

# THÈSE DE DOCTORAT

## Reconstruction sans-grille de courbes : théorie et applications en microscopie de fluorescence.

**Bastien Laville**

Équipe Morpheme (Centre Inria Université Côte d'Azur, CNRS, Université Côte d'Azur)

**Présentée en vue de l'obtention du grade de docteur en automatique, traitement du signal et des images d'Université Côte d'Azur**

**Dirigée par** : Laure Blanc-Féraud  
Directrice de recherche, CNRS, I3S

**Co-dirigée par** : Gilles Aubert  
Professeur des universités, CNRS, LJAD

**Soutenue le** : 15 septembre 2023

**Devant le jury, composé de :**

**Rapporteurs :**

Gabriel Peyré, Directeur de recherche, DMA, ENS ULM

Pierre Weiss, Directeur de recherche, Institut de Mathématiques de Toulouse

**Examineurs :**

Habib Ammari, Professeur, ETH Zurich

Vincent Duval, Chargé de recherche, Inria Paris (Mokaplan)

Juliette Leblond, Directrice de recherche, Inria (FACTAS)



# **Reconstruction sans-grille de courbes : théorie et applications en microscopie de fluorescence.**

Bastien Laville

## **Direction de thèse :**

### **Directrice :**

Mme Laure Blanc-Féraud (Directrice de recherche, CNRS I3S)

### **Co-directeur :**

M. Gilles Aubert (Professeur des universités, CNRS LJAD)

## **Membres du jury :**

### **Rapporteurs :**

M. Gabriel Peyré (Directeur de recherche, DMA, ENS Ulm)

M. Pierre Weiss (Directeur de recherche, Institut de Mathématiques de Toulouse)

### **Examineurs :**

M. Habib Ammari (Professeur, ETH Zurich)

M. Vincent Duval (Chargé de recherche, Inria Paris, Mokaplan)

Mme Juliette Leblond (Directrice de recherche, Inria UCA, FACTAS)

xvi+128 pages.

*Achevé d'impression : output.tex – le 17 août 2023, 13 h 30 +00:00*

## Résumé

Cette thèse de doctorat se propose de définir et d'expliciter les outils pour la reconstruction de courbes par des méthodes variationnelles sans-grille dans une perspective d'applications en imagerie biomédicale, et plus précisément en microscopie de fluorescence. Plusieurs algorithmes numériques convexes ou non-convexes ont déjà été développés au sein de l'équipe de recherche Morpheme qui a accueilli cette thèse, dans le cadre de l'optimisation non convexe parcimonieuse. Ces techniques ont été développées dans le cas discret, sur grille, ce qui signifie que l'image super-résolue est définie sur une grille plus fine que les images observées. *A contrario*, dans cette thèse de doctorat, nous développons des méthodes sans grille, en travaillant sur l'espace des mesures bornées. Ce choix apporte plusieurs avantages : la source n'est pas limitée par une grille fine et peut être localisée continûment, des bornes quantitatives sur l'écart entre la reconstruction et la source peuvent être énoncées, et ces approches permettent d'encoder une information structurelle ou géométrique dans le problème variationnel, de sorte à reconstruire une source d'une certaine forme, telle qu'un point ou une courbe.

Dans un premier temps, nous rappellerons l'état de l'art des méthodes sans-grille, qui porte sur la reconstruction de sources ponctuelles. Nous illustrerons en particulier son intérêt pratique en microscopie de fluorescence, notamment face à des modèles classiques : les reconstructions proposées sont aussi convaincantes que celles produites par des modèles d'apprentissage profond, et ce, sans nécessiter d'ensemble d'apprentissage ou d'hyperparamètres à régler. Puis, nous introduirons une méthode sans-grille dans le cadre de la microscopie de fluorescence par fluctuation. Cette dernière repose sur l'hypothèse d'indépendance des fluctuations de luminosité des fluorophores, ce qui permet en pratique de faire l'économie de matériel d'acquisition spécifique et coûteux. Les limites de ce modèle, formulé pour la reconstruction de sources ponctuelles, nous conduiront à nous intéresser sur la formulation d'une nouvelle méthode sans-grille, cette fois-ci pour la reconstruction de sources courbes. Nous nous pencherons sur la conception d'une méthode sans-grille, qui répond à la problématique de fluorophores le long de filaments qui forment de courbes dans les images, et qui n'est à notre connaissance pas traitée dans la littérature. En particulier, nous définirons l'espace d'optimisation impliqué, à savoir l'espace des mesures vectorielles à divergence finie, aussi appelé espace des

charges. Nous poursuivrons en explicitant la manière de définir des mesures portées par des courbes, et en écrivant une nouvelle fonctionnelle appelée CROC (pour *Curves Represented On Charges*). Nous finirons en montrant qu'il existe une solution de cette énergie qui s'écrit comme une combinaison linéaire finie de mesures portées par des courbes. Finalement, nous étudierons la mise en œuvre numérique de la minimisation de cette énergie, à l'aide d'une adaptation d'un algorithme glouton répandu dans la littérature. Nous discuterons des propriétés de notre algorithme, appelé *Charge Sliding Frank-Wolfe*, des difficultés rencontrées dans la définition du modèle d'observation, dans le choix de la paramétrisation discrète des courbes, etc. Le propos sera clôturé avec une conclusion qui rappellera les principaux points de chaque chapitre, ainsi qu'une ouverture, récapitulant les pistes de recherche possibles que nous avons dégagées au fil des chapitres.

**Mots-clefs :** méthodes sans-grille, imagerie biomédicale, mathématiques appliquées, estimation de mesures de Dirac.

# Abstract

This PhD thesis aims to define and explain the tools for curve reconstruction using off-the-grid variational methods, as part of a range of applications in biomedical imaging, and more specifically in fluorescence microscopy. Several convex or non-convex numerical algorithms have already been developed within the Morpheme research team, where this thesis was conducted, in the context of non-convex sparse optimisation. These techniques were developed in the discrete case, on a grid, which means that the super-resolved image is defined on a finer grid than the observed images. In contrast, in this PhD thesis, we are developing off-the-grid also called gridless methods, while working on the space of bounded measures. This choice brings several advantages : the source is not limited by a fine grid and can be located continuously, quantitative bounds on the discrepancy between the reconstruction and the source can be written, and these approaches allow encoding structural/geometric information in the variational problem, to recover a source of a certain shape, such as a point or a curve.

We will begin by reviewing the state of the art in off-the-grid methods, focusing on the reconstruction of point sources. In particular, we will illustrate its practical interest in fluorescence microscopy, especially compared with classical models : the reconstructions proposed are as convincing as those produced by deep learning models, and this without requiring a learning set or hyperparameters tuning. We will then introduce a gridless method for fluctuation-based fluorescence microscopy. This method exploits the independence of fluctuations in fluorophore luminosity, which means that specific and expensive acquisition equipment is not required. The limitations of this model, formulated for the reconstruction of point sources, will lead us to define a new off-the-grid method, this time for the reconstruction of curve. We will further study the design of an off-the-grid method, addressing the struggle of fluorophore along filaments that form curves in the images ; a problem which is not taken care of in the literature. In particular, we will define the involved optimisation space, namely the space of vector measures with finite divergence, also known as the space of charges. We will further explain how to define curve measures, and we will write a new functional called CROC (for Curves Represented On Charges). We will finish by showing that there is a solution to this energy that can be written as a finite linear combination of

measurements carried by curves. Finally, we will study the numerical implementation of this problem, using an adaptation of the greedy algorithm widely used in the literature. We will discuss the properties of our algorithm, called Charge Sliding Frank-Wolfe, and the difficulties encountered in the definition of the observation model, in choosing the discrete parametrisation of the curves, and so on. The thesis will be concluded with a reminder of the main points encountered in each chapter, and an opening section summing up the possible avenues of research that we have identified over the course of the chapters.

**Keywords:** off-the-grid methods, biomedical imaging, applied mathematics, Dirac measures estimation.

## Liste des symboles

$\mathbb{R}$	Ensemble des réels
$\mathbb{T}^d$	Tore en dimension $d \in \mathbb{N}^*$
$\mathbb{C}$	Ensemble des complexes
$\mathcal{X}$	Sous-ensemble de $\mathbb{R}^d$ dans lequel vivent les positions des mesures.
$\delta_x$	Mesure de Dirac à la position $x \in \mathcal{X}$ , aussi appelé $\delta$ -pic
$y$	Observation
$\Phi$	Opérateur <i>forward</i> , dit d'acquisition ou d'observation
$\eta$	Certificat dual d'un problème variationnel
$\mathcal{M}(\mathcal{X})$	Ensemble des mesures de Radon sur $\mathcal{X}$
$\ \cdot\ _{\text{TV}}$	Norme de la variation totale, notée TV
$\lambda, \alpha, \beta$	Paramètres de régularisation
$\partial f(x)$	Sous-différentiel de $f$ au point $x$
$R_y$	Covariance spatiale de la pile $y$
$\bar{y}$	Moyenne de la pile $y$
$\Lambda$	Opérateur de covariance
$\mathcal{L}^d$	Mesure de Lebesgue en dimension $d \in \mathbb{N}^*$
$\mathcal{H}^p$	Mesure de Hausdorff en dimension $p \in \mathbb{R}_+^*$
$\mathcal{V}$	Ensemble des mesures vectorielles à divergence finie, aussi appelées charges
div	Divergence au sens des distributions
$\mu_\gamma$	Mesure portée par une courbe $\gamma$ , élément de $\mathcal{V}$
$\nu_\gamma$	Mesure normalisée et portée par une courbe $\gamma$ , en particulier $\nu_\gamma = \mu_\gamma / \ \mu_\gamma\ _{\mathcal{V}}$

**viii** LISTE DES SYMBOLES

$\mathfrak{G}$  Ensemble des mesures normalisées portées par des courbes

Ext Ensemble des points extrémaux

$\mathcal{B}_\gamma^1$  Boule unité de la norme  $\|\cdot\|_\gamma$ , en particulier  $\text{Ext } \mathcal{B}_\gamma^1 = \mathfrak{G}$

LASSO *Least Absolute Shrinkage and Selection Operator*

BLASSO *Beurling LASSO*

CROC *Curves Represented On Charges*

SFW *Sliding Frank-Wolfe*

CPGD *Conic Particle Gradient Descent*

CSFW *Charge Sliding Frank-Wolfe*

# Introduction

## Avant-propos

L'imagerie biomédicale cherche à capturer des structures biologiques, notamment de petites tailles, à des échelles si petites que le phénomène de diffraction pour la lumière visible devient prépondérant dans l'image effectivement obtenue par le microscope optique. Le flou et les bruits qui en découlent parasitent alors l'image, empêchant d'en retirer toutes les informations. Afin de dépasser cette limite posée par la diffraction, nous pourrions travailler dans le cadre de la microscopie électronique, qui utilise un faisceau d'électrons pour illuminer un échantillon et en extraire une image de meilleure résolution. Les grossissements atteints en microscopie électronique sont ainsi plus forts qu'en microscopie optique, grâce à la longueur d'onde de De Broglie de l'électron bien plus petite que celles du visible. Néanmoins, les conditions d'acquisition en microscopie électronique (échantillon placé sous vide, refroidissement) ne permettent pas l'imagerie de cellules vivantes qui constituent pourtant l'un des cœurs de l'imagerie biomédicale. Dans cette perspective, il est donc nécessaire d'exploiter le cadre de la microscopie optique, travaillant dans les longueurs d'ondes du visible et devant composer avec les effets parasites de la diffraction.

Plusieurs approches biologiques se sont efforcées de contourner cette limite physique en exploitant les propriétés de différentes techniques d'imagerie, parmi lesquelles nous pouvons citer les méthodes de l'imagerie de fluorescence qui consiste à déposer des molécules fluorescentes sur la structure biologique à étudier. Ces composés appelés fluorophores émettent dans les longueurs d'onde du visible, et plusieurs protocoles d'acquisition ont été mis au point pour réaliser certaines propriétés (parcimonie des molécules dans une image, fluctuation temporelle d'amplitude du fluorophore, *etc.*) dont nous pouvons mettre à profit dans un post-traitement en aval, qui permet *in fine* d'améliorer la résolution des images produites.

Les méthodes variationnelles dites *sans-grille* peuvent en particulier tirer parti de ces procédés d'acquisitions, en permettant par exemple la localisation fine des fluorophores utilisés dans ce type d'imagerie. Elles sont appelées sans-grille en référence au caractère continu de l'estimation, qui ne repose pas sur une grille. De manière plus générale,

les méthodes sans-grille permettent d'incorporer un *a priori* structurel sur la source à localiser et s'affranchissent des problèmes de discrétisation rencontrés en méthode variationnelles discrètes.

Cette thèse est le résultat d'un travail s'efforçant de mêler ces approches pour construire de nouvelles méthodes sans-grille, que ce soit du point de vue numérique ou du point de vue pratique. Ces démarches ont été développées dans le contexte de l'imagerie biologique, mais elles restent valides dans le contexte plus général des problèmes inverses.

## Contexte général

Cette thèse de doctorat a été financée par la bourse EUR-DS4H, gérée par l'Agence Nationale pour la Recherche (ANR) avec la référence [ANR-17-EURE-0004](#). Elle a été préparée au sein de l'équipe de recherche Morpheme, équipe commune entre l'Inria Sophia-Antipolis, les laboratoires CRNS I3S (Informatique, Signaux et Systèmes de Sophia Antipolis, UMR 7271) et IBV (Institut de biologie de Valrose, UMR 7277) ainsi que l'Université Côte d'Azur.

## Brève introduction des problématiques biologiques

L'imagerie par microscopie de fluorescence offre plusieurs types d'approches permettant de passer au-delà de la limite de diffraction, afin de reconstruire des détails fins sur les structures biologiques étudiées. Comme évoqué plus haut, elle repose sur l'utilisation de fluorophores (des molécules fluorescentes) qui recouvrent la structure biologique d'intérêt. Plusieurs déclinaisons de ce principe ont été proposées, selon différentes propriétés des molécules exploitées. Nous cherchons dans cette thèse à imager des échantillons biologiques de structure curviligne, filamenteuse : ces formes sont particulièrement répandues dans le vivant, à toutes les échelles (vaisseaux sanguins, fibres, polymères du cytosquelette).

Dans cette optique, plusieurs nouvelles pistes de recherche se sont dégagées dans les objectifs de l'équipe Morpheme grâce aux travaux de l'Institut de biologie de Valrose. Les objets d'études portent donc sur l'imagerie d'organismes vivants de structure filamenteuse et susceptibles de se déplacer pendant le processus d'imagerie, ce qui impose des contraintes sur les temps d'acquisition. On peut par exemple citer la recherche sur *Candida Albicans*, un fungi généralement inoffensif dans le corps humain, qui peut devenir un agent responsable d'infection mortelle. Cette transformation se traduit par le passage de sa forme ovoïde naturelle à une forme filamenteuse : l'étude de la région apicale dans sa forme morbide pourrait permettre de comprendre le métabolisme à l'œuvre et par conséquent d'aider la recherche sur les traitements adéquats.

Dans cette même perspective, nous pouvons également relever le cas des algues du genre *Ostreopsis* : espèces tropicales proliférant maintenant en périodes de fortes chaleurs sur les plages de la Méditerranée ou sur la côte basque, elles peuvent relâcher des toxines qui posent de sérieux problèmes sanitaires et économiques, que ce soit pour la pêche

ou les activités balnéaires. L'étude des filaments contenus dans leur cytosquelette est essentielle dans la lutte contre ces espèces envahissantes.

## Objectifs et motivations

À travers les exemples mentionnés précédemment, il est clair que les contraintes sont ici de deux ordres : organisme vivant se déplaçant et structure filamenteuse. L'imagerie de fluorescence par fluctuation permet de répondre à cette première problématique, mais elle ne dispose pas d'une approche sans-grille pour l'exploiter. La seconde contrainte n'a pas non plus d'équivalent dans la littérature : les approches variationnelles sans-grille ne sont pas capables d'incorporer un *a priori* structurel de courbes dans leur reconstruction.

À ce titre, nous nous proposons d'abord dans cette thèse de dégager de l'état de l'art sans-grille les principaux résultats et méthodes utiles pour nos problèmes d'imagerie. L'accent est mis sur la qualité de la reconstruction, en un temps réduit. Puis, nous introduirons un nouveau cadre mathématique pour exploiter de manière sans-grille le protocole de microscopie de fluctuations. Les limites de cette modélisation nous conduiront dans un second temps à proposer une nouvelle formulation sans-grille permettant d'incorporer l'*a priori* structurel sur les courbes. Cette méthode sans-grille est d'abord formulée et prouvée théoriquement, avant d'être illustrée en pratique sur de premiers exemples.

## Organisation de la thèse

### Contributions principales

Les contributions de cette thèse de doctorat s'articulent autour de trois axes majeurs : la mise en œuvre numérique efficace des méthodes sans-grille reconstruisant des sources ponctuelles, le développement de méthodes sans-grille pour la microscopie de fluctuation avec des sources ponctuelles, et la construction d'une approche sans-grille permettant de reconstruire des sources courbes (ou linéiques, curvilignes, *etc.*) justifiée numériquement et pour laquelle un algorithme de reconstruction numérique est proposée. Plus précisément, cette thèse propose les apports à la littérature suivants :

- une revue détaillée des principaux résultats théoriques des méthodes sans-grille. Dans le cadre des mesures de Radon, les fonctionnelles en jeu ainsi que les garanties de reconstruction sont énoncées. Ces fondements théoriques seront illustrés par leur mise en œuvre en pratique. Des tests sur des données issues de l'imagerie de fluorescence sont effectués afin de montrer l'intérêt des méthodes sans-grille par rapport aux approches discrètes (reconstruction de pixels);
- un cadre de travail dynamique, afin de modéliser l'observation dans un cadre de microscopie de fluctuation de luminosité. À l'aide de l'hypothèse d'indépendance statistique des fluctuations de chaque fluorophore, un opérateur d'observation ainsi qu'un problème variationnel sans-grille sont formulés. Cette optimisation est mise en œuvre numériquement et ses résultats sont discutés;

- concernant la reconstruction de courbes sans-grille, quelques pistes de recherche pertinentes, mais infructueuses, sont présentées. En particulier, les raisons de leur échec sont explicitées;
- un espace de mesures  $\mathcal{V}$ , muni d'une norme  $\|\cdot\|_{\mathcal{V}}$  est proposé comme espace vectoriel modélisant les courbes. Quelques propriétés algébriques qu'il exhibe sont prouvées. Une nouvelle fonctionnelle appelée CROC qui repose sur une régularisation par la norme  $\|\cdot\|_{\mathcal{V}}$  est introduite. Les conditions d'optimalité ainsi que les certificats associés sont explicités;
- un nouveau théorème énonçant la structure géométrique des points extrémaux de la boule unité de  $\|\cdot\|_{\mathcal{V}}$  est introduit. Ce théorème est prouvé, et par là même la démarche pour reconstruire des mesures portées par des courbes par optimisation de la fonctionnelle CROC est validée;
- un algorithme appelé *Charge Sliding Frank-Wolfe* est proposé pour mettre en œuvre la résolution numérique de CROC. Les difficultés rencontrées dans la modélisation de l'observation, de la discrétisation et dans l'élaboration de garanties de reconstruction sont finalement énoncées et discutées.

### Organisation du manuscrit

Le manuscrit peut se découper en deux grandes parties. La première traite des problèmes inverses cherchant à estimer une mesure source ponctuelle. Dans un premier temps, un état de l'art des méthodes sans-grille est proposé, ce qui permet l'élaboration de nouveaux résultats numériques illustrant les capacités de ces méthodes en microscopie de fluorescence ([chapitre 1](#)). Dans cette perspective, un nouveau cadre d'application est considéré, à savoir la microscopie de fluorescence par fluctuations de luminosité ([chapitre 2](#)). La modélisation et les résultats obtenus ne semblent pas totalement convenir, car la méthode proposée postule la reconstruction de sources ponctuelles, là où les échantillons biologiques nécessitent de modéliser des sources courbes. Ainsi, dans un second temps, un nouveau cadre de travail est défini afin de permettre la reconstruction sans-grille de source courbe. Un espace variationnel, une énergie, des conditions d'optimalité et un théorème énonçant la structure courbe du minimiseurs sont énoncés ([chapitre 3](#)). La mise en œuvre numérique ainsi que les difficultés qui affleurent sont détaillées et discutées ([chapitre 4](#)).

## Liste de publications

Journaux à comité de lecture :

- « *Off-the-grid curve reconstruction through divergence regularisation : an extreme point result.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. **SIIMS, SIAM**. Volume 16, Issue 2, pp. 867-885. Juin 2023.
- « *Off-the-grid variational sparse spike recovery : methods and algorithms.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. **Journal of Imaging**, Volume 7, Issue 12. Décembre 2021.

Conférences internationales :

- « *Off-the-grid charge algorithm for curve reconstruction in inverse problems.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. Volume 14009 of the **Lecture Notes in Computer Science series, Springer**. Mai 2023.
- « *Off-the-grid covariance-based super-resolution microscopy.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. International Conference on Acoustics, Speech, & Signal Processing (**IEEE ICASSP**). Mai 2022.

Conférences nationales :

- « *Reconstruction de courbes en super-résolution sans-grille.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. **GRETSI**, Grenoble, France. Août 2023.
- « *Super-résolution sans-grille pour l'imagerie de fluctuation.* » Bastien Laville, Laure Blanc-Féraud et Gilles Aubert. **GRETSI**, Nancy, France. Septembre 2023.

Les travaux menés au cours de cette thèse ont également été présentés au sein de ces conférences/séminaires :

- « *An extreme point result for off-the-grid curve reconstruction.* » Invitation à présenter à la onzième édition de *Applied Inverse Problems Conference* à Göttingen, Germany. Septembre 2023.
- « *Off-the-grid curve reconstruction : bridge the gap between 0 and 2-rectifiable measures.* » Journée thématique MIA, Bordeaux, France. Mars 2023.
- « *Prix Pierre Laffitte : super-résolution sans-grille et retombées pour l'industrie.* » Mines Paris-PSL, Biot, France, finaliste pour le prix Pierre Laffitte. Octobre 2022.
- « *New off-the-grid approaches.* » Invitation à présenter au *symposium* sur la variation totale à la conférence de mathématiques appliquées CANUM (Évian-les-Bains, France). Juin 2022.
- « *Off-the-grid static and dynamic super-resolution.* » Invitation à présenter dans le cadre de l'ANR MicroBlind, au Centre International de Rencontres Mathématiques (Marseille, France). Décembre 2021.

Un article « *Dynamic off-the-grid reconstruction of curves in the roto-translation space* » est également en préparation avec le concours de M. Théo Bertrand et M. João Pinto Anastacio Machado (CNRS Ceremade, Paris).

# Table des matières

<b>Résumé</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Liste des symboles</b>	<b>vii</b>
<b>Introduction</b>	<b>ix</b>
Avant-propos . . . . .	ix
Contexte général . . . . .	x
Brève introduction des problématiques biologiques . . . . .	x
Objectifs et motivations . . . . .	xi
Organisation de la thèse . . . . .	xi
Contributions principales . . . . .	xi
Organisation du manuscrit . . . . .	xii
Liste de publications . . . . .	xiii
<b>Table des matières</b>	<b>xiv</b>
<b>I Reconstruction de mesures ponctuelles pour des problèmes en microscopie de fluorescence</b>	<b>1</b>
<b>1 Reconstruction sans-grille de sources ponctuelles, théorie et applications.</b>	<b>3</b>
1.1 Présentation . . . . .	4
1.2 Contexte théorique de la reconstruction sans-grille de pics . . . . .	10
1.2.1 Qu'est-ce qu'une mesure? . . . . .	10
1.2.2 Observations . . . . .	12
1.2.3 Une fonctionnelle sans-grille : le BLASSO . . . . .	14
1.2.4 Problèmes duaux et certificats . . . . .	15
1.2.5 Garanties de reconstruction de la mesure source . . . . .	17
1.3 Stratégies numériques pour résoudre le BLASSO . . . . .	18

1.3.1	Programmation semi-définie et flot de gradient . . . . .	19
1.3.2	Algorithme glouton : la méthode du gradient conditionnel . . . . .	21
1.4	Applications et résultats en imagerie SMLM . . . . .	26
1.4.1	Métrique quantifiant la qualité de la reconstruction . . . . .	26
1.4.2	Résultats sur une pile d'images SMLM . . . . .	28
1.5	Conclusion du chapitre . . . . .	32
<b>2</b>	<b>Fluctuations de luminosité : reconstruction de mesures ponctuelles</b>	<b>35</b>
2.1	Présentation . . . . .	36
2.2	Modélisation temporelle . . . . .	37
2.2.1	Positionnement du problème . . . . .	37
2.3	Un problème variationnel sur la covariance . . . . .	40
2.4	Résultats numériques . . . . .	46
2.4.1	Données simulées . . . . .	47
2.4.2	Données expérimentales . . . . .	48
2.4.3	Des reconstructions en pointillé . . . . .	50
2.5	Conclusion partielle . . . . .	51
<b>II</b>	<b>Reconstruction sans-grille de mesures courbes : théorie et applica- tions</b>	<b>53</b>
<b>3</b>	<b>Un nouveau régulariseur pour les mesures courbes</b>	<b>55</b>
3.1	Preliminaires . . . . .	56
3.1.1	Dérivées fractionnaires . . . . .	57
3.1.2	Transport optimal, <i>shortest paths</i> , ... . . . .	59
3.1.3	Théorème de la représentation : régulariseur et points extrémaux	59
3.2	L'espace des charges ou des champs de vecteurs à divergence finie . . . .	62
3.3	Un problème variationnel sur l'espace des charges . . . . .	66
3.4	Les points extrémaux de la boule unité de la norme $\mathcal{V}$ . . . . .	71
3.5	Conclusion partielle . . . . .	78
<b>4</b>	<b>Reconstruction sans-grille numérique de courbes</b>	<b>79</b>
4.1	Introduction . . . . .	80
4.2	L'algorithme du <i>Charge (Sliding) Frank-Wolfe</i> pour la reconstruction de courbes . . . . .	81
4.2.1	Rappels . . . . .	81
4.2.2	Un algorithme dans le cadre général . . . . .	82
4.3	Illustration numérique pour le problème de super-résolution . . . . .	84
4.3.1	Discrétisation des courbes . . . . .	84
4.3.2	Définition d'un opérateur d'observation . . . . .	85
4.3.3	Déroulement de l'algorithme : cas polygonal . . . . .	86
4.3.4	Discussion sur le cas Bézier . . . . .	92

4.3.5	Des pistes d'amélioration . . . . .	95
4.4	Conclusion . . . . .	96
<b>5</b>	<b>Conclusion et perspectives</b>	<b>97</b>
5.1	Conclusions et contributions . . . . .	97
5.2	<i>The Off-the-grid Cookbook</i> : des perspectives en sans-grille . . . . .	99
	<b>Précisions sur le chapitre 1</b>	<b>103</b>
	Définitions et notions utiles . . . . .	103
	Détails d'analyse fonctionnelle . . . . .	103
	Preuve de la dualité de Fenchel . . . . .	104
	Précisions sur les algorithmes par flot de gradient . . . . .	106
	Détails sur l'autodérivation . . . . .	114
	Détails sur le transport optimal . . . . .	116
	<b>Précisions sur le chapitre 3</b>	<b>117</b>
	<b>Bibliographie</b>	<b>121</b>

PARTIE I

---

RECONSTRUCTION DE MESURES PONCTUELLES POUR DES PROBLÈMES  
EN MICROSCOPIE DE FLUORESCENCE



## Chapitre 1

# Reconstruction sans-grille de sources ponctuelles, théorie et applications.

### Points clefs – les méthodes sans-grille sont pertinentes en problème inverse.

1. Ces travaux de recherche sont motivés par l'**imagerie biomédicale**, où l'on se propose de reconstruire des structures biologiques à partir d'acquisition floutée, bruitée, dégradée, *etc.* Ce travail s'inscrit dans le cadre plus général des **problèmes inverses**.
2. Au sein des méthodes résolvant les problèmes inverses de sources ponctuelles, on distingue plusieurs catégories et sous-catégories, dont les méthodes dites **sans-grille** ou **off-the-grid**. S'affranchissant des limites intrinsèques aux formulations matricielles, elles ont émergé dans la dernière décennie et font depuis l'objet d'**un intérêt certain**.
3. Les résultats tant théoriques que pratiques qu'elles produisent pour la reconstruction de  $\delta$ -pics ont suscité un engouement qui a donné lieu à de nombreux articles, essentiellement du point de vue des résultats mathématiques, mais également en pratique, avec des **algorithmes efficaces** et des **résultats convaincants**.

### Contributions – un état de l'art des méthodes sans-grille et des mises en œuvres.

4. Nous proposons dans ce chapitre **une revue** non exhaustive, mais suffisamment large, des **principaux résultats théoriques** des méthodes sans-grille. Dans le cadre des mesures  $\mathcal{M}(\mathcal{X})$ , les fonctionnelles en jeu ainsi que les garanties de reconstruction seront énoncés.
5. Ces fondements théoriques seront illustrés par leur mise en œuvre en **pratique**. Les **algorithmes** les plus courants seront détaillés et **testés** sur des exemples illustratifs.
6. Enfin, des tests sur des données issues de l'imagerie de fluorescence seront effectués afin de montrer l'intérêt des méthodes sans-grille par rapport à d'autres approches. Ces résultats sans-grille au niveau du reste de l'état de l'art seront également appréciés à l'aune des avantages que ces méthodes proposent, en ce qu'elles ne nécessitent **pas d'ensemble d'apprentissage**, ne disposent que d'**un seul paramètre à régler** et sont **simples d'utilisation** pour l'opérateur.

**Chapitre 1 – Reconstruction sans-grille de sources ponctuelles :**


---

1.1	Présentation . . . . .	4
1.2	Contexte théorique de la reconstruction sans-grille de pics . . . . .	10
1.2.1	Qu'est-ce qu'une mesure? . . . . .	10
1.2.2	Observations . . . . .	12
1.2.3	Une fonctionnelle sans-grille : le BLASSO . . . . .	14
1.2.4	Problèmes duaux et certificats . . . . .	15
1.2.5	Garanties de reconstruction de la mesure source . . . . .	17
1.3	Stratégies numériques pour résoudre le BLASSO . . . . .	18
1.3.1	Programmation semi-définie et flot de gradient . . . . .	19
1.3.2	Algorithme glouton : la méthode du gradient conditionnel . . . . .	21
1.4	Applications et résultats en imagerie SMLM . . . . .	26
1.4.1	Métrique quantifiant la qualité de la reconstruction . . . . .	26
1.4.2	Résultats sur une pile d'images SMLM . . . . .	28
1.5	Conclusion du chapitre . . . . .	32

---

**1.1 Présentation**

Dans cette partie, nous proposons une revue détaillée des méthodes *variationnelles sans-grille* aussi dites *off-the-grid* ou *gridless*, en particulier dans le cadre du problème de la reconstruction de  $\delta$ -pics/sources ponctuelles qui a en effet été le premier à susciter l'intérêt des chercheurs et qui constitue l'essentiel de la littérature de ce domaine. Nous présenterons les principaux résultats théoriques et numériques de la littérature, soulignant l'intérêt de ces méthodes dans de nombreux domaines gravitant autour des problèmes inverses. Dans le cadre de cette revue et de notre travail sur les méthodes sans-grille, nous avons développé une implémentation des méthodes numériques les plus performants, en mettant l'accent sur l'efficacité et la réduction du temps de calcul. À travers cette mise en œuvre, nous avons pu appliquer une méthode sans-grille au problème de la super-résolution en microscopie de fluorescence. Les codes et les résultats obtenus constituent un apport à la littérature sur les méthodes sans-grille, et sont une démonstration concrète de la pertinence de ces méthodes dans le domaine des problèmes inverses.

