stability of our algorithm, we also did the experiment with a noise two times larger ($\sigma_\theta = 10\,\text{deg}$ and $\sigma_d = 0.7$ Å). We just find four additional matches preceding the beginning of the HTH motif, from (7 LYS - 61 LEU) to (11 ILE - 64 GLY). The two other clusters now score 14 and 9 matches.

This shows that the plain use of frames improves the robustness of motifs detection (the previous algorithm using only points was far more sensitive to the noise model). Indeed, the orientation of an amino acid is crucial to determine the position of collateral chains and most interactions of the protein happen within these side atoms. The position of these atoms is then not only determined by the position of the backbone but also its orientation and using just points to represent amino acids generally leads to a significant amount of additional matches with non compatible orientations. This implies a drastic reduction of selectivity for the matching process. In this case, frames bring rather more selectivity than more precision.

11.4.2 Detection of a binding site : the heme pocket

The globin family collect proteins of many different organisms. Their amino acid homology can be as low as 16%. However, their 3D structure is still related and constitute the globin fold. It is mainly composed of α helices that form a pocket for the active site which in myoglobins and hemoglobins binds a heme group. The motif constituted by the amino acids binding the heme in 9 globins was extensively studied by Lesk *et al.* (Lesk et Chothia, 1980).

We define the motif by 15 of the 19 non sequential positions of amino acids that make contact with the heme in the α subunit of human hemoglobin (PDB code 4HHB, chain α (Fermi et al., 1984)). We choose the 15 positions given by Lesk *et al.* as being present in 7 or more globins. We search for it within the β subunit of the same protein (chain β), horse hemoglobin (PDB code 2DHB (Perutz et al., 1973), chain α and β), myoglobin (PDB code 4MBN (Takano, 1984)), and sea lamprey cyanohemoglobin (PDB code 2LHB (Honzatko et al., 1985)). The resulting matches are given in table (11.2). We present in figure (11.10), (11.11) and (11.12) images of the motif, the β chain of human hemoglobin, and the registration of the two structures.

The matching with horse hemoglobin (2DHB) was done with the two chains together and the two matches were perfectly identified. For the sea lamprey cyanohemoglobin (2LHB), the coordinates reported by (Honzatko et al., 1985) are not the same as those used by to Lesk *et al.*: a SER is deleted after 96-SER, and 98-LEU, 99-ARG were inserted (see remark 6 of the PDB entry). Hence residue 101 from PDB correspond to residue 100 of Lesk *et al.* and so on.

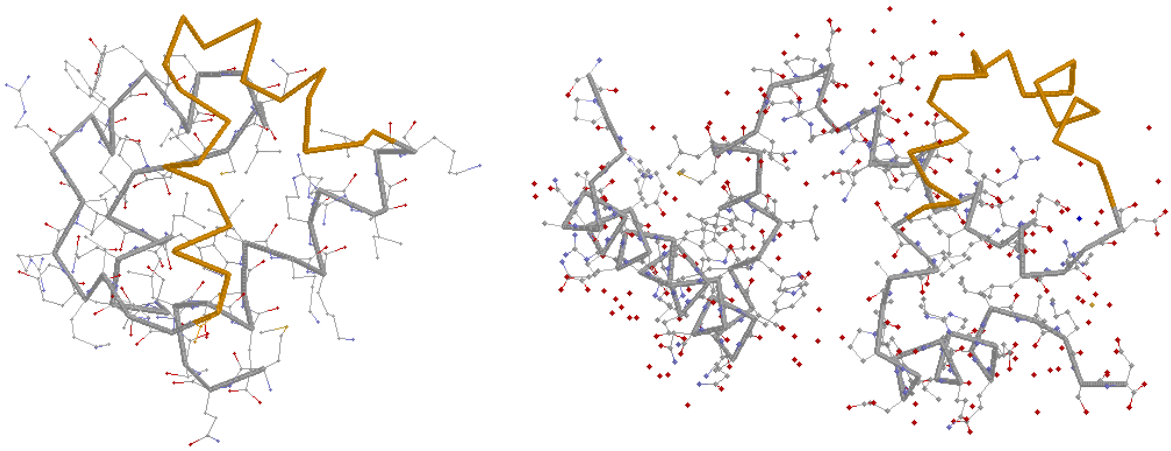


FIG. 11.7 – *The CRO protein (2CRO) of phage 434 on the left and the tryptophan repressor of E. Coli (2WRP) on the right. The matched part (the HTH motif) is displayed in black.*

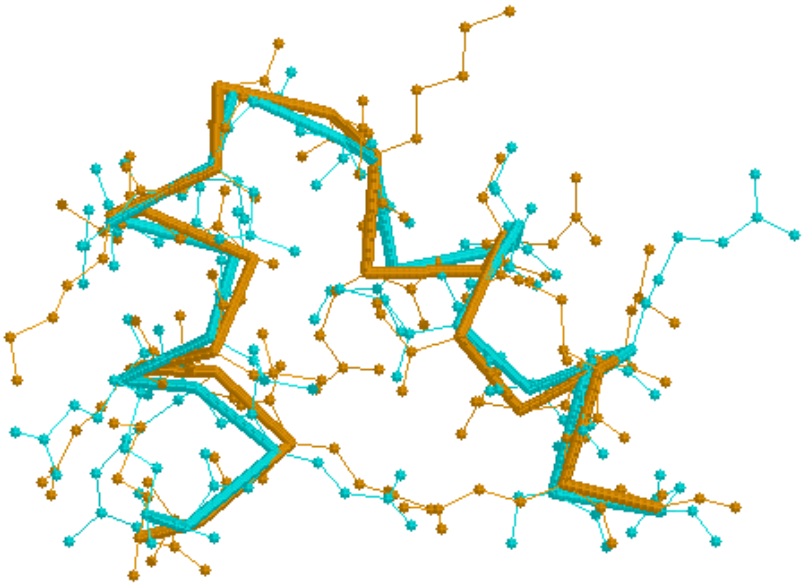


FIG. 11.8 – *Registration found between the HTH motifs of 2CRO and 2WRP. We can see that not only the backbone is very well matched, but also collateral chains are pretty well conserved.*

4HHB α (motif)	4HHB β	2DHB α	2DHB β	4MBN	2LHB
42 TYR	41 PHE	42 TYR	41 PHE	42 LYS	51 PHE
43 PHE	42 PHE	43 PHE	42 PHE	43 PHE	52 PHE
58 HIS	63 HIS	58 HIS	63 HIS	64 HIS	73 HIS
61 LYS	66 LYS	61 LYS	66 LYS	67 THR	76 ARG
62 VAL	67 VAL	62 VAL	67 VAL	68 VAL	77 ILE
65 ALA	70 ALA	65 GLY	70 SER	71 ALA	80 ALA
66 LEU	71 PHE	66 LEU	71 PHE	72 LEU	81 VAL
83 LEU	88 LEU	83 LEU	88 LEU	89 LEU	101 LEU
86 LEU	91 LEU	86 LEU	91 LEU	92 SER	104 LYS
87 HIS	92 HIS	87 HIS	92 HIS	93 HIS	105 HIS
91 LEU	96 LEU	91 LEU	96 LEU	97 HIS	109 PHE
93 VAL	98 VAL	93 VAL	98 VAL	99 ILE	111 VAL
97 ASN	102 ASN	97 ASN	102 ASN	103 TYR	115 TYR
98 PHE	103 PHE	98 PHE	103 PHE	104 LEU	116 PHE
101 LEU	106 LEU	101 LEU	106 LEU	107 ILE	119 LEU
T_x	0.142694	0.493322	-0.04313	3.44391	10.1853
T_y	-1.59334	1.36633	-0.39684	14.5187	18.8873
T_z	2.60057	0.162643	2.35737	-6.03178	-14.3859
R_x	-3.14764	0.004991	3.11788	1.13209	1.73283
R_y	-0.007530	0.033864	-0.03635	-0.672907	1.58384
R_z	0.0802843	-0.016262	-0.127903	0.015026	0.113322

TAB. 11.2 – *Matches and transformations resulting from the algorithm. The transformations map the motif onto the scanned structures.*

In these experiments, no other match scoring more than 5 was found. This tends to show that the correct recognition greatly emerges from the noise of false positives, and hence our scheme appears to be very robust.

11.4.3 Accuracy of frames

With the result of these two experiments, we can study the accuracy of our modeling with respect to the variability of amino acids. In figure (11.9) we draw the distribution of distance and angular error between matched frames in 2CRO and 2WRP (HTH experiment).

We can see that if two matched amino acids can be as far as 1.2 Å, their orientation differs from a maximum of 25^o, with a RMS of 16^o. As far as the heme experiments are concerned, the maximum and RMS distances are generally larger, but the angular error does not grow (table 11.3).

		4HHB	2DHB α	2DHB β	4MBN	2LHB
Distance	RMS	0.9	0.3	0.8	0.75	0.8
	Max	1.3	0.6	1.2	1.4	1.3
Angular error	RMS	$\pi/13$	$\pi/12$	$\pi/10$	$\pi/11$	$\pi/13$
	Max	$\pi/9$	$\pi/6$	$\pi/6$	$\pi/5$	$\pi/6$

TAB. 11.3 – *Error values for the heme experiment.*

This study mainly shows that the orientation of amino acids is at least as stable as their position in a motif. This was in fact predictable since the orientation of an amino acid determines in a certain

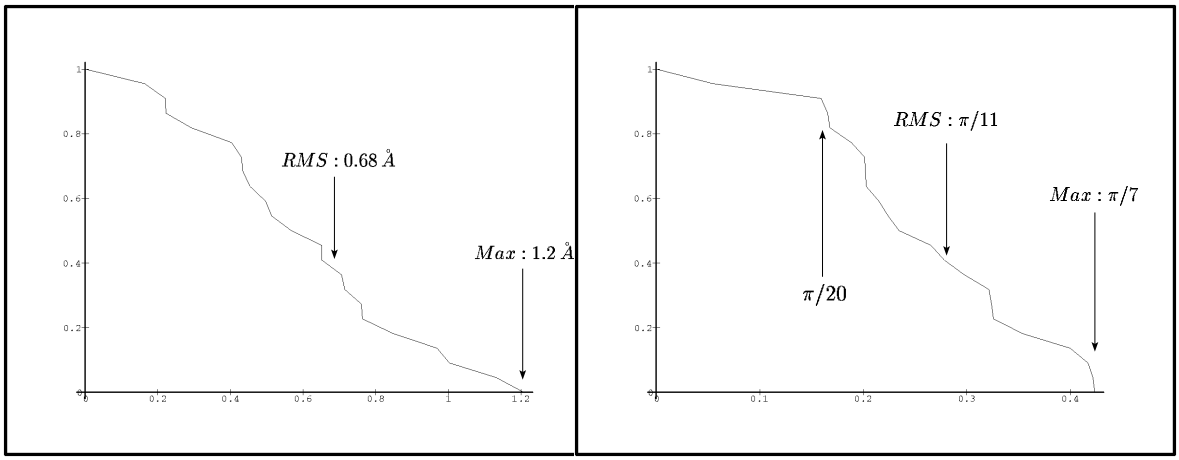


FIG. 11.9 – *Distribution of distance (left) and angular error (right) between matched amino acids in 2CRO and 2WRP.*

way the position of its residue. Our modeling with frames is then justified and allows to use this reliable information.

11.5 Conclusion

Modeling amino acids by the 3 atoms of their backbone allows to define a complete and unique associated reference frame, which turns out to be very stable. Each couple of amino acids has hence 6 invariants for rigid transformations that we use in a geometric hashing scheme to discover initial matches. These are clustered, verified and extended. The error inherent to the problem is integrated in the process, thanks to an error analysis and Extended Kalman Filter. Experiments confirm the validity, efficiency and robustness of our approach.

This algorithm is also currently working for substructure matching in volume images (medical images) with frames extracted from surfaces (extremal points). This stresses the analogy between 3D matching problems and points out the fact that frames can, in numerous cases, advantageously replace points.

Future work will be articulated upon three axes. We plan first to use a probabilistic scheme for geometric hashing, for instance (Rigoutsos et Hummel, 1993). A second direction would be the extension of our algorithm for multiple alignments research and last but not least, we hope to demonstrate the possibility of automatically discovering and extracting the model of an unknown motif common to a given group of proteins.

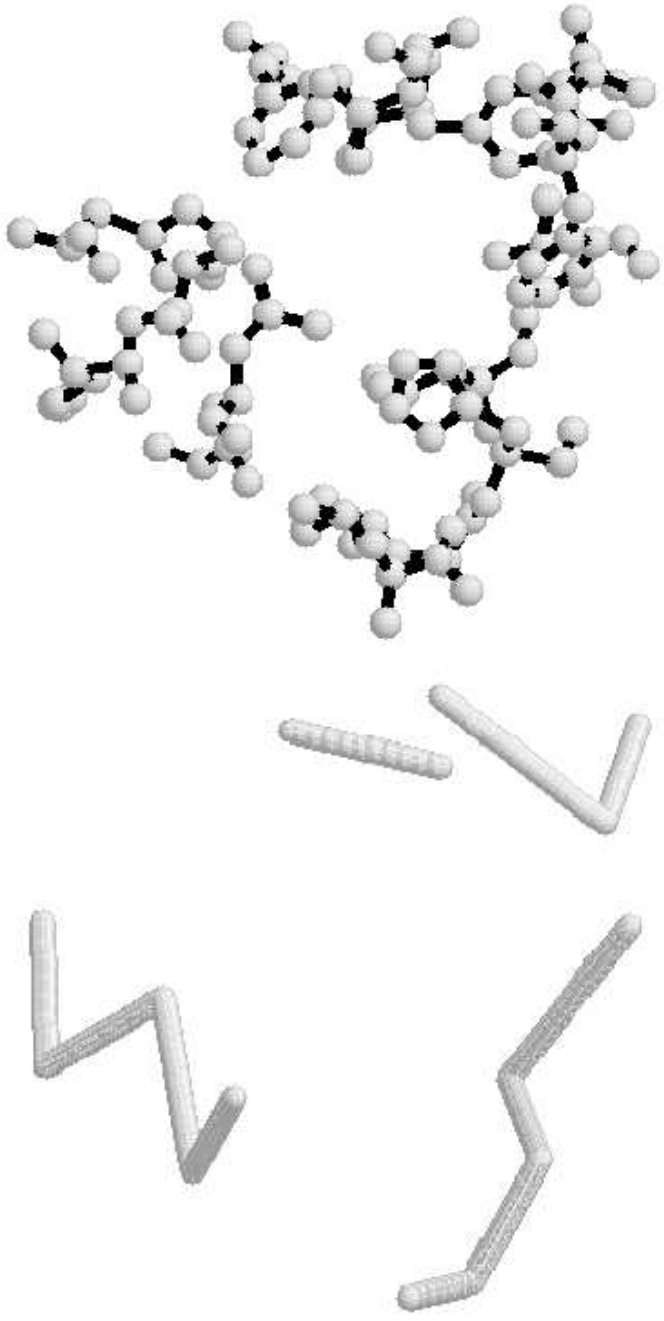


FIG. 11.10 – *The motif extracted from the α chain of human hemoglobin (4HHB) displayed in balls and sticks (left) and its backbone (right)*

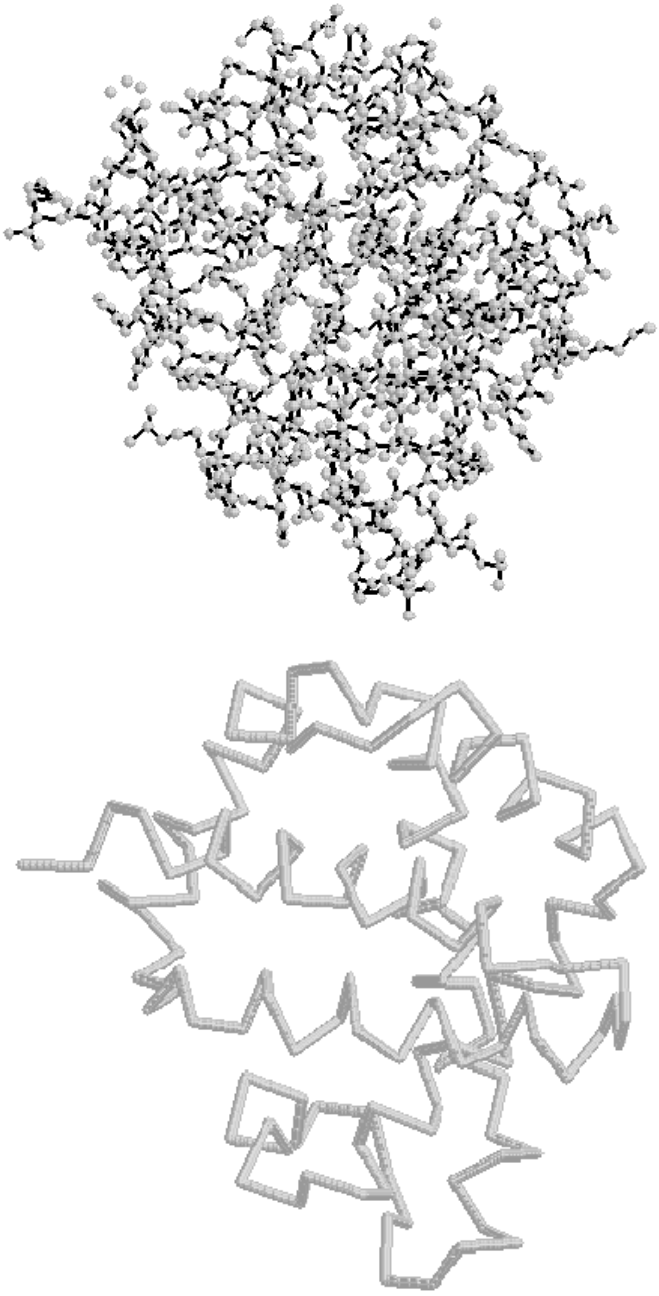


FIG. 11.11 – *The β chain of human hemoglobin (4HHB) displayed in balls and sticks (left) and its backbone (right)*