Il convient tout d'abord de définir ce que sont les problèmes inverses. De manière formelle, ces problèmes consistent à reconstruire les causes à partir des conséquences. Ils modélisent précisément de nombreux problèmes physiques, notamment en ce qui concerne l'acquisition par des capteurs : on cherche ainsi à reconstruire une certaine quantité physique telle que la position, intensité, conductivité, *etc.*, à partir de mesures d'une autre quantité qui lui est liée. Ce lien s'écrit concrètement comme une équation dite *d'observation*, qui peut être en pratique une équation matricielle, fonctionnelle, aux dérivées partielles, *etc.*

Le problème est généralement mal posé, ce qui signifie que l'existence, l'unicité et la stabilité d'une ou de plusieurs solutions du problème inverse ne sont pas garanties.

Le *sparse spikes problem*, que l'on pourrait traduire par « problème de pics parcimonieux », est un exemple pratique de problème inverse rencontré dans de nombreux domaines tels que le traitement des images ou des signaux, les télécommunications, l'apprentissage automatique, la super-résolution, etc. Il s'agit de reconstruire des sources ponctuelles, dites  $\delta$ -pics, situées sur un domaine  $\mathcal{X} \subset \mathbb{R}^d$  en dimension  $d \in \mathbb{N}^*$  à partir d'une acquisition  $y$ . On utilise pour ce faire un *a priori* sur la source, que l'on suppose parcimonieuse : nous savons en effet que la source doit être composée de seulement quelques sources ponctuelles, et le modèle doit le prendre en compte pour « mieux poser » le problème inverse. Cela inclut des sources telles que les étoiles en astronomie, les fractures en sismologie, etc.

Un pic est typiquement modélisé par une mesure de Dirac  $a\delta_x$  avec une amplitude  $a \in \mathbb{C}$  et une position  $x \in \mathcal{X}$ . Toute la difficulté réside dans l'estimation du nombre  $N$  de  $\delta$ -pics, de leurs amplitudes  $(a_i)_{i=1}^N$  et de leurs positions  $(x_i)_{i=1}^N$ . L'objectif est donc de reconstruire la mesure  $m = \sum_{i=1}^N a_i \delta_{x_i}$  uniquement à partir d'un petit nombre d'observations  $y$  dans un espace de Hilbert  $\mathcal{H}$  (typiquement  $L^2(\mathcal{X})$ ) lié à  $m$  par l'intermédiaire d'un opérateur  $\Phi$  qui tient compte de la détérioration de l'entrée (flou, dégradation de l'échantillonnage ou *downgrading*) telle que  $y \stackrel{\text{def.}}{=} \Phi m + w$  où  $w \in \mathcal{H}$  est un bruit additif. La reconstruction des pics est dite *sans-grille* (ou « *off-the-grid* ») si les positions  $(x_i)_{i=1}^N$  ne sont pas contraintes sur une grille et donc  $(x_i)_{i=1}^N$  ne sont pas limitées à un ensemble fini de valeurs<sup>1</sup> : ceci offre de nouvelles perspectives mathématiques intéressantes, avec par exemple des garanties pour la reconstruction ; au détriment d'une difficulté mathématique accrue (perte de certaines propriétés algébriques, abstraction supplémentaire, etc.).

**Attention** — Le terme *sans-grille* peut être trompeur : le lecteur non averti pourrait en effet se demander comment la reconstruction peut-être sans-grille alors même que l'on travaille avec des images, donc des acquisitions munies d'une grille, et qu'il faut bien discrétiser la position des pics pour la mise en œuvre numérique. L'expression anglaise *off-the-grid* est plus parlante : elle incorpore mieux l'idée que le modèle sans-grille est la limite formelle (*off*, comme *par-delà la grille*) d'un modèle avec grille, dont le pas de discrétisation serait de plus en plus fin. On cherche ainsi à estimer continûment les valeurs des coordonnées des pics, justement pour éviter de raffiner la grille comme le font les méthodes discrètes. Ces deux modèles sont rediscutés un peu plus loin dans ce chapitre.

Le problème des pics parcimonieux est rencontré dans de nombreuses situations, telles que :

- l'acquisition comprimée [15], où l'on cherche à reconstruire un vecteur qui est *s*-parcimonieux,  $v \in \mathbb{C}^N$  d'après  $M$  mesures physiques  $Av_0$  où  $A \in \mathbb{C}^{M \times N}$  ;
- le *machine learning*, les « *mixtures* », etc. Si nous souhaitons par exemple régresser une distribution de probabilité à des données, il s'agira d'estimer les paramètres

1. La grille est un ensemble discret, d'où le terme de méthode *discrète*.

- $(a_i) \in \mathbb{R}^N$  et  $(x_i) \in \mathcal{X}^N$  d'un mélange  $\sum_{i=1}^N a_i \varphi(x_i)$  de  $N$  distributions de base décrites par  $\varphi$ . On pourrait ainsi retrouver les moyennes  $\mu_i \in \mathbb{R}$  et les écarts-types  $\sigma_i \in \mathbb{R}^+$  d'une mixture gaussienne, voir [29] pour plus de détails sur ce sujet ;
- l'apprentissage profond tel que l'entraînement des réseaux de neurones avec une seule couche cachée [20] ;
  - le traitement du signal, par exemple avec la décomposition tensorielle de bas rang pour l'estimation de la direction d'arrivée, plus connue comme la *Direction of Arrival*, à travers un réseau de capteurs (donc multiples échantillonnages) ;
  - la super-résolution, un problème central dans le traitement des images. Il s'agit ainsi de reconstruire des détails fins à partir d'un signal (ou d'une image) altéré. Elle inclut les opérateurs physiques classiques d'acquisition tels que l'échantillonnage de Fourier, la transformée de Laplace ou la convolution gaussienne.

C'est ce dernier point qui nous intéresse dans le cadre du *sparse spike problem*. Toute la difficulté procède de la dégradation induite par le processus d'acquisition, qui implique en particulier deux choses :

- une détérioration par le système d'acquisition, typiquement modélisée par la fonction d'étalement du point, aussi appelée PSF pour *point spread function* en imagerie, qui agit comme un filtre passe-bas ;
- l'acquisition du capteur, qui entraîne un échantillonnage et une pollution par des bruits de différents types, caractérisés par des densités telles que gaussienne, de Poisson, etc.

En résumé, nous souhaitons reconstruire le nombre correct de  $\delta$ -pics avec des amplitudes et des positions dans le cadre continu à partir d'une acquisition discrète filtrée et bruitée. D'un point de vue théorique, cette question peut être abordée soit par l'approche variationnelle, soit par la méthode de Prony :

- La méthode de Prony [55] et ses variantes telles que MUSIC (*MUltiple SIgnal Classification*), ESPRIT (*Estimation of Signal Parameters by Rotational Invariance Techniques*) ou *Matrix Pencil*, qui reconstruisent le signal source à partir de  $N$  échantillons uniformément répartis dans le spectre de Fourier, en 1D et sans bruit. Il s'agit de décomposer le signal sur une base d'exponentielles complexes, paramétrées par l'amplitude  $A_i$ , le facteur d'amortissement  $\sigma_i$ , la fréquence  $\omega_i$  ainsi que le déphasage  $\phi_i$  ; soit en notant  $\hat{f}$  le signal reconstruit :

$$\hat{f}(t) = \sum_{i=1}^N A_i e^{\sigma_i t} \cos(\omega_i t + \phi_i).$$

Les résultats sont convaincants dans le cas 1D sans bruit et peuvent être étendus à un contexte multivarié et bruité ; mais ces méthodes manquent encore de polyvalence (1D, opérateur nécessairement type Fourier, échantillonnage uniforme, etc.) puisqu'elles ne peuvent pas être étendues au contexte qui nous intéresse. C'est pourquoi nous n'examinerons pas cette approche dans le présent document ;

- l’approche variationnelle qui n’impose aucune structure particulière à l’opérateur d’acquisition, qui peut être adaptée à n’importe quel type de bruit et qui ne nécessite aucun a priori sur le nombre de sources ponctuelles. L’idée principale est de résoudre le problème inverse en trouvant, parmi toutes les sources de signaux possibles, celle qui minimise une fonction objective appelée *l’énergie*, formulée comme un compromis entre un terme de fidélité des données et un terme de régularisation, qui applique un a priori sur la forme de la source à reconstruire.

Il existe en particulier deux types d’approches variationnelles : les approches discrètes et les approches sans-grille. Dans le cadre discret, on cherche à reconstruire les pics sur une grille fine prescrite, typiquement avec plus de points que la grille qui munit l’image d’acquisition. Nous appelons ainsi grille grossière celle qui équipe l’acquisition à faible résolution, et grille fine la grille plus précise (par un facteur dit de super-résolution  $q \in \mathbb{N}^*$ ) de la reconstruction. Il s’agit donc d’un problème de dimension finie, où les positions des sources ponctuelles doivent nécessairement se trouver sur une grille  $\mathcal{G}$  de  $L$  points maillant le domaine  $\mathcal{X}$ . Il s’agit ainsi d’un problème de reconstruction de vecteurs parcimonieux, et il peut être traité en imposant l’hypothèse de parcimonie par la minimisation de la norme  $\ell_1$  du vecteur inconnu. Ce problème est connu sous le nom de LASSO [88] ou de poursuite de base (*basis pursuit*), défini comme un problème variationnel avec un paramètre de régularisation  $\lambda > 0$  contrôlant le compromis entre la fidélité aux données et l’application de l’a priori :

$$\min_{a \in \mathbb{R}^L} \underbrace{\|y - \Phi_L a\|_{\mathcal{H}}^2}_{\text{attache aux données}} + \underbrace{\lambda \|a\|_1}_{\text{a priori parcimonieux}} \quad (\text{LASSO})$$

où  $\Phi_L : \mathbb{R}^L \rightarrow \mathcal{H}$  est l’opérateur d’acquisition avec un vecteur de taille  $L$  comme entrée et  $\mathcal{H}$  est un espace de Hilbert. Une grille est utile pour illustrer le concept de parcimonie dans le cas des pics : en effet, la parcimonie exprime simplement le fait que seuls quelques points de la grille  $L$  sont d’amplitudes non-nulles. De plus, comme un ordinateur ne peut stocker que des tableaux et des vecteurs, il est naturel de travailler avec des problèmes de dimension finie, même pour l’analyse théorique. Ceci étant, on pourrait se demander comment choisir la discrétisation. Une grille avec un pas trop petit produit des instabilités numériques [42], là où un pas trop grand conduit à des erreurs d’arrondi. On aimerait donc localiser les pics le plus précisément possible sans avoir recours à une grille : une discrétisation des positions entraînerait nécessairement une approximation sur les positions. La formalisation mathématique appropriée pour évacuer ces inconvénients de discrétisation consiste à représenter une collection de  $\delta$ -pics par des mesures de Dirac, un élément de l’espace des mesures de Radon  $\mathcal{M}(\mathcal{X})$  : on peut pour l’instant voir cet espace comme une généralisation de la notion de fonction. L’opérateur d’acquisition est maintenant  $\Phi : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}$ , l’a priori géométrique est ainsi imposé par une norme sur  $\mathcal{M}(\mathcal{X})$  appelée la norme TV. Ce problème variationnel est appelé BLASSO (pour Beurling LASSO) :

$$\min_{m \in \mathcal{M}(\mathcal{X})} \underbrace{\|y - \Phi m\|_{\mathcal{H}}^2}_{\text{attache aux données}} + \underbrace{\lambda \|m\|_{\text{TV}}}_{a \text{ priori parcimonieux}}. \quad (\text{BLASSO})$$

**Attention** — La norme TV dont il est question ici ne doit pas être confondue avec la *régularisation TV*, un concept du traitement d’images qui fait cette fois-ci intervenir la régularisation du gradient d’une fonction intégrable. Nous aurons l’occasion de revenir sur ce point et d’en discuter plus en détail dans la suite du propos.

Dans le cadre sans-grille, les pics peuvent se déplacer continûment sur le domaine  $\mathcal{X}$  : une comparaison entre la reconstruction discrète et la reconstruction sans-grille est donnée dans la figure 1.1. Le cadre sans-grille peut être considéré comme la limite du cas discret avec une grille de plus en plus fine [41].

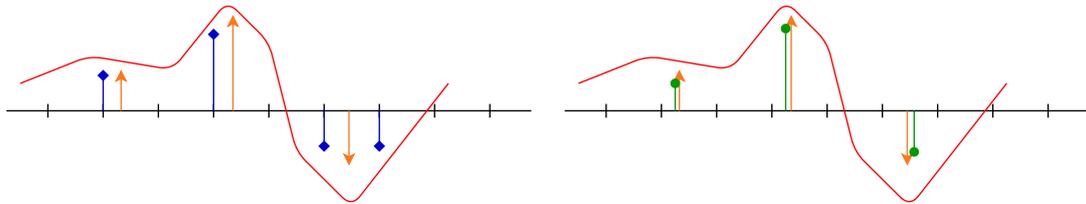


FIGURE 1.1 – (a) Reconstruction discrète, qui peut être entendue comme des pics avec un support contraint sur une grille dite fine (b) Reconstruction sans-grille, les pics peuvent se déplacer continuellement sur domaine, ici la ligne. La ligne rouge est l’acquisition  $y$ , les pics oranges sont la source (la cause que nous voulons retrouver), les pics bleus sont la reconstruction discrète contrainte sur une grille et les verts peuvent se déplacer librement puisqu’ils sont sans-grille. Notons que lorsqu’un pic de la source se trouve entre deux points de la grille, deux pics peuvent être retrouvés dans la reconstruction discrète.

Ce passage du domaine discret au domaine continu appelé sans-grille ou « *off-the-grid* » conduit à des travaux théoriques importants, en particulier avec la définition d’un critère précis dépendant du signal pour la récupération des pics stables [43], et des résultats sur la distance de séparation minimale (voir la section suivante). Évidemment, certaines difficultés inhérentes à la dimension infinie et au manque de propriétés algébriques de l’ensemble d’optimisation surviennent ; nous discuterons dans la suite des stratégies permettant de les prendre en charge. La comparaison entre les cadres discrets et sans-grille peut être résumée comme suit :

- le problème discret est modélisé par le (LASSO), c’est-à-dire par la minimisation d’une fonction convexe définie sur une grille fine dans un espace de Hilbert de dimension finie  $\mathbb{R}^L$ . La norme  $\ell_1$  peut ne pas imposer suffisamment l’*a priori* de parcimonie : on peut alors la remplacer par la pseudo-norme non-continue  $\ell_0$ , mais on aboutit alors à un problème non-convexe combinatoire NP-difficile. Il

existe des relaxations continues de  $\ell_0$  telles que CEL0 [84], mais, eu égard à son caractère non-convexe, le problème reste difficile tant d'un point de vue théorique que numérique. Malgré le manque de garanties, il existe de nombreux algorithmes (ISTA, FISTA, IRL1 [84], *etc.*) pour calculer la solution de LASSO ou de sa variante relaxée  $\ell_0$  ;

- le problème sans-grille est traité par le (BLASSO), une fonctionnelle convexe définie sur  $\mathcal{M}(\mathcal{X})$ . La propriété convexe est essentielle d'un point de vue théorique car elle permet d'obtenir des informations cruciales sur l'existence, l'unicité et l'estimation du support selon le bruit. En revanche,  $\mathcal{M}(\mathcal{X})$  est un espace de Banach sans structure hilbertienne, ce qui complique fortement la formulation explicite d'algorithme proximal, il est de surcroît de dimension infinie et non réflexif pour la topologie forte : les résultats de convergence portent alors essentiellement sur la topologie faible-\*. Malgré ce manque de propriétés algébriques, nous disposons aujourd'hui d'un large éventail d'algorithmes pour traiter ce problème, tels que les algorithmes de recherche de racines [64, 30], les flots de gradient [20, 21] ou les algorithmes gloutons [49, 13, 35].

**Attention** — Ce chapitre se concentre sur les résultats de la littérature, portant pour l'essentiel sur la question des sources ponctuelles. Tout un pan des derniers travaux en sans-grille étudient la reconstruction de sources exhibant d'autres géométries : c'est entre autres ce que nous nous proposons de faire au chapitre 3. La bonne compréhension des notions fondamentales en reconstruction de sources ponctuelles est impérative dans la suite de nos travaux, et c'est pourquoi nous ne traiterons dans ce chapitre que de la reconstruction de mesures de Dirac, que nous mettrons en perspective avec d'autres structures de mesures dans le chapitre 3.

Une reconstruction sans-grille peut ensuite être évaluée à l'aide de métriques appropriées, par exemple la *flat metric* [33] exploitant la théorie du transport optimal des mesures. Cette métrique évalue la qualité de la reconstruction et peut être appliquée directement aux résultats de la reconstruction sans-grille et même discrète.

Dans ce qui suit, nous passons en revue les résultats importants du domaine des méthodes variationnelles sans-grille. Le document est organisé en trois sections, à savoir :

- l'analyse variationnelle de l'espace  $\mathcal{M}(\mathcal{X})$ , les propriétés et les garanties de reconstruction concernant le *sparse spike problem* aujourd'hui bien documentées [16, 13, 17, 43] seront rappelées dans la section théorique 1.2 ;
- de nombreuses stratégies ont été envisagées pour mettre en œuvre numériquement la résolution du (BLASSO), les plus pertinentes seront présentées et mises en contexte dans la section numérique 1.3 ;
- des applications pratiques et de nouvelles métriques ont été envisagées pour les méthodes sans-grille, comme la super-résolution SMLM ; ces résultats sont présentés et discutés dans la section 1.4.

**Résumé** — À la fin de chaque paragraphe, un encadré gris intitulé « résumé » comme celui-ci rappellera les principaux résultats mis en évidence dans la section. Nous invitons le lecteur à s’y référer pour un résumé rapide de la section.

## 1.2 Contexte théorique de la reconstruction sans-grille de pics

Dans la suite du propos,  $\mathcal{X}$  désigne l’espace ambiant dans lequel vivent les positions des pics. Nous supposons que  $\mathcal{X}$  est un sous-ensemble de  $\mathbb{R}^d$  tel que son intérieur  $\overset{\circ}{\mathcal{X}}$  est une sous-variété de dimension  $\mathbb{N}^*$  [35]. Ce cadre englobe  $\mathcal{X} = \mathbb{R}^d$ , le tore  $\mathcal{X} = \mathbb{T}^d \stackrel{\text{def.}}{=} \mathbb{R}^d / \mathbb{Z}^d$ , tout compact avec un intérieur non vide, *etc.*

### 1.2.1 Qu’est-ce qu’une mesure ?

Comme nous l’avons indiqué dans la section précédente, la mesure de Dirac est l’objet approprié pour décrire un pic qui n’est pas contraint sur un ensemble fini de positions. Cet objet n’est pas une fonction, car il est impossible d’exhiber une classe d’équivalence intégrable satisfaisant les propriétés de la mesure de Dirac. Il faut donc considérer la notion de mesure de Radon, qui peut s’interpréter comme une généralisation des fonctions classiques.

**Attention** — Le terme de mesure est assez ambiguë dans un cadre de problème inverse, on pourrait ainsi confondre l’objet mathématique *mesure (de Radon)* avec la grandeur physique qui est mesurée (comme le nombre de photons dans le capteur de l’image, *etc.*). Une mesure de Radon n’est *a fortiori* pas un relevé de la quantité de Radon dans l’observation : dans la suite du document, nous commettrons un abus de langage en omettant le substantif *de Radon*, mais il doit bien être clair qu’une mesure est un cas particulier de distribution ; nous parlerons d’observation ou d’acquisition pour lever toute ambiguïté sur la mesure physique.

D’un point de vue fonctionnel, c’est un sous-ensemble de l’espace de distribution  $\mathcal{D}'(\mathcal{X})$ , à savoir l’espace des formes linéaires sur l’espace des fonctions de test  $\mathcal{D}(\mathcal{X})$  c’est-à-dire fonctions lisses (dérivées continues à tout ordre) à support compact. Cette approche fonctionnelle<sup>2</sup> consiste à définir une mesure comme une forme linéaire sur un espace de fonctions, à savoir :

2. On peut en réalité définir de manière équivalente l’espace des mesures de Radon, soit par une approche liée aux ensembles, soit par une approche d’analyse fonctionnelle (grâce au théorème de Riesz-Markov). Dans l’approche liée aux ensembles, une *mesure* est un objet qui prend ses valeurs sur des ensembles. Une *mesure de Borel* est une mesure définie sur tous les ensembles ouverts de  $\mathcal{X}$ , et une *mesure de Radon* est une mesure de Borel telle qu’elle est finie sur tous les ensembles compacts de  $\mathcal{X}$  (à un isomorphisme près). Le point de vue des distributions et le point de vue ensemblistes représentent deux visions différentes pour décrire le même objet.

**Définition 1.1** (Fonctions continues évanescentes sur  $\mathcal{X}$ ). Nous appelons  $\mathcal{C}_0(\mathcal{X}, \mathcal{Y})$  l'ensemble des fonctions continues nulles à la limite/au bord (aussi dites évanescentes), à savoir toutes les applications continues  $\psi : \mathcal{X} \rightarrow \mathcal{Y}$  telles que :

$$\forall \varepsilon > 0, \exists K \subset \mathcal{X} \text{ compact}, \quad \sup_{x \in \mathcal{X} \setminus K} \|\psi(x)\|_{\mathcal{Y}} \leq \varepsilon.$$

Lorsque  $\mathcal{Y} = \mathbb{R}$  nous écrirons simplement  $\mathcal{C}_0(\mathcal{X})$ . Puisque nous disposons d'un espace de fonctions de test approprié, nous pouvons préciser la dualité en jeu :

**Définition 1.2** (Dual topologique). Si  $E$  est un espace vectoriel topologique, on note  $E^*$  son dual topologique c'est-à-dire l'espace de toutes les formes linéaires continues  $\psi : E \rightarrow \mathbb{R}$ . Le produit entre un élément  $\phi \in E$  et une application  $\psi \in E^*$  est appelée crochet de dualité et se note avec l'application bilinéaire  $\langle \phi, \psi \rangle_{E \times E^*} \stackrel{\text{def.}}{=} \psi(\phi)$ .

Nous pouvons alors définir l'espace des mesures de Radon avec cette notion de dualité :

**Définition 1.3** (Ensemble des mesures de Radon). On note  $\mathcal{M}(\mathcal{X})$  l'ensemble des mesures de Radon réelles signées sur  $\mathcal{X}$  de masses finies. C'est le dual topologique<sup>3</sup> de  $\mathcal{C}_0(\mathcal{X})$  avec la norme supremum  $\|\cdot\|_{\infty, \mathcal{X}}$  par le théorème de représentation de Riesz-Markov [46]. Ainsi, une mesure de Radon  $m$  est une forme linéaire continue sur les fonctions  $f \in \mathcal{C}_0(\mathcal{X})$ , et pour  $m \in \mathcal{M}(\mathcal{X})$  on définit le crochet de dualité  $\langle m, f \rangle_{\mathcal{M}} \stackrel{\text{def.}}{=} \int_{\mathcal{X}} f \, dm$ .

Le terme « signé » fait référence à la généralisation du concept de mesure positive, en autorisant la quantité  $\langle f, m \rangle_{\mathcal{M}}$  à être négative. Nous pouvons définir de la même manière l'espace des mesures de Radon réelles positives  $\mathcal{M}^+(\mathcal{X})$  dual de  $\mathcal{C}_0(\mathcal{X}, \mathbb{R}^+)$  et l'espace des mesures de Radon complexes  $\mathcal{M}_{\mathbb{C}}(\mathcal{X})$  dual de  $\mathcal{C}_0(\mathcal{X}, \mathbb{C})$ .

Les exemples classiques de mesures de Radon sont les suivants :

- la mesure de Lebesgue en dimension  $d \in \mathbb{N}^*$  ;
- la mesure de Dirac  $\delta_z$  centrée en  $z \in \mathcal{X}$ , aussi appelé  $\delta$ -pic. Pour toute fonction  $f \in \mathcal{C}_0(\mathcal{X})$ , on a  $\langle f, \delta_z \rangle_{\mathcal{M}} = f(z)$  ;
- les mesures discrètes  $m_{a,x} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$  où  $N \in \mathbb{N}$ ,  $\mathbf{a} \in \mathbb{C}^N$ ,  $\mathbf{x} \in \mathcal{X}^N$ .

Puisque  $\mathcal{C}_0(\mathcal{X})$  est un espace de Banach,  $\mathcal{M}(\mathcal{X})$  est complet [13] en le munissant de sa norme duale appelée norme de la variation totale (TV), définie pour  $m \in \mathcal{M}(\mathcal{X})$  par :

$$|m|(\mathcal{X}) \stackrel{\text{def.}}{=} \|m\|_{\text{TV}} \stackrel{\text{def.}}{=} \sup \left( \int_{\mathcal{X}} f \, dm, f \in \mathcal{C}_0(\mathcal{X}), \|f\|_{\infty, \mathcal{X}} \leq 1 \right).$$

La norme TV d'une mesure est aussi appelée sa *masse*. Il convient de noter que pour une mesure discrète, définie plus haut comme  $m_{a,x} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$ , il vient  $|m_{a,x}|(\mathcal{X}) = \|\mathbf{a}\|_1$ .

3. il peut également être défini comme le dual topologique de l'espace des fonctions continues  $\mathcal{C}(\mathcal{X})$  si  $\mathcal{X}$  est compact.

Cette norme généralise donc la norme  $\ell_1$  aux mesures, il est intéressant au passage de noter que l'espace des fonctions intégrables s'injecte continûment<sup>4</sup> dans l'espace des mesures :  $L^1(\mathcal{X}) \hookrightarrow \mathcal{M}(\mathcal{X})$ .

**Attention** — On note indifféremment la norme TV de  $m$  avec  $|m|(\mathcal{X})$  ou  $\|m\|_{\text{TV}}$ . La première écriture permet de rappeler le caractère *mesure* que porte  $m$ , donc d'insister sur sa capacité à *mesurer* un borélien, ici  $\mathcal{X}$ .

Le lecteur intéressé pourra consulter l'annexe 5.2 pour plus de détails sur certaines notions et résultats de l'analyse fonctionnelle.

**Résumé** — On modélise une source ponctuelle par une mesure de Dirac, un élément de l'espace des mesure de Radon  $\mathcal{M}(\mathcal{X})$ . Cet espace est défini par dualité, il est complet s'il est muni de la norme TV. Il est cependant de dimension infinie et non réflexif (voir 5.2), ce qui pose des difficultés supplémentaires à prendre en compte dans l'optimisation.

### 1.2.2 Observations

Introduisons l'espace où vivent les données acquises. Nous désignerons par  $\mathcal{H}$  cet espace de Hilbert; par exemple des images  $\mathcal{H} = L^2(\mathcal{X})$ . Si  $m \in \mathcal{M}(\mathcal{X})$  est la mesure source, nous appelons *observation* ou *acquisition*  $y \in \mathcal{H}$  le résultat de l'opérateur d'acquisition  $\Phi : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}$  évalué sur  $m$ , avec le noyau d'observation/d'acquisition  $\varphi : \mathcal{X} \rightarrow \mathcal{H}$  :

$$y \stackrel{\text{def.}}{=} \Phi m = \int_{\mathcal{X}} \varphi(x) dm(x). \quad (1.1)$$

Cette dernière intégrale ne doit pas être confondue avec le crochet de dualité  $\langle f, m \rangle_{\mathcal{M}} = \int_{\mathcal{X}} f(x) dm(x)$  mentionné dans la définition 1.3 ci-dessus. En effet, si  $f(x) \in \mathbb{R}$  pour  $x \in \mathcal{X}$ , on a  $\varphi(x) \in \mathcal{H}$  : l'intégrale dans (1.1) est alors une *intégrale de Bochner* [22], qui est la notion appropriée pour traiter les applications à valeurs vectorielles. Elle est bien définie pour peu que  $\varphi$  soit continue et bornée [20, 35]. Pour résumer :

- pour une fonction test  $f \in \mathcal{C}_0(\mathcal{X})$ , on a logiquement la dualité  $\int_{\mathcal{X}} f(x) dm(x) \in \mathbb{R}$ ;
- en ce qui concerne l'opérateur  $\Phi$ , il faut bien remarquer que  $\int_{\mathcal{X}} \varphi(x) dm(x) \in \mathcal{H}$ !

**Remarque.** *Les mesures sont des objets qui généralisent les fonctions, au prix de certaines de leurs propriétés. Ainsi, on ne peut pas définir un produit de mesures : que serait alors le carré d'une mesure Dirac ? Plus généralement pour les distributions, on ne peut pas disposer tout à*

4. On dit qu'un espace vectoriel normé (EVN)  $E$  s'injecte continûment dans un autre EVN  $F$ , qu'on note  $E \hookrightarrow F$ , si pour tout  $u \in E$ , il existe  $C$  tel que  $\|u\|_F \leq C\|u\|_E$ . Dit autrement, toute suite qui converge dans  $E$  convergera également dans  $F$ , vers la même limite.

la fois des propriétés de dérivation, de multiplication entre distributions et du  $\delta$  de Dirac [80]. Il faut également prendre garde dans la définition des fonctions de mesure : ces fonctionnelles doivent être au plus (sous-)linéaires pour être bien définies [87].

Dans la suite, on supposera  $\varphi \in \mathcal{C}^2(\mathcal{X}, \mathcal{H})$ . Définissons l'opérateur adjoint de  $\Phi : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}$  pour la topologie faible-\*, à savoir l'application  $\Phi^* : \mathcal{H} \rightarrow \mathcal{C}_0(\mathcal{X})$ . Il est défini pour  $x \in \mathcal{X}$  et  $p \in \mathcal{H}$  par  $\Phi^*(p)(x) = \langle p, \varphi(x) \rangle_{\mathcal{H}}$ . Le choix de  $\varphi$  et de  $\mathcal{H}$  dépend du processus physique d'acquisition, les noyaux les plus classiques étant :

- le noyau de convolution, typiquement  $\mathcal{H} = L^2(\mathcal{X})$  et  $\forall x \in \mathcal{X}, \varphi(x) \stackrel{\text{def.}}{=} (s \mapsto \tilde{\varphi}(s - x)) \in \mathcal{H}$ , pour la PSF  $\tilde{\varphi} \in \mathcal{C}_0^2(\mathbb{R}^d)$ . On utilise par exemple le noyau gaussien, centré en  $c \in \mathcal{X}$  et d'écart-type  $\sigma > 0$ , défini par  $s \mapsto \tilde{\varphi}(s - c) \stackrel{\text{def.}}{=} 1 / \sqrt{2\pi\sigma^2} e^{-\|s-c\|_2^2 / 2\sigma^2}$  ;
- le noyau de Fourier avec la fréquence de coupure  $f_c \in \mathbb{N}$  et  $\mathcal{H} = \mathbb{C}^{2f_c+1}$ , pour  $x \in \mathcal{X} = \mathbb{T}$  en 1D :

$$\varphi(x) = \left( e^{2i\pi kx} \right)_{|k| \leq f_c} ;$$

- le noyau de Laplace [32] pour une fonction de pondération positive  $\xi \in \mathcal{C}(\mathcal{X})$  propre au processus d'acquisition et  $\mathcal{H} = L^2(\mathbb{R}^+)$  :  $\forall x \in \mathcal{X}, \varphi(x) \stackrel{\text{def.}}{=} (s \mapsto \xi(x)e^{-sx}) \in \mathcal{H}$ .

Ces trois noyaux correspondent à différents contextes physiques de l'imagerie, et sont donc rencontrés dans de multiples processus d'acquisition, tels que la spectroscopie par résonance magnétique nucléaire (Fourier), la super-résolution SMLM (convolution), MA-TIRF (Laplace), *etc.*

Nous utiliserons dans la suite la notation suivante pour l'opérateur discret : soit  $x = (x_1, \dots, x_N)$  et  $a \in \mathbb{R}^N : \Phi_x(a) \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i \varphi(x_i)$ .

**Résumé** — Une observation, ou acquisition, vivant dans l'espace de Hilbert  $\mathcal{H}$  d'une mesure  $m$  est la quantité  $\Phi m$ .  $\Phi$  est appelé l'opérateur d'acquisition, il est totalement défini par son noyau  $\varphi$  spécifique au contexte physique de l'imagerie.