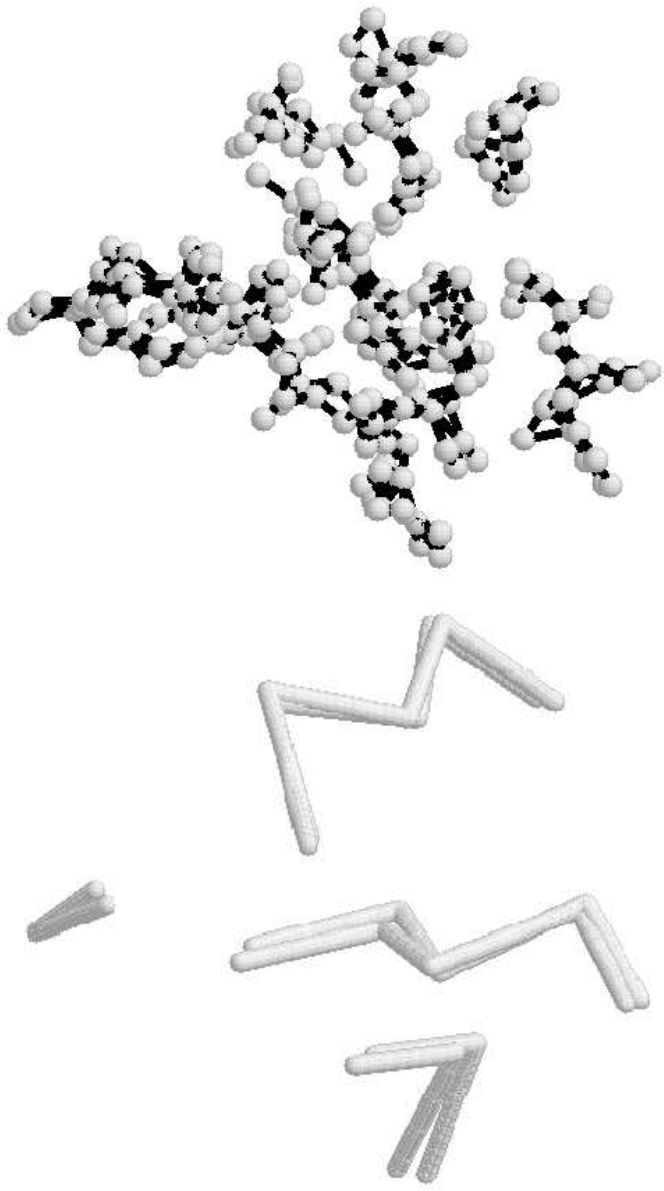


FIG. 11.12 – *Motif and chain 4HHB β registered. For the sake of visibility, we only show the matched residues. Balls and sticks (left) and backbone (right)*

Chapitre 12

Modélisation et recalage multiple

Ce chapitre a été publié dans *Image Fusion and Shape Variability Techniques* (Pennec, 1996). L'idée générale est d'étudier les techniques de recalage multiple et d'aborder la théorie des formes sur les points pour pouvoir généraliser ensuite ces méthodes à des primitives quelconques.

Summary

We present in this article a characterization of the mean rigid shape in any dimension, and propose an algorithm to compute it after a multiple registration step. In order to cope with the missing point problem (« occlusion »), we propose another iterative algorithm that alternates the computation of the mean shape and the registration of all objects to it. Experiments with synthetic and real data show a good convergence of both algorithms, the iterative one being much faster. We present two applications in 3D modeling with the mean shape of the heme binding motif (molecular biology) and the model of the head of one patient based on extremal points from 24 MRI images (medical imaging).

12.1 Introduction

To study the shape of an object, it is common to describe the object by a fixed number k of characteristic features in order to obtain a k-tuple $(x_1, \dots x_k)$. Two object are said to have the same form if there exists a transformation that can superimpose them, and thus identify their "characteristic" k-tuples.

General studies focuses on k-tuples of points and their form under rigid transformation with possibly scaling (Kendall, 1989; Goodall, 1991; Le et Kendall, 1993). There are fewer works dealing with more general transformations, apart from (Ambartzumian, 1982) for affine shapes, and most applications concern 2D landmarks (Bookstein, 1991; Bookstein, 1986; Small, 1988). In computer vision and medical imaging, however, some more complex features can appear, like lines, frames... and we need to work in full 3D. This type of constraints is also shared in structural biology, when it comes to the tertiary structure of the proteins.

In this framework, several problems can be identified. Firstly, the identification and measurement of meaningful features, or landmarks, is much more difficult in 3D than in 2D and can become a bottleneck of the shape studies if done by hand. Then, There is a need for an automatic extraction of characteristic features: the C_α position of an amino acid is usually considered as representative for the amino acid in the protein, and extremal points (Thirion, 1994) can be good candidates for landmarks. The remaining problem is how to discriminate meaningful features from spurious (or more often uninteresting) ones? In the medical imaging field, one can think of a scale-space approach to characterize features by their stability, but eventually, the saliency of features is their stability across a wide range of observations for a given "object". We are here confronted to a multiple matching and registration problem in the presence of noise and outliers.

Once meaningful features have been identified and matched in different observations, it is interesting to merge their observations in order to obtain a model that incorporate all meaningful information and has a reduced noise. The shape of this model can then be studied, compared with others, used to determine if a similar shape is present in an observation... The set of models in the world we consider will then constitute an atlas.

Algorithms for recognition and registration of two observations have already been developed for 3D substructure matching of proteins (Pennec et Ayache, 1998) and registration of medical images based on extremal points ((Thirion, 1994) and (Pennec et Thirion, 1997) for a validation). We are interested in this article in multiple registration and how to define a mean shape based on point features under rigid transformations.

We present in the first part a characterization of expected and mean rigid shapes in the Fréchet sense (Fréchet, 1948) which is basically the generalization to nD of some 2D result of Ziezold (Ziezold, 1994; Ziezold, 1989). We show that the mean rigid shapes can be computed independently after a multiple registration step of the k-tuples, and we propose an iterative method to solve the multiple registration problem. Experiments on synthetic data show that the algorithm behave well and reasonably fast.

In the real world, however, there are often occlusion and thus incomplete k-tuples: some usually stable features can be mismatched, highly displaced or simply forgotten in the extraction scheme in one or a few observations. It would be inauspicious to simply dismiss these features since they can bring useful information in most of the cases. We extend then in the second part the multiple registration and mean shape estimation scheme in order to cope with incomplete shapes.

In the third part, we present an application in molecular biology: the globin family collects proteins of many different organisms. Their amino acid homology can be as low as 16%. However, their 3D structure is still related and constitute the globin fold. It is mainly composed of α helices that form a pocket for the active site which in myoglobins and hemoglobins binds a heme group. The motif constituted by the amino acids binding the heme in 9 globins was extensively studied by Lesk *et al.* (Lesk et Chothia, 1980). In this example, the correspondences between amino acids implied in the heme binding are given, and we want to determine the mean shape of the binding motif. A collection of models of motifs would allow to routinely scan each newly determined structure of protein or help protein design by providing reliable parts of the structure.

The last part concern medical imaging, and more especially extremal points. This is more complex since we have to find out automatically the multiple correspondences between images. We register thus each pair of images and look for maximal cliques of correspondences between the whole set of images. This process is rather time consuming and can certainly be improved, but we focus here on the following step: multiple registration and fusion. The experiment is performed on 24 3D MRI images of the head of the same patient, and we analyze the resulting (rigid) shape model based on stable extremal point. The next step in such an experiment would be to repeat this opera-

tion on several patients independently, and then look for the similar or affine shape based on these models. We could then characterize extremal points that are anatomically stable and incorporate them into an atlas.

12.2 Rigid shapes in $\mathbb{R}^m$

We consider here that characteristic features are points in $\mathbb{R}^m$ and investigate the shape of a k-tuple of such landmarks under rigid transformations. In order to simplify the notations, we note $\mathcal{M} = \mathbb{R}^m$ the manifold of features and $\mathcal{G}$ the rigid motion group.

A rigid transformation of $\mathbb{R}^m$ is represented by a translation vector $t \in \mathbb{R}^m$ and a rotation matrix $(m \times m)$ R satisfying $R.R^{\mathrm{T}} = R^{\mathrm{T}}.R = Id$ and $\det(R) = +1$ in order to exclude reflections. We note $f = (R\,;\,t)$ such a transformation. The two basic operations inside the group are

- the composition: $f_1 \circ f_2 = (R_1.R_2\,;\,R_1.t_2 + t_1)$,
- the inversion: $f^{(-1)} = (R^{\mathrm{T}}\,;\,-R^{\mathrm{T}}.t)$.

The action of the group on the features is simply

- the application: $f \star x = R.x + t$.

An object can be represented by a k-tuple $X = (x^{(1)}, \dots x^{(k)}) \in \mathcal{M}^k$ of features. The action of the transformation group $\mathcal{G}$ on k-tuples is $f \star X = (f \star x^{(1)}, \dots f \star x^{(k)})$. The *rigid shape* of this object is characterized by the relative position between features of the k-tuple, i.e. the quotient space $\mathcal{M}^k/\mathcal{G}$: two objects have the same shape if they are congruent modulo $\mathcal{G}$ (i.e. if there exist a rigid transformation $g \in \mathcal{G}$ such that $Y = g \star X$).

12.2.1 Distances on points, k-tuples and rigid shapes

The Euclidean distance on points comes from the euclidean scalar product:

$$d(x^{(1)}, x^{(2)})^2 = \|x^{(1)} - x^{(2)}\|^2 = < x^{(1)} - x^{(2)} | x^{(1)} - x^{(2)} > \qquad \text{with} \qquad < x|y >= x^{\mathrm{T}}.y = \mathrm{Tr}(x.y^{\mathrm{T}})$$

We note that, by definition of *rigid* transformations, the Euclidean distance is invariant under such a transformation: $d(x, y) = d(f \star x, f \star y)$, whereas the scalar product (and thus the norm) are invariant under rotation: $< R.x|R.y >=< x|y >$.

Let now $X = (x^{(1)}, \dots x^{(k)})$ and $Y = (y^{(1)}, \dots y^{(k)})$ be two k-tuples of points (X and Y are "points" in $\mathcal{M}^k = \mathbb{R}^{m \times k}$). We define the distance between these k-tuples as

$$D(X, Y)^2 = \sum_{h=1}^{k} d(x^{(h)}, y^{(h)})^2$$

A normalization factor $1/k$ can be used if we want to compare distances between k-tuple independently of their number of points k. This distance is obviously invariant under rigid transformation: $D(X, Y) = D(f \star X, f \star Y)$. If we consider the k-tuples as vectors ($X^{\mathrm{T}} = (x^{(1)\top}, \dots x^{(k)\top})$), then the distance D is the Euclidean distance of $\mathbb{R}^{k \times m}$.

In order to obtain a distance on the shape space, we need to minimize this distance on k-tuples over all possible transformations:

$$d_k(\tilde{X}, \tilde{Y}) = \inf_{f \in \mathcal{G}} \left(D(f \star X, Y)\right)$$

Since D is invariant by rigid transformations, we have $D(f \star X, Y) = D(X, f^{(-1)} \star Y)$ and thus $\inf_f \left(D(f \star X, Y)\right) = \inf_g \left(D(X, g \star Y)\right)$ which means that d_k is symmetric: $d_k(\tilde{X}, \tilde{Y}) = d_k(\tilde{Y}, \tilde{X})$.

The triangular inequality is respected by the infimum, and thus the only point remaining to prove that d_k is a distance on shapes is the definiteness: $\left(d_k(\tilde{X},\tilde{Y})=0\right) \Rightarrow \left(\tilde{X}=\tilde{Y}\right)$.

In fact, a stronger property that implies the definiteness can be shown in this case: the above infimum is reached for any k-tuples X and Y (the proof closely follows the one of theorem 12.2). The definiteness of d_k is then easily derived: $d_k(\tilde{X},\tilde{Y})=0$ means that there exists $f \in G$ such as $D(f \star X, Y)=0$, which implies (since D is a distance) that $Y = f \star X$. This is exactly our definition of equality for shapes: we have thus $\tilde{X}=\tilde{Y}$, and d_k is a distance on rigid shapes.

12.2.2 Mean shapes

Extending the formalism of the Fréchet expectation (section 4.2.1) to rigid shapes, we define the set of mean shapes of the set of shapes $\{\tilde{X}_i\}_{i=1}^n$ as

$$\tilde{\mathbb{M}}(\{\tilde{X}_i\}) = \arg \inf_{\tilde{m} \in \mathcal{M}^k \backslash \mathcal{G}} \left\{ \sum_{i=i}^{n} d_k(\tilde{m}, \tilde{X}_i)^2 \right\} \tag{12.1}$$

In terms of k-tuples, we can write this equation:

$$\mathbb{M}(\{X_i\}) = \arg \inf_{m \in \mathcal{M}^k} \left\{ \inf_{(f_1,\dots f_n) \in \mathcal{G}^k} \left\{ \sum_{i=1}^{n} D(m, f_i \star X_i)^2 \right\} \right\} \tag{12.2}$$

We note that if $m \in \mathbb{M}(\{X_i\})$, then any k-tuple of the form $m' = f \star m$ also represents of the mean shape. The problem we are confronted with is thus two-folded: find both a k-tuple m representing the mean shape and a set of transformations f_i that register the k-tuple X_i with it. In fact, we can extend the results of Ziezold (Ziezold, 1994; Ziezold, 1989) and solve first for the multiple registration problem, and then for the mean shape.

A central notion to decouple the multiple registration from the mean shape problem is the optimal position. The k-tuples $X_1, \dots X_n$ are in optimal position to each other if their positions realize the minimum of their relative distances:

$$\sum_{1 \le i < j \le n} D(X_i, X_j)^2 = \inf_{(f_1,\dots f_n) \in \mathcal{G}^k} \left\{ \sum_{1 \le i < j \le n} D(f_i \star X_i, f_j \star X_j)^2 \right\}$$

Théorème 12.1 *$M \in \mathbb{M}$ represents a mean shape if and only if*

- *there exists a set of rigid transformations $(f_1, \dots f_n)$ such as the k-tuples $f_1 \star X_1, \dots f_n \star X_n$ are in optimal position and*
- *considering the k-tuples as vectors: $M = \dfrac{1}{n} \sum_{i=1}^{n} f_i \star X_i$*

In this case, the variance is $s^2 = \dfrac{1}{n^2} \sum_{1 \le i < j \le n} D(f_i \star X_i, f_j \star X_j)^2$.

Proof: In this part, we consider the k-tuples as points of $\mathbb{R}^{mk}$, and the distance D on these points is induced by the Euclidean scalar product. Let $(f_1, \dots f_n)$ be a set of transformations, M a k-tuple and G be the barycenter of the transformed k-tuples X_i:

$$G = \frac{1}{n} \sum_{i=1}^{n} f_i \star X_i$$

We have:

$$\begin{aligned}
\sum_{1\le i<j\le n} & D(f_i\star X_i, f_j\star X_j)^2 = \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} D(f_i\star X_i, f_j\star X_j)^2 \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|f_i\star X_i - f_j\star X_j\|^2 \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left(\|f_i\star X_i - M\|^2 + \|f_j\star X_j - M\|^2 - 2 < f_i\star X_i - M | f_j\star X_j - M >\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\left(n\|f_i\star X_i - M\|^2 + \sum_{j=1}^{n}\|f_j\star X_j - M\|^2 - 2n < f_i\star X_i - M | G - M >\right) \\
&= n\sum_{j=1}^{n}\|f_j\star X_j - M\|^2 - n^2\|G-M\|^2
\end{aligned}$$

This is summarized in: $\displaystyle\sum_{i=1}^{n} D(f_i\star X_i, M)^2 = nD(G,M)^2 + \frac{1}{n}\sum_{1\le i<j\le n} D(f_i\star X_i, f_j\star X_j)^2$

Since the infima of the above expressions is reached (see theorem (12.2) below), M represents a mean feature if and only if the k-tuples $f_i\star X_i$ are in optimal position, and M is the barycenter G of the k-tuples in this optimal position. The empirical variance s^2 is given by the value of the normalized criterion at this minimum:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} D(f_i\star X_i, M)^2 = \frac{1}{n^2}\sum_{1\le i<j\le n} D(f_i\star X_i, f_j\star X_j)^2$$

■

12.2.3 Characterization of the optimal positions

Let $\{X_i = (x_i^{(1)}, \dots x_i^{(k)})\}_{i=1}^n$ be a set of n k-tuples. We call centroid of a k-tuple X_i the barycenter of its points $\bar{x}_i = \frac{1}{k}\sum_{h=1}^{k} x_i^{(h)}$. The corresponding centered k-tuples is then the set the sets $Y_i = (y_i^{(1)}, \dots y_i^{(k)})$ with $y_i^{(h)} = x_i^{(h)} - \bar{x}_i$.