**Remarque.** L'absence de réflexivité de  $\mathcal{M}(\mathcal{X})$  peut se révéler problématique : si  $\mathcal{M}(\mathcal{X})$  est le dual topologique de  $\mathcal{C}(\mathcal{X})$ , le dual de  $\mathcal{M}(\mathcal{X})$  pour la topologie forte est strictement plus gros que  $\mathcal{C}_0(\mathcal{X})$ . Ceci pourrait s'avérer gênant lorsque l'on est amené à travailler avec la notion de certificat, défini à partir de  $\Phi$  et qui vit dans l'adjoint de  $\mathcal{M}(\mathcal{X})$  (donc potentiellement autre chose qu'une fonction continue évanescence). Pour éviter ce problème, on peut supposer [13, Remarque 2] que l'opérateur d'acquisition s'écrit comme l'adjoint  $K^* : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}$  d'un opérateur  $K : \mathcal{H} \rightarrow \mathcal{C}(\mathcal{X})$ . Ceci étant, comme l'opérateur d'acquisition est faible-\* à faiblement continu, on pourra faire l'économie de cette hypothèse dans la suite du document [43].

### 1.2.3 Une fonctionnelle sans-grille : le BLASSO

Soit  $m_{a_0, x_0} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_{0,i} \delta_{x_{0,i}}$  la mesure source (aussi appelée mesure de la vérité-terrain) d'amplitude  $\mathbf{a}_0 \in \mathbb{R}^N$  et de positions  $\mathbf{x}_0 \in \mathcal{X}^N$ , le *sparse spike problem* consiste à reconstruire cette mesure à partir de l'observation/l'acquisition  $y \stackrel{\text{def.}}{=} \Phi m_{a_0, x_0} + w$  où  $w \in \mathcal{H}$  est un bruit additif, typiquement du bruit blanc gaussien. Pour résoudre ce problème, nous utilisons la fonctionnelle convexe suivante, appelée BLASSO pour Beurling-LASSO :

$$\operatorname{argmin}_{m \in \mathcal{M}(\mathcal{X})} T_\lambda(m) \stackrel{\text{def.}}{=} \frac{1}{2} \|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{X}) \quad (\mathcal{P}_\lambda(y))$$

avec un paramètre de régularisation  $\lambda > 0$  qui traduit le compromis entre la fidélité et l'*a priori* de parcimonie sur la reconstruction. En pratique, il faudra estimer ce paramètre pour trouver un équilibre dans le nombre de sources reconstruites : ceci a fait l'objet de plusieurs travaux pour le LASSO, mais également dans le cadre sans-grille [26].

**Attention** — L'*a priori* de parcimonie fait référence aux travaux sur la minimisation du LASSO : on sait que la solution doit comporter peu de points sur la grille/pixels non-nuls, et c'est ce que la norme  $\ell_1$  pénalise. En réalité, la variation totale impose un *a priori* structurel sur la solution : elle doit être composée d'une somme *finie* de mesures ponctuelles. Le lien entre régulariseur et *a priori* géométrique sur la solution préside à de nombreux travaux récents en sans-grille, et sera rediscuté plus en détail dans le chapitre 3.

Le nom BLASSO a été introduit dans les travaux de [2, 30] en raison du lien entre le problème de *Generalised Minimal Extrapolation* (GME) où l'on cherche à reconstruire une mesure de Radon à partir de plusieurs observations sur ses coefficients de Fourier, et les travaux [5] du mathématicien norvégien Beurling<sup>5</sup> qui coïncide avec le problème de GME dans le cas d'un opérateur de Fourier.

Le BLASSO écrit dans un contexte sans bruit s'écrit :

$$\operatorname{argmin}_{\Phi m = y_0} |m|(\mathcal{X}) \quad \text{avec } y_0 = \Phi m_{a_0, x_0}. \quad (\mathcal{P}_0(y_0))$$

Le BLASSO est lié à son homologue discret, le (LASSO) : on peut formellement voir cette énergie comme la limite fonctionnelle (au sens de la  $\Gamma$ -limite) du LASSO sur une grille de plus en plus fine [42]. Si le problème LASSO assure l'existence et l'unicité de sa solution, que peut-on dire de son équivalent sans-grille ? Tout d'abord, observons que :

—  $m \mapsto |m|(\mathcal{X})$  est semi-continue inférieurement ;

5. Pour être précis, A. Beurling a étudié la fonction de variation totale minimale parmi toutes les fonctions à variations bornées dont la transformée de Fourier est contrainte sur un domaine donné.

—  $\Phi$  est continue pour  $\mathcal{M}(\mathcal{X})$  muni de la topologie faible-\* et  $\mathcal{H}$  muni de sa topologie faible.

Ainsi, on peut établir l'existence de solutions à  $(\mathcal{P}_\lambda(y))$  grâce à des résultats d'analyse convexe, comme prouvé dans [13].

**Résumé** — Le problème *sparse spikes*, ou problème de reconstruction des sources ponctuelles, peut être résolu grâce à la fonctionnelle convexe sur  $\mathcal{M}(\mathcal{X})$  appelée BLASSO, que l'on note  $(\mathcal{P}_\lambda(y))$ . Pour  $m \in \mathcal{M}(\mathcal{X})$  une mesure, cette énergie consiste en un terme d'attache aux données comparant les données observées à  $\Phi m$ , et une régularisation qui consiste en la norme TV de  $m$  pour imposer *a priori* de parcimonie. L'existence de solutions du BLASSO est un fait bien connu et prouvé.

Les difficultés résident alors dans les questions qui suivent :

1. Quelles sont les conditions de reconstruction d'une mesure discrète, pour un régime de bruit donné? La mesure reconstruite est-elle unique?
2. Dans quelles conditions pouvons-nous récupérer le nombre exact de  $\delta$ -pics, leurs amplitudes et leurs positions; est-ce que nous disposons de la stabilité du support?
3. Comment pouvons-nous résoudre numériquement le BLASSO, malgré la dimension infinie et la non-réflexivité de  $\mathcal{M}(\mathcal{X})$ ?

Afin de répondre à ces différentes questions, nous devons introduire quelques notions d'analyse convexe dans la sous-section qui suit.

#### 1.2.4 Problèmes duaux et certificats

Le BLASSO dans l'équation  $(\mathcal{P}_\lambda(y))$  ci-dessus est un problème de minimisation d'une fonctionnelle convexe.

Nous pouvons alors appliquer les résultats d'Ekeland-Temam [45, Remarque 4.2]<sup>6</sup> et définir un problème dual qui s'écrit pour  $p \in \mathcal{H}$  (voir l'Annexe 5.2 pour la preuve) :

$$\operatorname{argmax}_{\|\phi^* p\|_{\infty, \mathcal{X}} \leq 1} \langle y, p \rangle_{\mathcal{H}} - \frac{\lambda}{2} \|p\|_{\mathcal{H}}^2 \quad (\mathcal{D}_\lambda(y))$$

qui peut être réécrit comme la projection sur un espace convexe fermé. [2, 13] :

$$\operatorname{argmax}_{\|\phi^* p\|_{\infty, \mathcal{X}} \leq 1} \left\| \frac{y}{\lambda} - p \right\|_{\mathcal{H}}^2 \quad (\mathcal{D}'_\lambda(y))$$

6. Une petite mise en garde s'impose quant à ces résultats : l'espace  $V$  doit être un espace réflexif, ce qui n'est manifestement pas le cas ici avec  $V = \mathcal{M}(\mathcal{X})$ . Cependant, comme indiqué aussi dans le chapitre 3, l'hypothèse de réflexivité n'est nécessaire que pour la preuve d'existence.

La dualité de Fenchel entre  $(\mathcal{P}_\lambda(y))$  et  $(\mathcal{D}_\lambda(y))$  est prouvée dans [13]. Ainsi, toute solution  $m_\lambda$  de  $(\mathcal{P}_\lambda(y))$  est liée à l'unique solution  $p_\lambda$  de  $(\mathcal{D}_\lambda(y))$  par les conditions d'extrémalité [43] :

$$\begin{cases} \Phi^* p_\lambda \in \partial|m_\lambda|(\mathcal{X}), \\ -p_\lambda = \frac{1}{\lambda}(\Phi m_\lambda - y) \end{cases} \quad (1.2)$$

où  $\partial|\cdot|(\mathcal{X})$  est le sous-différentiel de la norme TV. En effet, puisque la norme TV n'est pas différentiable (tout comme la norme  $\ell^1$ ) mais semi-continue inférieurement pour la topologie faible-\* nous utilisons son sous-différentiel qui s'écrit, pour  $m \in \mathcal{M}(\mathcal{X})$  :

$$\partial|m|(\mathcal{X}) = \left\{ \eta \in \mathcal{C}_0(\mathcal{X}) ; \|\eta\|_{\infty, \mathcal{X}} \leq 1 \text{ et } \int_{\mathcal{X}} \eta \, dm = |m|(\mathcal{X}) \right\}. \quad (1.3)$$

Les éléments de ce sous-différentiel sont appelés des *certificats*. Grâce à la dualité forte, nous pouvons définir des certificats particuliers appelés *certificats duaux*. [16].

**Définition 1.4.** Nous appelons  $\eta_\lambda \stackrel{\text{def.}}{=} \Phi^* p_\lambda$  un certificat dual de  $m_\lambda$ , pour  $p_\lambda$  qui satisfait aux conditions d'optimalité de (1.2).

C'est un certificat, puisque  $\Phi^* p_\lambda \in \partial|m_\lambda|(\mathcal{X})$  et il est dit *dual* puisqu'il vérifie la deuxième condition d'optimalité de (1.2) : il est donc défini par la solution duale  $p_\lambda$ . Dit autrement, un certificat dual  $\eta_\lambda$  est associé à la mesure  $m_\lambda$  et il *certifie* que la mesure  $m_\lambda$  est bel et bien un *minimum* du BLASSO. Ainsi, s'il existe des solutions de  $(\mathcal{P}_\lambda(y))$  de la forme  $m_\lambda \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i \delta_{x_i}$ , le support vérifie [43] pour  $0 \leq i \leq N$  :  $|\eta_\lambda|(x_i) = 1$ .

On peut établir, de la même manière, un tel lien entre la solution  $m_0$  du BLASSO sans-bruit  $(\mathcal{P}_0(y_0))$  et ses certificats  $\eta_0$ , qui ne sont du reste pas uniques dans le cas général. Nous appellerons alors dans le reste du chapitre  $\eta_0$  le certificat de norme dual minimale, c'est-à-dire le certificat dual  $\eta_0$  avec la norme *supremum* la plus petite  $\|\eta_0\|_{\infty, \mathcal{X}}$  parmi tous les certificats duaux. Il est montré dans [43] que ce certificat de norme minimale  $\eta_0$  possède des propriétés importantes, puisqu'il détermine en quelque sorte la stabilité des positions des pics reconstruits lorsque le bruit additif est faible, en particulier leur proximité avec les positions de la mesure source  $m_{a_0, x_0}$  : voir la définition 1.6 dans la section ci-dessous.

**Résumé** — Nous avons défini le problème primal du BLASSO dans la section précédente, nous pouvons définir son problème dual grâce à la convexité. Une solution  $m_\lambda$  du BLASSO et une solution  $p_\lambda$  du problème dual sont liées par une condition d'extrémalité. La solution duale  $p_\lambda$  définit le certificat dual, un élément du sous-gradient spécifié par  $\eta_\lambda = \Phi^* p_\lambda$  : le certificat dual  $\eta_\lambda$  *certifie* que  $m_\lambda$  est une solution de la BLASSO. Nous pouvons alors établir des conditions plus précises sur l'unicité ou la reconstruction du support de la source.

### 1.2.5 Garanties de reconstruction de la mesure source

Nous discuterons dans cette section des deux premières questions que nous avons formulées un peu plus haut ; à savoir l'existence, l'unicité et les conditions de reconstruction du support. Un outil classique pour établir certaines propriétés de reconstruction se trouve dans la notion de distance de séparation *minimum*.

**Définition 1.5** (Distance de séparation *minimum*). *La distance de séparation minimale est une caractérisation du support de la mesure discrète  $m_{a_0, x_0}$  selon :*

$$\Delta(m_{a_0, x_0}) \stackrel{\text{def.}}{=} \min_{i \neq j} |x_{0,i} - x_{0,j}|.$$

La condition de reconstruction est déterminée par cette distance de séparation minimale, elle-même déterminée par le type de mesures de Radon (complexe, réelle, réelle et positive) et le type d'opérateur d'acquisition. En particulier :

- si l'opérateur est une acquisition du spectre de Fourier entre  $[-f_c, f_c]$  pour la fréquence de coupure  $f_c$  avec  $\mathcal{X} = \mathbb{T}^d$  le tore en dimension  $d$  dans un cadre sans bruit, il est nécessaire que  $\Delta(m_{a_0, x_0}) \gtrsim \frac{2}{f_c}$  si la mesure source est complexe [16]. Avec quelques hypothèses en plus [47] nous pouvons préciser le résultat en  $\Delta(m_{a_0, x_0}) \gtrsim \frac{1.26}{f_c}$ , et  $\Delta(m_{a_0, x_0}) \gtrsim \frac{1.87}{f_c}$  si la mesure source est réelle [16] ;
- quel que soit l'opérateur  $\Phi$  [30, 34] dans un cadre sans-bruits, il n'y a pas de condition sur la distance de séparation pour reconstruire la mesure source si celle-ci est **positive**, néanmoins la constante de stabilité explose quand  $\Delta(m_{a_0, x_0}) \rightarrow 0$ .

Ces résultats sont importants, mais ne disent rien de la reconstruction en présence de bruit : nous nous attendons pourtant à trouver du bruit dans les images que nous traitons et *a priori* à être limités par ce régime de bruit. Pour prendre en charge cette nouvelle difficulté, nous devons ajouter certaines conditions à la mesure vérité terrain, d'après les travaux de [43] que nous introduisons :

**Définition 1.6** (Condition de source non-dégénérée). *La source  $m_{a_0, x_0}$  vérifie la NDSC (Non-Degenerate Source Condition) si :*

- il existe  $\eta \in \text{Im } \Phi^*$  tel que  $\eta \in \partial |m_{a_0, x_0}|(\mathcal{X})$  ;
- $\forall s \in \mathcal{X} \setminus \cup_{i=1}^N \{x_{0,i}\}, |\eta_0(s)| < 1$  ;
- $\forall i \in \llbracket 1, N \rrbracket$ , la matrice Hessienne  $\nabla^2 \eta_0(x_{0,i}) \in \mathbb{R}^{d \times d}$  est inversible.

La première condition revient à supposer que  $m_{a_0, x_0}$  est une solution du  $(\mathcal{P}_0(y_0))$  et qu'il existe une solution au problème dual. On dit que  $\eta_0$  est *non dégénéré* si ces trois conditions sont vérifiées. Nous pouvons alors écrire l'une des garanties théoriques les plus importantes, à savoir :

**Théorème 1.1** (Robustesse au bruit [43]). *Soit  $\Gamma_{x_0}$  la matrice  $N \times N$  définie par  $\Gamma_{x_0} \stackrel{\text{def.}}{=} (\varphi(\cdot - x_{0,i}), \varphi'(\cdot - x_{0,i}))_{i=1}^N$ , avec  $\varphi'$  la dérivée du noyau évoqué en 1.2.2. Supposons que les*

vecteurs colonnes de  $\Gamma_{x_0}$  forment une famille libre, et que  $m_{a_0, x_0}$  vérifie la condition NDSC évoquée en 1.6. Alors, il existe  $\alpha > 0, \lambda_0 > 0$  tel que pour tout  $0 \leq \lambda \leq \lambda_0$  et  $w$  tel que  $\|w\| \leq \alpha\lambda$ ; il existe  $N$  couples  $(a_{\lambda, i}, x_{\lambda, i})$  tel que  $m_\lambda \stackrel{\text{def.}}{=} \sum_{i=1}^N a_{\lambda, i} \delta_{x_{\lambda, i}}$  soit l'unique solution du  $(\mathcal{P}_\lambda(y))$  composée d'exactement  $N$  pics. En particulier pour  $\lambda = 1/\alpha \|w\|_{\mathcal{H}}$  nous disposons d'un contrôle sur la différence en termes de positions entre la vérité-terrain et la reconstruction :

$$\forall i \in \llbracket 1, N \rrbracket : \|x_{\lambda, i} - x_{0, i}\| = \mathcal{O}(\|w\|_{\mathcal{H}}) \text{ et } |a_{\lambda, i} - a_{0, i}| = \mathcal{O}(\|w\|_{\mathcal{H}}).$$

Sous la condition de source non-dégénérée et pour  $\lambda$  et  $\|w\|_{\mathcal{H}}^2/\lambda$  suffisamment petit, on peut reconstruire une mesure de Radon avec exactement le même nombre de  $\delta$ -pics que  $m_{a_0, x_0}$ . En outre, la mesure reconstruite converge (faiblement-\*) vers la mesure vérité-terrain quand le niveau de bruit tend vers 0. Les auteurs de [43] introduisent également la notion de précertificats dérivés évanescents. Le certificat  $\eta_0$  est en effet difficile à calculer à partir du problème dual de  $(\mathcal{P}_0(y_0))$ , notamment à cause de la contrainte  $\|\eta_0\|_{\infty, \mathcal{X}} \leq 1$ , le précertificat permet de simplifier ce calcul, en résolvant en lieu et place un simple système linéaire. Le lecteur intéressé pourra jeter un œil à cet article parmi tant d'autres [43, 34] pour ces concepts novateurs.

**Résumé** — Le critère de la distance de séparation minimale est utilisé pour évaluer les possibilités de reconstruction, d'abord dans un environnement sans bruit. Sous un régime de faible bruit, un théorème précise que la mesure source  $m_{a, x}$  composée de  $N$   $\delta$ -pics peut être récupérée par BLASSO, avec un contrôle sur les écarts (en termes d'amplitudes et de positions) entre la mesure reconstruite et la mesure source.

Nous avons jusqu'ici donné les garanties établies dans le domaine sans-grille sur la reconstruction d'une mesure source, en particulier en présence de bruits. Dans la section suivante, nous nous proposons de répondre à la troisième question et de discuter la manière de calculer la solution numérique du problème inverse; une tâche relativement difficile qui nécessite de prendre en compte les difficultés de l'espace d'optimisation.

### 1.3 Stratégies numériques pour résoudre le BLASSO

Le problème  $\mathcal{P}_\lambda(y)$  est une optimisation sur l'ensemble des mesures de Radon, un espace de dimension infinie et non réflexif. Nous rappelons qu'il s'écrit :

$$\operatorname{argmin}_{m \in \mathcal{M}(\mathcal{X})} T_\lambda(m) \stackrel{\text{def.}}{=} \frac{1}{2} \|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{X}). \quad (\mathcal{P}_\lambda(y))$$

Une approche naïve consisterait à contraindre  $m$  sur la grille fine  $(p_i)_i^L$ , ce qui reviendrait à résoudre le problème LASSO :

$$\min_{a \in \mathbb{R}^L} \|y - \Phi_L a\|_{\mathcal{H}}^2 + \|a\|_1$$

avec l'opérateur discret  $\Phi_L a \stackrel{\text{def.}}{=} \sum_{i=1}^L a_i \varphi(p_i)$  et  $\varphi$  le noyau de l'opérateur. Cette approche comporte de nombreux inconvénients : par exemple, la solution du LASSO en faible régime de bruit et lorsque la taille du pas de discrétisation tend vers 0, contient des paires de  $\delta$ -pics autour du vrai  $\delta$  de Dirac [41, 42]. De plus, le raffinement du pas conduit à une dégradation du conditionnement de l'opérateur d'acquisition, ce qui entraîne des difficultés numériques. Il faudrait donc plutôt travailler sur  $\mathcal{M}(\mathcal{X})$ , mais les difficultés sont de plusieurs ordres :

- c'est un espace de dimension infinie. Ceci étant, nous verrons dans un prochain chapitre que la solution est nécessairement une somme *finie* de mesures, ce qui réduit fortement l'espace dans lequel on vient optimiser le BLASSO ;
- c'est un espace de Banach, mais pas de Hilbert. Il n'y a donc pas de structure de produit scalaire<sup>7</sup> donc pas d'algorithme proximal explicite ;
- c'est un espace non réflexif, tout du moins pour la topologie forte. Les résultats de convergence des algorithmes ne pourront être écrits qu'en topologie faible\*.

Nous présentons brièvement dans la suite quelques méthodes numériques pour résoudre le BLASSO [2, 35, 21]. Avant de décrire ces méthodes, remarquons que des travaux récents ont pu dégager des axes de travail prometteurs, comme la descente de gradient projeté [89, 90]. Elle s'appuie sur une initialisation sur-paramétrée : une mesure discrète avec de nombreux  $\delta$ -pics par rapport à la vérité terrain est initialisée, puis on applique une descente de gradient sur les amplitudes et les positions de la mesure sur paramétrée, combinée à chaque étape avec une projection sur un ensemble de contraintes de positions pour forcer la séparation des pics. Cette projection peut être remplacée par une « heuristique » qui se résume à la fusion des  $\delta$ -pics qui ne sont pas suffisamment séparés [90]. L'approche de descente est à rapprocher des gradients conditionnels *polyatomiques* [59], utilisant une estimation polyatomique (ou sur paramétrée) des positions ainsi qu'une étape corrective pour obtenir une convergence plus rapide – d'un facteur estimé en pratique de 1,5 à 20 – vers le *minimum*.

### 1.3.1 Programmation semi-définie et flot de gradient

#### Méthode du point intérieur, hiérarchie

La programmation semi-définie a constitué l'un des premiers schémas de résolution du BLASSO, dans le cas spécifique d'une acquisition de Fourier sur le tore 1D  $\mathbb{T}^1$  [16, 30, 2, 17].

Soit  $d = 1$  la dimension de l'intérieur de  $\mathcal{X}$ , l'opérateur d'acquisition noté  $\mathcal{F}_n$ <sup>8</sup> est une observation des coefficients de Fourier jusqu'à  $f_c \in \mathbb{N}$ , avec  $n = 2f_c + 1$  le

7. que serait le produit de mesures ? voir plus haut.

8. et pas  $\Phi$  pour éviter une ambiguïté avec l'opérateur d'acquisition général

nombre de mesures **physiques**. Ainsi  $\mathcal{F}_n : \mathcal{M}_C(\mathcal{X}) \rightarrow \mathbb{C}^n$  et pour une mesure discrète  $m_{a,x} \stackrel{\text{def.}}{=} \sum_{j=1}^N a_j \delta_{x_j}$  il vient  $\mathcal{F}_n(m_{a,x}) = \left( \sum_j a_j e^{2i\pi k x_j} \right)_{|k| \leq f_c}$ .

Cette méthode utilise la programmation semi-définie (SDP) pour calculer efficacement les *minima* du BLASSO. Elle participe de l'approche de Hilbert [57], qui consiste à décomposer globalement la fonction objectif en fonctions simples. La solution du problème dual de  $(\mathcal{P}_\lambda(y))$  est un polynôme  $p$  lié à un certificat par  $\mathcal{F}_n^* p$  : l'idée est alors de reconstruire le certificat dual comme une combinaison linéaire de polynômes trigonométriques [2], ce qui suffit à trouver la mesure associée à ce certificat reconstruit. Cette mesure associée est une solution du BLASSO : le problème dual peut quant à lui être résolu grâce par une approche de programmation semi-définie. Cette dernière peut typiquement être réalisée avec un algorithme du point intérieur (*Interior Point Method*) comme proposé en [16].

Notons que cette stratégie ne peut être adoptée que pour  $d = 1$ . Le cas en dimension supérieure nécessite de faire appel à la hiérarchie de Lasserre [64]. Pour le dire simplement, la hiérarchie de Lasserre produit une suite de problèmes SDP imbriqués dans le cône des moments de la mesure positive.

Ces méthodes sont prouvées asymptotiquement exactes et ont été appliqués avec succès au problème de super-résolution [17]. On ne sait néanmoins pas si l'algorithme a une convergence finie en général, on ne sait donc *a fortiori* pas quand arrêter la hiérarchie<sup>9</sup> pour obtenir une solution du BLASSO [20]. De plus, ces approches par SDP sont applicables uniquement à une certaine classe d'opérateur d'acquisition  $\Phi$ , typiquement l'opérateur de Fourier, ou des filtres à supports compacts dans le domaine de Fourier, ce qui exclue d'emblée la convolution gaussienne (pourtant essentielle en super-résolution et en imagerie biomédicale). La classe d'algorithme qui suit cherche à mieux tirer parti du cadre continu tout en évacuant les limitations sur l'opérateur d'acquisition.

**Résumé** — Cette approche se résume à la résolution du problème dual, à la reconstruction du support de la mesure grâce au certificat associé à la solution duale, et enfin à la résolution d'un problème linéaire pour obtenir les amplitudes estimées correspondantes. Cette stratégie peut être étendue à un contexte multivarié : elle reste en revanche assez restrictive sur le type d'opérateur direct admissible et ne propose pas de convergence finie dans le cas général.

## Flot de gradient

Cette approche nécessite un peu plus de mise en contexte pour bien la comprendre. Nous reportons donc en annexe 5.2 les explications détaillées sur son fonctionnement. Nous pouvons ceci dit la résumer simplement : la reconstruction est initialisée par une *sur-paramétrisation*, c'est-à-dire que l'on place sur le domaine  $\mathcal{X}$  un très grand nombre de

9. on peut l'expliquer par le fait que les polynômes trigonométriques positifs en dimension  $d > 1$  ne sont pas nécessairement des fonctions *sums of square*

mesures de Dirac. On procède ensuite à un flot de gradient, dont le flot limite est, sous certaines hypothèses, une solution du BLASSO. Ce type d’algorithmes est applicable à tous les opérateurs rencontrés dans les problèmes inverses, avec un temps de calcul rapide et seulement trois paramètres à régler

**Résumé** — Le flot de gradient  $(v_t)_t$  est calculé par l’algorithme du *Conic Particle Gradient Descent*, qui consiste en une estimation par descente de gradient (par rapport à la métrique du cône) sur les amplitudes et les positions d’une mesure sur-paramétrée, c’est-à-dire une mesure avec un nombre de  $\delta$ -pics bien plus grand que celui de la mesure source. Les itérations sont moins coûteuses que les algorithmes gloutons (présentés plus bas), mais le CPGD manque de garanties de reconstruction pour des régimes de bruits faibles.

### 1.3.2 Algorithme glouton : la méthode du gradient conditionnel

L’algorithme du gradient conditionnel, aussi appelé l’algorithme de Frank-Wolfe (FW) [70, 49] cherche à résoudre le problème  $\min_{m \in C} f(m)$  pour  $C$  n’importe quel convexe faiblement compact d’un espace vectoriel topologique, et  $f$  une fonction convexe et différentiable ; la différentielle est ainsi notée  $df$ . Cet algorithme consiste en une minimisation itérative d’une version linéarisée de  $f$  : il ne nécessite donc que les dérivées directionnelles de  $f$  et se passe de toute structure hilbertienne, contrairement à un algorithme proximal classique, formulé en termes de distance euclidienne. Une définition du gradient conditionnel est donnée dans le pseudo-code 1, dans le cadre très général de la minimisation d’une fonction  $f$  quelconque.

---

#### Algorithme 1 : Frank-Wolfe.

---

```

1 pour  $0 \leq k \leq K$  faire
2    $s^k \in \operatorname{argmin}_{s \in C} f(m^k) + df(m^k)(s - m^k)$ .
3   si  $df(m^k)(s^k - m^k) = 0$  alors
4     |  $m^k$  est une solution. Stop.
5   sinon
6     | Étape de recherche :  $\gamma^k \leftarrow$  soit  $\frac{2}{k+2}$  soit
7       |  $\operatorname{argmin}_{\gamma \in [0,1]} f(m^k + \gamma(s^k - m^k))$ .
8       | Mise à jour :  $m^{k+1} \leftarrow m^k + \gamma^k(s^k - m^k)$ .
9   fin
10 fin

```

---

Ceci appelle déjà à quelques remarques :

- l’hypothèse de compacité sur  $C$  assure que l’argmin de l’étape 2 est bien non-vide ;
- en ligne 7, il est possible de remplacer  $m^{k+1}$  par n’importe quel élément  $\hat{m} \in C$  tel que  $f(\hat{m}) \leq f(m^{k+1})$  sans affecter les propriétés de convergence de l’algorithme ;

Cette dernière propriété est intéressante en pratique puisqu'elle permet de recourir à des heuristiques (telles que l'approche polyatomique évoquée plus haut) sans pour autant invalider les résultats théoriques prouvés dans la littérature (tels que la convergence énoncée plus bas). Deux problèmes nous empêchent cependant d'appliquer directement cet algorithme au BLASSO :  $T_\lambda$  n'est pas différentiable, et l'ensemble d'optimisation  $\mathcal{M}(\mathcal{X})$  n'est pas borné. Il est donc nécessaire d'effectuer un *epigraphical lift*<sup>10</sup> [53, 35] pour obtenir une fonctionnelle différentiable qui partage les mêmes mesures *minimum* que  $T_\lambda$  :

$$\min_{(r,m) \in C} \tilde{T}_\lambda(m,r) \stackrel{\text{def.}}{=} \frac{1}{2} \|y - \Phi(m)\|_{\mathcal{H}}^2 + \lambda r \quad (\tilde{\mathcal{P}}_\lambda(y))$$

avec le borné  $C = \{(r,m) \in \mathbb{R}^+ \times \mathcal{M}(\mathcal{X}); |m|(\mathcal{X}) \leq r \leq M\}$  et  $M \stackrel{\text{def.}}{=} \frac{\|y\|_{\mathcal{H}}^2}{2\lambda}$ . Bien que  $C$  ne soit pas faiblement compact, il est compact pour la topologie faible-\* et les hypothèses de l'algorithme 1 sont toujours vérifiées. L'algorithme du Frank-Wolfe est donc bien définie pour l'énergie  $\tilde{T}_\lambda$ , différentiable au sens de Fréchet sur le Banach  $\mathbb{R} \times \mathcal{M}(\mathcal{X})$ . Cette différentielle s'écrit d'ailleurs :

$$d\tilde{T}_\lambda : (r', m') \mapsto \int_{\mathcal{X}} \Phi^*(\Phi m - y) dm' + \lambda r'.$$

Finalement, il vient que  $m^*$  est un *minimum* de  $T_\lambda$  si et seulement si  $(|m^*|(\mathcal{X}), m^*)$  minimise  $(\tilde{\mathcal{P}}_\lambda(y))$ , et  $T_\lambda(m^*) = \tilde{T}_\lambda(|m^*|(\mathcal{X}), m^*)$ . Dans le reste du document, pour des raisons de lisibilité et de clarté, nous ne mentionnerons pas la composante en  $r$ , et nous nous référerons à la quantité  $(|m^*|(\mathcal{X}), m^*)$  uniquement par  $m^*$ .

**Remarque.** *Le lecteur curieux pourrait se demander d'où provient ce  $M$  qui permet d'avoir un ensemble d'optimisation borné et par conséquent d'appliquer l'algorithme du gradient conditionnel sur  $C$ . Soit  $m$  une mesure minimum de  $\tilde{T}_\lambda$ , alors :*

$$\lambda |m|(\mathcal{X}) \leq \tilde{T}_\lambda(m) \leq \tilde{T}_\lambda(0) = \frac{1}{2} \|y\|_{\mathcal{H}}^2 \stackrel{\text{def.}}{=} \lambda M.$$

Nous avons noté précédemment que la mise à jour de l'itérée  $m^{k+1}$  dans la ligne 7 peut être remplacée par n'importe quel point  $\hat{m}$  améliorant la fonction objective, et comme nous l'avons observé cette remarque est assez intéressante, car elle peut améliorer considérablement la propriété de convergence de l'algorithme [13, 7]. Une amélioration intéressante de l'algorithme de Frank-Wolfe repose ainsi sur la modification de l'étape d'optimisation convexe finale par une optimisation non convexe à la fois sur les amplitudes et les positions des  $\delta$ -pics reconstruits, de manière simultanée. Cette modification est présentée dans l'algorithme 2.

10. Un *lift* peut grossièrement se définir comme l'ajout d'une dimension au problème pour en dégager une propriété supplémentaire.