Théorème 12.2 *Optimal positions exist and the centroids of the k-tuples correspond in such a position. The remaining criterion to maximize for rotations is then:*

$$F = \sum_{i,j} Tr\left(R_i.H_{ij}.R_j^T\right) \qquad with \qquad H_{i,j} = \begin{cases} \displaystyle\sum_{h=1}^{k} y_j^{(h)}.y_i^{(h)\,T} & if\ i\neq j \\ 0 & if\ i=j \end{cases} \tag{12.3}$$

Proof: The criterion to be minimized for the multi-registration problem is written:

$$\begin{aligned}
C &= \sum_{1\le i<j\le n} D(f_i\star X_i, f_j\star X_j)^2 \\
&= \sum_{h=1}^{k}\sum_{1\le i<j\le n} \|R_i.x_i^{(h)} + t_i - R_j.x_j^{(h)} - t_j\|^2
\end{aligned}$$

An optimum of this criterion is characterized by a null derivative. In particular, we get for the translation t_l:

$$\frac{\partial C}{\partial t_l}^{\mathrm{T}} = 0 = \sum_{h=1}^{k}\sum_{i=1}^{n} 2\left(R_l.x_l^{(h)} + t_l - R_i.x_i^{(h)} - t_i\right)$$

and thus $\quad R_l.\bar{x}_l + t_l = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(R_i.\bar{x}_i + t_i).$

Since an optimal position is defined up to a global rigid transformation (see above) we can fix for

instance $t_1 = -R_1.\bar{x}_1$, and we obtain $R_1.\bar{x}_1 + t_1 = 0 = \frac{1}{n}\sum_{i=1}^{n}(R_i.\bar{x}_i + t_i)$. which means that each translation t_l is written: $t_l = -R_l.\bar{x}_l$. All the centroids are thus superimposed.
Now, let $y_i^{(h)} = x_i^{(h)} - \bar{x}_i$ be the barycentric "coordinates" of each k-tuple. The criterion may be written:

$$C = \sum_{h=1}^{k} \sum_{1 \le i < j \le n} \|R_i.y_i^{(h)} - R_j.y_j^{(h)}\|^2$$

Developing it, we obtain:

$$C = \frac{1}{2}\sum_h \sum_i \sum_{j \neq i} \left({y_i^{(h)}}^{\mathrm{T}}.y_i^{(h)} + {y_j^{(h)}}^{\mathrm{T}}.y_j^{(h)} - 2.{y_i^{(h)}}^{\mathrm{T}}.R_i^{\mathrm{T}}.R_j.y_j^{(h)} \right)$$

The minima of C are thus given by the maxima of the following criterion F:

$$F = \sum_h \sum_i \sum_{j \neq i} {y_i^{(h)}}^{\mathrm{T}}.R_i^{\mathrm{T}}.R_j.y_j^{(h)} = \mathrm{Tr}\left(\sum_i \sum_{j \neq i} R_i \left(\sum_h y_j^{(h)}.{y_i^{(h)}}^{\mathrm{T}} \right) R_j^{\mathrm{T}} \right)$$

Let $H_{i,j}$ be the correlation matrix $H_{i,j} = \sum_{h=1}^{k} y_j^{(h)}.{y_i^{(h)}}^{\mathrm{T}}$ with the convention that $H_{i,i} = 0$ (we note that $H_{i,j} = H_{j,i}^{\mathrm{T}}$). The criterion to maximize can be written:

$$F = \sum_{i,j} \mathrm{Tr}\left(R_i.H_{i,j}R_j^{\mathrm{T}} \right)$$

Since the minimization occurs on $(SO_3)^n$, which a compact set, the minimum of this criterion is reached for at least one set $(R_1, \ldots R_n) \in (SO_3)^n$.

■

12.2.4 A closed form solution for the simple registration problem

Since $H_{1,2} = H_{2,1}^{\mathrm{T}}$, the criterion reduces in the case of two k-tuples to $F = 2.\mathrm{Tr}\left(R_1.H_{1,2}.R_2^{\mathrm{T}}\right) = 2.\mathrm{Tr}\left(R_2^{\mathrm{T}}.R_1.H_{2,1}^{\mathrm{T}}\right)$. As the solution is defined up to a rotation (if R_1 and R_2 are solutions, $R_0.R_1$ and $R_0.R_2$ are also), we are usually looking for the transformation $R = R_2^{\mathrm{T}}.R_1$ from the first k-tuple to the second one. A closed form solution is well known using the singular value decomposition (Arun et al., 1987; Umeyama, 1991) and recalled in section (8.1.1.3) (quaternions are only valid in 3D). We just recall here the results.

Théorème 12.3 *Let $U.D.V^T = K$ be a singular value decomposition of K. Assuming that the singular values are sorted by decreasing value, the maximum of the criterion $G = Tr(R.K^T)$ is reached over all proper rotations for*

$$R = U.S.V^T \qquad with \qquad S = diag(1, \ldots 1, \det(U).\det(V))$$

with the value $G_{opt} = Tr(D.S)$. The maximum is unique if $rank(K) \geq m - 1$.

12.2.5 An iterative scheme for the multiple registration problem

As in the simple registration case, we observe first that if the set of rotations $\{R_1, \ldots R_n\}$ maximize the criterion $F = \sum_{i,j} \mathrm{Tr}\left(R_i.H_{ij}.R_j^{\mathrm{T}}\right)$, then the set $\{R_0.R_1, \ldots R_0.R_n\}$ also maximize the criterion with the same value. The second step is to isolate in the criterion the influence of a single rotation R_l: we have indeed

$$F = 2F_l + \sum_{i \neq l} \sum_{j \neq l} \mathrm{Tr}(R_i.H_{i,j}.R_j) \qquad \text{with} \qquad F_l = \sum_i \mathrm{Tr}(R_l.H_{l,i}.R_i^{\mathrm{T}})$$

The basic idea is to maximize iteratively the criteria F_l according to the rotation R_l only. Since the second part of the global criterion F does not depend upon R_l, its value is fixed during the optimization of F_l, whereas the value of F_l is increasing. The value of the global criterion F is thus always increasing, and since we optimize on a compact set (SO_m^n), it converges to a local maximum.

In order to speed up the process, we propose to start by registering independently each k-tuple Y_i (for $i > 1$) with the first one Y_1: we fix $R_1 = Id$ and R_i maximizes the criterion $\mathrm{Tr}(R_i.(H_{1,i})^{\mathrm{T}})$ (the criterion F_1 is thus maximum with respect to R_1). Then, we iteratively maximize over R_l the criterion $F_l = \mathrm{Tr}(R_l.(\sum_i R_i.H_{i,l})^{\mathrm{T}})$. We consider that the maximum is reached when the value of the global criterion increases less than a given threshold ε in one pass on each index.

The globality of the maximum reached can be verified by several optimizations with different starting point (start to register with Y_i instead of Y_1) and/or a randomization of the order of maximization. In practice, we observed that a sufficient number of points well localized on a non symmetric shape will lead to a single maximum that can be reached easily.

12.2.6 Extension to incomplete k-tuples

In the real world, however, there are often missing data and thus incomplete k-tuples: some usually stable features can be mismatched, highly displaced or simply forgotten in the extraction scheme in one or a few observations. It would be inauspicious to simply dismiss these features since they can bring useful information in most of the cases.

We consider thus that a missing point in a k-tuple X can correspond to any point location in another k-tuple Y: if there are p missing points, the k-tuple can be viewed as a p-plane in $\mathcal{M}^k$. In order to keep a standard representation, we associate to the k-tuple $X = (x^{(1)}, \dots x^{(h)})$ a binary vector ε_X such that $\varepsilon_X(h)$ is 1 if the point exists and 0 if the point is missing. The "distance" between incomplete k-tuples can thus be written:

$$D(X,Y)^2 = \sum_{h=1}^{k} \varepsilon_X(h).\varepsilon_Y(h).d(x^{(h)}, y^{(h)})^2$$

This is not really a distance since the triangular inequality is not verified and this is null if the associated p-plane and q-plane intersect. However, this corresponds to the distance on k-tuples we define above if there is no missing data, and the distance of a full k-tuple (the mean one for instance) to an incomplete one is the Euclidean distance from a point to a p-plane. We can thus define the set of mean k-tuples as above (equation 12.2).

Unfortunately, if we could theoretically separate the multiple registration from the mean shape computation, we cannot separate the rotation and translation parts in the multiple registration criterion. Since we now have to optimize on a non compact set, we prefer to use another (simpler) iterative scheme: assuming that we have an initial multiple registration, we can compute a mean k-tuple M (with $m^{(h)} = \{\sum_{i=1}^{n} \varepsilon_{X_i}(h).(R_i.x_i^{(h)} + t_i)\}/\{\sum_{i=1}^{n} \varepsilon_{X_i}(h)\}$), which can now be used to minimize criterion (12.2) for the simple registration of each k-tuple with M. We iterate this process until convergence or a maximum number of iterations.

The initial registration can be obtained by using the previous algorithm with sub-k-tuples that are present in all observation or more easily with a "naive" registration of all k-tuples with the first one.

12.3 Experiments

Synthetic experiments showed that the two iterative process converge (to the numerical precision of the machine) in about 5 iterations. The incomplete scheme does not give a sensible improvement of the mean shape precision if only a few points are missing, but turns out to be more and more interesting as the number of missing points increases and/or the number n of k-tuples increases. Indeed, the first method is not any more usable if the number of *globally* common points is too small (we need for instance at least 3 points to register in 3D), and its computational complexity increases in n^2 whereas it is linear for the second method.

12.3.1 Mean shape of the heme binding motif

We present here an application in molecular biology: the globin family collects proteins of many different organisms. Their amino acid homology can be as low as 16%. However, their 3D structure is still related and constitutes the globin fold. It is mainly composed of α helices that form a pocket for the active site which in myoglobins and hemoglobins binds a heme group. The motif constituted by the amino acids binding the heme in 9 globins was extensively studied in (Lesk et Chothia, 1980). We used the correspondences given in this study and the atoms coordinates from the Protein Data Bank to determine the mean shape of the binding motif.

A collection of such models of motifs would allow to routinely scan each newly determined structure of protein or help protein design by providing reliable parts of the structure.

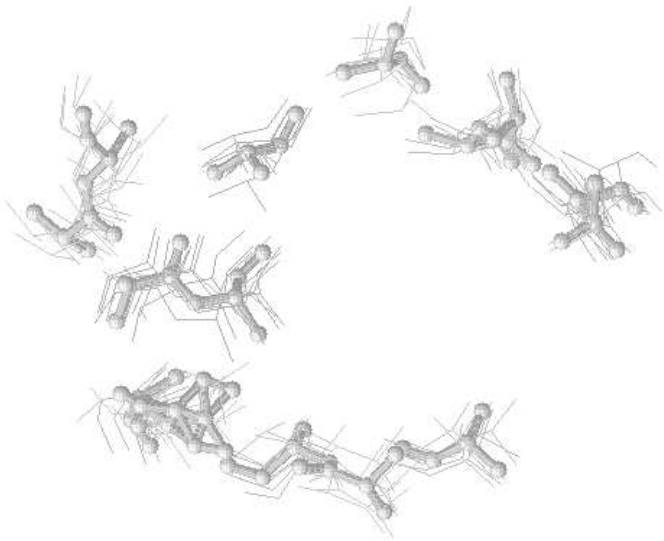


FIG. 12.1 – *The 16 amino acids from 9 globins of the active site are represented by their 4 backbone atoms (bindings are lines). The mean shape of the motif is displayed in balls and sticks. On the bottom-left of the figure, a compact group of atoms represents two alternative (and exclusive) positions of an amino-acid: k-tuples are incomplete.*

12.3.2 Model of a single patient based on extremal points

In 3D medical imaging, the identification and measurement of meaningful features, or landmarks, is much more difficult than in 2D and can become a bottleneck of the shape studies if done by hand.

we believe that extremal points are good landmarks candidates thanks to their automatic extraction. The remaining problem is to discriminate meaningful features from spurious or uninteresting ones. In fact, the saliency of features is their stability across a wide range of observations for a given "object". We are thus confronted to a multiple matching and registration problem in the presence of noise and outliers. Once meaningful features have been identified and matched in different observations, it is interesting to merge their observations in order to obtain a model that incorporate all meaningful information and has a reduced noise.

We presented in (Pennec et Thirion, 1995) an algorithm for the registration based on 3D frames which also quantifies the uncertainty on both the data and the transformation. We used it to register medical images and showed that the accuracy of the registration is far below both the voxel size and the uncertainty of the individual features. In this method, only the most stable features are used to compute the registration, and a lot of matches are discarded due to their large uncertainty.

The aim of the present experiments is to fuse together several registered images of a single patient in order to construct an average model based on extremal points. We are mainly interested here in the topological stability of the features, i.e. determine which features can be reliably matched in most images, even with large geometric deviations. The selectivity of the features (chap. 10) therefore very important. Such a study on several patients will eventually lead to identify the most stable anatomical features (landmarks), and will allow us to better reduce the complexity while increasing the robustness of the registration task.

In 3D medical imaging, the identification and measurement of meaningful features, or landmarks, is much more difficult and can become a bottleneck of the shape studies if done by hand. we believe that extremal points are good landmarks candidates thanks to their automatic extraction. The remaining problem is to discriminate meaningful features from spurious or uninteresting ones. In fact, the saliency of features is their stability across a wide range of observations for a given "object". We are thus confronted to a multiple matching and registration problem in the presence of noise and outliers. Once meaningful features have been identified and matched in different observations, it is interesting to merge their observations in order to obtain a model that incorporate all meaningful information and has a reduced noise. The shape of this model can then be studied, compared with others, used to determine if a similar shape is present in an observation (among other objects or outliers for instance), etc.

The experiment is performed on 24 3D MRI images of the head of the same patient (courtesy of Pr. Ron Kikinis). We register each pair of images and look for maximal cliques of correspondences between the whole set of images. This process is rather time consuming and can certainly be improved, but we focus here on multiple registration and fusion. The 24 sets of highly incomplete k-tuples where registered and merged using the second method within a few seconds. In order to have a more complete model, we add to each extremal point the standard deviation of the observations, which gives an idea of the geometrical stability or accuracy of this feature, and its probability of observation (the percentage of images where it was observed).

We present in figure (12.3) the surface of the brain and the crest lines extracted from the first image and the most stable extremal points from the computed model. We observed that about 30 frames are extremely well preserved, both geometrically and topologically, and 70 to 200 others are observed in more than 80% of the images, depending on the error bound we consider (see also figure 12.2). We plan to adapt automatically the error bound for each feature in order to be able to maximize for each model feature the number of non ambiguous matches. This should allow to compute more robustly the stable features. An interesting result is that the most stable extremal points are located on the surface of the brain and not on the skin or on the skull. Since the images comes from the magnetic resonance modality, the skull is indeed not very visible. This points out

the fact that the registration is mainly done on the surface of the brain, which was expected.

The next step in such an experiment would be to repeat this operation on several patients independently, and then look for the similar or affine shape based on these models. We could then characterize extremal points that are anatomically stable and incorporate them into an atlas, as described for instance in (Subsol et al., 1995).

12.4 Conclusions

We observe in this study that multiple registration and mean shape estimation problems can be solved efficiently in any dimension with a reasonable computational cost. Several tracks are left open, the first being the extension of this scheme to similarities and affine transformations. Another very interesting point would be to generalize the theory to other types of geometric features, such as oriented points, lines, frames... Last but not least, an automatic analysis of shapes should include a powerful and reliable *multiple matching* algorithm in order to automatically determine the landmarks correspondences.

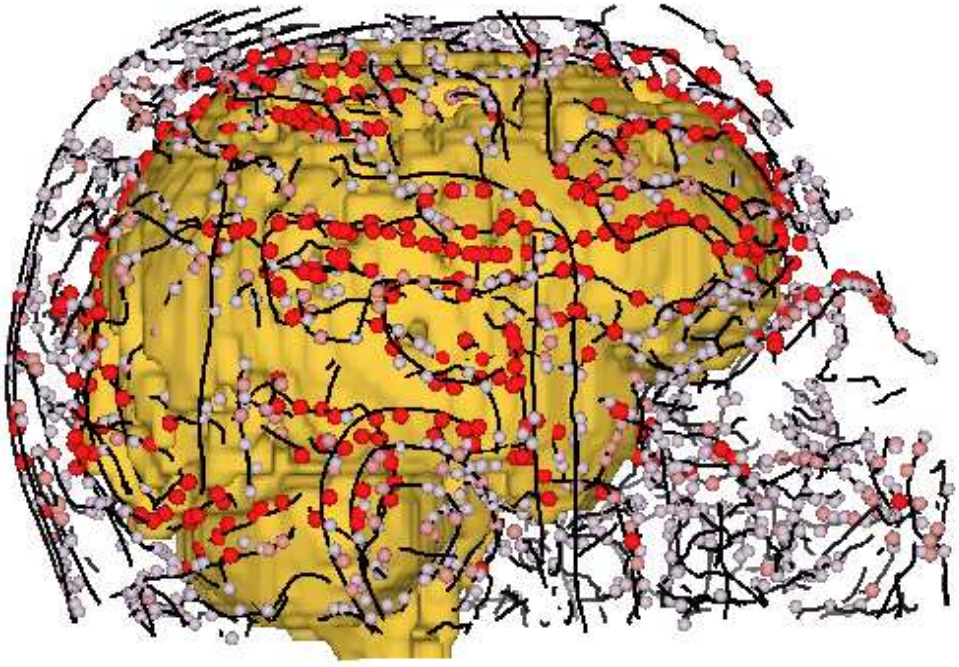


FIG. 12.2 – *Extremal points matched in more than 80% of the 24 images with a large error bound.*

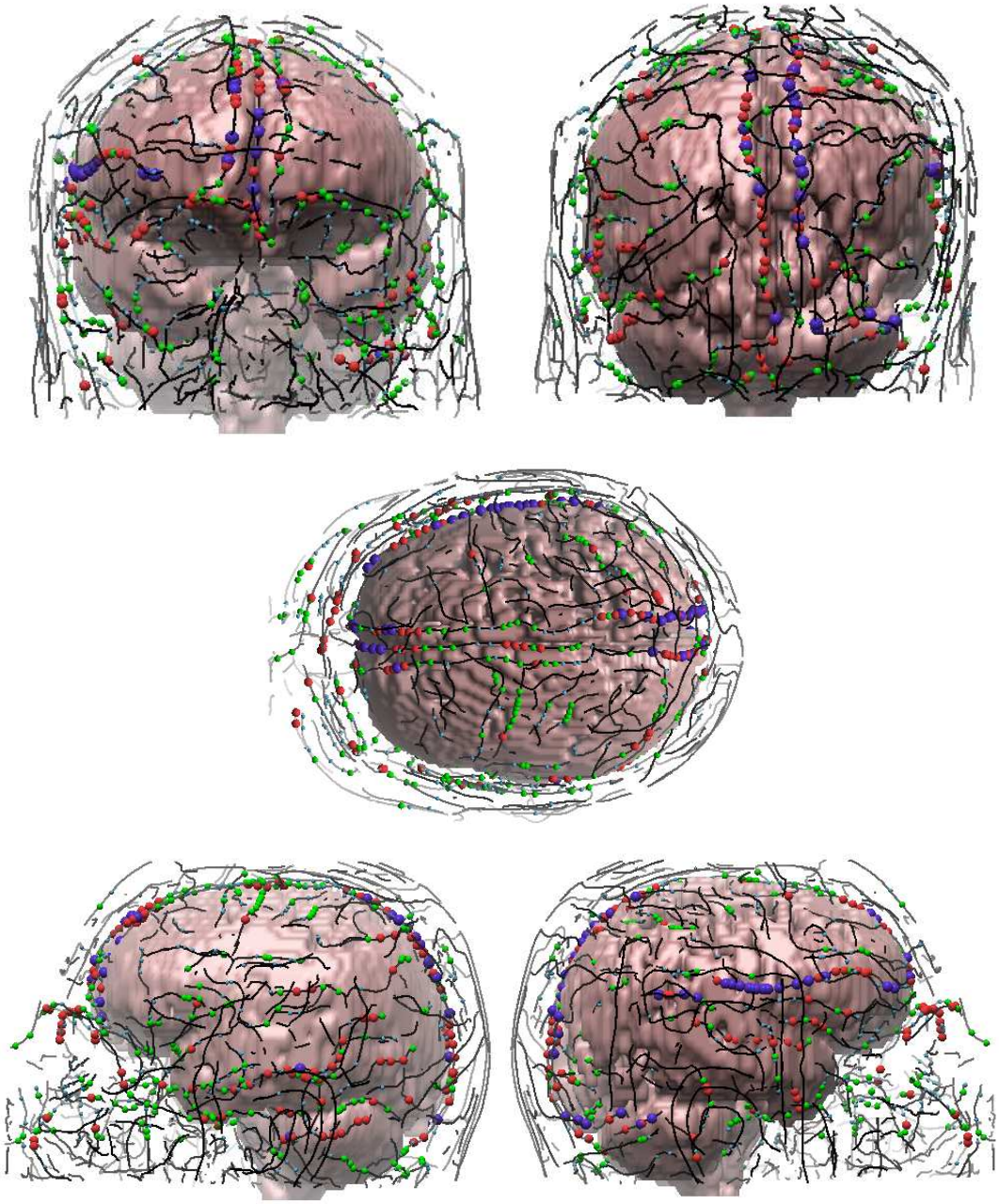


FIG. 12.3 – *The surface of the brain is displayed along with the crest lines extracted from the first image. The most stable frames of the model are represented by spheres colored from decreasing stability order (and size) in purple, red, green and blue. Top line from left to right: front and rear views of the head. Middle: top view. Bottom line: left and right views.*

Troisième partie

Conclusion

Chapitre 13

Conclusion et perspectives

« The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them. »
Sir William Bragg

13.1 Conclusion

Nous avons cherché dans cette thèse à développer des méthodes génériques pour gérer l'incertitude sur les primitives géométriques, afin de pouvoir résoudre des problèmes appliqués en imagerie médicale et en biologie moléculaire.

Cette étude est basée sur l'observation que, en traitement d'image, on peut souvent concentrer la plus grande partie de l'information dans un ensemble restreint de caractéristiques localisées, dont la nature est principalement géométrique. Ces primitives forment une variété différentielle qui n'est généralement pas un espace vectoriel et sur laquelle agit un groupe de transformation qui modélise les différentes « prises de vue » possibles de l'images ainsi que les « déformations » admissibles des objets observés. La seconde observation sur laquelle repose ce travail est que la mesure de ces primitives est intrinsèquement incertaine à cause de l'accumulation des erreurs de mesure et des imprécisions au cours de la chaîne de traitement.

Le traitement rigoureux de ces deux points est indispensable pour obtenir des algorithmes de reconnaissance et de recalage capables de déterminer eux-mêmes leur fiabilité et leur précision. La donnée simultanée des estimations et de leur fiabilité est en effet un pré-requis pour la validation de ces algorithmes.

13.1.1 Le côté théorique

Nous avons développé dans la première partie de ce manuscrit les bases d'une **théorie géométrique de l'incertitude** sur les variétés, en constituant un ensemble complet et cohérent d'opérations pour traiter nos problèmes applicatifs, développés en second partie. Nous replaçons tout d'abord les variétés de primitives dans le cadre de la géométrie riemannienne, les transformations constituant un groupe de Lie. Ceci nous permet de définir une mesure invariante, puis une métrique

riemannienne invariante sur ces espace, en montrant au passage les conditions d'existence. L'analyse des propriétés de ces structures nous mène à une représentation tout à fait centrale de ces variétés que nous appelons carte exponentielle.

Nous construisons sur ces bases une théorie des **probabilités sur les primitives géométriques** en définissant de manière intrinsèque la densité de probabilité d'une primitive aléatoire, puis la moyenne au sens de Fréchet et enfin la matrice de covariance. Si nous avons pu caractériser de manière exacte la propagation de ces caractéristiques dans certaines opérations de base, nous nous sommes contentés d'approximations dans les autres cas et de nombreux résultats théoriques restent encore à établir.

Nous avons aussi étudié l'**aspect statistique** du problème en définissant les notions de bruit homogène et isotrope, puis en construisant une généralisation de la loi gaussienne aux variétés et en introduisant une distance statistique entre primitives aléatoires : la distance de Mahalanobis. Cette partie nécessiterait encore de nombreux développements, en particulier pour analyser d'autres généralisations possibles de la loi gaussienne basées sur d'autres propriétés et pour développer des estimateurs et des tests statistiques adaptés aux variétés.

Enfin, nous résumons les principaux résultats de cette partie en sélectionnant, dans une optique orientée-objet, les **opérations de base** sur lesquelles on peut définir des algorithmes généraux. Ce chapitre constitue en quelque sorte l'articulation entre la théorie et son implantation informatique.

13.1.2 Le côté applicatif

Parmi les techniques d'acquisition d'**images médicales**, certaines sont relativement anciennes comme la radiographie, et d'autres beaucoup plus récentes comme l'IRM, le scanner X ou l'échographie. Avec ces dernières, l'imagerie médicale a été révolutionnée par deux facteurs : l'apport de données tridimensionnelles, et l'introduction de l'informatique non seulement dans le système d'acquisition, mais aussi dans le traitement des images. Le traitement automatique de données aussi volumineuses permet en effet d'une part de mieux cerner les paramètres importants et d'autre part de fournir des mesures quantitatives et non plus simplement qualitatives. Nous nous intéressons ici au recalage de structures anatomiques dans les images médicales, problème dans lequel il est parfois vital de connaître la qualité du résultat obtenu, par exemple pour la planification d'une opération chirurgicale ou la mesure de l'évolution d'une tumeur, à partir de laquelle on va juger de l'efficacité d'un nouveau médicament.

Un autre domaine utilise des techniques d'acquisition relativement similaires, mais à une autre échelle, pour déterminer la structure atomique tridimensionnelle des protéines. Il s'agit de la **biologie moléculaire**. Parmi les problèmes géométriques liés à l'étude de ces structures, on peut citer la recherche de structures communes, indicatives de fonctionnalités similaires, et permettant d'identifier la configuration des sites actifs. Cette connaissance peut ensuite être utilisée pour imaginer et tester de nouvelles substances susceptibles d'interagir avec les protéines concernées, ou servir d'amers pour la modélisation de la structure d'une autre protéine présentant ce motif. Là encore, il est important de savoir si les structures communes détectées sont dues au hasard ou sont significatives d'une configuration très bien conservée. Pour toutes ces raisons, la gestion correcte de l'incertitude est indispensable et justifie l'utilisation de l'arsenal théorique présentée en première partie.

Les primitives qui nous intéressent ici sont principalement des **repères**, des repères semi-orientés et des point 3D, les transformations considérées étant **rigides**. Nous étudions donc les principales propriétés des rotations sous forme de matrices, de quaternions unitaires et de vecteurs rotation. Cela nous permet d'illustrer sur un exemple utile les notions théoriques de la première partie du manuscrit

(géodésiques, mesure et métrique invariante...) et de déterminer la carte principale, fondamentale dans notre théorie. Il ne nous reste plus alors qu'à détailler les opérations de base avec lesquelles nous construisons des algorithmes génériques de plus haut niveau.

Nous nous intéressons ainsi au problème du **recalage**, d'abord à partir d'appariements de points, puis de primitives génériques, en insistant particulièrement sur l'estimation de l'incertitude sur la transformation déterminée. Nous introduisons pour cela deux algorithmes directement issus des propriétés théoriques, basés sur la minimisation de distance riemannienne et de Mahalanobis, le filtrage de Kalman permettant une autre approche de ce dernier critère. Nous analysons et comparons les performances de ces trois algorithmes pour mieux pouvoir les combiner dans un seul algorithme pratique et générique. La précision de cet algorithme est également un point important puisque des expériences de recalage d'images IRM du cerveau d'un même patient ont montré qu'on atteignait une précision de l'ordre de $1/10^{\text{e}}$ de voxel.

L'étape suivante est la **validation** du recalage, qui se traduit dans notre cas par la vérification de l'estimation d'incertitude sur la transformation. Nous développons une méthode statistique qui valide pleinement nos algorithmes à la fois sur les données synthétiques et sur des données réelles et nous permet de détecter le point faible de la chaîne de traitement : l'estimation du modèle de bruit sur les primitives. Cette méthode peut également être utilisée comme un test statistique pour décider si deux mouvements sont identiques, ce que nous illustrons avec une tentative d'analyse du mouvement relatifs des os dans le bassin. Le test montre ici que, contrairement à la première impression, les primitives sont trop incertaines pour que l'on puisse conclure qu'il y a mouvement.

La **mise en correspondance** et la **reconnaissance** sont des problèmes très complexes et la gestion correcte de l'erreur dans ces algorithmes a été l'une des principales motivations de cette thèse, en particulier à cause des problèmes rencontrés dans l'algorithme de reconnaissance de sous-structures dans les protéines. Si nous avons pu formaliser relativement facilement ces algorithmes dans le cadre générique des primitives, de nombreux problèmes restent à résoudre, autant en ce qui concerne la complexité et l'efficacité, comme pour l'indexation et la recherche, qu'au niveau théorique, avec la détermination et la gestion des espaces d'invariants n-aires. Enfin, nous avons abordé de problème de la validation de la mise en correspondance avec le calcul du nombre moyen de faux positif, calcul qui serait à intégrer dans les algorithmes eux-mêmes en déterminant les règles de propagation de cette probabilité dans les opérations utilisées.

Pour finir, nous avons abordé le problème de la **modélisation** automatique d'objets à partir de donnés multiples par le biais du calcul des formes moyennes à base de points. Cette expérience a été très profitable, en particulier pour déterminer parmi les techniques existantes celles qui sont généralisables, à défaut d'être génériques. Les applications potentielles sont importantes puisqu'il s'agit de la création d'atlas anatomiques que nous qualifierons de géométriques dans le cas de l'imagerie médicale, et de la constitution de bases de données géométriques de sites actifs ou de motifs structuraux en biologie moléculaire.

13.1.3 Au centre : l'informatique

Au cours de cette thèse, les aspects théoriques et pratiques se sont constamment rejoints dans l'idée qu'une modélisation théorique rigoureuse était le meilleur moyen d'arriver à résoudre des problèmes pratiques et appliqués. Nous avons dès lors mené une recherche en partant à la fois du problème théorique de base (comment faire des probabilités et des statistiques sur une variété ?) et des applications que nous voulions réaliser, les deux approches se rejoignant au niveau algorithmique dans le concept d'une structure de programmation orientée objet pour gérer l'information géométrique.

Ainsi, le développement de **la théorie a été guidé par les applications** étudiées et par la volonté d'obtenir des résultats algorithmiquement implantables. Dans un premier temps, cette théorie ne concernait que les repères et les points, mais l'utilisation des points extrémaux, qui ne sont que des repères semi-orientés, nous a conduit à faire des développements importants. En effet, les points et les repères sont identifiables à des groupes de transformation, en l'occurrence les translations et les transformations rigides, ce qui amène des simplifications considérables qui ne sont plus valables dans le cas de primitives seulement homogènes comme les points extrémaux.

A l'opposé, nous avons cherché à **formaliser les applications dans des termes génériques** et les développements théoriques nous ont permis d'identifier les points clés de cette organisation. Les algorithmes de descente de gradient proposés pour minimiser les critères de recalage ou de fusion sont ainsi directement issus de l'écriture du chapitre théorique sur l'espérance de Fréchet, et n'étaient pas concevables sans leur dérivation théorique.

La **structure algorithmique** résultant de cette double approche constitue à la fois un outil de validation de la théorie et une plate-forme de développement générique pour de nouvelles applications, solidement ancrés sur les bases théoriques. Au cours des trois années consacrées à la préparation de cette thèse, cette structure a subi plusieurs fois des mutations majeures, et l'écriture de ce manuscrit devrait permettre une évolution prochaine vers une implantation en C^{++}, combinant à la fois les avantages d'un calcul matriciel et vectoriel très rapide en C avec les avantages conceptuels de la programmation orientée objet pour le développement d'algorithmes génériques de haut niveau.

13.2 Perspectives

13.2.1 Théorie

Statistiques D'un point de vue théorique, de nombreux points restent ouverts. Par exemple, nous n'avons pas de formule exacte pour la propagation des moments des primitives aléatoires dans un certain nombre d'opérations. La généralisation de la loi gaussienne aux variétés que nous proposons n'est sans doute pas la meilleure : les paramètres sont relativement difficiles à calculer et cette loi est de plus de dérivée discontinue sur le lieu de coupure de sa moyenne.

Le choix de cette loi « normale » est important car il conditionne tous les développements statistiques que l'on peut imaginer ensuite. Une autre manière de définir cette loi pourrait ainsi être liée à la loi des grand nombres. Mais celle-ci existe-t-elle sur une variété ?

Il serait également intéressant de développer des estimateurs et des tests adaptés adaptés aux caractéristiques que nous avons introduites (moyenne et covariance), mais aussi de concevoir d'autres caractéristiques et des statistiques robustes... Nous pensons qu'il y a en fait matière à développer une théorie des statistiques sur les variétés.

Invariants et espaces de formes Un champ de recherche connexe, que nous avons à peine effleuré dans la section (10.4.3) et le chapitre (12), concerne le lien entre les invariants n-aires et les espaces de formes définis par Kendall.

Dans cette optique, un premier point difficile consiste à construire ces espaces, puis à les munir d'une métrique riemannienne compatible avec les métriques de la variété et du groupe. Il reste alors à trouver un moyen de gérer algorithmiquement ces objets. En effet, il n'y a plus ici de groupe agissant sur ces variétés et nous n'avons plus ni fonction de placement ni carte principale : les cartes exponentielles peuvent être différentes et non identifiables en tout point. De plus, ces variétés peuvent présenter des singularités (Kendall, 1989) et ne sont pas forcément géodésiquement complètes.

Là encore, il y a un travail de recherche théorique intéressant à faire et menant à une application directe : des algorithmes de reconnaissance corrects et efficaces. Ce travail a été amorcé par D. Kendall et H. Li pour les espaces de forme basés sur les points et soumis à l'action des transformations rigides et des similitudes.

Variétés « non rigides » L'évocation des similitudes nous amène au troisième point des perspectives théoriques : un certain nombre de résultats que nous avons développé dans ce manuscrit ne tiennent que si la variété de nos primitives possède une métrique invariante par l'action du groupe de transformation.

Cette hypothèse n'est plus vraie pour les points dès que l'on considère les similitudes ou les transformations affines. Les points projectifs soumis aux homographies n'ont pas non plus de métrique invariante. Toutefois, les groupes (localement compacts) ont toujours une métrique invariante à gauche et seule la variété pose problème. Il faut avouer qu'une bonne partie du chapitre 3 visait à mieux cerner les éléments de géométrie différentielle et riemannienne mis en cause afin de voir comment lever cette restriction.

Nous pensons que le problème peut être abordé en utilisant la connexion invariante qui permet de définir les géodésiques sans faire intervenir de métrique. La connexion étant invariante, les géodésiques restent globalement inchangées sous l'action du groupe et les cartes exponentielles sont donc simplement soumises à des transformations linéaires. En choisissant une métrique riemannienne compatible avec la connexion et en gardant la trace des modifications apportées à cette métrique, nous pensons qu'il est possible de généraliser bon nombre de nos résultats à ces variétés de type « non rigide ».

Ces extensions permettraient de traiter un nombre considérablement élargi d'applications, en particulier touchant à la vision par ordinateur.

Approximations Enfin, même si l'on peut contourner ce problème, il n'est pas toujours possible de résoudre explicitement les équations des géodésiques sur un groupe ou une variété, et nous ne connaissons pas dans ce cas-là les cartes exponentielles. Ainsi, si nous avons pu déterminer les géodésiques pour les similitudes, nous ne savons déjà plus le faire pour les transformation affines.