**Algorithme 2 : Sliding Frank-Wolfe.**


---

**Entrées :** Acquisition  $y \in \mathcal{H}$ , nombre d'itérations  $K$ ,  $\lambda > 0$ .

1 Initialisation :  $m^{[0]} = 0$ ,  $N^{[k]} = 0$ .

2 **pour**  $k$ ,  $0 \leq k \leq K$  **faire**

3     Pour  $m^{[k]} = \sum_{i=1}^{N^{[k]}} a_i^{[k]} \delta_{x_i^{[k]}}$  tel que  $a_i^{[k]} \in \mathbb{R}$ ,  $x_i^{[k]} \in \mathcal{X}$ , trouver  $x_*^{[k]} \in \mathcal{X}$  tel que :

$$x_*^{[k]} \in \operatorname{argmax}_{x \in \mathcal{X}} \left| \eta^{[k]}(x) \right| \quad \text{où} \quad \eta^{[k]}(x) \stackrel{\text{def.}}{=} \frac{1}{\lambda} \Phi^* (\Phi m^{[k]} - y),$$

4     **si**  $\left| \eta^{[k]}(x_*^{[k]}) \right| \leq 1$  **alors**

$m^{[k]}$  est la solution du BLASSO. Stop.

5     **sinon**

6     Calculer  $m^{[k+1/2]} = \sum_{i=1}^{N^{[k]}} a_i^{[k+1/2]} \delta_{x_i^{[k+1/2]}} + a_{N^{[k]}+1}^{[k+1/2]} \delta_{x_*^{[k]}}^{[k+1/2]}$  tel que :

$$a_i^{[k+1/2]} \in \operatorname{argmin}_{a \in \mathbb{R}^{N^{[k]}+1}} \frac{1}{2} \|y - \Phi_{x^{[k+1/2]}}(a)\|_{\mathcal{H}}^2 + \lambda \|a\|_1$$

      pour  $x^{[k+1/2]} \stackrel{\text{def.}}{=} (x_1^{[k]}, \dots, x_{N^{[k]}}^{[k]}, x_*^{[k]})$ .

7     Calculer  $m^{[k+1]} = \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \delta_{x_i^{[k+1]}}$  tel que :

$$(a_i^{[k+1]}, x_i^{[k+1]}) \in \operatorname{argmin}_{(a,x) \in \mathbb{R}^{N^{[k]}+1} \times \mathcal{X}^{N^{[k]}+1}} \frac{1}{2} \|y - \Phi_{x^{[k+1/2]}}(a)\|_{\mathcal{H}}^2 + \lambda \|a\|_1.$$

8     **fin**

9 **fin**

---

**Résultat :** Mesure discrète  $m^{[k]}$  où  $k$  est l'itération d'arrêt.

---

Elle permet d'obtenir une convergence théorique vers la solution unique de BLASSO en un nombre fini d'itérations, qui est même empiriquement une convergence en  $N$  étapes. Cette version est appelée algorithme du *Sliding Frank-Wolfe* [35], car les positions des pics sont glissantes sur le domaine continu  $\mathcal{X}$ ; là où un Frank-Wolfe classique se réduirait à une étape convexe d'optimisation des positions à amplitudes fixées. Les auteurs ont également prouvé dans le même article que la suite de mesures produite  $m^{[k]}$  converge vers le *minimum* pour la topologie faible-\*

Le Frank-Wolfe est par nature un algorithme itératif : en pratique, il est donc possible d'interrompre le calcul, ou tout du moins de choisir de manière appropriée le  $\lambda$  de sorte à ne pas reconstruire de *spurious spikes*, des mesures de Dirac fallacieuses ou dit autrement des mesures qui correspondent en réalité à du bruit.

**Attention** — Nous verrons plus en détail dans le chapitre 3 que l’algorithme du gradient conditionnel n’est pas seulement limité à la reconstruction de sources ponctuelles : il est même tout à fait indiqué pour des sources qui ne sont **pas ponctuelles** !

Comme dans la méthode SDP, le choix du régularisateur  $\lambda$  est crucial : s’il est choisi trop grand, il produira une solution avec moins de  $\delta$ -pics que nécessaire, s’il est choisi trop bas, il reconstruira une solution avec des pics parasites.

La ligne 3 de l’algorithme 2 est généralement résolue par une recherche sur la grille grossière (*grid search*), l’étape convexe de la ligne 5 peut utiliser un solveur convexe comme FISTA [4] et l’étape non convexe dans la ligne 6 peut être traitée par une modification de la méthode Broyden-Fletcher-Goldfarb-Shann (BFGS, aussi en version à mémoire limitée L-BFGS et bornée BFGS-B) [14]. Les reconstructions dans le cadre 2D avec un noyau de convolution gaussien, similaires aux conditions SMLM discutées plus loin, sont présentées dans la figure 1.2. La luminosité étant toujours une quantité positive, on peut contraindre [32] le SFW pour construire une mesure positive du cône  $\mathcal{M}^+(\mathcal{X})$ , en modifiant :

- la condition d’arrêt  $\eta^{[k]}(x_*^{[k]}) \leq 1$  ;
- l’étape du LASSO est résolue sur  $a \in \mathbb{R}_+^{N^{[k]+1}}$  ;
- l’étape non-convexe travaille sur  $\mathbb{R}_+^{N^{[k]+1}} \times \mathcal{X}^{N^{[k]+1}}$ .

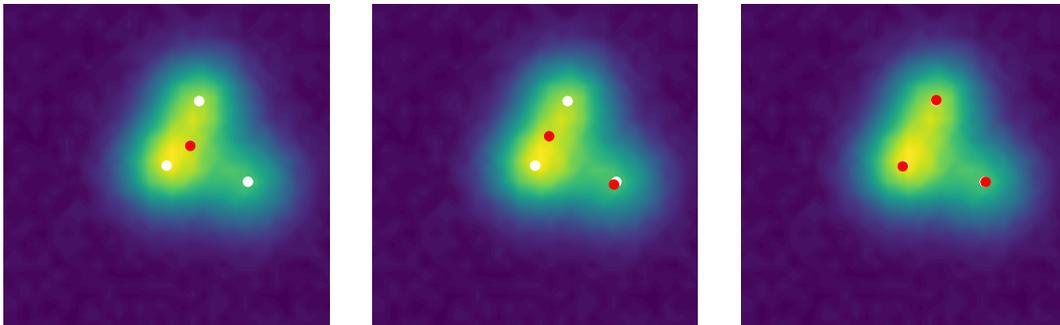


FIGURE 1.2 – (a) Première itération  $k = 0$  (b) Première itération du *Sliding Frank-Wolfe*  $k = 1$  (c) Fin du calcul pour  $k = 2$ , résultats d’une reconstruction par le *Sliding Frank-Wolfe* sur le domaine  $\mathcal{X} = [0, 1]^2$  pour le noyau gaussien d’écart-type  $\sigma = 0, 1$  et un bruit gaussien additif de variance 0, 1. Tous les  $\delta$ -pics sont correctement reconstruits, et ce, uniquement à partir de l’acquisition, en régularisant avec  $\lambda = 3 \times 10^{-2}$ .

Cet algorithme modifié offre ainsi un bon compromis entre la précision et les garanties théoriques. Cependant, il souffre d’une charge de calcul élevée pour chaque itération, ce qui le rend lent à utiliser. Nous proposons dans la suite une amélioration pratique pour fluidifier le calcul du *Sliding Frank-Wolfe* (sans toutefois changer sa complexité algorithmique) et en particulier de l’étape non-convexe, à travers la notion d’auto-différentiation. On rappelle les définitions suivantes :

### Autodérivation

Commençons déjà par discuter de deux approches répandues pour calculer le gradient d'une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  numériquement :

- la dérivation par différences finies, avec une complexité en  $\mathcal{O}(n)$ . Elle est entachée d'erreurs numériques, plus précisément à cause des erreurs d'arrondi (erreurs dues à l'utilisation de représentations en virgule flottante à bits finis) et des erreurs de troncature (différence entre les gradients analytiques et les gradients numériques). Pour des petits pas de dérivation, les erreurs d'arrondi dominent, à l'inverse pour de grands pas ce sont les erreurs de troncature qui vont poser quelques difficultés. Ces erreurs peuvent être conséquentes dans les problèmes mal posés et sont source d'instabilités numériques ;
- la dérivation symbolique, où l'on calcule explicitement l'expression du gradient. Elle devient vite rédhibitoire pour des fonctions aux définitions « exotiques ».

L'autodérivation, ou dérivation automatique, se propose de calculer le gradient de manière efficace, à savoir en un nombre d'opérations égal (à une multiplication près) à l'évaluation de la fonction. Elle a pris son essor dans les dernières années [54], notamment avec l'introduction du calcul sur carte graphique (GPU) et d'interface (*wrapper*) tel que CUDA qui permettent de délivrer le plein potentiel de cette méthode. Nous donnons en annexe 4 quelques précisions sur l'auto-dérivation [74, 48].

La dérivation automatique pour trouver le gradient de  $f$  (*backward*) ne demande en pratique [48] que 4 à 5 fois plus d'opérations que le calcul de  $f$  (*forward*). L'utilisation de cette approche mise en œuvre avec PyTorch [74] permet un calcul efficace du gradient de la fonction non-convexe du *Sliding Frank-Wolfe* en un temps convenable pour des utilisations pratiques. Si la littérature mentionne l'algorithme de L-BFGS(-B) pour le calcul de cette étape 7, nous lui préférons l'algorithme ADAM [63] qui permet un calcul plus rapide empiriquement d'un facteur 10, et évite en comparaison certains *minimum* locaux.

Pour conclure cette sous-section, nous rappelons dans la table 1.1 les avantages et les inconvénients de chacun des algorithmes. Compte tenu du manque de garanties théoriques de l'optimum évalué par le CPGD, et de la limitation des méthodes SDP aux opérateurs type Fourier, nous utiliserons dans la section qui suit l'algorithme du gradient conditionnel et en particulier le *Sliding Frank-Wolfe* pour résoudre un problème de super-résolution en SMLM.

**Résumé** — La méthode du gradient conditionnel est un algorithme glouton qui consiste en la minimisation itérative d'une version linéarisée de la fonction convexe à optimiser. Cet algorithme peut être appliqué à n'importe quel opérateur direct sans restriction sur l'espace  $\mathcal{X}$ . À une modification près (SFW), l'algorithme de Frank-Wolfe propose une convergence finie, empiriquement une convergence en  $N$  étapes pour une mesure de source composée de  $N$   $\delta$ -pics. Les itérations sont cependant coûteuses en temps de calcul.

TABLE 1.1 – Avantages et inconvénients des différentes stratégies d’algorithmes sans-grille, programmation semi-définie (SDP) contre algorithme du *Sliding Frank-Wolfe* (SFW) contre *Conic Particle Gradient Descent* (CPGD).

Algorithme	Opérateur	Espace $\mathcal{X}$	Taux de convergence	Temps de calcul	Nombre de paramètres
SDP [16]	Fourier	Tore $\mathbb{T}^d$	Asymptotique	Moyen	$\lambda$
SFW [13, 35]	Tous	Compact	sous-linéaire	Long	$\lambda$
CPGD [21]	Tous	Tore $\mathbb{T}^d$	$\log(\varepsilon)$	Rapide	$\lambda, \alpha, \beta$

## 1.4 Applications et résultats en imagerie SMLM

Afin d’illustrer l’intérêt des méthodes sans-grille en pratique, nous nous proposons dans la suite de résoudre un problème de super-résolution, afin de reconstruire des structures biologiques à de très petites échelles.

### 1.4.1 Métrique quantifiant la qualité de la reconstruction

Si l’on a accès à la vérité-terrain, c’est-à-dire les positions réelles des sources ponctuelles, on peut évaluer la qualité de la reconstruction avec :

- des métriques de détection, telles que l’indice de Jaccard ;
- des métriques de qualité de reconstruction, telles que la norme  $L^2$  dans le cas discret.

Les métriques de détection peuvent s’appliquer directement à la reconstruction sans-grille. Nous discuterons plutôt dans cette partie de la deuxième classe de métrique, celles quantifiant donc la « qualité de reconstruction ». Les algorithmes que nous avons évoqués plus haut produisent une liste de mesures de Dirac, que l’on peut comparer avec la mesure source  $m_{a_0, x_0}$ , aussi appelée mesure vérité-terrain. Cette comparaison ne peut manifestement pas être menée avec des outils discrets, telle que la norme  $\ell_2$ , puisque l’on ne peut pas directement comparer un élément de  $\mathcal{M}(\mathcal{X})$  avec celui de  $L^2(\mathcal{X})$ . On pourrait alors penser à comparer  $\Phi m_{a_0, x_0}$  et la mesure reconstruite passée dans l’opérateur  $\Phi$ , mais il se poserait toujours la question des erreurs induites par la discrétisation de  $\Phi$ .

L’examen de la norme 2 du vecteur  $x_i$  des positions reconstruites par rapport au vecteur  $x_{0,i}$  n’est du reste pas plus instructive : nous avons besoin du même nombre d’éléments pour  $x$  et  $x_0$ , nous devons trier le vecteur des positions et nous n’avons aucune garantie que la correspondance d’une position de  $x$  avec une autre de  $x_0$  soit la bonne.

Il faudrait donc plutôt considérer une distance sur l’espace de mesure. Nous utiliserons dans la suite la 1-distance de Wasserstein  $\mathcal{W}_1$  [75] : nous invitons le lecteur à se pencher sur l’annexe 4 pour quelques rappels sur des définitions utiles et plus de détails

sur le cadre de transport optimal utilisé dans cette section. La distance de Wasserstein avec des mesures de même masse<sup>11</sup> est définie<sup>12</sup> comme :

**Définition 1.7** (Transport optimal à masse égale). *Pour  $0 \leq p < +\infty$  et  $m_1, m_2 \in \mathcal{M}(\mathcal{X})$  tel que  $|m_1|(\mathcal{X}) = |m_2|(\mathcal{X})$ , la  $p$ -distance de Wasserstein s'écrit :*

$$\mathcal{W}_p(m_1, m_2) \stackrel{\text{def.}}{=} \left( \min_{\gamma \in \Gamma(m_1, m_2)} \int_{\mathcal{X} \times \mathcal{X}} |u - v|^p d\gamma(u, v) \right)^{1/p}. \quad (1.4)$$

$\Gamma(m_1, m_2)$  est l'ensemble des applications de transport entre  $m_1$  et  $m_2$ , on peut consulter l'annexe 4 pour plus d'informations à ce sujet. Cette notion n'est néanmoins pas suffisante pour notre application puisque la métrique ne peut prendre en entrée que des mesures de masses égales. Dans le cas d'une mesure discrète, nous rappelons que la masse est simplement la somme du module des amplitudes individuelles : nous ne pouvons donc généralement pas comparer une mesure source et une mesure reconstruite avec des amplitudes différentes. La solution classique consiste donc à distribuer la masse unitaire, divisée par le nombre de  $\delta$ -pics, uniformément sur chaque mesure de Dirac composant la mesure discrète. Il serait cependant beaucoup plus pratique d'incorporer le cas des masses différentes dans la métrique, d'autant plus que les méthodes sans-grille appliquées dans le cadre de l'imagerie produiront de toute évidence des reconstructions avec des mesures de Dirac d'amplitudes différentes. La métrique appropriée pour comparer deux mesures de masses différentes s'appelle la norme de Kantorovitch-Rubinstein, permettant de définir la *flat metric* [69, 75, 33]. Soit  $\mathcal{C}_b(\mathcal{X})$  l'espace des fonctions continues sur  $\mathcal{X}$  de norme *supremum* inférieure à 1 :

**Définition 1.8** (Transport optimal à masses différentes). *Soit  $m \in \mathcal{M}(\mathcal{X})$  à premier moment fini et  $\tau > 0$ , on appelle norme de Kantorovitch-Rubinstein la quantité :*

$$F_\tau(m) \stackrel{\text{def.}}{=} \sup_{f \in \mathcal{C}_b(\mathcal{X})} \left( \int_{\mathcal{X}} f dm, \|f\|_{\infty, \mathcal{X}} \leq \tau, f \text{ Lipschitz}, \|f\|_{\text{Lip}} \leq 1 \right)$$

où  $\|f\|_{\text{Lip}}$  est la constante de Lipschitz de  $f$ . On définit ensuite la *flat metric*  $d_\tau$  pour  $m_1, m_2 \in \mathcal{M}(\mathcal{X})$  à premiers moments finis par :

$$d_\tau(m_1, m_2) \stackrel{\text{def.}}{=} F_\tau(m_1 - m_2).$$

La paramètre  $\tau$  est homogène à une distance, et il doit être compris dans le cadre du transport optimal comme le coût de création/destruction d'une mesure de Dirac. La *flat metric* coïncide avec la 1-distance de Wasserstein, pour  $m_1, m_2$  de masses égales, lorsque

11. c'est-à-dire de même norme TV.

12. En fait, elle est bien définie sur le sous-ensemble  $X \stackrel{\text{def.}}{=} \left\{ m \in \mathcal{M}(\mathcal{X}), |m|(\mathcal{X}) \leq \|y\|_{\mathcal{H}}^2 / 2\lambda \right\}$  parce que seul le sous-ensemble borné de  $\mathcal{M}(\mathcal{X})$  est métrisable pour la topologie faible-\*, nous devons donc restreindre l'ensemble des mesures à  $X$  afin d'obtenir un espace polonais. Le cadre pratique de cette métrique exploitant le transport optimal est présenté dans l'annexe 4. Comme toutes les solutions du BLASSO appartiennent à  $X$  [43], nous conserverons ce léger abus de notation dans le reste du document.

$\tau \rightarrow +\infty$  [69]; elle coïncide également avec la norme de variation totale de  $m_1 - m_2$  lorsque  $\tau \rightarrow 0$ . Elle peut alors être entendue comme une interpolation entre la norme de la variation totale et la « 1-norme » de Wasserstein. En outre, lorsque le nombre de  $\delta$ -pics est correctement estimé, la *flat metric* représente l'erreur moyenne en termes de localisation et se trouve être similaire à l'erreur quadratique moyenne [33]. Finalement, la *flat metric* peut être étendue à la reconstruction discrète, c'est-à-dire aux images sur une grille fine; cette métrique est alors une méthode applicable à la reconstruction discrète. Il est alors tout à fait possible de comparer une reconstruction discrète (obtenue par un LASSO sur une grille plus fine par exemple) et une reconstruction sans-grille.

Pour résumer, nous disposons lorsque l'on cherche à comparer la mesure reconstruite et la mesure réelle de cette alternative :

- si la mesure source est composée de  $N$  pics, nous fixons l'amplitude de chaque  $\delta$ -pic à  $1/N$ . Nous appliquons la même procédure à la mesure reconstruite (avec un nombre différent ou non de  $\delta$ -pics), divisant ainsi uniformément la masse unitaire sur toutes les mesures de Dirac de la mesure considérée. La luminosité reconstruite n'est ainsi pas considérée comme pertinente et n'est pas prise en compte dans le calcul de la métrique : nous pouvons ensuite calculer directement la distance 1-Wasserstein, puisqu'elle est égale à la *flat metric* dans ce cas;
- nous voulons tenir compte de l'information de luminosité, et nous utilisons la *flat metric* pour comparer la mesure reconstruite et la mesure vérité-terrain.

**Résumé** — Les métriques classiques de la qualité de la reconstruction, telles que la norme  $\ell_2$ , ne peuvent pas être appliquées directement à la reconstruction sans-grille. Il est possible d'utiliser un score de transport optimal tel que la *flat metric* : elle prend en compte à la fois les reconstructions en amplitude et en position, et elle peut être facilement étendue à la reconstruction discrète (images sur une grille fine).

#### 1.4.2 Résultats sur une pile d'images SMLM

Dans le cadre de la super-résolution pour l'imagerie biomédicale, on souhaite reconstruire des détails fins dans des structures biologiques d'intérêt afin de mieux les étudier. En effet, les corps observés sont généralement plus petits que la limite de Rayleigh, longueur physique d'environ 200 nm qui correspond à l'échelle à laquelle le phénomène de diffraction de la lumière entre en jeu. Cette diffraction provoque un flou de l'image, qui peut être décrit comme une convolution de l'image par la fonction d'étalement du point, ou PSF (pour *Point Spread Function*) mentionnée plus haut. Nous voulons donc effectuer une *déconvolution* afin d'éliminer le flou dû à la diffraction et d'obtenir une image super-résolue. Il convient de noter qu'il existe d'autres systèmes d'imagerie pour lesquels les opérateurs en jeu sont légèrement différents de la déconvolution : *Nuclear Magnetic Resonance spectroscopy* avec un opérateur d'acquisition Fourier [38], MA-TIRF avec Laplace [35].

**Attention** — Le terme de *déconvolution* peut être ambigu pour la communauté du traitement du signal : il peut ainsi prendre le sens d'une méthode rehaussant une gamme de fréquence dans le spectre de Fourier de l'acquisition. La *super-résolution* cherche au contraire à reconstruire des fréquences supérieures à la fréquence de coupure, dans notre cas, c'est ce qui nous permet de reconstruire des détails fins de l'image. Les deux termes peuvent être utilisés indifféremment dans la littérature sans-grille.

Afin d'améliorer la résolution spatiale par rapport aux techniques de microscopie standard limitées par la diffraction et de permettre l'imagerie de structures biologiques plus petites que la limite de Rayleigh, il est possible d'utiliser l'imagerie SMLM pour « *Single Molecule Localisation Microscopy* » (microscopie de localisation de molécules uniques). Il s'agit d'une technique de microscopie de fluorescence qui a fait ses preuves pour la super-résolution [78]. Elle nécessite des fluorophores photoactivables ayant, grosso modo, deux états, par exemple « *On* » et « *Off* ». Ces molécules ne sont donc visibles sur les acquisitions que dans le cas « *On* », et l'idée est alors d'éclairer seulement quelques molécules dans l'échantillon pour produire une image et pouvoir les localiser précisément ; les molécules fluorescentes sont liées à la structure biologique et comme seules quelques molécules émettent dans une image, l'image obtenue est parcimonieuse<sup>13</sup>, ce qui permet une localisation précise. Ce processus est répété jusqu'à ce que toutes les molécules aient été éclairées et imagées. Toutes les positions des molécules imagées, image par image, peuvent alors être combinées pour former une image super-résolue qui permet de dépasser la limite de la diffraction, du reste en se débarrassant des dégradations dues au processus d'acquisition (flou, bruit, etc.). La qualité de la reconstruction est naturellement limitée par le nombre d'images nécessaires pour reconstruire l'image finale super-résolue, ce qui implique un coût en temps (précieux dans la mesure où l'organisme étudié peut se déplacer), en mémoire physique et par la densité de fluorophores allumés à chaque étape. En effet, il existe un risque de chevauchement gênant la localisation des molécules puisque le critère de séparation n'est pas respecté.

Les méthodes sans-grille peuvent être appliquées à n'importe quelle pile SMLM avec seulement la connaissance de l'opérateur d'acquisition, la PSF du système d'imagerie dans le cas présent. Elle est typiquement modélisée par le disque d'Airy, ou pour simplifier par une gaussienne d'écart-type  $\sigma$ . Dans ce chapitre, une méthode sans-grille exploitant le *Sliding Frank-Wolfe* est testée sur une pile d'acquisition SMLM 2D du Challenge<sup>14</sup> de l'École Polytechnique Fédérale de Lausanne (EPFL)<sup>15</sup>. Considérons donc la première image de la pile, on localise les sources ponctuelles et stocke les coordonnées des points reconstruits. Puis, nous passons à la deuxième image, nous localisons les

13. comme expliqué plus haut, cela signifie qu'il n'y a que quelques pixels de luminosité non nulle dans l'image.

14. Souvent référencé par *Challenge SMLM 2013* dans la littérature.

15. <https://srm.epfl.ch/DatasetPage?name=MT0.N1.HD>

sources, et ainsi de suite. Notons que les méthodes sans-grille dans ce cadre variationnel ne sont pas les seules à tirer parti du domaine continu, nous pourrions par exemple citer la régression de PSF utilisée au sein de DAOSTORM [58], *etc.*

Outre la convolution, nous sommes aussi confrontés aux bruits qui entachent l'observation, plus précisément à trois types de bruits principaux :

- le bruit de grenaille (également appelé bruit de Schottky ou bruit quantique) dû à la nature quantique de la lumière. Il provient du fait que les fluorophores émettent des photons de manière aléatoire, de sorte qu'entre  $t$  et  $t + 1$  (temps d'exposition), un nombre variable de photons a été émis, et donc un nombre variable de photons a été relevé par le capteur. Ainsi, l'amplitude du signal électrique généré dans le capteur (à chaque pixel) fluctue selon une statistique de Poisson ;
- le courant d'obscurité est un phénomène dû à l'agitation naturelle des électrons. Cette agitation naturelle est suffisante pour éjecter occasionnellement un électron de la bande de valence vers la bande de conduction sans effet photoélectrique. Des charges supplémentaires sont donc créées et interfèrent avec le signal. Le nombre d'électrons générés par l'agitation thermique suit une distribution de Poisson ;
- le bruit d'amplification et de lecture. Ce bruit est produit par le circuit électronique qui amplifie et convertit les paquets d'électrons en tension. Il est généralement modélisé par un bruit gaussien.

Nous avons donc plusieurs bruits qui polluent chacune des images observées. Pour traiter ce problème inverse mal posé, nous utilisons les résultats sur le BLASSO, avec le terme de moindres carrés comme terme d'attache aux données et la norme TV comme régulariseur, imposant *l'a priori* de parcimonie. Dans l'approche bayésienne, le terme des moindres carrés correspond au maximum de vraisemblance lorsque l'acquisition est polluée par un bruit gaussien : notre modèle fait ainsi l'approximation d'un bruit gaussien. Les mesures physiques sont discrètes, de sorte qu'à chaque image, il faut traiter une matrice composée de  $N_1 \times N_2$  pixels, chacun d'eux ayant une taille de  $(b_1, b_2)$ . Soit  $(c_{i,1}, c_{i,2})$  le centre du  $i$ -ème pixel, nous désignons les  $i$ -ème pixels du capteur de la caméra par

$$\Omega_i \stackrel{\text{def.}}{=} (c_{i,1}, c_{i,2}) + \left[ -\frac{b_1}{2N_1}, \frac{b_1}{2N_1} \right] \times \left[ -\frac{b_2}{2N_2}, \frac{b_2}{2N_2} \right].$$

Nous pouvons alors clarifier l'opérateur d'acquisition  $\Phi : m \mapsto \mathbb{R}^{N_1 N_2}$  qui contient l'opération d'intégration sur les pixels de la caméra [35], en effet avec l'évaluation du noyau gaussien discret  $\varphi$  avec des écarts types  $\sigma$ , pour  $i \in \{1, \dots, N_1 N_2\}$  il vient :

$$[\varphi(x)]_i \stackrel{\text{def.}}{=} \frac{1}{2\pi\sigma^2} \int_{\Omega_i} e^{-\left(\frac{(x_1-s_1)^2}{2\sigma^2} + \frac{(x_2-s_2)^2}{2\sigma^2}\right)} ds_1 ds_2.$$

Dans l'ensemble de données du SMLM, on a l'écart-type de la PSF  $\sigma = 149.39\text{nm}$  et  $b_1 = b_2 = 100\text{nm}$ . La reconstruction est effectuée par notre implémentation du

*Sliding Frank-Wolfe* en python<sup>16</sup> dans la mesure où il s’agit de la méthode la plus robuste disponible : en effet, elle fonctionne avec un noyau gaussien, elle a des résultats prouvés dans un régime de bruit, *etc.* Les résultats sont présentés dans la figure 1.3. La pile de 2500 images de  $64 \times 64$  est dite de fortes densités avec un SNR élevé : le nombre de fluorophores activés est assez important, et le bruit n’est pas négligeable<sup>17</sup>.

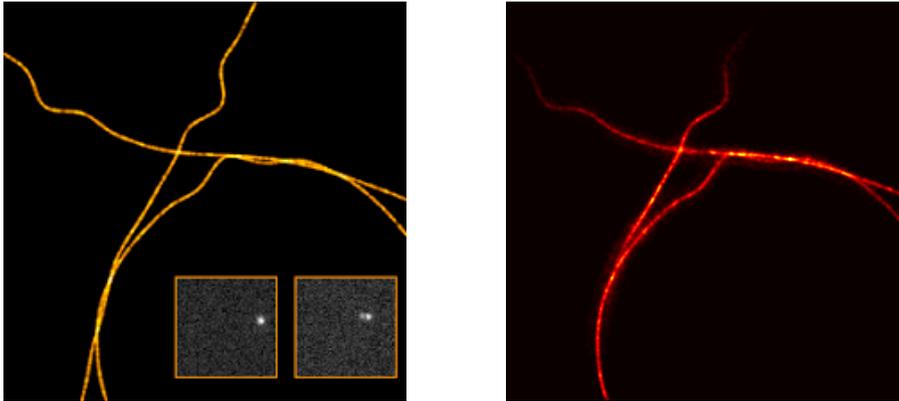


FIGURE 1.3 – (a) Filaments constituant la vérité-terrain, deux extraits de la pile dans le carré ci-dessous : convolution + tous les bruits décrits précédemment. (b) Mesure reconstruite par *Sliding Frank-Wolfe* visualisée par un noyau gaussien avec un  $\sigma$  plus petit (voir texte).

La *flat metric* entre la mesure reconstruite  $m_{a,x}$  et la mesure vérité-terrain  $m_{a_0,x_0}$  est alors calculée, et elle vaut  $d_\tau(m_{a,x}, m_{a_0,x_0}) = 1,7 \times 10^{-2}$ . La reconstruction est convaincante et retrouve bien les détails fins des structures biologiques, on peut clairement voir les filaments entrelacés dans la partie droite de l’image.

Notons qu’une caractéristique intéressante de la reconstruction sans grille est qu’une fois la mesure de Radon calculée, il est facile de la représenter par n’importe quel opérateur sur une grille fine de son choix. En effet, comme on ne peut pas représenter une mesure discrète  $m$ , nous traçons plutôt  $\Phi m$  où  $\Phi$  est la PSF avec un écart-type sensiblement plus petit, afin de voir clairement la déconvolution. Dans toutes nos reconstructions, nous convolons la mesure reconstruite avec la PSF pour une variance de  $\sigma/6$  et nous la représentons sur une grille 32 fois plus fine. À titre de comparaison, les méthodes discrètes sont exécutées pour une grille fine qui est fixe, et si l’on souhaite une reconstruction plus fine, il faut refaire tout le calcul sur cette nouvelle grille.

Nous testons enfin la reconstruction sans-grille sur un ensemble de données réelles avec un haut SNR et une grande densité, images de filaments fournies par le défi SMLM 2013 de l’EPFL. Dans cette pile de 500 images de taille  $128 \times 128$  pixels, la FWHM (« *full width at half maximum* ») du système d’acquisition est estimée à 351,8 nm. Nous rappelons que la FWHM est la largeur de la courbe gaussienne mesurée entre les points de l’axe  $y$  qui correspondent à la moitié de l’amplitude maximale. Notons

16. Voir notre dépôt GitHub pour notre implémentation PyTorch : <https://github.com/XeBasTeX>

17. Voir <https://srm.epfl.ch/DatasetPage?name=MT0.N1.HD> pour plus d’informations.

également qu'elle est liée à la variance  $\sigma$  par  $\text{FWHM} = 2\sqrt{2 \ln 2} \times \sigma$ . Nous comparons la reconstruction de la méthode sans-grille avec la sortie de l'algorithme Deep-STORM [73], qui constitue l'un des algorithmes de l'état de l'art offrant les résultats les plus convaincants sur le plan visuel. Les reconstructions de la méthode sans-grille et de l'algorithme Deep-STORM sont présentées dans la figure 1.4, où l'on peut apprécier la reconstruction par la méthode sans grille sur des détails fins. La reconstruction semble un peu floue par rapport à Deep-STORM, en raison de la convolution par un noyau gaussien, même d'écart-type faible. Cependant, il est intéressant de noter que les deux comparaisons permettent de retrouver les filaments, en particulier dans les cercles jaunes qui grossissent l'image : la reconstruction sans-grille semble mieux préserver la structure que le résultat approximatif de Deep-STORM. La qualité de la reconstruction est particulièrement intéressante pour la reconstruction sans-grille car elle ne nécessite pas de jeux de données tests pour produire cette reconstruction, contrairement à Deep-STORM. La seule donnée nécessaire est la connaissance (ou tout du moins une estimation) de l'opérateur direct, la reconstruction sans-grille peut alors être effectuée à partir de n'importe quelle entrée sans avoir à entraîner le modèle sur différents types et niveaux de bruit.