Nous entrevoyons deux solutions à ce problème. L'une est d'utiliser une approximation des cartes exponentielles correspondant aux géodésiques sur certains axes. Cette solution doit être utilisable pour les transformations affines puisque l'on connaît les géodésiques pour le cas rigide et des similitude. Par contre, pour des groupes de transformation plus généraux, il faut sans doute passer à une solution basée sur les générateurs infinitésimaux, qui permet encore de gérer une bonne approximation des géodésiques dans un voisinage faible autour d'un point du groupe. Il reste à savoir comment cela peut s'appliquer aux variétés.

13.2.2 Algorithmes

Reconnaissance D'un point de vue informatique, nous avons déjà évoqué dans la section (10.4) plusieurs points clés. Il s'agit d'étudier

- les algorithmes de classification (clustering), en particulier ceux pouvant inclure une information d'incertitude,
- des algorithmes de recherche, soit du plus proche voisin, soit des primitives compatibles.

Pour ce dernier type de problèmes, il nous semble difficile de concevoir des techniques de hachage de l'espace qui puissent à la fois gérer correctement l'erreur et être algorithmiquement efficaces.

Par contre, les techniques de subdivision adaptative de l'espace nous paraissent intéressantes et certains algorithmes doivent pouvoir être généralisés à nos variétés homogènes en utilisant, pour couper l'espace, le fait que les géodésiques passant par un point sont des lignes droites dans la carte exponentielle en ce point.

La résolution de ces problèmes permettrait de construire des algorithmes génériques de reconnaissance dont on pourrait étudier les performances grâce aux calculs de robustesse (nombre de faux positifs). Un travail important consisterait également à intégrer ce calcul des faux positifs dans les opérations de base pour la mise en correspondance et de regarder les liens avec les critères de matching probabilistes, qui présentent généralement des dégradations de performances plus souples avec l'accroissement du bruit de mesure ou du nombre de primitives.

D'un point de vue applicatif, cette étude de la reconnaissance est particulièrement importante : nous avons vu en effet le gain de sélectivité énorme que pouvaient apporter des primitives plus complexes que des points. Malheureusement, à l'heure actuelle, la complexité théorique des algorithmes est noyée par la masse de calculs supplémentaires nécessaires pour traiter l'erreur, ne serait-ce qu'approximativement. Rappelons qu'à l'origine de la plus grande partie de ce travail de thèse, il y avait la volonté de pouvoir gérer correctement l'erreur sur les repères dans l'algorithme de reconnaissance proposé pour les protéines.

Modélisation Un problème encore plus complexe mais excessivement intéressant concerne la reconnaissance et le recalage multiple, dans un but de modélisation. Nous avons abordé le recalage multiple au chapitre 12 pour les points et il semble possible d'utiliser ces techniques sur les primitives sans trop de problème.

Par contre, la mise en correspondance simultanée de N images soulève beaucoup de questions. Comment faire pour ne pas privilégier une image par rapport à une autre ? Dans le cas de deux images, nous avons proposé d'imposer une contrainte de symétrie sur les appariements. Il semble raisonnable de vouloir imposer une contrainte similaire sur les appariements dans toutes les images, mais la symétrie simple n'est pas suffisante : si a est une primitive de l'image A appariée symétriquement à b dans l'images B, celle ci étant elle-même appariée symétriquement à c dans C, rien ne nous impose que a ne soit pas apparié symétriquement à une autre primitive c' de C. Il s'agirait donc de trouver la clique maximale dans ce graphe des d'appariements, algorithme qui est NP-complet.

La résolution de ces problèmes ouvre la voie à la modélisation géométrique automatique, qui possède un champ d'applications relativement vaste. Citons pas exemple la comparaison de la banque de donnée des structures de protéines (PDB) avec elle même, qui permettrait la création d'une base de données des motifs structuraux et des motifs de liaisons. La modélisation pose par contre d'autre questions : dans le chapitre 12, nous avons associé aux primitives des modèles une stabilité géométrique, mais aussi une probabilité d'observation. De nombreuses observations supplémentaires, en particulier sur le couplage des observations, seraient ainsi de la plus grande utilité. Par ailleurs, la conception même de la structure de ces modèles va influencer les algorithmes de reconnaissance les utilisant.

Courbes et surfaces Enfin, le travail que nous avons effectué dans ce manuscrit repose entièrement sur des primitives isolées et identifiables. Comment peut-on développer le même type de résultats, en particulier en ce qui concerne l'incertitude du recalage, si l'on considère maintenant des courbes et des surfaces au lieu de points isolés ? En général, on considère les courbes et les surfaces comme un ensemble de points, éventuellement munis d'une orientation, et on calcule la mise en correspondance et le recalage par ICP. On pourrait donc penser à utiliser directement les

techniques de recalage que nous avons développées dans ce manuscrit et prédire ainsi l'incertitude de la transformation estimée.

Le problème principal que nous voyons à cette généralisation directe est le suivant : l'incertitude sur le recalage varie grossièrement comme $1/\sqrt{n}$ où n est le nombre de points utilisés. Si maintenant nous échantillonnons notre courbe avec quatre fois plus de points, on divise l'incertitude par deux : à la limite, avec un échantillonage infini, l'incertitude sur la transformation est nulle ! En fait, nous pensons qu'il faudrait définir une distance entre courbes (ou surfaces) continues à l'aide d'une intégration, puis discrétiser cette intégrale pour obtenir un critère discret. Dans cette optique, l'élément de longueur ou de surface ds se transforme en une pondération δs qui prend effectivement en compte la proportion de la courbe représentée par chaque point. Avec un échantillonage infini, l'incertitude sur la transformation tend maintenant vers une limite non nulle.

Il reste quand même quelques problèmes de taille dans cette modélisation : on suppose pour le recalage classique que les points ou les primitives ont des bruits indépendants. Ce n'est clairement pas le cas pour des points voisins sur une courbe : comment représenter et estimer le bruit sur les points constituant le représentation discrète de la courbe ou de la surface ? Ceci soulève également le problème de la représentation à utiliser.

Ce type d'extension serait toutefois très valable pour de nombreuses applications, y compris la reconnaissance, où les courbes peuvent relier des points particuliers et ainsi diminuer encore la complexité du problème.

13.2.3 Applications

Imagerie Médicale Nous n'avons pour l'instant validé nos algorithmes de recalage que d'un point de vue théorique. Pour qu'ils soient vraiment utiles et crédibles en imagerie médicale, l'étape suivante consistera à les valider sur des fantômes, c'est-à-dire des acquisitions réelles d'objets dont on connaît précisément la géométrie, puis à les intégrer dans les logiciels utilisables par des praticiens en vue d'une validation clinique. Les cadres stéréotaxiques sont des fantômes du plus haut intérêt, puisqu'ils sont ensuite utilisés cliniquement. On peut ainsi vérifier par des acquisitions en tant que fantôme la précision qu'ils apportent et la validité de notre prédiction, puis vérifier lors de l'utilisation clinique que les résultats restent cohérents.

Un autre point qui nous semble prometteur concerne la détection et la quantification du mouvement tridimensionnel. Nous étudions actuellement la possibilité que mesurer précisément le mouvement des fémurs à partir d'images IRM. Cette étude remet en question les techniques d'extraction de primitives géométriques habituellement utilisées car l'os est une structure très peu visible dans ce type d'images. De plus, elle nécessite des algorithmes de reconnaissance permettant de trouver plusieurs structures rigides de faible importance dans des scènes très denses.

Biologie Moléculaire L'algorithme de reconnaissance de sous-structures que nous avons présenté dans ce manuscrit n'a pas suivi pour l'instant les évolutions de notre cadre algorithmique. Il s'agit dans un premier temps de mettre à jour certaines procédures utilisées pour la reconnaissance afin de diminuer le nombre de paramètres et d'accroître la fiabilité et la rapidité. Ce dernier point est en effet fondamental pour envisager de comparer, en un temps acceptable, une structure de protéine à la PDB tout entière. Rappelons que cette base de donnée s'accroît très rapidement : 3200 structures en avril 1995, environ 5000 en novembre 1996 !

L'étape suivante consiste à comparer la PDB avec elle-même pour construire une nouvelle base de donnée contenant la modélisation géométrique des motifs observés. Cette comparaison nécessite un algorithme de reconnaissance entièrement automatique, rapide et permettant des comparaisons

incrémentales pour mettre à jour la base de motifs avec l'évolution de la PDB. D'autres applications sont très prometteuses et font intervenir une large composante géométrique :

- Le docking : il s'agit de trouver parmi un ensemble de structures celles qui peuvent se lier à un site actif, c'est-à-dire de trouver les clefs adaptées à une serrure. Pour l'industrie pharmaceutique, ce crochetage informatique des protéines est d'une importance capitale pour réduire les expériences *in vivo*.
- La modélisation : si l'on ne connaît que 5000 structures tridimensionnelles de protéines, on connaît environ 52000 séquences. L'enjeu est de construire des modèles 3D de leur structure pour pouvoir déterminer les fonctionnalités et les modes de fonctionnement. Une base de donnée de motifs géométriques peut permettre de réduire notamment la complexité de ce problème en fournissant des amers géométriques très précisément localisés et qui découpent littéralement le problème du déploiement tridimensionnel de la protéine.

Vers d'autres horizons géométriques Nous avons cherché dans ce manuscrit à construire les bases d'une « algèbre d'opérations » générique pour manipuler informatiquement les données géométriques incertaines. Dans cette optique, et même si cette algèbre n'est encore qu'embryonnaire, nous pensons que nos travaux peuvent trouver une application dans les nombreux domaines (robotique, vision...) qui reposent sur une modélisation géométrique de leur monde.

En espérant que ce nouveau-né soit porteur d'un peu de flou dans ces mondes géométriques si cartésiens...

Quatrième partie

Appendices

Annexe A

Jacobiens des fonctions scalaires et vectorielles

Soient $x = [x_1, \ldots, x_n]^{\mathrm{T}}$ et $y = [y_1, \ldots, y_n]^{\mathrm{T}}$ deux vecteurs de $\mathbb{R}^n$, λ un scalaire, $f(x)$ une fonction scalaire du vecteur x ($f : \mathbb{R}^n \to \mathbb{R}$) et $g(y) = [g_1(y), \ldots, g_m(y)]^{\mathrm{T}}$ une fonction vectorielle ($g : \mathbb{R}^n \to \mathbb{R}^m$). Les jacobiens de ces fonctions sont définis par les fonction matricielles :

$$J_f = \frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

$$J_g = \frac{\partial g}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

de telle sorte que l'élément (i, j) de J_g est $[J_g]_{ij} = \dfrac{\partial g_i}{\partial x_j}$. On a alors

$$f(x + dx) = f(x) + J_f(x).dx \qquad \text{et} \qquad g(x + dx) = g(x) + J_g(x).dx$$

où le point (.) représente la multiplication matricielle [1]. Les fonctions scalaires sont donc un cas particulier des fonctions vectorielles (espace de dimension 1). On pourrait de même remarquer que les vecteurs sont un cas particulier des matrices et étendre le formalisme aux fonctions matricielles mais cela nécessiterait l'introduction de tenseurs d'ordre 3 (3 indices) et complique nettement les notations puisque l'on a alors des produits tensoriels différents suivant les indices considérés. Nous nous limiterons donc aux fonctions vectorielles.

A.1 Composition et multiplication de fonctions

Soient f et g deux fonctions vectorielles de dimensions compatibles pour la composition ($f : \mathbb{R}^n \to \mathbb{R}^m$ et $g : \mathbb{R}^m \to \mathbb{R}^p$) et $h(x) = g(f(x))$ la fonction composée ($h = g \circ f : \mathbb{R}^n \to \mathbb{R}^p$). La

1. Les scalaires sont classiquement identifiés à l'espace vectoriel unidimensionnel $\mathbb{R}^1$ et a l'espace des matrices réelles 1×1.

composition des jacobiens s'exprime par:

$$J_h(x) = J_g(f(x)).J_f(x)$$

Soit maintenant f une fonction scalaire et $h = f.g$ la fonction multiplicative ($h(x) = f(x).g(x)$). Le jacobien de h est:

$$J_h(x) = f(x).J_g(x) + g(x).J_f(x)$$

A.2 Fonction de deux variables

Soit $\alpha(x, x)$ une fonction vectorielle de deux variables vectorielles et J_{α_1} et J_{α_2} ses deux jacobiens :

$$J_{\alpha_1}(x, y) = \frac{\partial(\alpha(x, y))}{\partial x} \qquad J_{\alpha_2}(x, y) = \frac{\partial(\alpha(x, y))}{\partial y}$$

Cette fonction peut, par exemple, représenter un produit vectoriel ou un produit scalaire. Soient f et g deux fonctions vectorielles et $h(x) = \alpha(f(x), g(x))$ la fonction composée. Le jacobien de h est donné par :

$$J_h(x) = J_{\alpha_1}\left(f(x), g(x)\right).J_f(x) + J_{\alpha_2}\left(f(x), g(x)\right).J_g(x)$$

A.3 Jacobiens de fonctions standard

Produit scalaire : $\langle x \mid y \rangle = \langle y \mid x \rangle = \sum x_i.y_i = x^{\mathrm{T}}.y = \mathrm{Tr}(x.y^{\mathrm{T}})$

$$J_{\langle . \mid . \rangle_1}(x, y) = \frac{\partial \langle x \mid y \rangle}{\partial x} = y^{\mathrm{T}}$$

$$J_{\langle . \mid . \rangle_2}(x, y) = \frac{\partial \langle x \mid y \rangle}{\partial y} = x^{\mathrm{T}}$$

Norme : $\|x\| = \sqrt{\sum x_i^2} = \sqrt{\langle x \mid x \rangle}$

$$J_{\|.\|}(x) = \frac{\partial(\|x\|)}{\partial x} = \frac{x^{\mathrm{T}}}{\|x\|} = (\mathrm{Normalisation}(x))^{\mathrm{T}}$$

Normalisation : $J_N(x) = \frac{\partial}{\partial x}\left(\frac{x}{\|x\|}\right) = \frac{\|x\|^2.I_n - x.x^{\mathrm{T}}}{\|x\|^3}$

Carré de la norme : $\frac{\partial \|x\|^2}{\partial x} = 2x^{\mathrm{T}}$

Multiplication par une matrice M (constante) : $\frac{\partial M.x}{\partial x} = M$

A.4 Formules et opérations particulières à $\mathbb{R}^3$

La particularité réside dans le fait que le produit extérieur de $\mathbb{R}^3$ ne fait intervenir que deux vecteurs. C'est ce que l'on appelle le produit vectoriel :

$$x \times y = \left| \begin{array}{l} x_2.y_3 - x_3.y_2 \\ x_3.y_1 - x_1.y_3 \\ x_1.y_2 - x_2.y_1 \end{array} \right.$$

Il est facile de vérifier que ce produit est anti-commutatif ($x \times y = -y \times x$) et s'annule si et seulement si les deux vecteurs sont colinéaires : $x \times y = 0 \Leftrightarrow x = \lambda.y$.

Matrice anti-symétrique associé au produit vectoriel : C'est la matrice anti-symétrique telle que, quel que soit y : $S_x.y = x \times y$. Cette matrice s'écrit :

$$S_x = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}$$

Jacobiens du produit vectoriel : $x \times y = S_x.y$

$$J_{\times_1}(x,y) = \frac{\partial(x \times y)}{\partial x} = -S_y$$

$$J_{\times_2}(x,y) = \frac{\partial(x \times y)}{\partial y} = S_x$$

Identité de Lagrange : $\|x \times y\|^2 = \|x\|^2\|y\|^2 - \langle x \mid y \rangle^2$

Formule de Gibbs : $x \times (y \times z) = \langle x \mid z \rangle .y - \langle x \mid y \rangle .z$

$$S_x.S_y = y.x^{\mathrm{T}} - \langle x \mid y \rangle .I_3$$

Identité de Jacobi : $x \times (y \times z) + y \times (z \times x) + z \times (x \times y) = 0$

$$S_{(x \times y)} = S_x.S_y - S_y.S_x$$

Propriété métrique : $\langle x \times y \mid z \times t \rangle = \langle x \mid z \rangle . \langle y \mid t \rangle - \langle x \mid t \rangle . \langle y \mid z \rangle$

Produit mixte :

$$(x,y,z) = \langle x \mid y \times z \rangle = \langle x \times y \mid z \rangle = \det([x,y,z]) = \left| \begin{array}{ccc} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{array} \right|$$

$$\det([x,y,z]) = x^{\mathrm{T}}.S_y.z = -x^{\mathrm{T}}.S_z.y = -y^{\mathrm{T}}.S_x.z = x^{\mathrm{T}}.S_y.z$$

Jacobien de la normalisation : il se simplifie en

$$J_N(x) = \frac{\partial}{\partial x}\left(\frac{x}{\|x\|}\right) = \frac{-S_x^2}{\|x\|^3}$$

Exemple de dérivation : (utile pour la dérivation des rotations)

$$\frac{\partial(S_r^2.x)}{\partial r} = \frac{\partial(r \times (r \times x))}{\partial r} = -S_{(r\times x)} + r.\frac{\partial(r \times x)}{\partial r} = S_x.S_r - 2S_r.S_x$$

Annexe B

Dérivation d'un scalaire par une matrice

Nous avons dit précédemment que l'extension du formalisme des dérivées aux matrices nécessitait l'emploi de tenseurs. Il est cependant un cas important où ce n'est pas nécessaire : la dérivation d'une fonction scalaire α d'une matrice A de dimension (m, n).

$$J_\alpha(A) = \frac{\partial \alpha(A)}{\partial A} = \begin{bmatrix} \frac{\partial \alpha(A)}{\partial a_{1,1}} & \cdots & \frac{\partial \alpha(A)}{\partial a_{1,n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \alpha(A)}{\partial a_{m,1}} & \cdots & \frac{\partial \alpha(A)}{\partial a_{m,n}} \end{bmatrix}$$

de telle sorte que l'élément (i, j) du jacobien soit $[J_\alpha]_{ij} = \frac{\partial \alpha(A)}{\partial a_{ij}}$. Notons bien que cette notation ne rentre pas dans le cadre précédent puisque les règles de composition ne s'appliquent plus. Ce n'est donc qu'une façon agréable et synthétique de noter cette dérivée.

B.1 Dérivation par une matrice générique

La méthode générale pour obtenir le jacobien est la suivante.

1/ Écrire la fonction sous la forme d'une somme sur les indices en n'utilisant pas i et j (par exemple $k, l, \ldots$).

2/ Dériver par rapport à a_{ij} : si a_{kl} apparaît dans l'expression sommée, la dérivée ne sera non nulle que lorsque $(k, l) = (i, j)$. On enlève donc a_{kl} et la sommation sur ces deux indices, que l'on remplace dans le reste de l'expression par les indices i et j. On répète bien sûr cette opération si d'autres indices sont utilisés.

3/ Réordonner les termes des sommations restantes de façon à ce que i soit le premier indice et j le dernier, en utilisant au besoin des transposées. Certains termes, voire certaines sommations, doivent parfois être factorisées pour former un scalaire qui sera placé devant le terme variable.

4/ Résumer le résultat sous forme matricielle.

Exemple 1 :

$$\frac{\partial (c^{\mathrm{T}}.A.b)}{\partial A} = c.b^{\mathrm{T}} \tag{B.1}$$

On a en effet : $\alpha(A) = c^{\mathrm{T}}.A.b = \sum_{kl} c_k.a_{kl}.b_l$ soit $\dfrac{\partial \alpha}{\partial a_{ij}} = c_i.b_j$ d'où le résultat.

$$\frac{\partial(\mathrm{Tr}(B.A.C))}{\partial A} = B^{\mathrm{T}}.C^{\mathrm{T}} \tag{B.2}$$

Exemple 2 :

La fonction scalaire s'écrit $\alpha(A) = \mathrm{Tr}(B.A.C) = \sum_{mkl} b_{mk}.a_{kl}.c_{lm}$

et la dérivée par rapport à un élément est donc:

$$\frac{\partial \alpha}{\partial a_{ij}} = \sum_m b_{mi}.c_{jm} = \sum_m [B^{\mathrm{T}}]_{im}.[C^{\mathrm{T}}]_{mj}$$

$$\frac{\partial(c^{\mathrm{T}}.A^{\mathrm{T}}.B.A.d)}{\partial A} = B.A.d.c^{\mathrm{T}} + B^{\mathrm{T}}.A.c.d^{\mathrm{T}} \tag{B.3}$$

Exemple 3 :

La fonction scalaire s'écrit $\alpha(A) = c^{\mathrm{T}}.A^{\mathrm{T}}.B.A.d = \sum_{klmn} c_k.a_{lk}.b_{lm}.a_{mn}.d_n$

et la dérivée par rapport à un élément est donc:

$$\frac{\partial \alpha}{\partial a_{ij}} = \sum_{m,n} c_j.b_{im}.a_{mn}.d_n + \sum_{k,l} c_k.a_{lk}.b_{li}.d_j = \left(\sum_{m,n} b_{im}.a_{mn}.d_n\right).c_j + \left(\sum_{k,l} b_{li}.a_{lk}.c_k\right).d_j$$

d'où le résultat. Si B est une matrice symétrique (que l'on note Q pour ne pas confondre), on obtient :

$$\frac{\partial(c^{\mathrm{T}}.A^{\mathrm{T}}.Q.A.d)}{\partial A} = Q.A.(c.d^{\mathrm{T}} + d.c^{\mathrm{T}}) \tag{B.4}$$

B.2 Dérivation par une matrice symétrique

Ces dérivations sont utiles pour estimer par exemple des matrices de covariance, mais les termes hors-diagonaux de la matrice sont maintenant liés : $a_{ij} = a_{ji}$ et la dérivation en est affectée : le terme a_{kl} dans une somme a maintenant une dérivée non nulle par rapport à a_{ij} si $(k,l) = (i,j)$ ou $(k,l) = (j,i)$ (si $i \neq j$). On doit donc doubler les dérivées précédemment obtenues en inversant les indices i et j, sauf sur la diagonale.

Notons $\mathrm{diag}(A)$ le vecteur des éléments diagonaux de la matrice carrée A et $\mathrm{DIAG}(x)$ la matrice diagonale ayant pour éléments les composantes du vecteur x:

$$\mathrm{diag}(A) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{nn} \end{bmatrix} \qquad \text{et} \qquad \mathrm{DIAG}(x) = \begin{bmatrix} x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_n \end{bmatrix}$$

Cette notation peut être étendue aux matrices avec la signification suivante : $\mathrm{DIAG}(A)$ est la matrice conservant les éléments diagonaux de la matrice carrée A en annulant tous les autres :

$$\mathrm{DIAG}(A) = \mathrm{DIAG}(\mathrm{diag}(A)) = \begin{bmatrix} a_{1,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{n,n} \end{bmatrix}$$

Avec ces notations et à partir de la dérivée $J_\alpha(A) = \partial\alpha/\partial A$ du scalaire α par une matrice générique A, on obtient la dérivée par une matrice symétrique Σ par la formule :

$$\frac{\partial\alpha(\Sigma)}{\partial\Sigma} = J_\alpha(\Sigma) + J_\alpha(\Sigma)^{\mathrm{T}} - \mathrm{DIAG}(J_\alpha(\Sigma)) \tag{B.5}$$

$$\frac{\partial(c^{\mathrm{T}}.\Sigma.b)}{\partial\Sigma} = c.b^{\mathrm{T}} + b.c^{\mathrm{T}} - \mathrm{DIAG}(c.b^{\mathrm{T}}) \tag{B.6}$$

Exemple 1 :

$$\frac{\partial(\mathrm{Tr}(B.\Sigma.C))}{\partial\Sigma} = C.B + B^{\mathrm{T}}.C^{\mathrm{T}} - \mathrm{DIAG}(C.B) \tag{B.7}$$

Exemple 2 :

Exemple 3 :

$$\begin{aligned}\frac{\partial(c^{\mathrm{T}}.\Sigma.B.\Sigma.d)}{\partial\Sigma} = \;& B.\Sigma.d.c^{\mathrm{T}} + B^{\mathrm{T}}.\Sigma.c.d^{\mathrm{T}} + c^{\mathrm{T}}.d\Sigma.B^{\mathrm{T}} + d^{\mathrm{T}}.c.\Sigma.B \\ & -\mathrm{DIAG}(B.\Sigma.d.c^{\mathrm{T}}) - \mathrm{DIAG}(B^{\mathrm{T}}.\Sigma.c.d^{\mathrm{T}})\end{aligned}$$

Si $B = Q$ est elle même une matrice symétrique, on obtient en notant $F = c.d^{\mathrm{T}} + d.c^{\mathrm{T}}$:

$$\frac{\partial(c^{\mathrm{T}}.\Sigma.Q.\Sigma.d)}{\partial\Sigma} = 2\,Q.\Sigma.F - \mathrm{DIAG}(Q.\Sigma.F)$$

B.3 Quelques dérivations de matrices

B.3.1 Inversion

Comme on ne peut pas calculer $\partial A^{(\text{-}1)}/\partial A$ (qui serait un tenseur), on se contente ici de calculer $J = \frac{\partial a_{kl}^{(\text{-}1)}}{a_{ij}}$ où $a_{kl}^{(\text{-}1)} = [A^{(\text{-}1)}]_{kl}$ est l'élément de la matrice inverse et non pas l'inverse de l'élément. Par définition, on a

$$\frac{\partial A^{(\text{-}1)}}{\partial a_{ij}} = \lim_{\varepsilon\leftarrow 0}\left(\frac{(A+\varepsilon.B)^{(\text{-}1)} - A^{(\text{-}1)}}{\varepsilon}\right) = A^{(\text{-}1)}.\lim_{\varepsilon\leftarrow 0}\left(\frac{(Id+\varepsilon.B.A^{(\text{-}1)})^{(\text{-}1)} - Id}{\varepsilon}\right)$$

où B est la matrice où seul l'élément (i,j) est non nul : $[B]_{m,n} = \delta_{m,i}.\delta_{n,j}$. Pour ε suffisamment faible, un développement limité donne :

$$(Id+\varepsilon.B.A^{(\text{-}1)})^{(\text{-}1)} = Id - \varepsilon.B.A^{(\text{-}1)} + O(\varepsilon^2)$$

et la limite ci dessus est donc :

$$\lim_{\varepsilon\leftarrow 0}\left(\frac{(Id+\varepsilon.B.A^{(\text{-}1)})^{(\text{-}1)} - Id}{\varepsilon}\right) = -B.A^{(\text{-}1)} \qquad \text{d'où} \qquad \frac{\partial A^{(\text{-}1)}}{\partial a_{ij}} = -A^{(\text{-}1)}.B.A^{(\text{-}1)}$$

En simplifiant la valeur de B, on a :

$$\frac{\partial a_{kl}^{(\text{-}1)}}{a_{ij}} = -\left[A^{(\text{-}1)}.B.A^{(\text{-}1)}\right]_{k,l} = -\sum_{m,n} a_{k,m}^{(\text{-}1)}.\delta_{i,m}.\delta_{j,n}.a_{n,l}^{(\text{-}1)}$$

soit

$$\frac{\partial a_{kl}^{(\text{-}1)}}{a_{ij}} = -a_{k,i}^{(\text{-}1)}.a_{j,l}^{(\text{-}1)} \tag{B.8}$$

B.3.2 Distance de Mahalanobis

On peut maintenant envisager la dérivation de

$$\alpha = x^{\mathrm{T}}.A^{(-1)}.y = \sum_{k,l} x_k.a_{k,l}^{(-1)}.y_l$$

On considère pour l'instant une matrice A générique : on a alors

$$\frac{\partial \alpha}{\partial a_{ij}} = \sum_{k,l} x_k.y_l.\frac{\partial a_{kl}^{(-1)}}{a_{ij}} = -\sum_{k,l} x_k.y_l.a_{k,i}^{(-1)}.a_{j,l}^{(-1)} = -\sum_{k,l} a_{k,i}^{(-1)}.x_k.y_l.a_{j,l}^{(-1)}$$

soit

$$\frac{\partial (x^{\mathrm{T}}.A^{(-1)}.y)}{A} = -A^{(-\mathrm{T})}.x.y^{\mathrm{T}}.A^{(-\mathrm{T})} \tag{B.9}$$

Dans le cas d'une distance de Mahalanobis, on a une matrice de covariance symétrique, et la dérivée est donc :

$$\frac{\partial (x^{\mathrm{T}}.\Sigma^{(-1)}.y)}{\Sigma} = \mathrm{DIAG}(\Sigma^{(-\mathrm{T})}.x.y^{\mathrm{T}}.\Sigma^{(-\mathrm{T})}) - \Sigma^{(-\mathrm{T})}.(x.y^{\mathrm{T}} + y.x^{\mathrm{T}}).\Sigma^{(-\mathrm{T})}$$

B.3.3 Déterminant

Développons le déterminant en cofacteurs de la i$^{\mathrm{e}}$ ligne :

$$\alpha = \det(A) = \sum_k a_{ik}.a_{ik}^*$$

où a_{ik}^* est le cofacteur de a_{ik}, qui ne dépend d'aucun élément de la i$^{\mathrm{e}}$ ligne, et en particulier pas de a_{ik} :

$$\frac{\partial \det(A)}{\partial a_{ik}} = a_{ik}^*$$

La dérivée cherchée est donc la matrice des cofacteurs, aussi connue sous le nom de matrice conjuguée et sa relation bien connue avec l'inverse nous permet d'écrire :

$$\frac{\partial \det(A)}{\partial A} = A^* = \det(A).A^{(-\mathrm{T})} \tag{B.10}$$

A partir de cette expression, on obtient facilement la dérivée du log du déterminant:

$$\frac{\partial(\log(\det(A)))}{\partial A} = \frac{1}{\det(A)}.\frac{\partial \det(A)}{\partial A} = A^{(-\mathrm{T})}$$

Annexe C

Filtrage de Kalman

Assume we have a set of measurements (or data) $\{\chi_i\}$ and we search state variable $\mathbf{g}$ such that, for each exact data χ_i, we have the vectorial relation

$$z_i(\chi_i, g) = 0$$

This is called the measurement equation. In our case, the state g is the sought rigid motion and the data are couples of matched points or frames, or simple measurements of this rigid motion (section 8.3). Since we are working with noisy data, we only know the measured values of data $\hat{\chi}_i = \chi_i + \omega_i$ (the observation). The additive noises ω_i are assumed to be independent, white and centered with a known covariance Ω_i:

$$E(\omega_i) = 0 \qquad E(\omega_i . \omega_i^{\mathrm{T}}) = \Omega_i \qquad \text{and} \qquad E(\omega_i . \omega_j^{\mathrm{T}}) = 0 \qquad \text{for } i \neq j$$

The measurement equations are generally not linear, but assuming we know a good estimate $\hat{g}$ of the state g, we can linearize them around the estimates and solve the problem with standard linear optimization techniques, namely here Kalman Filtering. This is the basis of the Extended Kalman Filtering technique. Since Kalman Filtering is a recursive filter, we assume that we have at each step i an estimate $\hat{g}_{i-1}$ of the state vector. We can then linearize the measurement equation around $(\hat{\chi}_i, \hat{g}_{i-1})$ with a first order Taylor series expansion. Taking the following notations:

$$\hat{z}_i = z_i(\hat{\chi}_i, \hat{g}_{i-1}) \qquad M_i = \widehat{\frac{\partial z_i}{\partial g}} = \left.\frac{\partial z_i}{\partial g}\right|_{(\hat{\chi}_i, \hat{g}_{i-1})} \qquad \widehat{\frac{\partial z_i}{\partial \chi}} = \left.\frac{\partial z_i}{\partial \chi}\right|_{(\hat{\chi}_i, \hat{g}_{i-1})}$$

the Taylor expansion of the measurement equation $z_i(\chi_i, g) = 0$ gives

$$\hat{z}_i + \widehat{\frac{\partial z_i}{\partial g}} . (g - \hat{g}_{i-1}) + \widehat{\frac{\partial z_i}{\partial \chi}} . (\chi_i - \hat{\chi}_i) \simeq 0$$

This equation can be re-written in the linear form $M_i . g = \gamma_i + \nu_i$ where γ_i is the linearized measurement $\gamma_i = M_i . \hat{g}_{i-1} - \hat{z}_i$ and ν_i is a centered noise

$$\nu_i = \widehat{\frac{\partial z_i}{\partial \chi}} . (\hat{\chi}_i - \chi_i)$$

of covariance $\Sigma_{\boldsymbol{ii}}$:

$$\Sigma_{\boldsymbol{ii}} = \left(\widehat{\frac{\partial z_i}{\partial \chi}}\right) .\Omega_i. \left(\widehat{\frac{\partial z_i}{\partial \chi}}\right)^{\mathrm{T}}$$

Assuming that an initial estimate $(\hat{g}_0, \hat{G}_0)$ of the state is given, a reasonable (weighted) least-squares criterion to minimize can be constructed on the basis of Mahalanobis distances:

$$C = \frac{1}{2}(\hat{g}_0 - g)^{\mathrm{T}}.\hat{G}_0^{-1}.(\hat{g}_0 - g) + \frac{1}{2}\sum_i (\gamma_i - M_i.g)^{\mathrm{T}}.\Sigma_{\boldsymbol{ii}}^{(-1)}.(\gamma_i - M_i.g)$$

The recursive solution for the minimization of this criteria is called the Kalman Filter (Jazwinsky, 1970; Ayache, 1991). At each step, the input is an estimate $(\hat{g}_{i-1}, \hat{G}_{i-1})$ of the state and the linearized measurement $(\gamma_i, \Sigma_{\boldsymbol{ii}})$, and the output is the updated estimation of the state $(\hat{g}_i, \hat{G}_i)$. We just recall here the equations of the filter:

$$\begin{array}{rl} K_i & = \hat{G}_{i-1}.M_i^{\mathrm{T}}.(\Sigma_{\boldsymbol{ii}} + M_i.\hat{G}_{i-1}.M_i^{\mathrm{T}})^{-1} \\ \hat{g}_i & = \hat{g}_{i-1} - K_i.\hat{z}_i \\ \hat{G}_i & = (\ Id - K_i.M_i).\hat{G}_{i-1} \end{array} \qquad \text{(C.1)}$$

For linear transformation of Gaussian variables, the Kalman Filter produces the optimal estimate of the state (in the sense of minimum error variance), which turns out to be also the maximum likelihood estimate. Moreover, it preserves the Gaussian nature of the random variables and thus there is no loss of information in keeping only the mean value and covariance matrix.

If the Gaussian assumption is removed, the Filter remains the best *linear* estimator, but is no longer the best one amongst all non-linear estimators. In the general case of non-linear transformations with any type of noise (which is our case in this article), the Extended Kalman Filter only represent a sub-optimal non linear estimator, but appears to provide accurate estimates in practice. However, some care has to be taken about the initial state and the order of the measurements.

Annexe D

Notations utilisées

Espaces

$\mathbb{R}$	Ensemble des nombres réel.
$\mathbb{R}^n$	Espace vectoriel euclidien de dimension n.
$\mathbb{Q}$	Espace des quaternions.
$\mathcal{M}$	Variété différentielle (manifold).
$\mathcal{G}$	Groupe de transformation (groupe de Lie).
$\mathcal{H}$	Sous-groupe de transformation.
$\mathcal{D}$, $\mathcal{X}$	Ensemble de définition ou autre.
$\mathbf{1}_{\mathcal{U}}$	Fonction indicatrice d'un ensemble $\mathcal{U}$.
$\mathcal{F}_{\mathrm{x}}$	Coset de la primitive x.