**Résumé** — Nous avons testé une méthode sans-grille sur les données synthétiques et expérimentales de SMLM. Le problème sans-grille peut être résolu grâce à l'algorithme du *Sliding Frank-Wolfe* qui produit des résultats très convaincants. Ces performances sont d'autant plus intéressantes qu'il n'y a qu'un seul paramètre, facile à régler et robuste au bruit. Il peut donc être facilement adapté à tout autre ensemble de données dont l'opérateur d'observation est connu.

## 1.5 Conclusion du chapitre

Ce chapitre s'est proposé de fournir un état de l'art sur les méthodes variationnelles sans-grille pour la reconstruction de sources ponctuelles. Cette approche *off-the-grid* s'entend comme la limite des méthodes discrètes parcimonieuses telles que le LASSO, quand le pas de discrétisation tend vers 0. Elle se traduit par l'utilisation d'une fonctionnelle convexe, appelée BLASSO, écrite sur l'espace des mesures de Radon  $\mathcal{M}(\mathcal{X})$  qui est un espace de Banach de dimension infinie. La littérature fournit de nombreux résultats théoriques qui exposent, entre autres, les garanties pour retrouver le bon nombre de sources ponctuelles en optimisant le BLASSO, des bornes sur l'écart entre la reconstruction et la source, *etc.* En outre, plusieurs travaux ont permis d'écrire différents algorithmes permettant de résoudre numériquement le problème variationnel, avec des méthodes de programmation semi-définie positive, des algorithmes gloutons ou des flots de gradient. Tous ces résultats présentent un réel intérêt pratique, qui est démontré par l'application de ces méthodes sans-grille sur des problèmes concrets tels que le problème de super-résolution. Nous avons ainsi montré l'utilité de l'approche sans-grille à travers un exemple dans le cadre de la microscopie de fluorescence : la reconstruction obtenue est

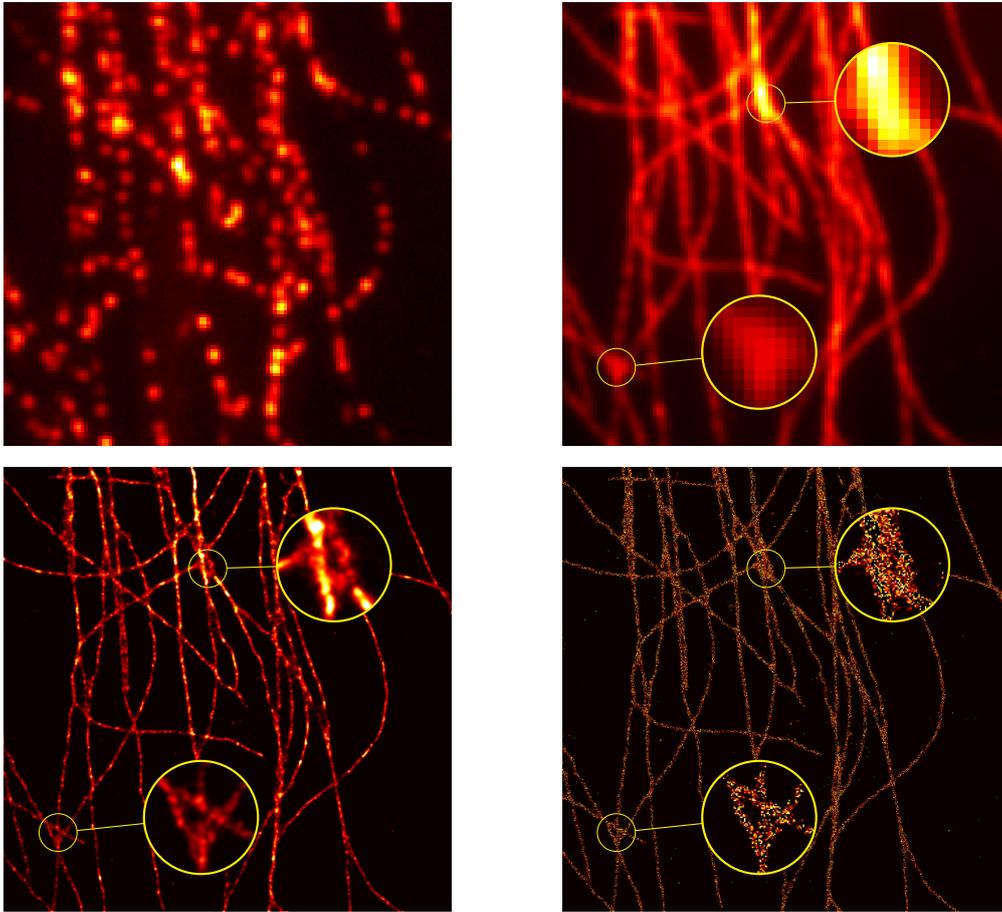


FIGURE 1.4 – (a) Extrait de la pile (b) Moyenne de la pile (c) Reconstruction sans-grille (d) Deep-STORM.

de qualité comparable aux productions de l'apprentissage profond, en ne nécessitant pas de jeux de données et avec la seule connaissance de l'opérateur d'observation.



## Chapitre 2

# Fluctuations de luminosité : reconstruction de mesures ponctuelles par la covariance spatiale.

**Points clefs – la microscopie de fluctuations constitue une bonne alternative aux méthodes de super-résolution traditionnelles :**

1. Les méthodes de super-résolution en microscopie de fluorescence telles que SMLM produisent de bons résultats, mais nécessitent un **nombre important d'images**, ce qui n'est compatible avec l'imagerie de cellules vivantes.
2. Elles peuvent nécessiter en outre l'utilisation de **fluorophores** et de **matériels spécifiques**, qui se révèlent coûteux.
3. La microscopie de fluctuations permet de **diminuer** fortement le **nombre d'images** nécessaires, de surcroît avec du **matériel conventionnel**. Néanmoins, la littérature ne propose **pas de formulation sans-grille** pour travailler dans ce cadre.

**Contributions – une méthode sans-grille pour la microscopie de fluctuations**

4. Nous nous proposons dans ce chapitre de formuler un cadre de travail **dynamique**, afin de modéliser l'observation dans un cadre de **fluctuations de luminosité**.
5. À l'aide de l'**hypothèse d'indépendance** statistique des **fluctuations** de chaque fluorophore, nous formulons un **opérateur** d'observation ainsi qu'un **problème** variationnel **sans-grille**.
6. La résolution de ce problème produit des résultats en demi-teinte : si des **détails fins** sont bien retrouvés par la reconstruction, la **solution est en pointillée** alors qu'on souhaiterait retrouver des structures courbes.

**Chapitre 2 – Fluctuations de luminosité : reconstruction de mesures ponctuelles :**

2.1	Présentation . . . . .	36
2.2	Modélisation temporelle . . . . .	37
2.2.1	Positionnement du problème . . . . .	37
2.3	Un problème variationnel sur la covariance . . . . .	40
2.4	Résultats numériques . . . . .	46
2.4.1	Données simulées . . . . .	47
2.4.2	Données expérimentales . . . . .	48
2.4.3	Des reconstructions en pointillé . . . . .	50
2.5	Conclusion partielle . . . . .	51

**2.1 Présentation**

Afin d'améliorer la résolution spatiale obtenue par des techniques de microscopie standard limitées par la diffraction et de permettre l'imagerie fine de structures biologiques, nous avons vu dans le chapitre 1 qu'il existait plusieurs techniques de microscopie de fluorescence pour effectuer la super-résolution d'un échantillon biologique. On peut ainsi relever des méthodes telles que *STimulated Emission Depletion* (STED) [56], qui est nocive pour l'échantillon et implique un équipement spécial, *Structured Illumination Microscopy* (SIM) [51] qui a une résolution spatiale limitée et l'imagerie *Single Molecule Localization Microscopy* (SMLM) [78]. Cette dernière couplée à une méthode sans-grille a produit de très bons résultats, mais sa résolution temporelle reste limitée, ce qui ne permet pas l'imagerie d'organismes vivants, elle nécessite de plus l'utilisation de fluorophores spécifiques (photoactivables).

Au fil des dernières années, plusieurs méthodes de microscopie de fluorescence pour la super-résolution ont été proposées pour contourner ces problèmes. En particulier, une méthode intéressante, qui exploite l'indépendance des fluctuations temporelles stochastiques de luminosité des colorants fluorescents attachés aux corps étudiés, s'est dégagée dans la littérature. La technique *Super-Resolution Optical Fluctuation Imaging* (SOFI) qui tire parti de cette idée utilise ainsi la pile d'images acquise en calculant ses cumulants statistiques d'ordre ( $n$ ) qui réduisent la fonction d'étalement du point (PSF) d'un facteur  $\sqrt{n}$ . Cette approche est également moins contraignante puisqu'elle peut être réalisée avec un système de microscope classique et des colorants conventionnels. Malgré une amélioration de la résolution temporelle, sa résolution spatiale reste un peu limitée par rapport à la SMLM. Une limitation similaire est perceptible dans la microscopie à super-résolution des fluctuations radiales (SRRF) [52] où les symétries radiales locales provenant de la symétrie de la PSF sont calculées dans chaque image. Nous pourrions également citer dans cette veine les algorithmes déroulant tels que *Sparsity Based Super-Resolution Correlation Microscopy* (SPARCOM) [83] qui exploitent à la fois la nature parcimonieuse de la distribution des fluorophores et l'*a priori* statistique sur la décorrélation des fluctuations.

Ces techniques de microscopie de fluctuations sont développées dans un cadre discret, ce qui signifie que l'optimisation est effectuée sur une image super-résolue définie sur une grille plus fine que les images observées. Au contraire, nous définissons dans ce chapitre une méthode sans-grille, dans la continuité des méthodes détaillés dans le chapitre 1.

Dans ce chapitre, nous proposons la définition d'un opérateur  $\Lambda$  et d'un problème variationnel sans-grille ( $\mathcal{Q}_\lambda(\mathbf{y})$ ) pour exploiter l'hypothèse de fluctuations des fluorophores; nous présentons ensuite la reconstruction produite par notre méthode sur des données simulées et expérimentales. Elle fournit des résultats convaincants avec un seul paramètre de réglage et sans recourir à un ensemble de données d'apprentissage, tout en étant adaptée à l'imagerie des cellules vivantes, aux microscopes courants et aux colorants conventionnels.

## 2.2 Modélisation temporelle

Comme dans le chapitre 1,  $\mathcal{X}$  est un compact de  $\mathbb{R}^d$ . Nous disposons d'une pile d'images  $\mathbf{y}$  acquise sur  $[0, T]$  où  $T \in \mathbb{R}_+^*$ , et nous allons chercher à exploiter une information temporelle dans cette observation : il est donc nécessaire pour notre modélisation de définir des quantités *dynamiques*, variant en fonction du temps. La source est composée de  $N$  fluorophores aux positions fixes  $x_i \in \mathcal{X}$  et d'amplitudes fluctuantes  $t \mapsto a_i(t)$  pour  $i$  allant de 1 à  $N$ . Nous souhaiterions pouvoir définir la pile d'acquisition comme l'équation qui suit pour  $x \in \mathcal{X}$  et  $t \in [0, T]$  p.p. :

$$\mathbf{y}(x, t) = \sum_{i=1}^N a_i(t) h(x - x_i) = \Phi \mu_t. \quad (2.1)$$

Un exemple de pile d'acquisition  $\mathbf{y}$  est donnée en figure 2.1, avec les positions de la source  $\mu$ . Il faut maintenant réfléchir à une manière d'incorporer une dimension temporelle et chercher à donner un sens à ces nouvelles quantités.

### 2.2.1 Positionnement du problème

De la même façon qu'en équation d'évolution, on veut maintenant tirer parti d'une composante temporelle, c'est-à-dire à un temps  $t$  associer une mesure de Radon. Nous pourrions ainsi penser à un espace tel que  $L^2([0, T]; L^2(\mathcal{X}))$  pour décrire les sources ponctuelles d'amplitudes fluctuantes, en nous inspirant du cadre de [12].

**Attention** — Les quantités dynamiques dont il va être question seront **Böchner** intégrables. L'intégrale de Böchner, d'abord formulée pour donner un sens à la mesurabilité et à l'intégration de fonctions à valeurs vectorielles, s'étend aux fonctions à valeurs dans un espace de Banach. Nous invitons le lecteur à se référer à [12] pour plus de détails sur ces notions appliquées au cadre sans-grille.

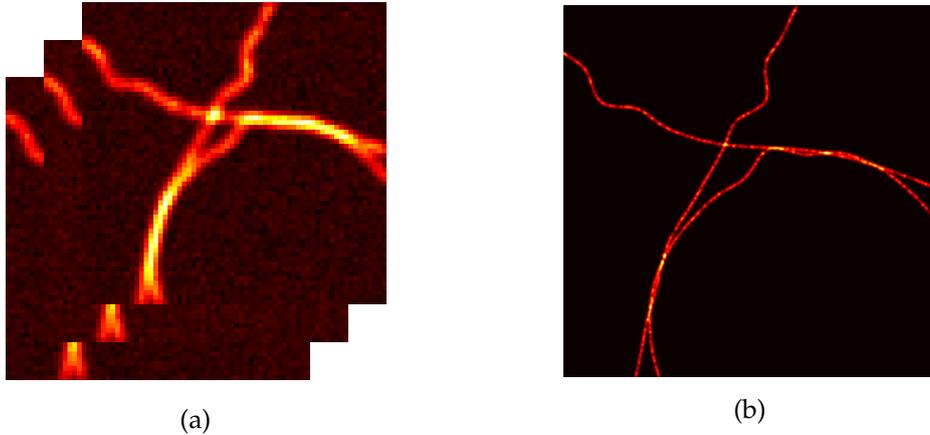


FIGURE 2.1 – (a) Trois extraits d’une pile d’images  $y$  en microscopie de fluctuation. (b) Les positions de la source  $\mu_t$  que l’on cherche à reconstruire.

À ce stade, il est déjà nécessaire de préciser ce qu’est une fonction mesurable dans un espace de Banach.

### Mesurabilité dans les espaces de Banach

Introduisons quelques éléments de la théorie des intégrales de Bochner [37, 39]. On note par  $(\Omega, \Sigma, \mu)$  l’espace métrique du temps, muni de la tribu borélienne et de la mesure de Lebesgue, typiquement  $\Omega = [0, T]$ . Les fonctions Bochner-intégrables avec lesquelles nous allons travailler seront à valeur dans l’espace de Banach  $E$ , par exemple  $\mathcal{M}(\mathcal{X})$ .

**Définition 2.1.** On dit que  $f : \Omega \rightarrow E$  est  $\mu$ -mesurable si elle est égale presque partout à la limite d’une suite de fonctions  $(f_n)_n : \Omega \rightarrow E$  étagées, c’est-à-dire :

$$\lim_{n \rightarrow +\infty} f_n(t) = f(t) \text{ pour } \mu\text{-presque tout } t \in \Omega.$$

Les fonctions  $\mu$ -mesurables sont aussi dites Bochner mesurables ou fortement mesurables.

**Définition 2.2.** Une fonction  $f : \Omega \rightarrow E$  est à valeurs séparées presque partout s’il existe un sous-ensemble  $N \subseteq \Omega$  de mesure nulle  $\mu(N) = 0$ , telle que  $f(\Omega \setminus N) \subseteq E$  est séparable.

**Définition 2.3.** Une fonction  $f : \Omega \rightarrow E$  est dite faiblement mesurable si l’application

$$t \mapsto \langle f, x^* \rangle (t) \stackrel{\text{def.}}{=} \langle f(t), x^* \rangle_{E, E^*}$$

est mesurable pour tout  $x^* \in E^*$ .

Finalement,

**Théorème 2.1** (Mesurabilité de Pettis [37]). *Une fonction  $f : \Omega \rightarrow E$  sur un espace mesuré  $(\Omega, \Sigma, \mu)$  et à valeurs dans un espace de Banach  $E$  est (Böchner-)mesurable si et seulement si elle est à la fois faiblement mesurable et à valeurs séparées.*

### Mesure dynamique et modélisation du problème

La notion de mesurabilité dans les espaces de Banach étant clarifiée, nous pouvons désormais parler de fonction *Böchner-intégrable*. La construction de l'intégrale de Böchner est similaire à celle de l'intégrale de Lebesgue avec les fonctions étagées. Dans la suite du propos, la notion d'intégrabilité au sens de Böchner sera définie grâce au théorème suivant :

**Théorème 2.2.** *Soit  $f : \Omega \rightarrow E$  une fonction Böchner mesurable.  $f$  est Böchner-intégrable si et seulement si  $\|f\|_E$  est Lebesgue-intégrable.*

De la même manière que pour les fonctions à valeurs scalaires, on définit l'espace  $L^p(\Omega; E)$  des fonctions Böchner-intégrables avec :

**Définition 2.4.** *Si  $1 \leq p \leq +\infty$ , l'espace  $L^p(\Omega; E)$  est l'ensemble des fonctions  $f$  mesurables telles que  $\|f\|_E \in L^p(\Omega; \mathbb{R})$ . La norme de cet espace s'écrit  $\|f\|_{L^p(\Omega; E)} = \|\|f\|_E\|_{L^p(\Omega; \mathbb{R})}$ .*

Plusieurs propriétés des espaces  $L^p$  scalaires sont conservées dans les espaces de Böchner  $L^p$ , notamment l'inégalité d'Hölder [39], ce qui entraîne l'inclusion des espaces  $L^p$ ; à savoir que pour  $1 \leq p \leq q \leq +\infty$ ,  $L^q(\Omega; E) \subset L^p(\Omega; E)$ . Nous aurons besoin de cette propriété pour donner un sens à nos quantités dans la suite de ce chapitre.

**Définition 2.5.** *Soit  $E$  un espace de Banach, on note :*

$$L^2(0, T; E) \stackrel{\text{def.}}{=} L^2([0, T]; E)$$

*l'ensemble des classes d'équivalence de carré (Böchner) intégrable.*

Posons maintenant  $\Omega = [0, T]$  et  $E = \mathcal{M}(\mathcal{X})$  : la source sera décrite par  $\mu \in L^2(0, T; \mathcal{M}(\mathcal{X}))$ . Cette mesure dynamique  $\mu$  peut s'interpréter comme une courbe paramétrée dont le graphe parcourt l'espace des mesures de Radon  $\mathcal{M}(\mathcal{X})$ . Nous pouvons préciser un peu plus sa forme : nous voulons modéliser des fluorophores, donc  $N \in \mathbb{N}^*$  sources ponctuelles dont les amplitudes fluctuent au cours du temps, mais où les positions sont fixes. Ainsi, on peut modéliser la mesure dynamique source par :

$$t \mapsto \mu(t) \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i(t) \delta_{x_i}, \quad (2.2)$$

pour des amplitudes  $a_i \in L^2(0, T; \mathbb{R})$  et des positions  $x_i \in \mathcal{X}$ . Nous cherchons donc à déterminer  $N$  le nombre de pics, les fonctions  $a_i$  de la luminosité et surtout les positions fixes  $x_i$  des mesures de Dirac.

### Opérateur d'acquisition

Soit la mesure de Radon temporelle  $\mu \in L^2(0, T; \mathcal{M}(\mathcal{X}))$  que l'on cherche à reconstruire. Dans l'optique d'appliquer les résultats [43] établis pour le BLASSO que nous avons présentés au chapitre 1 ; il est nécessaire que le noyau  $\varphi$  soit tel que  $\varphi \in \mathcal{C}^2(\mathcal{X}; \mathcal{H})$  pour définir l'opérateur linéaire borné  $\Phi$ , avec  $\mathcal{H}$  l'espace de Hilbert dans lequel notre opérateur prendra ses valeurs. Pour presque tout  $t \in [0, T]$ , avec la notation  $\mu_t \stackrel{\text{def.}}{=} \mu(t)$  on a :

$$\Phi(\mu_t)(x) \stackrel{\text{def.}}{=} \langle \varphi(x), \mu_t, = \rangle_{\mathcal{M}} \int_{\mathcal{X}} \varphi(x)(s) d\mu_t(s).$$

On note  $\mathbf{y} \in L^2(0, T, \mathcal{H})$  l'acquisition obtenue par l'opérateur d'observation, défini<sup>1</sup> pour tout  $x \in \mathcal{X}$  et  $t \in [0, T]$  p.p. par :

$$\mathbf{y}(x, t) \stackrel{\text{def.}}{=} \Phi(\mu_t)(x).$$

Dans le cas de la microscopie de fluctuations,  $\mathbf{y}$  correspond à la pile d'acquisitions et l'opérateur  $\Phi$  réalisera la convolution **spatiale** ; pour  $h$  la PSF du système d'acquisition, si le noyau de l'opérateur s'écrit  $\forall s \in \mathcal{X}, \varphi(s) = h(\cdot - s)$  [35], on aura pour presque tout  $t \in [0, T]$ , avec  $\mu_t$  défini en (2.2) :

$$\begin{aligned} \mathbf{y}(x, t) &= \sum_{i=1}^N a_i(t) h(x - x_i) \\ &= (\Phi\mu(t))(x). \end{aligned}$$

**Résumé** — La vérité-terrain peut être modélisée par une mesure dynamique  $t \mapsto \mu_t \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i(t) \delta_{x_i(t)}$ , somme pondérée par  $a_i(t)$  de mesures de Dirac aux positions fixes  $x_i$ , mais aux amplitudes fluctuantes. La pile d'acquisition se définit comme la quantité  $\mathbf{y}(x, t) = \Phi(\mu_t)(x)$ .

### 2.3 Un problème variationnel sur la covariance

Le processus d'observation ainsi modélisé, nous allons maintenant chercher à exploiter l'*a priori* physique dont nous disposons. Plus précisément, les fluctuations temporelles entre les luminosités  $a_i$  des fluorophores sont indépendantes [36], ce qui se traduit par :

$$\begin{aligned} \text{Cov}(a_i, a_j) &\stackrel{\text{def.}}{=} \frac{1}{T} \int_0^T a_i(t) a_j(t) dt - \frac{1}{T^2} \int_0^T a_i(t) dt \int_0^T a_j(t) dt \\ &= 0 \text{ si } i \neq j. \end{aligned} \tag{2.3}$$

1. Avec l'abus de notation  $\mathbf{y}(x, t)$  au lieu de  $y_t(x)$

Nous allons travailler dans ce chapitre de manière variationnelle, en essayant d'exploiter une covariance en continu, dans l'esprit de Colorme [86] pour définir l'énergie à minimiser. Dans cette partie (et cette partie seulement), on considèrera un problème sans bruits, l'acquisition  $\mathbf{y}$  sera alors définie pour presque tout  $t \in [0, T]$  :

$$\mathbf{y}_t = \Phi \mu_t$$

Nous pouvons définir les cumulants temporels en commençant par le cumulant d'ordre un.

**Définition 2.6** (Moyenne temporelle). *On appelle moyenne temporelle, ou cumulant statistique d'ordre un, la quantité  $\bar{\mathbf{y}}$  définie pour  $x \in \mathcal{X}$  par :*

$$\bar{\mathbf{y}}(x) = \frac{1}{T} \int_0^T \mathbf{y}(x, t) dt.$$

Cette intégrale est bien définie, puisque  $\mathbf{y} \in L^2(0, T, \mathcal{H})$  et  $L^2(0, T, \mathcal{H}) \subset L^1(0, T; \mathcal{H})$ .

**Attention** — La normalisation par  $\frac{1}{T}$  sert essentiellement à assurer l'homogénéité de l'expression, et peut être omise par le lecteur dans les calculs.

Soit  $\bar{a}_i = \frac{1}{T} \int_0^T a_i(t) dt$ , on définit  $m_{a,x} \stackrel{\text{def.}}{=} \sum_{i=1}^N \bar{a}_i \delta_{x_i}$ . Il est clair que :

$$\bar{\mathbf{y}} = \Phi m_{a,x}.$$

En posant un BLASSO, pour l'observation  $\bar{\mathbf{y}}$  et l'opérateur  $\Phi$ , on peut reconstruire  $m_{a,x}$  et par extension le nombre de fluorophores ainsi que leurs positions. Dit autrement, on peut faire le lien grâce à  $\bar{\mathbf{y}}$  entre  $\mu = \sum_{i=1}^N a_i(t) \delta_{x_i}$  que l'on cherche à reconstruire et  $m_{a,x} \stackrel{\text{def.}}{=} \sum_{i=1}^N \bar{a}_i \delta_{x_i}$  que l'on peut estimer par une méthode sans-grille. On peut dès lors définir le cumulant d'ordre deux et observer si l'on peut obtenir de la même manière une relation de cet acabit.

**Définition 2.7** (Covariance spatiale). *Soient  $(u, v) \in \mathcal{X}^2$ , la covariance spatiale  $\mathbf{R}_y \in L^2(\mathcal{X}^2)$  de la pile d'acquisition se définit par :*

$$\mathbf{R}_y(u, v) \stackrel{\text{def.}}{=} \frac{1}{T} \int_0^T (\mathbf{y}(u, t) - \bar{\mathbf{y}}(u)) (\mathbf{y}(v, t) - \bar{\mathbf{y}}(v)) dt.$$

Voyons maintenant comment transposer ce résultat à la covariance spatiale  $\mathbf{R}_y$ . Puisque  $\mathbf{y} \in L^2(0, T, \mathcal{H})$ , on peut écrire pour tout  $u, v \in \mathcal{X}$  :

$$\begin{aligned}
\mathbf{R}_y(u, v) &= \frac{1}{T} \int_0^T (\mathbf{y}(u, t) - \bar{\mathbf{y}}(u)) (\mathbf{y}(v, t) - \bar{\mathbf{y}}(v)) dt \\
&= \frac{1}{T} \int_0^T \mathbf{y}(u, t) \mathbf{y}(v, t) dt - \bar{\mathbf{y}}(u) \bar{\mathbf{y}}(v) - \bar{\mathbf{y}}(u) \bar{\mathbf{y}}(v) + \bar{\mathbf{y}}(u) \bar{\mathbf{y}}(v) \\
&= \frac{1}{T} \sum_{1 \leq i, j \leq N} h(u - x_i) h(v - x_j) \int_0^T a_i(t) a_j(t) dt - \bar{\mathbf{y}}(u) \bar{\mathbf{y}}(v).
\end{aligned}$$

En particulier, comme  $\mathbf{y}(\cdot, t) = \sum_{i=1}^N a_i(t) h(\cdot - x_i)$ , il vient :

$$\begin{aligned}
\bar{\mathbf{y}}(u) &= \frac{1}{T} \int_0^T \mathbf{y}(u, t) dt \\
&= \sum_{i=1}^N h(u - x_i) \frac{1}{T} \int_0^T a_i(t) dt.
\end{aligned}$$

On en déduit que

$$\bar{\mathbf{y}}(u) \bar{\mathbf{y}}(v) = \sum_{i, j} h(u - x_i) h(v - x_j) \left( \frac{1}{T} \int_0^T a_i(t) dt \frac{1}{T} \int_0^T a_j(t) dt \right).$$

Finalement, on peut factoriser les deux termes pour obtenir :

$$\begin{aligned}
\mathbf{R}_y(u, v) &= \frac{1}{T} \sum_{1 \leq i, j \leq N} h(u - x_i) h(v - x_j) \int_0^T a_i(t) a_j(t) dt \\
&\quad - \sum_{1 \leq i, j \leq N} h(u - x_i) h(v - x_j) \left( \frac{1}{T} \int_0^T a_i(t) dt \frac{1}{T} \int_0^T a_j(t) dt \right) \\
&= \sum_{1 \leq i, j} h(u - x_i) h(v - x_j) \left( \frac{1}{T} \int_0^T a_i(t) a_j(t) dt - \frac{1}{T^2} \int_0^T a_i(t) dt \int_0^T a_j(t) dt \right).
\end{aligned}$$

En utilisant l'hypothèse physique d'indépendance des fluctuations temporelles entre les molécules de (2.3) on obtient :

$$\begin{aligned}
\mathbf{R}_y(u, v) &= \sum_{1 \leq i, j} h(u - x_i) h(v - x_j) \left( \frac{1}{T} \int_0^T a_i(t) a_j(t) dt - \frac{1}{T^2} \int_0^T a_i(t) dt \int_0^T a_j(t) dt \right) \\
&= \sum_{i=1}^N h(u - x_i) h(v - x_i) \underbrace{\left( \frac{1}{T} \int_0^T a_i(t)^2 dt - \frac{1}{T^2} \left( \int_0^T a_i(t) dt \right)^2 \right)}_{\stackrel{\text{def.}}{=} M_i} \\
&= \sum_{i=1}^N M_i h(u - x_i) h(v - x_i).
\end{aligned}$$

Cette dernière équation donne l'intuition que  $R_y$  est le résultat d'un opérateur évalué sur une mesure discrète  $m_{M_0, x_0} \stackrel{\text{def.}}{=} \sum_{i=1}^N M_i \delta_{x_i}$ ; en particulier une mesure *stationnaire*, qui n'a donc plus de dépendance temporelle.

**Résumé** — À partir d'une acquisition spatio-temporelle  $y(x, t)$  on peut calculer la covariance spatiale  $R_y$ . Grâce à l'hypothèse d'indépendance des fluctuations de luminosité des fluorophores, cette quantité peut aussi être vue comme le résultat d'un opérateur  $\Lambda$  de  $\mathcal{M}(\mathcal{X})$  dans  $L^2(\mathcal{X}^2)$ .

### Opérateur de covariance pour mesure stationnaire

On va noter  $\Lambda$  l'opérateur de covariance qui, à une mesure de Radon, associe une classe d'équivalence de  $L^2(\mathcal{X}^2)$  :

$$\begin{aligned} \Lambda : \mathcal{M}(\mathcal{X}) &\longrightarrow \mathcal{H} = L^2(\mathcal{X}^2) \\ m &\longmapsto R_y. \end{aligned}$$

Dans le contexte de la microscopie de fluorescence, le noyau d'acquisition de  $\Phi$  est précisément la PSF  $\varphi(x) = h(\cdot - x)$ . On définit de la même manière le noyau de  $\Lambda$  avec  $\phi$ , qui doit donc incorporer la convolution et représenter la covariance spatiale. Son expression analytique évaluée en  $m \in \mathcal{M}(\mathcal{X})$  est donnée pour tout  $u, v \in \mathcal{X}$  par :

$$\begin{aligned} \Lambda(m)(u, v) &\stackrel{\text{def.}}{=} \langle \phi(u, v), m \rangle_{\mathcal{M}(\mathcal{X})} \\ &= \int_{\mathcal{X}} h(u - x)h(v - x) dm(x). \end{aligned}$$

Dans le cas d'une mesure discrète stationnaire  $m = \sum_{i=1}^N M_i \delta_{x_i}$ , on retrouve bien  $\Lambda m = R_y$  le cumulante établi un peu plus haut. Prouvons quelques propriétés de  $\Lambda$  nécessaires à l'adaptation au cadre de BLASSO.

**Proposition 2.1.** *L'application de covariance  $\Lambda : \mathcal{M}(\mathcal{X}) \rightarrow L^2(\mathcal{X}^2)$  est un opérateur linéaire, continu de la topologie faible-\* à la topologie faible. Son adjoint  $\Lambda^* : L^2(\mathcal{X}^2) \rightarrow \mathcal{C}_0(\mathcal{X})$  s'écrit  $\Lambda^*(p)(x) = \langle \phi(x), p \rangle_{L^2(\mathcal{X}^2)}$  pour tout  $p \in L^2(\mathcal{X}^2)$ ,  $x \in \mathcal{X}$ .*

*Démonstration.* Par continuité et bilinéarité du crochet de dualité, on a :

$$\forall p \in L^2(\mathcal{X}^2), \quad \langle p, \Lambda m \rangle = \int_{\mathcal{X}} \langle p, \phi(x) \rangle dm(x).$$

Comme  $x \mapsto \langle p, \phi(x) \rangle$  est dans  $\mathcal{C}(\mathcal{X})$ , on obtient la continuité faible-\* à faible et l'expression de l'opérateur adjoint. □

On a  $\phi : X \rightarrow L^2(\mathcal{X}^2)$  le noyau de  $\Lambda$ . Son adjoint  $\Lambda^* : L^2(\mathcal{X}^2) \rightarrow \mathcal{C}(X)$  va s'écrire pour tout  $x \in X$  et tout  $p \in L^2(\mathcal{X}^2)$  :

$$\begin{aligned} \Lambda^*(p)(x) &= \langle \phi(x), p \rangle_{L^2(\mathcal{X}^2)} \\ &= \int_{\mathcal{X}^2} [\phi(x)(s)] p(s) ds. \end{aligned}$$

Dans le cas où l'opérateur d'acquisition  $\Phi$  est une convolution, on a  $\phi(x) = h(\cdot - x) \times h(\cdot - x)$ , pour  $s = (s_1, s_2) \in \mathcal{X}^2$  et ainsi pour tout  $p \in L^2(\mathcal{X}^2)$  :

$$\begin{aligned} \Lambda^*(p)(x) &= \int_{\mathcal{X}^2} h(s_1 - x)h(s_1 - x)p(s_1, s_2) d(s_1, s_2) \\ &= (\tilde{\phi} * p)(x, x). \end{aligned}$$

En pratique,  $R_y$  sera une matrice de covariance : ce dernier résultat précise que le certificat sera (à une convolution près) la diagonale mise dans l'ordre lexicographique de la matrice de covariance. Un exemple de matrice de covariance et de dual de cette matrice de covariance sont donnés en figure 2.2.

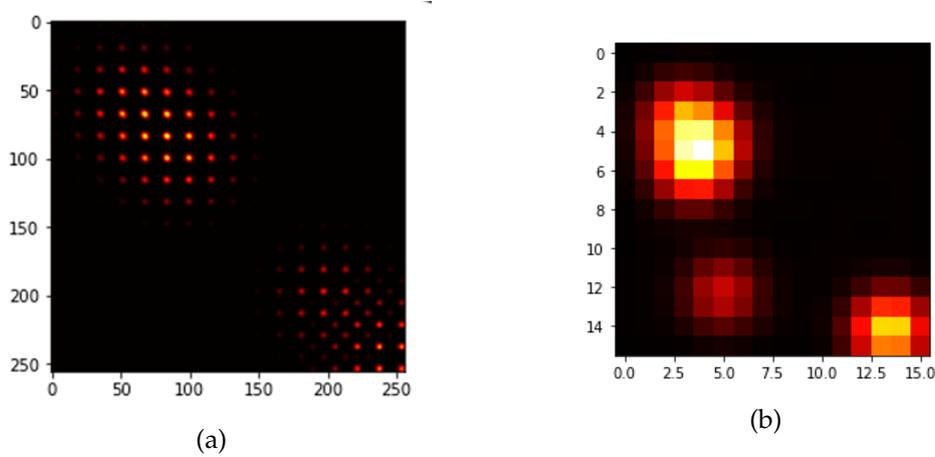


FIGURE 2.2 – (a) Matrice de covariance  $R_y$  en  $16^2 \times 16^2$  issue d'une source composée de trois mesures de Dirac. (b) Dual  $\Lambda^* R_y$  en  $16 \times 16$ .

**Attention** — On pourrait aussi définir des opérateurs de cumulants d'ordre  $n$  c'est-à-dire  $\Lambda^n(m) = \langle \phi^n, m \rangle_{\mathcal{M}(X)}$  avec le noyau  $\phi^n((x_1, \dots, x_n)) = h(\cdot - x^1) \dots h(\cdot - x^n)$ . Il convient ceci étant de garder à l'esprit que la taille de l'espace d'observation augmente considérablement avec l'ordre  $n$ . Si la pile est constituée d'images  $L \times L$ , la covariance spatiale sera de taille  $L^2 \times L^2$ . Par extension, le cumulatif d'ordre  $n$  sera lui de taille  $L^n \times L^n$ , ce qui peut vite devenir rédhibitoire pour un calcul en un temps « raisonnable

». Le cumulants d'ordre 2 semble offrir un bon compromis entre résolution spatiale et temps de calcul.

On rappelle que  $\mathbf{R}_y$  se définit pour l'acquisition dynamique  $\mathbf{y}$  par :

$$\begin{aligned} \mathbf{R}_y : L^2(0, T, L^2(\mathcal{X})) &\longrightarrow L^2(\mathcal{X}^2) \\ \mathbf{y} &\longmapsto \frac{1}{T} \int_0^T (\mathbf{y}(u, t) - \bar{\mathbf{y}}(u)) (\mathbf{y}(v, t) - \bar{\mathbf{y}}(v)) dt. \end{aligned}$$

Dans le cas sans-bruit, il est évident que l'on peut faire une équivalence  $\Lambda(m_{M_0, x_0}) = \mathbf{R}_y$ . Malheureusement, dans la réalité, la pile d'acquisition va être bruitée, notamment par un bruit de fond (qui modélise les fluorophores *out-of-focus*) et un bruit gaussien. Le bruit de fond sera supposé constant dans la suite de ce chapitre. Considérons maintenant la fonctionnelle qui va permettre de retrouver les positions des fluorophores, en reconstruisant la mesure  $m_{M_0, x_0}$  à partir de  $\mathbf{R}_y$ .

### Formulation variationnelle sur covariance

On s'inspire maintenant de la formulation de Colorme pour écrire notre formulation variationnelle. Pour  $\lambda > 0$  paramètre de régularisation :

$$\operatorname{argmin}_{m \in \mathcal{M}(\mathcal{X})} T_\lambda(m) \stackrel{\text{def.}}{=} \frac{1}{2} \|\mathbf{R}_y - \Lambda(m)\|_{\mathcal{H}}^2 + \lambda |m|(\mathcal{X}) \quad (\mathcal{Q}_\lambda(\mathbf{y}))$$

avec  $\mathbf{R}_y$  l'observation, qui est la covariance spatiale de la pile d'acquisition  $\mathbf{y} \in L^2(\mathcal{X}^2)$  obtenue expérimentalement. En minimisant cette fonctionnelle, nous allons obtenir une mesure stationnaire  $m_{M, x} = \sum_{i=1}^N M_i \delta_{x_i}$  : nous retrouverons donc les positions et les variances temporelles de luminosité  $M_i$ . De toute évidence, il n'est pas possible de reconstruire les luminosités  $a_i(t)$  avec cette méthode, nous ne pouvons retrouver que la variance temporelle  $M_i$  pour chaque amplitude  $a_i(t)$ . Nous pourrions à la rigueur envisager une étape de post-traitement où l'on viendrait estimer les luminosités, à positions des  $\delta$ -pics fixés.

L'existence de solutions et l'application des résultats détaillés au chapitre 1 est une adaptation de [43, 35]. Plus généralement, les résultats concernant les mesures de Dirac que nous avons décrits au chapitre 1 sont transposables à  $(\mathcal{Q}_\lambda(\mathbf{y}))$ , grâce à la régularité de  $\Lambda$ . En outre, nous pouvons définir des certificats duaux pour le problème  $(\mathcal{Q}_\lambda(\mathbf{y}))$  : comparé à un BLASSO formulé sur la moyenne  $\bar{\mathbf{y}}$ , on constate empiriquement que la PSF est réduite (d'un facteur  $\sqrt{2}$  [36]). De surcroît,  $\mathbf{R}_y$  est défini en retranchant la moyenne temporelle, ce qui permet d'évacuer la contribution du bruit de fond à  $\mathbf{R}_y$  et au certificat dual.

**Attention** — Si  $\mathbf{y}$  est une pile d’images  $L \times L$ , la covariance spatiale  $\mathbf{R}_y$  sera en  $L^2 \times L^2$ . Un certificat dual de  $(\mathcal{Q}_\lambda(\mathbf{y}))$  vit dans  $\mathcal{C}(\mathcal{X})$ , il pourrait donc être calculé sur une grille arbitraire. On prendra en général la même grille que celle des images de la pile, de taille  $L \times L$ .

Pour synthétiser les relations entre les quantités que nous avons définies plus haut, nous présentons un diagramme en figure 2.3.

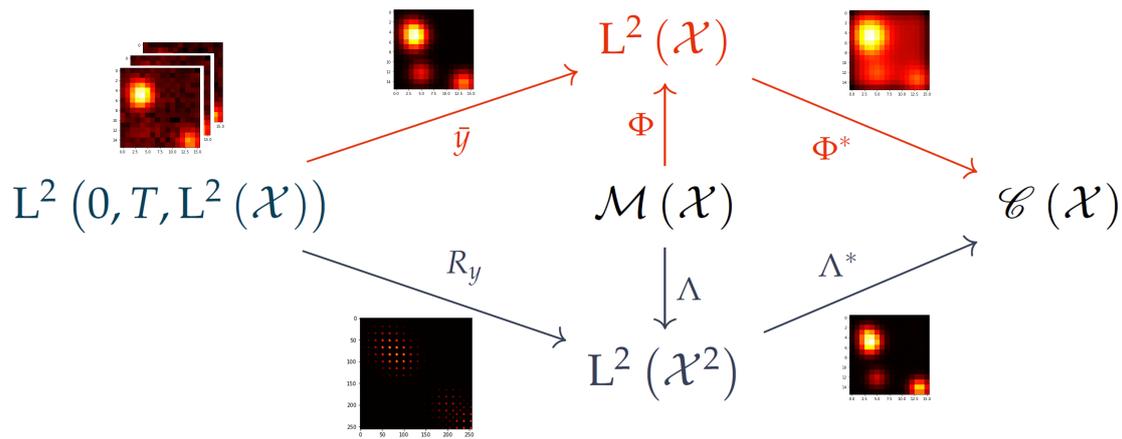


FIGURE 2.3 – Schéma synthétisant les quantités et relations entre les ensembles de travail. À partir d’une pile d’acquisition  $\mathbf{y}$ , on peut calculer sa moyenne temporelle  $\bar{\mathbf{y}}$  ou sa covariance  $\mathbf{R}_y$ , et relier ces deux quantités à des mesures de Radon « classiques » vivant dans  $\mathcal{M}(\mathcal{X})$ . On peut dans les deux cas calculer un certificat, on observe notamment que celui associé à la covariance est moins flouté et bruité.

**Résumé** — Une adaptation du BLASSO peut être formulée pour l’observation  $\mathbf{R}_y$  et l’opérateur  $\Lambda$ . Les garanties théoriques obtenues au chapitre précédent peuvent ainsi être appliquées dans ce cadre de travail en microscopie de fluctuations.

Le problème variationnel ainsi défini, nous pouvons chercher à mettre en œuvre numériquement l’optimisation de  $(\mathcal{Q}_\lambda(\mathbf{y}))$ . Ce problème variationnel étant en pratique un BLASSO formulé avec l’observation  $\mathbf{R}_y$  et l’opérateur  $\Lambda$ , la mise en œuvre numérique sera sensiblement similaire à celles présentées dans le chapitre 1. En particulier, nous utiliserons le *Sliding Frank-Wolfe* présenté en algorithme 2.

## 2.4 Résultats numériques

Dans cette section, nous présentons dans un premier temps les résultats obtenus sur des données synthétiques, pour pouvoir les comparer avec la vérité-terrain. Nous testerons ensuite notre méthode sur des données réelles, à savoir une pile d’images de microscopie de fluorescence produite pour un défi EPFL.

### 2.4.1 Données simulées

Considérons d’abord le bruit à prendre en compte dans notre modèle d’acquisition : dans ce contexte optique, l’observation s’écrirait  $t \in [0, T]$  p.p. :  $\mathbf{y}_t \stackrel{\text{def.}}{=} \mathcal{P}(\Phi\boldsymbol{\mu} + \mathbf{b}) + \mathbf{w}(t)$  pour  $t \in [0, T]$  p.p., avec  $\mathcal{P}$  un bruit multiplicatif de Poisson (qui ne sera pas pris en compte dans notre reconstruction),  $\mathbf{w}$  un bruit blanc gaussien et  $\mathbf{b}$  le bruit de fond.

Nous réalisons un premier test de notre méthode sur des structures filamenteuses, issues de la pile d’images du Challenge SMLM 2016<sup>2</sup>. Plus précisément, nous utilisons les positions de la vérité-terrain de ce *Challenge* pour initialiser le logiciel `SOFITool` [50] qui permet de simuler le processus d’acquisition : fluctuation de luminosité, flou, bruits, etc. Nous générons  $T = 1000$  images en  $64 \times 64$ , pour 100 images acquises par seconde simulées avec `SOFITool` : il permet de modéliser efficacement le phénomène de fluctuations de luminosité des fluorophores à partir d’une liste de position des émetteurs. Les réglages sont ainsi :

- 8700 fluorophores recouvrant les structures tubulaires ;
- la largeur à mi-hauteur de la PSF, ou FWHM<sup>3</sup> (*Full Width Half Maximum*), est de 229 nm ;
- la densité d’émetteurs est fixée à 10,7 émetteurs/pixel/image ;
- les paramètres de fluctuations dans `SOFITool` sont choisis à 5120ms pour le temps d’éclairage du fluorophore, 40 ms pour la simulation du photo-blanchiment et 20 s en moyenne avant que le fluorophore se décolore complètement [86].

La pile d’acquisition inclut aussi en moyenne 1000 photons/image par molécule et 100 photons/image par pixel, afin de simuler les molécules qui ne sont pas dans le foyer de la lentille du microscope, donc floues : ces émetteurs seront traités comme du bruit de fond. Le bruit de fond est simulé à 20 db ; pour conclure le SNR est d’environ 10 db. La solution obtenue en minimisant  $(Q_\lambda(\mathbf{y}))$  est comparée à la vérité-terrain, ainsi qu’à la première image de la pile simulée et à la reconstruction proposée par SRRF [27].

On peut observer que, contrairement à SRRF, notre méthode est capable de récupérer des détails très fins sur l’entrelacement des filaments, malgré le niveau de bruit, ce qui est assez difficile à réaliser par les méthodes de l’état de l’art. De plus, elle ne crée pas d’artefacts (contrairement au fond rouge de SRRF), calcule l’image super-résolue même sur les bords (comparer la (c) et (d) de la figure 2.4), s’avère proche du tracé des structures tubulaires de la vérité-terrain et ne nécessite qu’un seul paramètre de réglage : le réel positif  $\lambda$ . Enfin, contrairement à des approches d’apprentissage profond, notre algorithme n’a pas besoin d’un ensemble de données d’apprentissage et peut-être appliqué directement à n’importe quelle pile dont l’opérateur direct est connu. Ceci étant, il sera assez clair dans la suite que les résultats obtenus par notre algorithme ne reproduisent pas la structure courbe de la vérité-terrain, ce que SRRF parvient par exemple mieux à rendre.

2. <https://srm.epfl.ch/Datasets>

3.  $\text{FWHM} \stackrel{\text{def.}}{=} 2\sqrt{2\ln 2}\sigma$  pour  $\sigma$  l’écart-type de la PSF.

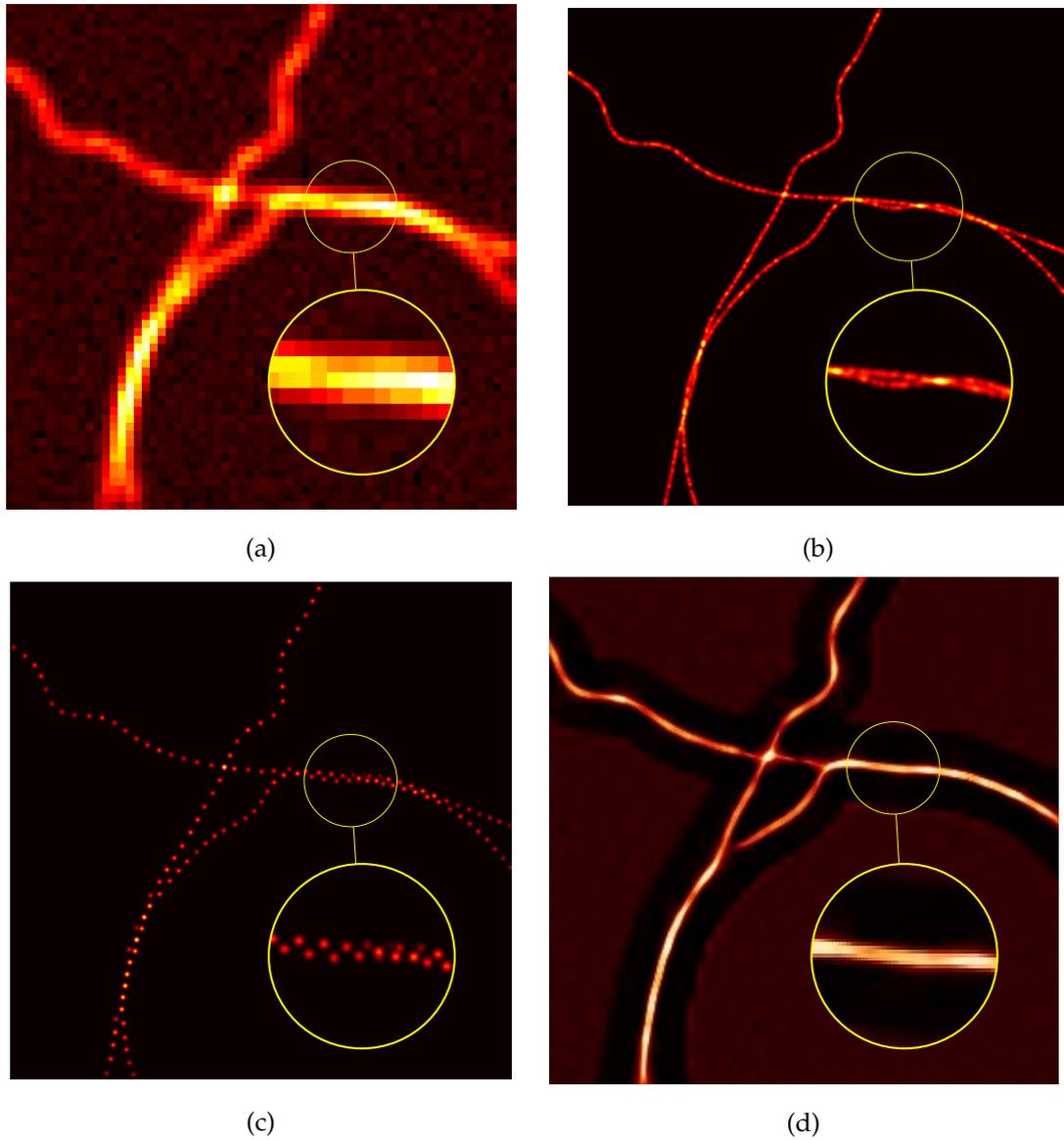


FIGURE 2.4 – Résultats pour une pile d’acquisition simulée  $T = 1000$  : (a) premier extrait de la pile  $y$  (b) vérité-terrain (c) Reconstruction par  $(\mathcal{Q}_\lambda(y))$  (d) SRRF.

Observons maintenant comment notre algorithme se comporte dans le cadre de données réelles.

### 2.4.2 Données expérimentales

En nous inspirant de SPARCOM [52], nous utilisons une pile d’images à haute densité de fluorophores, acquise dans le cadre d’un protocole SMLM. Bien que les luminosités des fluorophores en SMLM ne présentent pas de fluctuation/clignotement (*blinking*),

mais plutôt un comportement *on* et *off*, nous pouvons prendre cette approximation en notant qu'il y a plusieurs molécules par pixel, ce qui se traduit concrètement par une fluctuation de luminosité dans chaque pixel. Il est donc possible de considérer la pile SMLM acquise comme une série temporelle de fluorophores aux amplitudes fluctuantes. La pile acquise est composée de  $T = 500$  images, floutées par une PSF de FWHM de 351,8 nm. Les résultats sont présentés dans la figure 2.5.

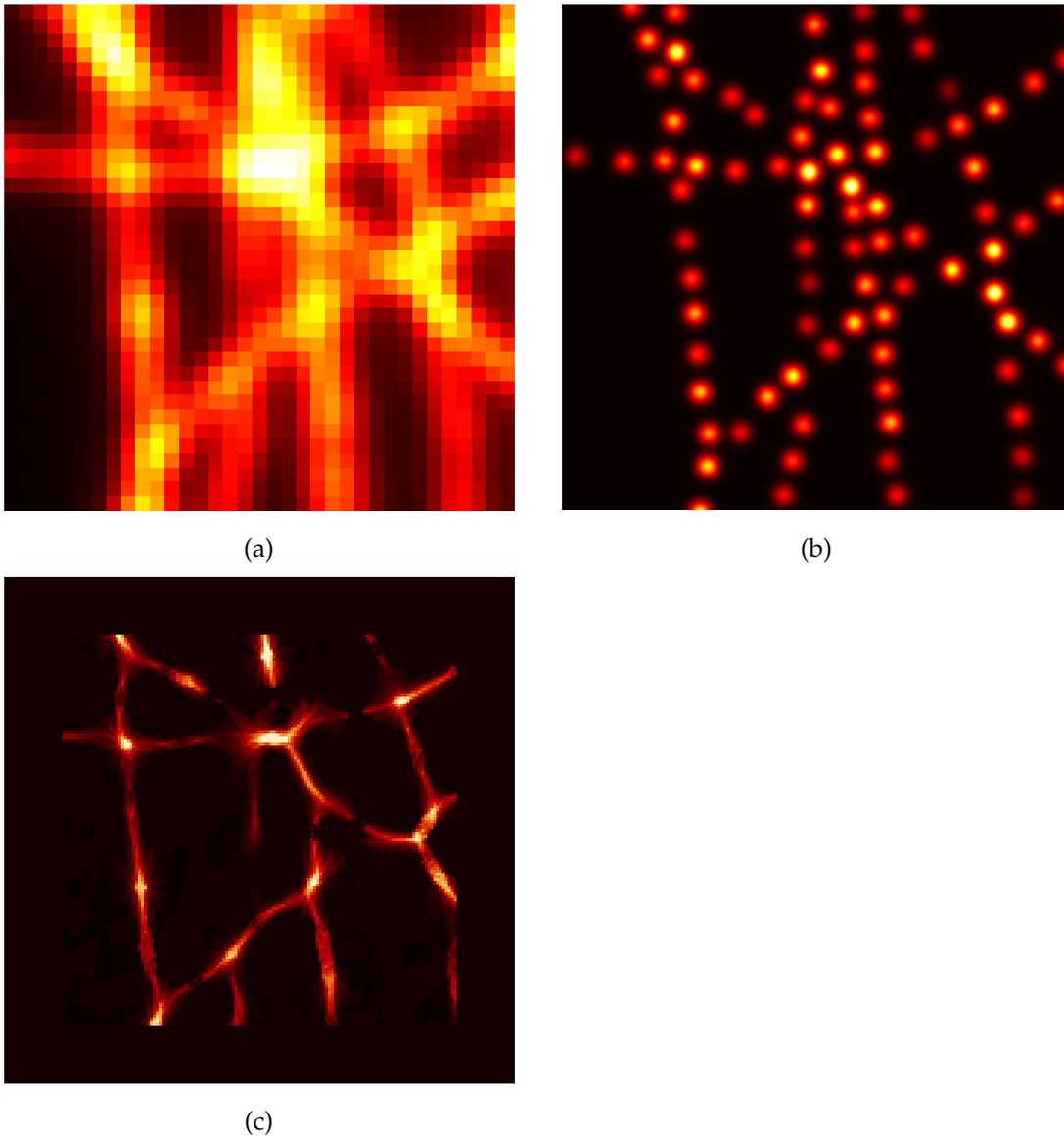


FIGURE 2.5 – Résultats pour des données réelles avec  $T = 500$  : (a) moyenne de la pile d'acquisition  $\bar{y}$  (b) solution de  $(Q_\lambda(y))$  (c) SRRF.

La reconstruction semble un peu en pointillé, ce qui est d'autant plus flagrant avec

la FWHM plus grande, ici à 351,8 nm contre 229 nm auparavant. Elle parvient cela dit à reconstruire une nouvelle fois la structure de l'entrelacement des filaments. Il convient de noter que nous reconstruisons des pics proches au maximum de  $\sigma/\sqrt{2}$  avec  $(\mathcal{Q}_\lambda(\mathbf{y}))$ , pour  $\sigma$  l'écart-type de la PSF.

### 2.4.3 Des reconstructions en pointillé

Même motif, même punition : les solutions de  $(\mathcal{Q}_\lambda(\mathbf{y}))$  sont toujours en pointillé, et ne collent pas avec la structure courbe des filaments que l'on cherche à reconstruire à cause de la modélisation choisie ici. C'est un problème récurrent dans d'autres types d'approches cherchant à exploiter ces fluctuations [83, 85].

**Attention** — Les structures imagées dans le cadre de la microscopie de fluorescence ont en pratique un aspect « guirlande de Noël ». Si la structure sous-jacente est effectivement courbe, les variations de luminosité produisent des « trous » dans le filament à imager, de sorte que les reconstructions produisent plutôt une courbe avec des discontinuités.

Ce résultat est à vrai dire plutôt cohérent : la fonctionnelle  $(\mathcal{Q}_\lambda(\mathbf{y}))$  est en fait une variante du BLASSO, adaptée à un opérateur  $\Lambda$  formulé sur un espace de matrices de covariance. Comme nous le verrons dans le chapitre 3 qui suit, la norme TV  $\|\cdot\|_{\text{TV}}$  qui régularise le BLASSO impose un *a priori* structurel sur la solution de cette énergie. En particulier, elle prescrit la géométrie de la reconstruction, ce qui est d'autant plus flagrant lorsqu'elle est couplée avec l'algorithme du gradient conditionnel, qui reconstruit itérativement une somme pondérée de points extrémaux. La solution finalement obtenue est donc *nécessairement* une combinaison linéaire de sources ponctuelles, ce qui donne cet aspect pointillé.

On peut d'ailleurs observer le même effet pour les méthodes discrètes. Comme nous l'avons évoqué plus haut,  $(\mathcal{Q}_\lambda(\mathbf{y}))$  est la version sans-grille de  $\text{Col}_0\text{rme}$  régularisé avec la norme  $\ell_1$  [86]. Une comparaison visuelle est ainsi proposée en figure 2.6.

Remarquons alors que :

- $(\mathcal{Q}_\lambda(\mathbf{y}))$  est un BLASSO et  $\text{Col}_0\text{rme}$  un LASSO pour les matrices de covariance ;
- le LASSO, en faisant tendre le nombre de points de discrétisation,  $\Gamma$ -converge vers le BLASSO [42] ;

La  $\Gamma$ -convergence implique la convergence des minimiseurs du LASSO, et par extension ceux de  $\text{Col}_0\text{rme}$ , vers ceux du BLASSO et en particulier de  $(\mathcal{Q}_\lambda(\mathbf{y}))$  : en retournant l'argument, on peut s'attendre à ce que les solutions de  $\text{Col}_0\text{rme}$  soient géométriquement proches de celles de notre énergie sans-grille, et c'est ce que l'on constate en pratique.

Il y a donc ici une barrière théorique qui contraint les solutions à un *a priori* ponctuel. Cette limitation a pu être dépassée dans le cas de  $\text{Col}_0\text{rme}$  par des approches *Plug & Play*, où l'on se propose d'apprendre le terme de régularisation à l'aide d'un ensemble d'apprentissage. Dans l'état de l'art des méthodes sans-grille, il n'existe pas encore de

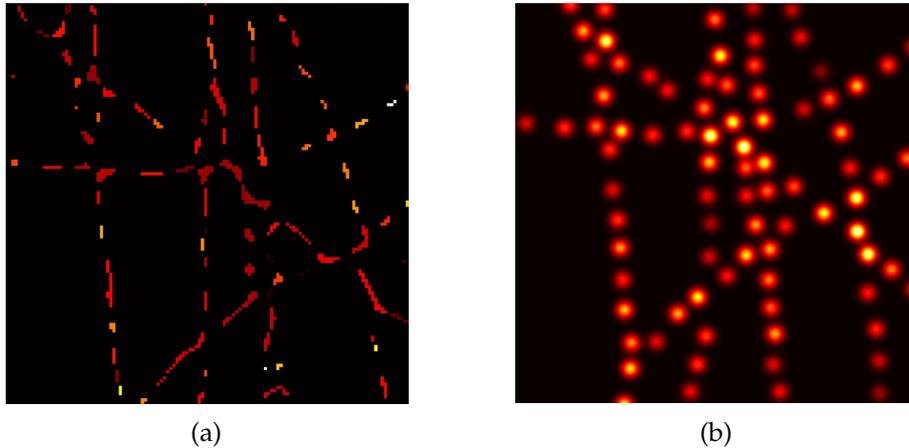


FIGURE 2.6 – Reconstructions produites par (a) Colorm (b) la minimisation de  $(\mathcal{Q}_\lambda(\mathbf{y}))$ .

telles adaptations, *a fortiori* il n’y a pas d’approche variationnelle sans-grille permettant de reconstruire une courbe à partir d’une observation, qu’elle soit une image ou une matrice de covariance.

**Résumé** — La reconstruction par  $(\mathcal{Q}_\lambda(\mathbf{y}))$  semble correctement exploiter des informations temporelles pour reconstruire des détails fins tel que l’enlacement des filaments. Cependant, elle n’est pas totalement satisfaisante car en pointillé : la modélisation choisie ne permet pas de reconstruire des courbes.

## 2.5 Conclusion partielle

Dans ce chapitre, nous avons formulé une modélisation temporelle du cadre d’acquisition en microscopie de fluorescence par fluctuation. La source est une mesure dynamique, combinaison linéaire de mesures de Dirac aux positions fixes mais aux amplitudes fluctuant au cours du temps. À l’aide de l’hypothèse d’indépendance des fluctuations de luminosité des fluorophores, il est possible de construire un opérateur qui fait le lien entre matrice de covariance spatiale et une mesure de Radon sans dépendance temporelle : cette mesure partage les mêmes positions que la mesure source. Cette méthode sans-grille n’est pas propre à la matrice de covariance, elle peut être naturellement étendue aux cumulants statistiques d’ordre supérieur. L’approche par covariance offre un bon compromis entre qualité de reconstruction et temps de calcul. Les solutions de  $(\mathcal{Q}_\lambda(\mathbf{y}))$  produites à partir de données synthétiques et expérimentales illustrent la capacité de reconstruction des détails fins par cette méthode sans-grille, permettant par exemple de déduire un enlacement de filaments. Les résultats sont néanmoins en pointillé, ce qui n’est que la conséquence logique de la régularisation, qui promeut la géométrie ponctuelle. Eu égard à cette limite théorique, il n’est pas possible en pratique

de reconstruire des solutions exhibant une géométrie courbe, qui constitue pourtant une part importante des structures à imager dans ce type de microscopie de fluorescence.

PARTIE II

---

RECONSTRUCTION SANS-GRILLE DE MESURES COURBES : THÉORIE ET  
APPLICATIONS



## Chapitre 3

# Un nouveau régulariseur pour les mesures courbes

**Points-clefs : la littérature ne traite pas la reconstruction de mesures 1-rectifiables.**

1. Le cas des mesures ponctuelles, ou **0-rectifiable** a été **abondamment étudié**, comme nous l'avons détaillé en chapitre 1.
2. De récents travaux ont proposé la reconstruction de mesures **2-rectifiables**, ou autrement dit portées par des surfaces. Des reconstructions de courbes, au sens de trajectoires suivies par des mesures de Dirac, ont aussi été étudiées. Elles impliquent de reconstruire des sources ponctuelles se déplaçant au cours du temps, donc une dimension temporelle **inadaptée à notre cadre** de travail.
3. La reconstruction à partir d'une seule image de mesures portées par des courbes, ou 1-rectifiable, restait encore à explorer.