Primitives

$\mathrm{f}, \mathrm{g}, \mathrm{h},\ \mathrm{x}, \mathrm{y}, \mathrm{z}$	Transformations et primitive : points d'une variété.
$f, g, h,\ x, y, z$	Vecteurs : représentations de ces objets.
$\vec{\mathrm{f}}, \vec{\mathrm{g}}, \vec{\mathrm{h}},\ \vec{\mathrm{x}}, \vec{\mathrm{y}}, \vec{\mathrm{z}}$	représentations de ces objets dans la *carte principale*.
$\mathbf{f}, \mathbf{g}, \mathbf{h},\ \mathbf{x}, \mathbf{y}, \mathbf{z}$	Transformations et primitives aléatoires.
$\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h},\ \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$	Vecteurs aléatoires correspondants.
$\mathrm{x}, \mathbf{x}, \hat{\mathrm{x}}$	Primitive exacte, primitive aléatoire modélisant la mesure, résultat (observation ou réalisation) d'une mesure.
o , o	Point choisi comme origine de la variété et sa représentation.
Id , Id	Origine du groupe de transformation : identité.
$\partial_v = \dot{\mathrm{x}}$, $v = \dot{x}$	Vecteur tangent en une primitive x (considérée comme une courbe $\mathrm{x}(t)$) et vecteur tangent correspondant dans la carte locale.

Opérateurs

$\mathrm{f} \circ \mathrm{g}$	Composition des transformations f et g.
$\mathrm{f}^{(\text{-}1)}$	Inversion de la transformation f.
$\mathrm{y} = \mathrm{f} \star \mathrm{x}$	Action de la transformation f sur la primitive x.
$\mathrm{f_x}$	Fonction de placement : $\mathrm{f_x}$ est *un* représentant de $\mathcal{F}_{\mathrm{x}}$.
$\overline{\mathrm{x}} \in \mathbb{E}[\,\mathbf{x}\,]$	Espérance au sens de Fréchet (ensemble).
$\overline{x} = \mathbf{E}[\,\boldsymbol{x}\,]$	Espérance au sens classique (vecteur ou réel).
$\mathbf{E}[\,\varphi(\mathbf{x})\,] = \mathbf{E}_{\mathbf{x}}[\,\varphi\,]$	Espérance d'une fonction réelle.
$\Sigma_{\boldsymbol{xx}}$	Matrice de covariance du vecteur aléatoire $\boldsymbol{x}$.
$\mathrm{dist}(\mathrm{x}, \mathrm{y})$	Distance entre primitives.

Calcul matriciel et vectoriel

$\langle x \mid y \rangle$	Produit scalaire euclidien de deux vecteurs.
M , J , R	Matrice, matrice jacobienne, matrice de rotation.
M^{T} , $M^{(\text{-}1)}$, , $M^{(\text{-T})}$	Transposée, inverse et inverse de la transposée de la matrice M.
$\mathrm{Tr}(M)$, $\det(M)$	Trace et déterminant de la matrice carrée M.
$m_{i,j} = [M]_{i,j}$	Élément (i, j) de la matrice M.
$x = \mathrm{diag}(M)$	Vecteur des valeurs diagonales de la matrice carrée M : $x_i = m_{i,i}$.
$M = \mathrm{DIAG}(x)$	Matrice diagonale des composantes du vecteur : $m_{i,j} = x_i.\delta_{i,j}$.
$M.A$, $M.x$	Multiplication matricielle.
I_n , I_3	Matrice identité en dimension n et 3.
$f(x)$, $g(y)$, $h(z)$	Fonctions scalaires ou vectorielles.
$J = \frac{\partial f}{\partial x} = \left[\frac{\partial f_i}{\partial x_j}\right]$	Jacobien ou différentielle de la fonction f. La notation mathématique usuelle correspondante serait plutôt Df.
$\lvert J \rvert = \lvert \det(J) \rvert$	Valeur absolue du déterminant du jacobien J.
$x \times y = S_x.y$	Produit vectoriel dans $\mathbb{R}^3$ et la matrice antisymétrique associée.
$q * p = Q_q.p$	Produit des quaternions et quaternion matriciel associé.
$(p * q)^{\mathrm{T}} = p^{\mathrm{T}}.P_q$	Produit des quaternions et anti-quaternion associé.

Bibliographie

Abola, E., Bernstein, F., Bryant, S., Koetzle, T., et Weng, J. (1987). « Protein Data Bank ». Dans Allen, F., Bergerhoff, G., et Sievers, R., éditeurs, *Crystallographic Databases - Information Contents, Software Systems, Scientific Applications*, pages 107–132, Bonn/Cambridge/Chester. Data Commission of the Int. Union of Crystallography.

Altmann, S. L. (1986). *Rotations, Quaternions, and Double Groups.* Clarendon Press - Oxford.

Ambartzumian, R. (1982). Random shapes by factorization. Dans Ranneby, B., éditeur, *Statistics in Theory and Practice.* Swedish Univ. Agric. Sci., Umea.

Anderson, T. (1958). *An Introduction to Multivariate Statistical Analysis.* J. Wiley, NY.

Arnaudon, M. (1994). Espérances conditionnelles et C-martingales dans les variétés. Dans J. Azema, P.A. Meyer, M. Y., éditeur, *Séminaire de probabilités XXVIII*, volume 1583 de *Lect. Notes in Math.*, pages 300–311. Springer-Verlag.

Arnaudon, M. (1995). Barycentres convexes et approximations des martingales continues dans les variétés. Dans J. Azema, P.A. Meyer, M. Y., éditeur, *Séminaire de probabilités XXIX*, volume 1613 de *Lect. Notes in Math.*, pages 70–85. Springer-Verlag.

Arun, K., Huang, T., et Blostein, S. (1987). « Least-Squares Fitting of Two 3-D Point Sets ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700.

Aurenhammer, F. (1991). « Voronoi diagrams: A survey of a fundamental geometric data structure ». *ACM Comput. Surv.*, 23:345–405.

Ayache, N. (1989). *Vision Stéréoscopique et Perception Multisensorielle.* Inter-Editions.

Ayache, N. (1991). *Artificial Vision for Mobile robots - Stereo-vision and Multisensor Perception.* MIT-Press.

Ayache, N. (1993). « Computer Vision Applied to 3D Medical Images: Results, Trends and Future Challenges ». Dans *Int. Symp. on Robotic Research, Hidden Valley, Pensylvania, USA*. Also as INRIA Research Report No 2050.

Ayache, N. et Faugeras, O. (1986). « HYPER : A new Approach for the Recognition and Positioning of Two-Dimensional Objects ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):44–54.

Bard, Y. (1974). *Nonlinear Parameter Estimation.* Academic Press.

Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T., et Tasumi, M. (1977). « The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures ». *J. of Mol. Bio.*, 112:535–542.

Besl, P. et Jain, R. (1985). « Three-dimensional object recognition ». *ACM Computing Surveys*, 17(1):75–145.

Besl, P. et McKay, N. (1992). « A Method for Registration of 3-D Shapes ». *PAMI*, 14(2):239–256.

Bingham, C. (1974). « An Antipodally Symmetric Distribution on the Sphere ». *The Annals of Statistics*, 2(6):1201–1225.

Boissonnat, J. et Yvinec, M. (1995). *Géométrie algorithmique*. Ediscience international, Paris. to appear in english at Cambridge University Press.

Bolles, R. et Cain, R. (1982). « Recognizing and locating partially visible objects, the local-feature-focus method ». *Int. J. Robotics Research*, 1(3):57–82.

Bookstein, F. (1986). « Size and shape spaces for landmark data in two dimensions (with discussion) ». *Statist. Sci.*, 1:181–242.

Bookstein, F. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge Univ. Press, Cambridge.

Bourguignon, J. (1990). *Calcul variationnel*. Cours de l'Ecole Polytechnique, Paris.

Branden, C. et Tooze, J. (1991). *Introduction to Protein Structure*. Garland Publishing.

Brennan, R. et Matthews, B. (1989). « The Helix-Turn-Helix DNA binding motif ». *J. Biol. Chem.*, 264:286–290.

Brown, L. (1992). « A survey of image registration techniques ». *ACM Computing Surveys*, 24(4):325–376.

Califano, A. et Mohan, R. (1991). « Multidimensionnal Indexing for Recognizing Visual Shapes ». Dans *Proc. CVPR 91*, pages 28–34.

Capolsini, P., Dalmas, S., et Papegay, Y. (1993). « "Quaterman" : Une bibliothèque maple de calculs et de manipulations sur l'algèbre des quaternions ». Rapport Technique 148, INRIA.

Carmo, M. d. (1992). *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin.

Casteljau, P. (1987). *Les quaternions*. Hermes.

Chin, R. et Dyer, C. (1986). « Model-based recognition in robot vision ». *ACM Computing Surveys*, 18(1):67–108.

Cohen, L. (1996). « Auxiliary variables and two-step iterative algoritms in computer vision problem ». *Journal Mathematical Imaging and Vision*. to apear. Also in Cahiers de Mathématiques de la Decision n^o 9511 and ICCV'95.

Csurka, G., Zeller, C., Zhang, Z., et Faugeras, O. (1995). « Characterizing the Uncertainty of the Fundamental Matrix ». Research Report 2560, INRIA.

Davies, E. (1991). « Alternative to abstract graph matching for locating objects from their salient features ». *Image and Vision Computing*, 9(4):252–261.

Doss, S. (1949). « Sur la moyenne d'un élément aléatoire dans un espace distancié ». *Bull. Sc. Math.*, 73:48–72.

Duda, R. et Heart, P. (1973). *Pattern Classification and Scene Analysis*. Wiley.

Durrant-Whyte, H. (1988). *Integration, Coordination and Control of Multi-Sensor Robot Systems*. Kluwer Academic Publishers.

EMERY, M. et MOKOBODZKI, G. (1991). Sur le barycentre d'une probabilité dans une variété. Dans J. AZEMA, P.A. Meyer, M. Y., éditeur, *Séminaire de probabilités XXV*, volume 1485 de *Lect. Notes in Math.*, pages 220–233. Springer-Verlag.

ESHERA, M. et FU, K. (1986). « An Image Understanding System Using Attributed Symbolic Representation and Inexact Graph-Matching ». *PAMI*, 8(5):604–618.

FAUGERAS, O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint.* MIT press.

FELDMAR, J. (1995). « *Recalage rigide, non rigide et projectif d'images médicales tridimensionnelles* ». PhD thesis, Ecole Polytechnique.

FELDMAR, J. et AYACHE, N. (1996). « Rigid, Affine and Locally Affine Registration of Free-Form Surfaces ». *IJCV*, 18(2):99–119.

FELDMAR, J., AYACHE, N., et BETTING, F. (1997). « 3D-2D Projective Registration of Free-Form Curves and Surfaces ». *Computer Vision and Image Understanding*, 65(3):403–424.

FERMI, G., PERUTZ, M., SHAANAN, B., et FOURME, R. (1984). « The crystal structure of human deoxyhaemoglobon at 1.74 angstroms resolution ». *J. Mol. Bio.*, 175:159.

FISCHER, D., BACHAR, O., NUSSINOV, R., et WOLFSON, H. (1992a). « An efficient automated computer vision based technique for detection of three dimensionnal structural motifs in proteins ». *J. of Biomolecular Structures and Dynamics*, 9(4):769–789.

FISCHER, O., NUSSINOV, R., et WOLFSON, H. (1992b). « 3D substructure matching in protein molecules ». Dans *Combinatorial Pattern matching 92 - LNCS 644*, pages 136–150. Spinger Verlag.

FRÉCHET, M. (1944). « L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque ». *Revue Scientifique*, pages 483–512.

FRÉCHET, M. (1948). « Les éléments aléatoires de nature quelconque dans un espace distancié ». *Ann. Inst. H. Poincaré*, 10:215–310.

FUNDA, J. et PAUL, R. (1988). « A Comparison of Transforms and Quaternions in Robotics ». Dans *Proc. IEEE International Conference on Robotics and Automation, Vol. 2*, pages 886–891. Computer Science Society.

GILL, P., MURRAY, W., et WRIGHT, M. (1981). *Practical optimization.* Academic Press, London and New York.

GOODALL, C. (1991). « Procrustes methods in the statistical analysis of shape (with discussion) ». *J. Roy. Statist. Soc. Ser. B*, 53:285–339.

GRENANDER, U. (1981). *Abstract inference.* John Wiley & Soons.

GRIMSON, W. (1990). *Object Recognition by Computer - The role of Geometric Constraints.* MIT Press.

GRIMSON, W. et HUTTENLOCHER, D. (1990a). « On the Sensitivity of Geometric Hashing ». Dans *Proc. third ICCV*, pages 334–338.

GRIMSON, W. et HUTTENLOCHER, D. (1990b). « On the sensitivity of the Hough transform for object recognition ». *IEEE PAMI*, 12(3):255–274.

GRIMSON, W. et HUTTENLOCHER, D. (1991). « On the verification of hypothesized matches in model-based recognition ». *IEEE PAMI*, 13(12):1201–1213.

Grimson, W., Huttenlocher, D., et Jacobs, D. (1991). « Affine Matching with Bounded Sensor Error : A Study of Geometric Hashing and Alignment ». Technical Report 1250, MIT AI Lab.

Grimson, W., Huttenlocher, D., et Jacobs, D. (1994). « A study of affine matching with bounded sensor error ». *Int. Journ. of Comput. Vision*, 13(1):7–32.

Guéziec, A. et Ayache, N. (1992). « Smoothing and Matching of 3D Space Curves ». Dans *Proceedings of the Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy.

Guéziec, A. et Ayache, N. (1993). « New Develoments on Geometric Hashing for Matching 3D Curves ». Dans *Proceedings of Int. Conf on Comput. Vis. and Pat. Recog*, pages 703–704. IEEE Computer Society Press.

Harrison, S. et Aggarwal, A. (1990). « DNA recognition by proteins with the Helix-Turn-Helix motif ». *Annu. Rev. Biochem.*, 59:933–969.

Herer, W. (1986). « Espérance mathématique au sens de Doss d'une variable aléatoire à valeur dans un espace métrique ». *C. R. Acad. Sc. Paris, Série I*, t.302(3):131–134.

Herer, W. (1988). « Espérance mathématique d'une variable aléatoire à valeur dans un espace métrique à courbure négative ». *C. R. Acad. Sc. Paris, Série I*, t.306:681–684.

Hochschild, G. (1965). *The structure of Lie groups*. Holden-Day.

Honzatko, R., Hendrickson, W., et Love, W. (1985). « Refinement of a molecular model for lamprey hemoglobin from petromyzon Marinus ». *J. Mol. Bio.*, 184:147.

Horaud, R. et Monga, O. (1993). *Vision par ordinateur : outils fondamentaux*. Hermes.

Horn, B. (1987). « Closed Form Solutions of Absolute Orientation Using Unit Quaternions ». *Journal of Optical Society of America*, A-4(4):629–642.

Horn, B., H.M., H., et Negahdaripour, S. (1988). « Closed Form Solutions of Absolute Orientation Using Orthonormal Matrices ». *Journal of Optical Society of America*, A-5(7):1127–1135.

Huber, P. (1981). *Robust Statistics*. John Wiley, New York.

Huttenlocher, D. et Ullman, S. (1987). « Object Recognition using Alignment ». Dans *Proc. of ICCV*, pages 72–78.

Huttenlocher, D. et Ullman, S. (1990). « Recognizing solid objects by alignment with an image ». *Int. Journ. Computer Vision*, 5(2):195–212.

Jain, A. et Dubes, R. (1988). *Algorithms for clustering data*. Englewood Cliffs, N.J.

Jaynes, E. (1996). « Probability Theory: The Logic Of Science ». Fragmentary edition of mai 1996 available at http://omega.math.albany.edu:8008/JaynesBook.html. Also available at ftp://bayes.wustl.edu/Jaynes.book/.

Jazwinsky, A. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Jolion, J.-M., Meer, P., et Bataouche, S. (1991). « Robust Clustering with Applications in Computer Vision ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(8):791–802.

Jupp, P. et Mardia, K. (1989). « A Unified View of the Theory of Directional Statistics, 1975-1988 ». *Int. Statistical Review*, 57(3):261–294.

Kanatani, K. (1990). *Group-Theoretical Methods in Image Understanding*. Numéro 20 dans Springer Series in Information Science. Springer Verlag.

KANATANI, K. (1993). *Geometric Computation for Machine Vision.* Numéro 37 dans The Oxford Engineering science series. Clarendon Press - Oxford.

KARCHER, H. (1977). « Riemannian center of mass and mollifier smoothing ». *Comm. Pure Appl. Math*, 30:509–541.

KENDALL, D. (1989). « A survey of the statistical theory of shape (with discussion) ». *Statist. Sci.*, 4:87–120.

KENDALL, M. et MORAN, P. (1963). *Geometrical probability.* Numéro 10 dans Griffin's statistical monographs and courses. Charles Griffin & Co. Ltd.

KENDALL, W. (1990). « Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence ». *Proc. London Math. Soc.*, 61(2):371–406.

KENT, J. (1992). « *The art of Statistical Science* », Chapitre 10 : New Directions in Shape Analysis, pages 115–127. John Wiley & Sons. K.V. Mardia, ed.

KLINGENBERG, W. (1982). *Riemannian Geometry.* Walter de Gruyter, Berlin, New York.

KOCH (1988). *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer, Berlin.

LACROIX, B. et CODANI, J. (1991). « Techniques Informatiques pour la Cartogrphie Physique du Génome Humain ». Research Report 1560, INRIA.