**Contributions – un cadre théorique pour la reconstruction sans-grille de courbes.**

4. Nous présentons **quelques pistes** de recherche que nous nous sommes proposés de suivre et qui **n'ont pas abouti**. Nous expliquons en particulier pourquoi elles se sont révélées infructueuses et en quoi elles ne sont pas pertinentes pour le problème des courbes sans-grille.
5. Nous définissons un **espace de mesures**  $\mathcal{M}$  dans lequel vivent nos **mesures portées par des courbes**, et nous prouvons quelques propriétés algébriques qu'il exhibe.
6. Nous présentons une **nouvelle fonctionnelle** appelée CROC, qui repose sur une régularisation par la norme  $\|\cdot\|_{\mathcal{M}}$ . Nous prouvons les **conditions d'optimalité**, ainsi que les **certificats associés**.
7. Nous introduirons un théorème énonçant la **structure géométrique des points extrémaux** de la boule unité de  $\|\cdot\|_{\mathcal{M}}$ . Nous prouvons ce théorème et par là même fournissons une démarche pour reconstruire des mesures portées par des courbes.

**Chapitre 3 – Un nouveau régulariseur pour les mesures courbes :**

3.1	Préliminaires . . . . .	56
3.1.1	Dérivées fractionnaires . . . . .	57
3.1.2	Transport optimal, <i>shortest paths</i> , . . . . .	59
3.1.3	Théorème de la représentation : régulariseur et points extrémaux	59
3.2	L'espace des charges ou des champs de vecteurs à divergence finie . .	62
3.3	Un problème variationnel sur l'espace des charges . . . . .	66
3.4	Les points extrémaux de la boule unité de la norme $\mathcal{V}$ . . . . .	71
3.5	Conclusion partielle . . . . .	78

**3.1 Préliminaires**

Dans le chapitre précédent, nous avons observé que l'utilisation du modèle sans-grille classique faisant intervenir le (BLASSO) était inadapté pour l'utilisation que nous en avons. En effet, contrairement au cadre SMLM où l'on cherchait à localiser des sources ponctuelles, la fluctuation d'amplitude nécessiterait plutôt de localiser des courbes. C'est même de manière générale une structure que nous sommes susceptibles de rencontrer en imagerie biomédicale : on peut penser naturellement aux vaisseaux sanguins, aux filaments situés dans le cytosquelette des organismes, *etc.* L'utilisation du BLASSO, et en particulier du régulariseur  $\|\cdot\|_{TV}$ , ne permet pourtant pas de reconstruire des courbes et ne peut produire qu'une combinaison de mesures de Dirac positionnées le long de la courbe à reconstruire : comme nous le verrons plus loin, cette limitation est inhérente à la régularisation  $\|\cdot\|_{TV}$ .

De récents travaux [28] ont remis au goût du jour les résultats obtenus en théorie géométrique de la mesure<sup>1</sup>, dans l'optique de la reconstruction sans-grille des mesures portées par des surfaces : comprendre, des mesures dont le support est 2-rectifiable. La méthode repose sur l'espace des fonctions à variation bornée  $BV(\mathcal{X})$  en dimension deux, à savoir les classes d'équivalence  $u$  telles que le gradient au sens des distributions  $Du$  soit une mesure :

$$BV(\mathcal{X}) \stackrel{\text{def.}}{=} \left\{ u \in L^2(\mathcal{X}), Du \in \mathcal{M}(\mathcal{X})^2 \right\}.$$

À l'aide de la régularisation par la variation totale du gradient, les auteurs ont été capables de reconstruire *de manière sans-grille* des mesures portées par des ensembles 2D avec la fonctionnelle suivante :

$$\operatorname{argmin}_{u \in BV(\mathcal{X})} \frac{1}{2} \|y - \Phi u\|_{\mathcal{H}}^2 + \lambda \|Du\|_{TV^2}. \quad (3.1)$$

Ainsi, l'état de l'art en sans-grille distingue :

1. Dans la suite du propos, nous utiliserons des notions issues de la théorie géométrique de la mesure, et nous invitons le lecteur à se reporter à l'annexe 4 pour plus de détails.

- la régularisation  $\|m\|_{\text{TV}}$  pour une mesure  $m$ , qui permet la reconstruction de sources ponctuelles, autrement dit de mesures 0-rectifiables ;
- la régularisation  $\|Du\|_{\text{TV}^2}$  pour  $u \in \text{BV}(\mathcal{X})$  qui fournit la reconstruction de sources portées par des ensembles, ou 2-rectifiables.

Il y a donc un fossé théorique entre les points et les ensembles qu'il faudrait combler pour pouvoir reconstruire des mesures 1-rectifiables, ou autrement dit portées par des courbes. Nous nous proposons dans la suite du propos de résumer brièvement les approches que nous avons essayées d'adopter pour reconstruire des courbes, et d'expliquer en particulier pourquoi elles se sont révélées infructueuses. Puis, nous nous attacherons à expliquer l'approche que nous avons finalement retenue, qui consiste en un nouvel espace de mesures, une nouvelle fonctionnelle et un nouveau régulariseur qui permettent effectivement de reconstruire des mesures portées par des courbes. Ce chapitre traitera uniquement du cadre théorique, les problématiques liées à la mise en œuvre concrète du problème seront détaillées dans le chapitre 4 qui suit.

### 3.1.1 Dérivées fractionnaires

L'idée de la régularisation par la dérivée fractionnaire provient d'une observation assez simple : le régulariseur  $\|m\|_{\text{TV}}$  produit un *minimum* composé de mesures ponctuelles, là où le régulariseur  $\|Du\|_{\text{TV}^2}$  permet d'obtenir un minimiseur formulé avec des ensembles 2D. Les deux font donc intervenir la variation totale<sup>2</sup>, et il paraît raisonnable de chercher le cas des courbes comme *un entre-deux*. Nous pouvons donc proposer la dérivée fractionnaire  $D^\alpha f$ , qui fait le pont entre la dérivée classique et la fonction identité, et qui pourrait donc combler le fossé théorique entre les deux régularisations.

Le réel  $\alpha$  est appelé l'ordre de dérivation, et il nous permet déjà de définir deux cas classiques :

**Cas  $\alpha = 0$**

L'espace d'optimisation est ici  $\mathcal{M}(\mathcal{X})$ . Le régulariseur s'écrit pour  $m \in \mathcal{M}(\mathcal{X})$  :

$$|m|(\mathcal{X}) \stackrel{\text{def.}}{=} \sup \left( \int_{\mathcal{X}} f \, dm, f \in \mathcal{C}_0(\mathcal{X}), \|f\|_{\infty, \mathcal{X}} \leq 1 \right).$$

**Cas  $\alpha = 1$**

L'espace d'optimisation est alors  $E = L^{d/(d-1)}(\mathcal{X})$ . Le régulariseur s'écrit pour  $u \in E$  :

$$\text{TGV}(u) \stackrel{\text{def.}}{=} \sup \left( \int_{\mathcal{X}} u \, \text{div}(\psi) \, dx, \psi \in \mathcal{C}_c^1(\mathcal{X}, \mathbb{R}^n), \|\psi\|_{\infty, \mathcal{X}} \leq 1 \right).$$

---

2. d'où l'ambiguïté pour les chercheurs des communautés des mesures et du traitement du signal.

**Cas de la dérivée fractionnaire**  $0 < \alpha < 1$ 

Soit  $\alpha \in \mathbb{R}^+$  l'ordre de dérivation, on suppose que  $\alpha \in [0, 1]$ . Il existe en réalité plusieurs manières de définir les dérivées fractionnaires (Riemann-Liouville, Grünwald-Letnikov, Caputo, by Fourier, *etc.*), nous utiliserons ici la notion développée dans les travaux de [25, 24], puisque le cadre qu'ils ont formulé présentent quelques similarités avec  $BV(\mathcal{X})$ . Une classe d'équivalence  $f \in L^1(\mathcal{X})$  appartient à l'espace  $BV^\alpha(\mathcal{X})$  si sa  $\alpha$ -dérivée fractionnaire :

$$|D^\alpha f|(\mathcal{X}) \stackrel{\text{def.}}{=} \sup \left( \int_{\mathcal{X}} f \operatorname{div}^\alpha \psi \, dx, \psi \in \mathcal{D}(\mathcal{X}), \|\psi\|_{\infty, \mathcal{X}} \leq 1 \right)$$

est finie.  $\operatorname{div}^\alpha$  est la divergence  $\alpha$ -fractionnaire, définie par

$$\operatorname{div}^\alpha f(y) \stackrel{\text{def.}}{=} a_{d, \alpha} \int_{\mathcal{X}} \frac{(y-x) \cdot (f(x) - f(y))}{|y-x|^{n+\alpha+1}} \, dx,$$

avec la constante

$$a_{d, \alpha} \stackrel{\text{def.}}{=} 2^\alpha \pi^{-n/2} \frac{\Gamma(\frac{n+\alpha}{2})}{\Gamma(\frac{-\alpha}{2})}$$

pour  $\Gamma$  la fonction gamma, qui étend la notion de factorielle. Comme nous l'avons évoqué, l'espace  $BV^\alpha(\mathcal{X})$  partage quelques propriétés avec l'espace  $BV(\mathcal{X})$  classique [24], c'est par exemple un espace de Banach muni de la norme :

$$\|u\|_{BV^\alpha(\mathcal{X})} \stackrel{\text{def.}}{=} \|u\|_{L^1(\mathcal{X})} + |D^\alpha u|(\mathcal{X}).$$

Cependant, des propriétés essentielles de  $BV(\mathcal{X})$  comme la formule de la co-aire ne sont plus valables dans ce cadre. Quoiqu'il en soit, il serait naturel de se demander si l'intuition suivante se vérifie en pratique : *Si l'on a un ensemble  $E$  de dimension d'Hausdorff  $2\alpha$ , peut-on le reconstruire grâce à la variation  $\alpha$ -fractionnaire ?* Nous pourrions ainsi espérer que le régulariseur  $|D^\alpha u|(\mathcal{X})$  ait pour *minimum* des fonctions indicatrices portées par un ensemble de même dimension d'Hausdorff  $\alpha$ .

Cette intuition est mise en défaut par un résultat de la littérature récente [23, Proposition 4.1], plus précisément si  $f \in BV^\alpha(\mathcal{X}) \cap \mathcal{L}^p(\mathcal{X})$  alors  $D^\alpha f$  a une décomposition de Lebesgue comme  $Df$ , à une transformation linéaire près appelée le *potentiel de Riesz*. Cette dernière propriété justifie l'absence d'une formule de la co-aire dans ce contexte. Dit autrement, l'impossibilité d'écrire une formule de la co-aire suggère que la reconstruction de fonctions à partir de leurs ensembles de niveaux ne fonctionnent pas pour la variation fractionnaire, au contraire de la régularisation dans  $BV(\mathcal{X})$ . Ainsi, même si une fonction dérivable  $\alpha$ -fractionnairement est similaire (dans une certaine mesure)

à une fonction BV, il est également possible que ses ensembles de niveaux aient un comportement « chaotique », compris au sens Cantor, c'est-à-dire être des fonctions caractéristiques avec un comportement critique.<sup>3</sup>

**Résumé** — Les dérivées fractionnaires ne sont visiblement d'aucune utilité pour la reconstruction des courbes. Elles ne font qu'imiter le comportement classique de la régularisation des fonctions à variations bornées  $\|D\cdot\|_{TV^2}$ .

### 3.1.2 Transport optimal, *shortest paths*, ...

De récents travaux [12, 11, 44] ont proposé la reconstruction sans-grille de mesures de Dirac dynamiques, comprises au sens de sources ponctuelles qui se déplaceraient au cours du temps. Cette approche permet effectivement de retrouver des *chemins*, dit autrement les courbes suivies par les mesures de Dirac, grâce au régulariseur formulé avec l'énergie de Benamou-Brenier [10]. Néanmoins, ces méthodes sans-grille reposent sur un ensemble de mesures sur le cylindre espace temps  $\mathcal{M}(\mathcal{X} \times [0, 1])$  : nous ne disposons pas d'une telle composante temporelle, dans le sens où nous cherchons pour le moment à reconstruire une courbe à partir **d'une image**, et pas d'une suite d'images. Le plongement proposé est ceci étant tout à fait pertinent, à savoir que la littérature met en évidence le lien entre  $\mathcal{M}(\mathcal{X} \times [0, 1])$  et  $\mathcal{M}(\Gamma)$ , pour :

$$\Gamma \stackrel{\text{def.}}{=} \{\gamma = (h, \xi); h \in \mathcal{C}([0, 1]), \xi : [0, 1] \rightarrow \mathcal{X} \text{ est continu}\}$$

les mesures sur des courbes d'amplitude  $h$  et de position  $\xi$ . Il n'y a en revanche pas de telle identification entre  $\mathcal{M}(\mathcal{X})$  et  $\mathcal{M}(\Gamma)$ , en tout cas en l'état actuel.

Les approches riemanniennes développées entre autres par [19, 71] sont également très intéressantes, notamment par leurs résultats pratiques convaincants. Elles posent en revanche quelques difficultés pour une adaptation sans-grille : il est par exemple nécessaire de définir les extrémités de la courbe, ce qui fait perdre un peu l'intérêt de l'approche variationnelle qui reconstruit les sources à partir d'une observation, avec seulement un paramètre de régularisation à régler.

Nous proposons dans la suite une autre manière de procéder, qui va nous permettre de définir un régulariseur qui prescrit la structure de courbes.

### 3.1.3 Théorème de la représentation : régulariseur et points extrémaux

Dans les sous-sections précédentes, nous avons cherché à formuler un nouveau problème variationnel *sans-grille* permettant de reconstruire des courbes : dit autrement, nous cherchons une nouvelle énergie dont (au moins) l'un des *minimum* serait une somme de courbes. Les approches que nous avons présentées plus haut jouent ainsi sur

3. D'après une correspondance avec les auteurs de [23] que nous remercions pour leurs apports au présent document.

le *régulariseur* pour modifier la structure du *minimum*, et pas sur le terme d'attache aux données.

Pour expliquer l'influence du régulariseur sur la géométrie des solutions, nous avons besoin de présenter quelques définitions et résultats concernant la notion de point extrémal.

**Définition 3.1** (Points extrémaux). *Soit  $X$  un espace vectoriel topologique et  $K \subset X$ . Un point extrémal  $x$  de  $K$  est un point tel que :*

$$\forall y, z \in K, \quad \forall \lambda \in (0, 1), \quad x = \lambda y + (1 - \lambda)z \implies x = y = z.$$

L'ensemble des points extrémaux de  $K$  est noté  $\text{Ext } K$ .

Nous présentons maintenant le théorème de Krein-Milman, qui précise que pour  $K$  convexe et compact, la fermeture de l'enveloppe convexe de  $\text{Ext } K$  coïncide avec  $K$ .

**Théorème 3.1** (Théorème de Krein-Milman). *Si  $K \subset X$  est un sous-ensemble compact, convexe et non-vide d'un espace d'Hausdorff localement convexe, alors  $K = \overline{\text{co}}(\text{Ext}(K))$ .*

Notons également le raffinement suivant du théorème de Krein-Milman en (3.1), énonçant la décomposition de tout point de l'espace en une combinaison convexe de points extrémaux, grâce à la notion [76] d'intégrale de Choquet.

**Théorème 3.2** (Théorème de Choquet). *Soit  $X$  un espace vectoriel topologique métrisable et  $K$  un sous-ensemble convexe, compact et non-vide. Alors, pour tout  $x \in K$  il existe une mesure de Borel  $\rho$  sur  $X$ , concentrée sur  $\text{Ext } K$  qui vérifie :*

$$x = \int_{\text{Ext } K} r \, d\rho(r)$$

*c'est-à-dire que pour tout  $\omega : X \rightarrow \mathbb{R}$  linéaire et continue, il vient  $\omega(x) = \int_{\text{Ext } K} \omega(r) \, d\rho(r)$ .*

Le théorème de Krein-Milman a suggéré le développement d'une catégorie de théorème dit *de la représentation*. Un théorème de la représentation fournit deux informations essentielles :

- même si l'énergie est formulée sur un espace de dimension infinie, on peut trouver un *minimum* qui est une somme **finie** d'éléments de l'espace d'optimisation ;
- ces éléments, appelés *atomes*, présentent une structure particulière dont la géométrie est totalement déterminée par le terme du régulariseur.

Plus précisément, dans le cadre variationnel, soit  $n \in \mathbb{N}^*$ ,  $\mathcal{H}_n$  l'espace des observations de dimension **finie**  $n$ ,  $G : \mathcal{H}_n \rightarrow \mathbb{R}$  une fonction, un terme arbitraire d'attache aux données. L'espace d'optimisation est dénoté  $E$ , supposons que  $\Lambda : E \rightarrow \mathcal{H}_n$  est linéaire et  $F : E \rightarrow \mathbb{R}$  est convexe. Considérons une fonctionnelle définie pour  $m \in E$  :

$$J(\mathbf{m}) \stackrel{\text{def.}}{=} G(\Lambda \mathbf{m}) + F(\mathbf{m}). \quad (3.2)$$

La boule unité du régulariseur  $F$  est notée  $\mathcal{B}_F \stackrel{\text{def.}}{=} \{\mathbf{u} \in E, F(\mathbf{u}) \leq 1\}$ . Le théorème qui suit, issu de deux articles proposant le même résultat mais par deux approches différentes [8, 9], établit le lien entre les minimiseurs de la fonctionnelle  $J$  dans (3.2) et les points extrémaux de la boule unité  $\mathcal{B}_F$  noté  $\text{Ext}(\mathcal{B}_F)$  :

**Théorème 3.3** (Théorème de la représentation). *Il existe  $\bar{\mathbf{u}} \in E$ , un minimiseur de (3.2) avec la représentation :*

$$\bar{\mathbf{u}} = \sum_{i=1}^p \alpha_i \mathbf{u}_i$$

où  $p \leq \dim \mathcal{H}_n$ ,  $\mathbf{u}_i \in \text{Ext}(\mathcal{B}_F)$  et  $\alpha_i > 0$  avec  $\sum_{i=1}^p \alpha_i = F(\bar{\mathbf{u}})$ .

Cela revient à caractériser les propriétés structurelles des minimiseurs, sans avoir réellement besoin de résoudre le problème variationnel [8]. De plus, cette description précise des points extrémaux de la boule unité du régularisateur est fondamentale pour l'implémentation numérique. En effet, l'algorithme de Frank-Wolfe construit une solution du problème variationnel en ajoutant itérativement des points extrémaux de la boule unité<sup>4</sup> de  $F$  à la mesure reconstruite : à chaque étape, l'algorithme ajoute un nouvel atome à la mesure reconstruite, ce nouvel élément a une structure prescrite puisqu'il s'agit d'un point extrémal du régularisateur.

Ainsi, si l'on reprend les énergies évoquées plus haut [9] :

— le BLASSO sur  $E = \mathcal{M}(\mathcal{X})$  est régularisé avec  $R = \|\cdot\|_{\text{TV}}$  et

$$\text{Ext}(\mathcal{B}_{\mathcal{M}}) = \{\pm \delta_x, x \in \mathcal{X}\}.$$

— le BLASSO adapté sur  $E = \text{BV}(\mathcal{X})$  est régularisé par  $R = \|\cdot\|_{\text{BV}}$  et :

$$\text{Ext}(\mathcal{B}_{\text{BV}}) = \left\{ \frac{1}{\text{Per}(E)} \chi_E, E \subset \mathcal{X} \text{ est un ensemble simple} \right\}.$$

— que peut-on dire d'une énergie sur un espace  $E$  avec un régulariseur  $R$  qui prescrive la structure de courbe ?

La littérature ne donne pas d'information sur ce dernier point : ceci étant, nous savons que nous cherchons à reconstruire des mesures portées par des courbes, donc une mesure 1-rectifiable c'est-à-dire une mesure absolument continue par rapport à  $\mathcal{H}_1$ . Un résultat de [81] est particulièrement éclairant à ce sujet. En introduisant l'espace des champs de vecteurs à divergence finie, l'article montre que toute mesure à divergence finie s'écrit comme une somme de trois mesures :

4. En réalité, l'algorithme ajoute des points extrémaux d'un ensemble de niveau de  $F$ , mais ce n'est pas problématique dans la mesure où les deux ensembles partageant la même structure de points extrémaux.

- la première mesure est portée par un ensemble 1-rectifiable, elle est absolument continue par rapport à la mesure d’Hausdorff  $\mathcal{H}_1$  avec une densité qui est dans l’espace tangent ;
- la deuxième mesure est la partie Cantor, singulière par rapport à la mesure de Lebesgue et qui s’évanouit sur tous les ensembles de mesure  $\mathcal{H}_1$  finie ;
- la troisième mesure est absolument continue par rapport à la mesure de Lebesgue.

Par conséquent, nous allons étudier l’espace de mesures en question, définir une énergie et montrer qu’un régulariseur sur cet espace prescrit la géométrie de courbes.

**Résumé** — La géométrie d’un minimiseur d’une fonctionnelle donnée est entièrement déterminée par le régulariseur. Cette structure peut être explicitée en déterminant les points extrémaux de la boule unité du régulariseur.

### 3.2 L’espace des charges ou des champs de vecteurs à divergence finie

On suppose dans ce chapitre que  $\mathcal{X}$  est un ouvert borné de  $\mathbb{R}^2$ . Soit l’espace des mesures de Radon *vectérielles* :

**Définition 3.2.** Nous définissons l’ensemble des mesures de Radon *vectérielles*  $\mathcal{M}(\mathcal{X})^2$  comme le dual topologique de l’espace des fonctions *vectérielles* continues  $\mathcal{C}_0(\mathcal{X})^2 \stackrel{\text{def.}}{=} \mathcal{C}_0(\mathcal{X}, \mathbb{R}^2)$ . Les propriétés du cas scalaire sont valables pour le cas *vectériel*, en effet  $\mathcal{M}(\mathcal{X})^2$  présente une norme TV naturelle notée par  $\|\cdot\|_{TV^2}$ , un crochet de dualité  $\langle \cdot, \cdot \rangle_{\mathcal{M}^2}$ , etc.

Nous ne discuterons que du cas en dimension  $d = 2$ , puisque des cas pathologiques surviennent dès que  $d > 2$ , voir [82, Section 1.3]. On note  $\text{div}$  l’opérateur de divergence, qui doit être compris au sens des distributions : pour  $\mathbf{m} \in \mathcal{M}(\mathcal{X})^2$  :

$$\forall \xi \in \mathcal{C}_0^\infty(\mathcal{X}), \quad \langle \text{div } \mathbf{m}, \xi \rangle_{\mathcal{D}'(\mathcal{X}) \times \mathcal{C}_0^\infty(\mathcal{X})} = -\langle \mathbf{m}, \nabla \xi \rangle_{\mathcal{M}^2}.$$

On dit qu’une mesure  $\mathbf{m}$  est à *divergence finie* si  $\text{div}(\mathbf{m}) \in \mathcal{M}(\mathcal{X})$ . Nous pouvons dès lors introduire l’espace suivant [81, 82] essentiel pour définir notre optimisation.

**Définition 3.3** (Espace des charges). On note  $\mathcal{V}$  l’espace des champs de vecteurs à divergence finie, aussi appelés charges<sup>5</sup>, à savoir l’espace des mesures *vectérielles* à divergence finie :

$$\mathcal{V} \stackrel{\text{def.}}{=} \left\{ \mathbf{m} \in \mathcal{M}(\mathcal{X})^2, \text{div}(\mathbf{m}) \in \mathcal{M}(\mathcal{X}) \right\}.$$

Une charge dont la divergence est nulle est appelée un solénoïde.

5. D’après une inspiration physique, Stasnilav Smirnov ayant en effet travaillé sur des problèmes liés à l’électromagnétisme et aux boucles de champs magnétiques.

Nous pouvons déjà préciser quelques propriétés de cet espace.

**Proposition 3.1** (Complétude).  *$\mathcal{V}$  est un espace de Banach muni de la norme  $\|\cdot\|_{\mathcal{V}} \stackrel{\text{def.}}{=} \|\cdot\|_{\text{TV}^2} + \|\text{div}(\cdot)\|_{\text{TV}}$ . Cette norme induit une topologie dite forte dans  $\mathcal{V}$ .*

*Démonstration.* Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de Cauchy de  $\mathcal{V}$ . Par définition, c'est aussi une suite de Cauchy dans  $\mathcal{M}(\mathcal{X})^2$ , qui est complet : par conséquent, elle admet une limite dans  $u \in \mathcal{M}(\mathcal{X})^2$ . Comme  $(u_n)_{n \in \mathbb{N}}$  est bornée (parce que c'est une suite de Cauchy) pour la topologie de la norme dans  $\mathcal{V}$  et par la semi-continuité inférieure de  $\|\text{div}(\cdot)\|_{\text{TV}}$ , alors :

$$\|\text{div} u\|_{\text{TV}} \leq \liminf_{n \rightarrow +\infty} \|\text{div} u_n\|_{\text{TV}} < +\infty$$

et comme  $u \in \mathcal{M}(\mathcal{X})^2$ , il vient  $\|u\|_{\mathcal{V}} = \|\text{div} u\|_{\text{TV}} + \|u\|_{\text{TV}^2} < +\infty$ . Ainsi  $u \in \mathcal{V}$ , étudions maintenant la convergence dans  $\mathcal{V}$ . Comme  $(u_n)_{n \in \mathbb{N}}$  est une suite de Cauchy, il existe  $v \in \mathcal{M}(\mathcal{X})$  tel que  $\text{div} u_n \xrightarrow[n \rightarrow +\infty]{} v$  pour la topologie de la norme TV. Ainsi, à un argument de densité près, pour tout  $\phi \in \mathcal{C}_0^\infty(\mathcal{X})$ , il vient :

$$\langle \text{div} u_n, \phi \rangle \xrightarrow[n \rightarrow +\infty]{} \langle v, \phi \rangle$$

mais par la dérivée au sens des distributions, on a  $\langle \text{div} u_n, \phi \rangle = \langle -u_n, \nabla \phi \rangle$ , et puisque  $\langle u_n, \nabla \phi \rangle \rightarrow \langle u, \nabla \phi \rangle$  il vient  $v = \text{div} u = \lim_{n \rightarrow +\infty} \text{div} u_n$  pour la topologie de la norme TV. Finalement,

$$\|u_n - u\|_{\mathcal{V}} = \|u_n - u\|_{\text{TV}^2} + \|\text{div} u_n - \text{div} u\|_{\text{TV}} \xrightarrow[n \rightarrow +\infty]{} 0,$$

ce qui garantit que  $\mathcal{V}$  est un espace de Banach, et ce qui termine la preuve.  $\square$

La topologie (forte) induite par cette norme de  $\mathcal{V}$  n'est manifestement pas adaptée pour établir des résultats de convergence, elle n'offre en effet aucune propriété sur les suites ou sur les ensembles bornés. Il est par conséquent nécessaire de choisir une topologie avec moins d'ouverts, mais plus de compacts, ce qui permettra d'écrire quelques résultats de compacité :

**Définition 3.4** (Convergence faible-\*).  *$\mathcal{V}$  peut être muni de la topologie faible-\*; on dit qu'une suite de charges  $(m_n)_{n \in \mathbb{N}}$  converge faiblement-\* vers  $m \in \mathcal{V}$ , que l'on note  $m_n \xrightarrow{*} m$ , si :*

$$\begin{aligned} \forall g \in \mathcal{C}_0^\infty(\mathcal{X})^2, \quad \int_{\mathcal{X}} g \, dm_n &\xrightarrow[n \rightarrow +\infty]{} \int_{\mathcal{X}} g \, dm, \\ \forall \psi \in \mathcal{C}_0^\infty(\mathcal{X}), \quad \int_{\mathcal{X}} \psi \, d(\text{div} m_n) &\xrightarrow[n \rightarrow +\infty]{} \int_{\mathcal{X}} \psi \, d(\text{div} m). \end{aligned}$$

Nous pouvons en profiter pour prouver le résultat suivant :

**Proposition 3.2.** *Toute suite bornée de  $\mathcal{V}$  a une sous-suite convergente au sens faible- $*$  dans  $\mathcal{V}$ .*

*Démonstration.* Soit  $(u_n)_{n \in \mathbb{N}}$  une suite bornée dans  $\mathcal{V}$ , donc dans  $\mathcal{M}(\mathcal{X})^2$ . D'après le théorème de Banach-Alaoglu-Bourbaki, il existe une sous-suite faiblement- $*$  convergente, à savoir qu'il existe un point  $u \in \mathcal{M}(\mathcal{X})^2$  et une fonction positive strictement croissante  $\varphi : \mathbb{N} \rightarrow \mathbb{N}$  tels que :

$$u_{\varphi(n)} \xrightarrow{*} u \text{ dans } \mathcal{M}(\mathcal{X})^2.$$

Comme  $(u_n)_{n \in \mathbb{N}}$  est bornée pour la topologie de la norme dans  $\mathcal{V}$  et par semi-continuité inférieure de  $\|\operatorname{div}(\cdot)\|_{\text{TV}}$ , il vient :

$$\|\operatorname{div} u\|_{\text{TV}} \leq \liminf_{n \rightarrow +\infty} \|\operatorname{div} u_{\varphi(n)}\|_{\text{TV}} < +\infty,$$

et l'on prouve que  $u \in \mathcal{V}$ , tandis que  $\operatorname{div} u_{\varphi(n)} \xrightarrow{*} \operatorname{div} u$  en utilisant les mêmes arguments que la preuve de la proposition précédente. Ainsi,  $u_{\varphi(n)} \xrightarrow{*} u$  dans  $\mathcal{V}$  ce qui conclut la preuve.  $\square$

Maintenant que nous avons discuté des quelques propriétés topologiques de  $\mathcal{V}$ , nous pouvons étudier plus en détail les éléments qui le compose.

**Propriétés.** *La mesure de Dirac vectorielle  $\delta \stackrel{\text{def.}}{=} (\delta, \delta)$  n'appartient pas à  $\mathcal{V}$ .*

*Démonstration.* Supposons par l'absurde que  $\delta$  appartienne à  $\mathcal{V}$ , donc que  $\operatorname{div} \delta \in \mathcal{M}(\mathcal{X})$ . Soit  $\zeta \in \mathcal{C}_0^\infty(\mathcal{X})$  tel que  $\partial_1 \zeta(0) = \partial_2 \zeta(0) = 1$  et la suite de fonctions :

$$\forall n \in \mathbb{N}^*, \forall x \in \mathcal{X}, \quad \zeta_n(x) = \frac{1}{n} \zeta(xn).$$

Alors, par le théorème de convergence dominée, pour toute mesure  $\mu \in \mathcal{M}(\mathcal{X})$  il vient  $\lim_{n \rightarrow +\infty} \int_{\mathcal{X}} \zeta_n d\mu = 0$ . Pourtant, pour tout  $n \in \mathbb{N}^*$  on a :

$$\int_{\mathcal{X}} \zeta_n d(\operatorname{div} \delta) = - \int_{\mathcal{X}} \nabla \zeta_n d\delta = -\partial_1 \zeta(0) - \partial_2 \zeta(0) = -2,$$

ce qui constitue donc une contradiction.  $\square$

Ce dernier résultat est particulièrement intéressant : les mesures de Dirac sont les points extrémaux de la boule unité de la norme TV ; ainsi le risque avec un nouvel espace, muni d'un raffinement d'une norme TV, serait de retomber sur la même structure ponctuelle prescrite par le régulariseur. Le problème est ici réglé en éliminant les mesures ponctuelles comme montré ci-dessous.

Nous donnons maintenant un exemple essentiel d'élément de  $\mathcal{V}$ , à savoir la *mesure portée par une courbe* c'est-à-dire une mesure dont le support est une courbe, ceci étant défini par intégration :

**Définition 3.5** (Mesure portée par une courbe). Soit  $\gamma : [0, 1] \rightarrow \mathcal{X}$  une courbe Lipschitz paramétrisée<sup>6</sup>, on dit que  $\mu_\gamma \in \mathcal{V}$  est une mesure portée par une courbe  $\gamma$  si :

$$\forall g \in \mathcal{C}_0(\mathcal{X})^2, \quad \langle \mu_\gamma, g \rangle_{\mathcal{M}^2} \stackrel{\text{def.}}{=} \int_0^1 g(\gamma(t)) \cdot \dot{\gamma}(t) dt.$$

On note  $\Gamma \stackrel{\text{def.}}{=} \gamma([0, 1])$  le support de cette courbe.