LAMDAN, Y. et WOLFSON, H. (1988). « Geometric Hashing : A General and Efficient Model-Based Recognition Scheme ». Dans *Proc. of Second ICCV*, pages 238–289.

LAMDAN, Y. et WOLFSON, H. (1989). « On the error analysis of Geometric Hashing ». Technical Report 467, Rob. Res. Lab., Courant Institute of Mathematical Science, N. Y. University. Robotics Report No 213.

LAMDAN, Y. et WOLFSON, H. (1991). « On the error analysis of Geometric Hashing ». Dans *IEEE Int. Conf. on Comput. Vis. and Patt. Recog.*, pages 22–27.

LATOMBE, J. (1991). « *Robot motion planning* », Chapitre Chap. 2: Configuration Space of a Rigid Object, pages 58–89. Kluwer Academic Publishers Groups.

LAWSON, C., ZHANG, R., SCHEVITZ, R., OTWINOWSKI, Z., JOACHIMIAK, A., et SIEGLER, P. (1988). « Flexibility of the DNA-Binding Domains of TRP Repressor ». *Proteins Struct., Funct., Genet.*, 3:18.

LE, H. (1995). « Mean size-and-shapes and mean shapes: a geometric point of view ». *Adv. Appl. Prob.*, 27(44–55).

LE, H. et KENDALL, D. (1993). « The riemannian structure of euclidean shape space: a novel environment for statistics ». *Ann. Statist.*, 21:1225–1271.

LE BORGNE, M. (1987). « Quaternions et Controle sur l'Espace des rotations ». Research Report 751, INRIA.

LEAVERS, V. (1993). « Survey : which Hough transform ? ». *CVGIP: Image Understanding*, 58(2):250–264.

LESK, A. et CHOTHIA, C. (1980). « How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins ». *J. Mol. Bio.*, 136:225–270.

LORUSSO, A., EGGERT, D., et FISHER, R. (1995). « A comparison of Four Algorithms for Estimating 3-D Rigid Transformation ». Dans *Proc. 6th British Machine Vision Conference (BMVC'95)*, pages 237–246.

Mardia, K. (1995). « Directional Statistics and Shape Analysis ». Research Report STAT95/24, University of Leeds, UK.

Marr, D. (1982). *Vision*. Freeman.

Maybeck, P. (1979). *Stochastic Models, Estimation and Control, Volume 1*, volume 141-1 de *Mathematics in Science and Engineering*. Academic Press, Inc.

Meer, P., Mintz, D., Kim, D. Y., et Rosenfeld, A. (1991). « Robust Regression Methods for Computer Vision: A Review ». *International Journal of Computer Vision*, 6(1):59–70.

Mondragon, A., Wolberger, C., et Harrison, S. (1989). « Structure of Phage 434 Cro Protein at 2.35 Angstroms Resolution ». *J. of Mol. Bio.*, 205:179.

Moore, R. (1966). *Interval Analysis*. Prentice Hall, Engelwood Cliffs, NJ.

Mundy, J. et Zisserman, A. (1992). *Geometric invariance in computer vision*. MIT Press.

Myers, E. (1991). « An overview of sequence comparison algorithms in molecular biology ». Technical Report TR 91-29, Univ. of Arizona, Dep. of Comp. Sci.

Neveu, J. (1990). *Introduction aux probabilités*. Ecole Polytechnique. Cours de l'Ecole Polytechnique.

Okabe, A., Boots, B., et Sugihara, K. (1992). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Chichester, England.

Olson, C. F. (1994). « Time and Space Efficient Pose Clustering ». Dans *Proc. Conf. Comp. Vision and Pattern Recognition (CVPR'94)*, pages 251–258, Los Alamitos, CA, USA. IEEE Computer Society Press. also as U. Cal. Berkley Research Report : UCB//CSD-93-755.

Orr, M., Fisher, R., et Hallam, J. (1991). « Uncertain reasoning: Intervals versus probabilities ». Dans *Proc. British Machine Vision Conference*, pages 351–354. Springer-Verlag.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc.

Paul, R. (1982). *Robot Manipulators*. MIT Press.

Pelat, D. (1992). *Bruits et Signaux*. Cours de l'Ecole doctorale d'astrophysique d'Ile de France.

Pennec, X. (1993a). « Correctness and Robustness of 3D Rigid Matching with Bounded Sensor Error ». Research report RR-2111, INRIA.

Pennec, X. (1993b). « Traitement de l'erreur dans les méthodes de Hachage géométrique et d'Alignement, une étude théorique ». Rapport de d.e.a., École Polytechnique, ENS Ulm. D.E.A. IMA.

Pennec, X. (1996). « Multiple Registration and Mean Rigid Shape - Application to the 3D case ». Dans Mardia, K., Gill, C., et I.L., D., éditeurs, *Image Fusion and Shape Variability Techniques (16th Leeds Annual Statistical Workshop)*, pages 178–185. University of Leeds, UK.

Pennec, X. et Ayache, N. (1994). « An $\mathcal{O}(n^2)$ Algorithm for 3D Substructure Matching of Proteins ». Dans Califano, A., Rigoutsos, I., et Wolson, H., éditeurs, *Shape and Pattern Matching in Computational Biology – Proc. First Int. Workshop, Seattle, Wash, June 20, 1994*, pages 25–40. Plenum Publishing. Published in Bioinformatics 14(6), 1998, p. 516-522.

Pennec, X. et Ayache, N. (1996a). « Quelques problèmes posés par la gestion des primitives géométriques en vision par ordinateur ». Dans *Journées ORASIS*, pages 111–116.

Pennec, X. et Ayache, N. (1996b). « Randomness and Geometric Features in Computer Vision ». Dans *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'96)*, pages 484–491, San Francisco, Cal, USA. Published in J. of Math. Imag. and Vision 9(1), july 1998, p. 49-67.

Pennec, X. et Ayache, N. (1998). « A geometric algorithm to find small but highly similar 3D substructures in proteins ». *Bioinformatics*, 14(6):516–522.

Pennec, X. et Thirion, J. (1995). « Validation of 3D Registration Methods based on Points and Frames ». Dans *Proc. of the 5th Int. Conf on Comp. Vision (ICCV'95)*, pages 557–562, Cambridge, Ma. Published in Int. J. of Comp. Vision 25(3), 1997, p. 203-229.

Pennec, X. et Thirion, J.-P. (1997). « A Framework for Uncertainty and Validation of 3D Registration Methods based on Points and Frames ». *Int. Journal of Computer Vision*, 25(3):203–229.

Perutz, M. et al. (1973). Private communication.

Picard, J. (1994). « Barycentres et martingales sur une variété ». *Annales de l'institut Poincaré - Probabilités et Statistiques*, 30(4):647–702.

Preparata, F. et Shamos, M. (1985). *Computational Geometry, an Introduction*. Springer Verlag.

Press, S. (1972). *Applied Multivariate Analysis*. Holt Rinehart and Winston, NY.

Press, W., Flannery, B., Teukolsky, S., et Vetterling, W. (1991). *Numerical Recipices in C*. Cambridge Univ. Press.

Reyes-Avila, L. (1990). « Les quaternions : une représentation paramétrique syst matique des rotations finies ». Rapport de Recherche 1303, INRIA.

Reyes-Avila, L. (1991). « Quaternion : une représentation paramétrique syst matique des rotations finies - Partie II : Quelques applications ». Rapport de Recherche 1454, INRIA.

Rigoutsos, I. (1992). « *Massively Parallel Baysian Object Recognition* ». PhD thesis, New York University.

Rigoutsos, I. et Hummel, R. (1991a). « Robust similarity invariant matching in the presence of noise ». Dans *Proc. 8th Israeli Conf. on Artificial Intelligence and Computer Vision*, pages 27–43.

Rigoutsos, I. et Hummel, R. (1991b). « Several Results on Affine Invariant Geometric Hashing ». Dans *Proc. 8th Israeli Conf. on Artificial Intelligence and Computer Vision*, pages 1–12.

Rigoutsos, I. et Hummel, R. (1993). « Distributed Bayesian Object Recognition ». Dans *Proceedings of Int. Conf on Comput. Vis. and Pat. Recog*, pages 180–186. IEEE Computer Society Press.

Rothe, I., Voss, K., Süsse, H., et Rothe, J. (1994). « A General Method to Determine Invariants ». Rapport Technique TR 503, University of Rochester, Computer Science Department.

Rousseeuw, P. et Leroy, A. (1987). *Robust Regression and Outliers Detection*. Wiley series in prob. and math. stat. J. Wiley and Sons.

Santalo, L. (1976). *Integral Geometry and Geometric Probability*, volume 1 de *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Company.

Sayle, R. et Bissel, A. (1992). « RasMol: A program for fast realistic rendering of molecular structures with shadows. ». Dans *Proceedings of the 10th Eurographics UK'92 Conference*, Universtity of Edinburg,Scotland.

Shapiro, L. et Haralick, R. (1981). « Structural description and inexact matching ». *PAMI*, 3:504–519.

Small, C. (1988). « Techniques of shape analysis on sets of points ». *Internat. Statist. Rev.*, 56:243–257.

SMITH, R. et CHEESEMAN, P. (1987). « On the representation and Estimation of Spatial Uncertainty ». *Int. Journ. of Robotics Research*, 5(4):56–68.

SMITH, R., SELF, M., et CHEESEMAN, P. (1988). « Estimating Uncertain Spatial Relationships ». Dans LEMMER, J. et KANAL, L., éditeurs, *Uncertainty in Artificial Intelligence 2*, pages 435–461. Elsevier. Proc. Second Work. on Uncertainty in Artif. Intell., Philadelphia, AAAI, August 1986.

SPIVAK, M. (1979). *Differential Geometry*, volume 1. Publish or Perish, Inc., 2nd edition.

STOCKMAN, G. (1987). « Object recognition and localization via pose clustering ». *Comp. Vision, Graphics, Image Processing*, 40(3):361–387.

STUELPNAGEL, J. (1964). « On the Parametrization of the Three Dimensional Rotation Group ». *SIAM Review*, 6(4):422–430.

SUBSOL, G., THIRION, J., et AYACHE, N. (1995). « A General Scheme for Automatically Building 3D Morphometric Anatomical Atlases: application to a Skull Atlas ». Dans *MRCAS'95*.

TAKANO, T. (1984). Refinement of myoglobin and cytochrome C. Dans HALL, S. et ASHSIDA, T., éditeurs, *Methods and Applications in Crystallographic Computing*, page 262. Oxford University Press, Oxford, England.

THIRION, J. (1993). « New Feature Points based on geometric invariants for 3D Image Registration ». Research Report 1901, INRIA. published in IJCV 18(2), 1996.

THIRION, J.-P. (1994). « Extremal Points : definition and application to 3D image registration ». Dans *IEEE conf. on Computer Vision and Pattern Recognition*, Seattle.

THIRION, J.-P. et GOURDON, A. (1993). « The 3D Marching Lines Algorithm : new results and proofs ». Rapport Technique 1881, INRIA. published in CVIU in 1995.

THIRION, J.-P. et GOURDON, A. (1995). « Computing the Differential Characteristics of Isointensity Surfaces ». *Computer Vision and Image Understanding*, 61(2):190–202.

TSAI, F. (1993). « A Probabilistic Approach to Geometric Hashing using Line Feature ». Technical Report 640, Robotic Research Laboratory, Courant Institute of Mathematical Science, N.Y. Univ.

UMEYAMA, S. (1991). « Least-Squares Estimation of Transformation Parameters Between Two Point Patterns ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380.

Van den ELSEN, P. (1993). « *Multimodality Matching of Brain Images* ». PhD thesis, Utrecht University, Netherland.

WALKER, M. et SHAO, L. (1991). « Estimating 3-D Location Paramters Using Dual Number Quaternions ». *CVGIP: Image Understanding*, 54(3):358–367.

WATSON, D. (1988). « Natural neighbor sorting on the n-dimensional sphere ». *Pattern Recognition*, 21(1):63–67.

WEINSHALL, D. (1993). « Model-based invariants for 3-D vision ». *International Journal of Computer Vision*, 10(1):27–42.

WELLS, W. (1993). « *Statistical Object Recognition* ». PhD thesis, MIT.

WEST, J., FITZPATRICK, J., et AL. (1996). « Comparison and evaluation of retrospective intermodality image registration techniques ». Dans *SPIE Medical Imaging conference (Medim'96)*. also submitted to J. of Assisted Tomography.

Wolfson, H. (1990). « Model-Based Recognition by Geometric Hashing ». Dans Faugeras, O., éditeur, *Proc. of 1st Europ. Conf. on Comput. Vision (ECCV 90)*, pages 526–536. Springer-Verlag. Lecture Notes in Computer Science 427.

Zhang, Z. (1993). « Le problème de la mise en correspondance : l'état de l'art ». Rapport de recherche 2146, INRIA.

Zhang, Z. (1994). « Iterative Point Matching for Registration of Free-Form Curves and Surfaces ». *Int. Journ. Comp. Vis.*, 13(2):119–152. Also Research Report No.1658, INRIA Sophia Antipolis, 1992.

Zhang, Z. et Faugeras, O. (1992). « *3D Dynamic Scene Analysis: a stereo based approach* », volume 27 de *Springer series in information science*, Chapitre 2: Uncertainty Manipulation and Parameter Estimation, pages 9–27. Springer Verlag.

Zhuang, X. et Huang, Y. (1994). « Robust 3-D 3-D Pose Estimation ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):818–824.

Ziezold, H. (1989). On expected figures in the plane. Dans Hübler, A., Nagel, W., Ripley, B., et Werner, G., éditeurs, *Geobild '89*, volume 51 de *Math. Res.*, pages 105–110. Akademie-Verlag, Berlin.

Ziezold, H. (1994). « Mean figures and mean shapes applied to biological figures and shape distributions in the plane ». *Biometrical J.*, 36:491–510.

Index

L'incertitude dans les problèmes de reconnaissance et de recalage
Application en imagerie médicale et biologie moléculaire

Xavier PENNEC
INRIA, B.P. 93, F-06902 Sophia-Antipolis Cédex, France.

Les primitives géométriques usuelles en traitement d'image tridimensionnel (points, points orientés, repères...) forment une variété qui n'est généralement pas un espace vectoriel, sur lequel agit un groupe de transformation qui modélise les différentes « prises de vue possibles » de l'image. Leur mesure est, de plus, intrinsèquement incertaine à cause de l'accumulation des erreurs et des imprécisions au cours de la chaîne de traitement. Le but de ce travail est de généraliser à ces primitives les notions de statistiques usuelles ainsi que les algorithmes de reconnaissance et de recalage utilisant normalement les points.

La première partie de la thèse est consacrée au développement d'outils mathématiques pour aborder ces problèmes. Nous montrons tout d'abord que l'on ne peut pas considérer ces primitives comme de simples vecteurs, puis, sur des bases de géométrie riemannienne, nous développons une notion de moyenne cohérente, puis de matrice de covariance. Nous généralisons alors d'autres opérations statistiques, telles que la distance de Mahalanobis et le test du χ^2. Enfin, nous montrons comment cette théorie peut être appliquée et implantée en machine dans une structure orientée objet générique.

Dans la seconde partie de la thèse, nous développons les calculs relatifs à ce formalisme pour les rotations et les transformations rigides agissant sur différents types de repères. Après analyse des algorithmes classiques sur les points, nous développons des algorithmes de recalage génériques basés sur les primitives, et nous proposons en particulier des méthodes pour estimer l'incertitude sur le recalage obtenu. Nous exposons parallèlement une méthode de validation statistique pour confirmer la précision de cette analyse, qui montre qu'un recalage d'une précision bien inférieure à la taille du voxel peut être obtenue dans le cas des images médicales 3D. Un deuxième volet de l'analyse statistique concerne la robustesse des algorithmes de reconnaissance.

Le champ applicatif que nous considérons va du recalage d'images médicales tridimensionnelles à la reconnaissance de sous-structures (3D) dans les protéines et souligne la validité de l'approche générique « orientée primitive » que nous avons choisie pour le traitement haut niveau de données géométriques.

Mots clés : *géométrie, probabilités, statistiques, incertitude, variété riemannienne, groupe de transformation, traitement d'image, recalage, reconnaissance, imagerie médicale, biologie moléculaire.*

Uncertainty in recognition and registration problems
Application in medical imaging and molecular biology

Usual geometric features in 3D image processing (like points, oriented points or frames) organize in a manifold which is generally not a vector space, onto which acts a transformation group that models the possible image view points. Their measure is moreover intrinsically corrupted by noise because of the combination of errors and inaccuracies in the processing system. The goal of this work is to generalize usual statistical notions to those features, as well as recognition and registration algorithms commonly used on points.

The first part concerns the development of mathematical tools to tackle those problems. We first show that we cannot simply consider these features as vectors and, on a Riemannian geometry basis, we develop a coherent definition of the expectation and the covariance matrix of random features. We then generalize other statistical operations, such as the Mahalanobis distance and the χ^2 test. Finally, we show how to apply and implement this theory in an object-oriented structure.

In the second part, we develop the calculus required to apply this framework to rotations and rigid transformations acting on different types of frames and points. We analyze classical algorithms for points registration and develop some new algorithms based on generic features. We also propose methods to estimate the uncertainty of the resulting registration. We then present a statistical validation method that corroborate the accuracy of this analysis. This shows that a registration accuracy far below the voxel size can be achieved for 3D medical images. Finally, we develop a statistical analysis related to the robustness of recognition algorithms.

The application field ranges from the registration of 3D medical images to the recognition of common substructures in proteins. The results emphasize the validity of the generic « oriented feature » approach we chose for processing high level geometric data.

Keywords: *geometry, probabilities, statistics, uncertainty, Riemannian manifold, transformation group, image processing, registration, recognition, medical imaging, molecular biology.*