Dit autrement, le crochet de dualité évalué en une mesure portée par une courbe est précisément la circulation [82] d'une fonction test vectorielle le long de la courbe  $\gamma$ . L'hypothèse de rectifiabilité peut être comprise comme un moyen de contraindre la longueur finie de la courbe, elle est en effet équivalente à  $\mathcal{H}_1(\Gamma) < +\infty$ . Pour des raisons de concision, nous nous référerons indifféremment à la mesure portée par une courbe  $\mu_\gamma$ , où la courbe  $\gamma$  sur laquelle la mesure est supportée. Nous présentons maintenant quelques propriétés qu'une courbe peut exhiber :

**Définition 3.6** (Quelques caractérisations de courbe). Une courbe est dite simple si la restriction de  $\gamma$  à  $[0, 1[$  est une application injective. Une courbe est fermée si  $\gamma(0) = \gamma(1)$ , elle est appelée boucle si elle est simple et fermée (elle est alors homotopique au cercle unité de  $\mathbb{R}^2$ ).

D'après le théorème de Sard pour les fonctions Lipschitz, on peut montrer que  $\mu_\gamma$  est indépendant de la paramétrisation de la courbe [3]; nous supposons donc dans la suite que  $\gamma$  est une paramétrisation à vitesse constante. De la même manière que pour les mesures ponctuelles dans le chapitre 1, on peut travailler avec une mesure portée par une courbe autant d'un point de vue fonctionnel (comme nous l'avons plus haut) que d'un point de vue ensembliste. En effet, pour tout ensemble de Borel  $B \subset \mathcal{X}$ , il vient :

$$\mu_\gamma(B) = \int_{\Gamma \cap B} \left( \sum_{t \in \gamma^{-1}(x)} \dot{\gamma}(t) \right) d\mathcal{H}_1(x). \quad (3.3)$$

Nous donnons finalement l'expression explicite de la divergence d'une mesure portée par une courbe.

**Proposition 3.3** (Divergence d'une courbe). Soit  $\mu_\gamma$  une mesure portée par la courbe  $\gamma$ , alors  $\text{div}(\mu_\gamma) = \delta_{\gamma(0)} - \delta_{\gamma(1)}$ . En particulier,  $\text{div}(\mu_\gamma) = 0$  si  $\gamma$  est une courbe fermée.

*Démonstration.* Soit  $\xi \in \mathcal{C}_0^\infty(\mathcal{X})$ , alors :

---

6. donc 1-rectifiable. Cette dernière propriété sera rappelée pour bien signifier au lecteur que la charge est « de longueur finie ».

$$\begin{aligned}
\langle \operatorname{div} \boldsymbol{\mu}_\gamma, \tilde{\xi} \rangle_{\mathcal{D}'(\mathcal{X}) \times \mathcal{C}_0^\infty(\mathcal{X})} &= - \langle \boldsymbol{\mu}_\gamma, \nabla \tilde{\xi} \rangle_{\mathcal{M}} \\
&= \int_0^1 \nabla \tilde{\xi}(\gamma(t)) \cdot \dot{\gamma}(t) \, dt \\
&= \int_0^1 \frac{d}{dt} (\tilde{\xi}(\gamma(t))) \, dt \\
&= \tilde{\xi}(\gamma(0)) - \tilde{\xi}(\gamma(1)) \\
&= \langle \delta_{\gamma(0)} - \delta_{\gamma(1)}, \tilde{\xi} \rangle_{\mathcal{D}'(\mathcal{X}) \times \mathcal{C}_0^\infty(\mathcal{X})}.
\end{aligned}$$

Donc  $\operatorname{div}(\boldsymbol{\mu}_\gamma) = \delta_{\gamma(0)} - \delta_{\gamma(1)}$  au sens des distributions.  $\square$

**Remarque** (Quelques liens avec le transport optimal). *Ces quelques propriétés appellent déjà à une question que l'on pourrait formuler ainsi : « est-ce qu'une mesure portée par une courbe est toujours définie par intégration, comme en définition (3.5) ? ». Ce problème est traité dans la théorie de la cohomologie de De Rham [72] par la notion de courant intégrable rectifiable. Une charge est un courant intégrable rectifiable, et s'écrit nécessairement à ce titre par une intégrale telle que celle en (3.5). Dit autrement, une charge possède de la régularité grâce à sa divergence finie, ce qui nous permet d'escamoter les cas pathologiques (non intégrables) de  $\mathcal{V}$ .*

**Résumé** — L'espace des charges  $\mathcal{V}$  est l'ensemble des mesures vectorielles à divergence finie, autrement dit les mesures vectorielles dont la divergence est une mesure. C'est un espace de Banach qui a notamment pour élément les mesure portées par une courbe Lipschitz  $\gamma$ , notée  $\boldsymbol{\mu}_\gamma$ .

### 3.3 Un problème variationnel sur l'espace des charges

Comme dans le cas scalaire, on voudrait définir un BLASSO sur  $\mathcal{V}$ . Pour  $\alpha > 0$ , nous présentons la fonctionnelle suivante que nous avons appelée CROC (pour *Courbes Représentées par les Charges*) :

$$\operatorname{argmin}_{\mathbf{m} \in \mathcal{V}} T_\alpha(\mathbf{m}) \stackrel{\text{def.}}{=} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{m}\|_{\mathcal{H}}^2 + \alpha \|\mathbf{m}\|_{\mathcal{V}}. \quad (\mathcal{Q}_\alpha(\mathbf{y}))$$

$\Phi : \mathcal{V} \rightarrow \mathcal{H}$  est linéaire et envoie un champ de vecteurs à divergence finie dans l'espace d'observation  $\mathcal{H}$ , supposé de dimension finie. L'observation  $\mathbf{y} \in \mathcal{H}$  vit dans ce même espace. On rappelle que  $\|\mathbf{m}\|_{\mathcal{V}} \stackrel{\text{def.}}{=} \|\mathbf{m}\|_{\text{TV}^2} + \|\operatorname{div} \mathbf{m}\|_{\text{TV}}$ , les différents termes de notre énergie peuvent être interprétés dans le sens suivant :

- $\frac{1}{2} \|\mathbf{y} - \Phi \mathbf{m}\|_{\mathcal{H}}^2$  est le terme d'attache aux données ;

- $\|\mathbf{m}\|_{\text{TV}^2}$  équivaut à la longueur de la courbe<sup>7</sup>, en effet  $\mathcal{H}_1(\Gamma)$  avec  $\Gamma$  l'image de l'application  $\gamma$ . Il permet d'éliminer les comportements oscillants à haute fréquence et d'adoucir la courbure de  $\gamma$ ;
- $\|\text{div } \mathbf{m}\|_{\text{TV}}$  est le terme de comptage de courbes (ouvertes). Ce terme exhibe alors deux propriétés importantes : d'une part, il régularise le problème et permet aux mesures portées par des courbes d'être des minimiseurs de CROC, comme cela est prouvé dans la prochaine section. D'autre part, si  $\gamma(0) \neq \gamma(1)$  alors  $\gamma$  est une courbe ouverte et :

$$\|\text{div } \boldsymbol{\mu}_\gamma\|_{\text{TV}} = \|\delta_{\gamma(0)} - \delta_{\gamma(1)}\|_{\text{TV}} = 2.$$

Une amélioration pratique de  $(\mathcal{Q}_\alpha(\mathbf{y}))$  pourrait consister à pénaliser les termes de longueur et de nombre de courbes avec des poids différents, c'est-à-dire prendre  $\alpha, \beta > 0$  et la régularisation  $\alpha\|\mathbf{m}\|_{\text{TV}^2} + \beta\|\text{div } \mathbf{m}\|_{\text{TV}}$ . Cette amélioration dépasse le cadre de ce chapitre, mais elle pourrait s'avérer utile dans la mise en œuvre numérique.

**Théorème 3.4.** *Le problème  $(\mathcal{Q}_\alpha(\mathbf{y}))$  admet au moins une solution.*

*Démonstration.* La fonctionnelle  $T_\alpha$  est propre et coercive sur  $\mathcal{V}$ . Soit  $(u_n)_{n \in \mathbb{N}}$  une suite minimisante de  $T_\alpha$ . Comme  $T_\alpha$  est coercive,  $(u_n)_{n \in \mathbb{N}}$  est bornée dans  $\mathcal{V}$  donc elle converge pour la topologie faible-\* vers une charge  $\mathbf{m} \in \mathcal{V}$ , à une sous-suite près (voir la proposition 3.2). Observons également que :

- $\|\cdot\|_{\mathcal{V}}$  est semi-continue inférieurement pour la topologie faible-\*;
- $\Phi : \mathcal{V} \rightarrow \mathcal{H}$  est continue pour  $\mathcal{V}$  munie de la topologie faible-\* et  $\mathcal{H}$  de la topologie faible.

Ainsi, par la méthode directe du calcul des variations, il est clair que  $\mathbf{m}$  est un minimiseur de  $T_\alpha$ , prouvant ainsi l'existence d'un minimiseur.  $\square$

**Remarque.** *Le problème  $(\mathcal{Q}_\alpha(\mathbf{y}))$  a une unique solution si  $\Phi$  est injective, puisqu'elle contraint  $T_\alpha$  à être strictement convexe. Il convient de noter qu'il s'agit ici d'une hypothèse assez forte qui n'est guère vérifiée en pratique, comme c'est le cas pour le problème super-résolution qui nous intéresse.*

Le théorème suivant donne les relations d'optimalité pour le minimiseur  $\bar{\mathbf{u}}$  :

**Théorème 3.5.** *Soit  $\mathbf{m}$  un minimiseur de  $(\mathcal{Q}_\alpha(\mathbf{y}))$ . Le problème dual admet un minimiseur  $q^* \in \mathcal{C}_0(\mathcal{X})$  optimal, la dualité forte est vérifiée et les conditions d'extrémalité s'écrivent :*

$$\begin{cases} -\nabla q^* & \in \partial\|\mathbf{m}\|_{\text{TV}^2} + \Phi^*(\Phi \mathbf{m} - \mathbf{y}) \\ -q^* & \in \partial\|\text{div } \mathbf{m}\|_{\text{TV}}. \end{cases} \quad (3.4)$$

7. En réalité  $\mathcal{H}_1(\Gamma) \leq \|\boldsymbol{\mu}_\gamma\|_{\text{TV}^2}$ , avec égalité pour des courbes simples (ce qui sera le cas en pratique).

*Démonstration.* Considérons la dualité au sens d'Ekeland-Temam [45, Remarque 4.2], avec une petite mise en garde : l'espace de Banach  $\mathcal{V}$  doit être réflexif, ce qui n'est clairement pas le cas ici. Cependant, l'hypothèse de réflexivité n'est nécessaire que pour la preuve d'existence. Puisque nous avons déjà prouvé l'existence de la solution, cette hypothèse de réflexivité n'est pas nécessaire dans notre cas. Revenons à la remarque 4.2 de [45] énonçant, pour  $\Lambda : \mathcal{V} \rightarrow Y$  linéaire,  $F : \mathcal{V} \rightarrow \mathbb{R}$  et  $G : Y \rightarrow \mathbb{R}$  convexe, que le problème primal

$$\inf_{u \in \mathcal{V}} F(u) + G(\Lambda u)$$

admet un problème dual s'écrivant :

$$\sup_{p^* \in Y^*} -F^*(\Lambda^* p^*) - G^*(-p^*). \quad (3.5)$$

Si  $m$  et  $p^*$  sont respectivement les solutions du problème primal et dual, les conditions d'extrémalité s'écrivent :

$$\begin{cases} \Lambda^* p^* & \in \partial F(m) \\ -p^* & \in \partial G(\Lambda m). \end{cases}$$

Dans notre cas, on choisit

$$\begin{aligned} \Lambda : \mathcal{V} &\longrightarrow \mathcal{M}(\mathcal{X}) \\ u &\longmapsto \operatorname{div} u \end{aligned}$$

et son dual

$$\begin{aligned} \Lambda^* : \mathcal{C}_0(\mathcal{X}) &\longrightarrow \mathcal{V}^* \\ p^* &\longmapsto -\nabla p^*. \end{aligned}$$

On choisit également, pour  $u \in \mathcal{V}$ ,  $F(u) \stackrel{\text{def.}}{=} \|u\|_{\text{TV}^2} + \frac{1}{2} \|y - \Phi u\|_{\mathcal{H}}^2$ , et  $G(q) \stackrel{\text{def.}}{=} \|q\|_{\text{TV}}$  pour  $q \in \mathcal{M}(\mathcal{X})$ . Le lecteur aura peut-être remarqué que  $G$  doit être continu en un certain  $\Lambda u_0$  où  $u_0 \in \mathcal{V}$  : cette condition est remplie pour la mesure triviale nulle. Le problème dual s'écrit alors :

$$\operatorname{argmax}_{q^* \in \mathcal{C}_0(\mathcal{X})} -F^*(-\nabla q^*) - G^*(-q^*).$$

Les conditions d'extrémalité se résument donc à  $q^*$  solution du problème dual (3.5) :

$$\begin{cases} \Lambda^* q^* & \in \partial F(m) \\ -q^* & \in \partial G(\Lambda m). \end{cases}$$

□

Dans la littérature sans-grille [43, 13, 35], le problème dual  $(\mathcal{D}_\lambda(y))$  du BLASSO  $(\mathcal{P}_\lambda(y))$  est essentiel pour établir l'existence des solutions et pour caractériser le certificat dual, que nous avons évoqué dans le chapitre 1. Ce dernier est fondamental pour la mise en œuvre numérique, puisqu'il intervient à la fois dans l'étape d'estimation du support de la mesure estimée et dans la condition d'arrêt des algorithmes gloutons de l'état de l'art. [13, 35].

**Corollaire 3.1.** *Soit  $\mathbf{m} \in \mathcal{V}$  un minimiseur de  $(\mathcal{Q}_\alpha(y))$ . Il existe alors  $\boldsymbol{\eta}_1 \in \partial\|\mathbf{m}\|_{\text{TV}^2}$  et  $\boldsymbol{\eta}_2 = \nabla q^*$  où  $q^* \in \partial\|\text{div } \mathbf{m}\|_{\text{TV}}$  avec  $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{C}_0(\mathcal{X})^2$  tel que :*

$$\langle \mathbf{m}, \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 \rangle_{\mathcal{M}^2} = \|\mathbf{m}\|_{\mathcal{V}}.$$

*Démonstration.* Examinons les deux conditions d'optimalité en étudiant un peu plus en détail les conditions d'extrémalité :

- considérons d'abord la première condition d'optimalité  $-\nabla q^* \in \partial F(\mathbf{m})$ . Dans ce cas, le sous-différentiel de la somme est la somme des sous-différentielles<sup>8</sup>, à savoir :

$$\partial F(\mathbf{m}) = \partial\|y - \Phi(\cdot)\|_{\mathcal{H}}^2(\mathbf{m}) + \partial\|\cdot\|_{\text{TV}^2}(\mathbf{m}).$$

Nous obtenons, en utilisant la forme fermée du terme des moindres carrés :

$$\partial\|y - \Phi(\cdot)\|_{\mathcal{H}}^2(\mathbf{m}) = \{-\Phi^*(y - \Phi \mathbf{m})\},$$

et donc :

$$\begin{cases} -\nabla q^* = \boldsymbol{\eta}_1 - \Phi^*(y - \Phi \mathbf{m}) \\ \boldsymbol{\eta}_1 \in \partial\|\mathbf{m}\|_{\text{TV}^2}. \end{cases}$$

Pour le certificat  $\boldsymbol{\eta}_1 \in \partial\|\mathbf{m}\|_{\text{TV}^2}$ , avec le résultat donné en (1.3) adapté au cas vectoriel, on a :

$$\begin{cases} \|\boldsymbol{\eta}_1\|_{\infty, \mathcal{X}} \leq 1 \\ \|\mathbf{m}\|_{\text{TV}^2} = \langle \mathbf{m}, \boldsymbol{\eta}_1 \rangle_{\mathcal{M}^2}. \end{cases}$$

- considérons ensuite la seconde condition d'optimalité  $-q^* \in \partial G(\text{div } \mathbf{m})$ . On note  $v \stackrel{\text{def.}}{=} \text{div } \mathbf{m}$ . Ensuite, comme indiqué précédemment dans l'équation (1.3) et puisque  $G \stackrel{\text{def.}}{=} \|\cdot\|_{\text{TV}}$  :

$$-q^* \in \partial G(v) \iff \begin{cases} q^* \in \mathcal{C}_0(\mathcal{X}) \\ \|q^*\|_{\infty, \mathcal{X}} \leq 1 \\ \langle -q^*, v \rangle_{\mathcal{M}} = \|v\|_{\text{TV}}. \end{cases}$$

8. Puisque la première sous-différentielle est réduite à un singleton, voir [45].

Mais, en notant  $\eta_2 \stackrel{\text{def.}}{=} \nabla q^*$  :

$$\|\operatorname{div} \mathbf{m}\|_{\text{TV}} = \langle -q^*, \operatorname{div} \mathbf{m} \rangle_{\mathcal{M}} = \langle \nabla q^*, \mathbf{m} \rangle_{\mathcal{M}^2} = \langle \eta_2, \mathbf{m} \rangle_{\mathcal{M}^2}.$$

Enfin, si l'on fusionne ces deux résultats, il vient :

$$\begin{aligned} \|\mathbf{m}\|_{\text{TV}^2} + \|\operatorname{div} \mathbf{m}\|_{\text{TV}^2} &= \langle \mathbf{m}, \eta_1 \rangle_{\mathcal{M}^2} + \langle \eta_2, \mathbf{m} \rangle_{\mathcal{M}^2} \\ &= \langle \mathbf{m}, \eta_1 + \eta_2 \rangle_{\mathcal{M}^2}, \end{aligned}$$

ce qui permet donc d'écrire  $\|\mathbf{m}\|_{\mathcal{V}} = \langle \mathbf{m}, \eta_1 + \eta_2 \rangle_{\mathcal{M}^2}$ . Finalement, on obtient :

$$\eta_1 + \eta_2 = \Phi^*(y - \Phi \mathbf{m}).$$

□

Nous avons donc détaillé les conditions d'optimalité, et défini le certificat  $\eta_\alpha = \eta_1 + \eta_2 \in \mathcal{C}_0(\mathcal{X})^2$ . Ces conditions entraînent plusieurs applications tant du point théorique que pratique : en effet, si l'on souhaite employer un algorithme dérivé de la méthode du gradient conditionnel [49], il est nécessaire de formuler une condition d'arrêt selon le critère «  $\|\eta_\alpha\|_{\infty, \mathcal{X}}$  est inférieur à une certaine quantité ». Dans le cas « classique » d'une somme pondérée de mesures de Dirac, le certificat dual  $\eta$  interpole simplement le signe des mesures ponctuelles [43], et le critère numérique revient à  $\|\eta\|_{\infty, \mathcal{X}} \leq 1$ . La propriété équivalente pour les mesures autres que 0-rectifiable n'est pas aussi évidente ; en effet, considérons le cas BV où la solution est une somme d'ensembles indicateurs  $\chi_E$  pour  $E$  un ensemble simple [28] : le critère se lit  $\langle \chi_E, \mathbf{m} \rangle_{\mathcal{M}} \leq P(E)$ . La condition d'arrêt sera liée à  $\langle \eta_\alpha, \mathbf{m} \rangle_{\mathcal{M}^2} \leq \|\mathbf{m}\|_{\mathcal{V}}$ , garantissant ainsi que  $\eta_\alpha$  est dans le sous-différentiel de l'énergie, à savoir la norme  $\mathcal{V}$ . Précisons également que la caractérisation précise d'un certificat dual d'une mesure portée par une courbe n'est pas évidente, et qu'il ne s'agit pas d'un simple doublon du cas ponctuel. Cependant, on peut noter que :

**Remarque.** Si  $\mathbf{m}$  est une mesure portée par une courbe  $\gamma$  telle que définie en définition (3.5) avec le support  $\Gamma = \gamma([0, 1])$ , alors :

- puisque  $\mathbf{m}$  est optimal,  $\eta_1$  vérifie la première condition  $\|\mathbf{m}\|_{\text{TV}^2} = \langle \mathbf{m}, \eta_1 \rangle_{\mathcal{M}^2}$ , et à l'aide de la formule de la co-aire en (3.3) couplé avec une égalité de Cauchy-Schwarz, il vient pour  $x \in \mathcal{X}$   $\mathcal{H}_1$ -p.p. :

$$\begin{cases} \eta_1(x) = \frac{\sum_{t \in \gamma^{-1}(x)} \dot{\gamma}(t)}{\left| \sum_{t \in \gamma^{-1}(x)} \dot{\gamma}(t) \right|} & \text{si } \sum_{t \in \gamma^{-1}(x)} \dot{\gamma}(t) \neq 0 \\ |\eta_1(x)| \leq 1 & \text{sinon.} \end{cases}$$

Si de surcroît  $\gamma$  est simple, cela revient à  $|\eta_1(x)| = 1$  si  $x$  appartient au support de la courbe  $\Gamma$ .

— En ce qui concerne  $\eta_2$ , l'analyse est plus ardue et, n'est du reste pas aussi intuitive. Supposons déjà  $\gamma$  ouvert ( $\eta_2$  serait nul autrement) et notons que  $\eta_2$  est le gradient d'une fonction  $q^*$ . La fonction  $q^*$  appartient au sous-différentiel de  $G$  et par conséquent de la norme TV, évaluée au point  $\text{div } \mathbf{m}$  qui est précisément une somme de mesures de Dirac. Comme nous l'avons déjà évoqué, un certificat, associé à la norme TV évaluée en une mesure discrète, interpole le signe des mesures de Dirac qui la compose [43]. Ainsi  $q^*$  peut s'interpréter comme une fonction interpolant le signe des mesures ponctuelles composant  $\text{div } \mathbf{m}$ , et  $\eta_2$  est le gradient de cette fonction.

**Résumé** — La fonctionnelle CROC, notée  $(\mathcal{Q}_\alpha(\mathbf{y}))$ , est une énergie convexe, régularisée par la norme  $\|\cdot\|_{\mathcal{V}}$ . Elle admet un minimiseur sur  $\mathcal{V}$ , et l'on peut définir des certificats duaux à l'aide des conditions d'extrémalité.

Nous avons donc décrit précisément les conditions d'optimalité ainsi que les certificats duaux liés au problème  $(\mathcal{Q}_\alpha(\mathbf{y}))$ . Ceci étant, nous manquons toujours d'une description plus précise de ses minimiseurs. Pour résoudre ce problème, nous invitons le lecteur à se rappeler du théorème de la représentation évoqué en (3.3) : la structure d'un des minimiseurs de CROC sera entièrement prescrite par le régulariseur.

Nous avertissons le lecteur que le choix de  $F$  et  $G$  dans la section qui suit est différent du choix précédent. En effet, le choix dans cette dernière section n'a été opéré que dans le but de faciliter le calcul du problème dual. Dans ce qui suit, il est plutôt logique de considérer  $F : \mathbf{m} \mapsto \|\mathbf{m}\|_{\mathcal{V}}$ , et évidemment le terme d'attache aux données  $G : p \mapsto \|y - p\|_{\mathcal{H}}^2$  ainsi que  $\Lambda = \Phi$ .

### 3.4 Les points extrémaux de la boule unité de la norme $\mathcal{V}$

Nous avons évoqué un peu plus haut en sous-section 3.1.3 et en définition (3.1) la notion de point extrémal de la boule unité du régulariseur, ainsi que la manière dont cette caractérisation géométrique contraignait la structure d'un des minimiseurs de l'énergie. Ces notions s'appliquent dans un cadre général, et en particulier dans le nôtre : nous allons dans cette section établir un résultat sur les points extrémaux de la boule unité de la norme de  $\mathcal{V}$ , ce qui permettra de caractériser finement la structure d'un minimiseur de  $(\mathcal{Q}_\alpha(\mathbf{y}))$ .

Dans cette perspective, il est nécessaire de recueillir quelques résultats de la littérature sur l'espace des charges pour former notre propre théorème. Considérons la définition [82, 62] :

**Définition 3.7** (Décomposable, complètement décomposable). Soit  $T \in \mathcal{V}$  et  $J \subset \mathcal{V}$ , on dit que  $T$  se décompose sur des charges de  $J$  s'il existe une mesure de Borel finie positive  $\rho$  sur  $J$  telle que :

$$\begin{aligned} \mathbf{T} &= \int_J \mathbf{R} \, d\rho(\mathbf{R}) \\ \|\mathbf{T}\|_{\text{TV}^2} &= \int_J \|\mathbf{R}\|_{\text{TV}^2} \, d\rho(\mathbf{R}). \end{aligned}$$

On dit que  $\mathbf{T}$  se décompose complètement sur des charges de  $J$  si cette dernière condition est vérifiée et si de plus :

$$\|\text{div } \mathbf{T}\|_{\text{TV}} = \int_J \|\text{div } \mathbf{R}\|_{\text{TV}} \, d\rho(\mathbf{R}).$$

Nous allons dans notre cas utiliser les ensembles suivants :

**Définition 3.8** (Ensembles des courbes). On note  $\mathfrak{S}$  l'espace des mesures portées par des courbes simples, qu'elles soient ouvertes ou fermées, muni de la topologie faible-\*

$$\mathfrak{S} \stackrel{\text{def.}}{=} \left\{ \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma}, \gamma \text{ est une courbe orientée simple Lipschitz} \right\}.$$

C'est un espace métrique (non-complet) pour la topologie faible-\*. On note également  $\mathfrak{S}_{\text{loop}}$  l'espace des mesures portées par des boucles :

$$\mathfrak{S}_{\text{loop}} \stackrel{\text{def.}}{=} \left\{ \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma}, \gamma \text{ est une boucle} \right\}.$$

Finalement, on note  $\mathfrak{S}_{\text{open}} \stackrel{\text{def.}}{=} \mathfrak{S} \setminus \mathfrak{S}_{\text{loop}}$  l'espace des courbes ouvertes.

Nous donnons dans la suite les résultats fondamentaux issus de [82, 62]. Rappelons qu'un solénoïde est une charge de  $\mathcal{V}$  de divergence nulle :

**Théorème 3.6** (Décomposition des solénoïdes [62, Proposition 1.17]). *Tout solénoïde  $\mathbf{T} \in \mathcal{V}$  peut être décomposé sur des éléments de  $\mathfrak{S}_{\text{loop}}$ .*

Soulignons que ce dernier théorème n'est valable que dans le cas  $d = 2$ ; il existe en effet plusieurs contre-exemples pour  $d > 2$  que nous avons évoqués plus haut. Le théorème qui suit est valable en toute dimension :

**Théorème 3.7** (Théorème C de Smirnov [82], raffiné avec [62, Proposition 1.17]). *Toute charge  $\mathbf{T} \in \mathcal{V}$  est une somme de deux charges  $\mathbf{P}$  et  $\mathbf{Q}$  tel que  $\text{div } \mathbf{P} = 0$  (et le théorème 3.6 peut ainsi être appliqué à  $\mathbf{P}$ ) et  $\mathbf{Q}$  est complètement décomposable sur les mesures portées par des courbes simples Lipschitz, c'est-à-dire les éléments de  $\mathfrak{S}$ .*

Nous présentons maintenant le résultat principal de ce chapitre, à savoir le théorème établissant un lien entre les mesures portées par des courbes et les points extrémaux de la boule unité du régulariseur, ici la norme de  $\mathcal{V}$ .

**Théorème 3.8** (Résultat principal). *Notons  $\mathcal{B}_{\mathcal{V}}^1 \stackrel{\text{def.}}{=} \{\mathbf{m} \in \mathcal{V}, \|\mathbf{m}\|_{\mathcal{V}} \leq 1\}$  la boule unité de la norme de  $\mathcal{V}$ , faiblement- $*$  compact. Alors :*

$$\text{Ext}(\mathcal{B}_{\mathcal{V}}^1) = \mathfrak{S}.$$

*Démonstration.* Commençons la preuve par l'inclusion

$$\text{Ext}(\mathcal{B}_{\mathcal{V}}^1) \supset \left\{ \frac{\boldsymbol{\mu}_{\gamma}}{\|\boldsymbol{\mu}_{\gamma}\|_{\mathcal{V}}}, \gamma \text{ est une courbe 1-rectifiable simple Lipschitz} \right\}.$$

Soit  $\gamma : [0, 1] \rightarrow \mathcal{X}$  une courbe Lipschitz<sup>9</sup> simple orientée de longueur  $\ell > 0$ , avec comme image  $\Gamma = \gamma([0, 1])$ , et  $\boldsymbol{\mu}_{\gamma}$  la mesure portée par cette courbe. Notons qu'il s'agit soit d'une courbe ouverte, soit d'une courbe fermée. Par l'absurde, supposons qu'il existe  $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{V}$  tels que  $\|\mathbf{u}_1\|_{\mathcal{V}} \leq 1$ ,  $\|\mathbf{u}_2\|_{\mathcal{V}} \leq 1$  et pour  $\lambda \in (0, 1)$  :

$$\frac{\boldsymbol{\mu}_{\gamma}}{\|\boldsymbol{\mu}_{\gamma}\|_{\mathcal{V}}} = \lambda \mathbf{u}_1 + (1 - \lambda) \mathbf{u}_2. \quad (3.6)$$

Pour tout  $A \subset \mathcal{X}$  Borel, en notant  $|\cdot|_{\mathcal{V}}(A) \stackrel{\text{def.}}{=} |\cdot|(A) + |\text{div} \cdot|(A)$ , il vient :

$$\frac{1}{\|\boldsymbol{\mu}_{\gamma}\|_{\mathcal{V}}} |\boldsymbol{\mu}_{\gamma}|_{\mathcal{V}}(A) = \lambda |\mathbf{u}_1|_{\mathcal{V}}(A) + (1 - \lambda) |\mathbf{u}_2|_{\mathcal{V}}(A). \quad (3.7)$$

En effet, en s'inspirant d'une preuve de [9, Théorème 4.7], s'il existe  $A \subset \mathcal{X}$  tel que  $\lambda |\mathbf{u}_1|_{\mathcal{V}}(A) + (1 - \lambda) |\mathbf{u}_2|_{\mathcal{V}}(A) > \frac{|\boldsymbol{\mu}_{\gamma}|_{\mathcal{V}}(A)}{\|\boldsymbol{\mu}_{\gamma}\|_{\mathcal{V}}}$  on peut alors écrire :

---

9. Le caractère 1-rectifiable est en fait automatique avec l'hypothèse Lipschitz. Ceci étant, pour des raisons pédagogiques et pour bien insister sur le caractère 1-rectifiable du support nous l'avons laissé dans le théorème. Les articles de Smirnov précisent la rectifiabilité certainement pour gérer les cas éventuels où  $\mathcal{X}$  n'est pas borné, typiquement  $\mathcal{X} = \mathbb{R}^d$ .

$$\begin{aligned}
 1 &= \frac{|\boldsymbol{\mu}_\gamma|_{\mathcal{Y}}(\mathcal{X})}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} \\
 &= \frac{|\boldsymbol{\mu}_\gamma|_{\mathcal{Y}}(A)}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} + \frac{|\boldsymbol{\mu}_\gamma|_{\mathcal{Y}}(A^c)}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} \\
 &< \lambda|\mathbf{u}_1|_{\mathcal{Y}}(A) + (1-\lambda)|\mathbf{u}_2|_{\mathcal{Y}}(A) + \frac{|\boldsymbol{\mu}_\gamma|_{\mathcal{Y}}(A^c)}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} \\
 &< \lambda|\mathbf{u}_1|_{\mathcal{Y}}(A) + (1-\lambda)|\mathbf{u}_2|_{\mathcal{Y}}(A) + \lambda|\mathbf{u}_1|_{\mathcal{Y}}(A^c) + (1-\lambda)|\mathbf{u}_2|_{\mathcal{Y}}(A^c) \\
 &< 1.
 \end{aligned}$$

Il vient ainsi  $1 < 1$  ce qui est absurde, par conséquent :

$$\lambda|\mathbf{u}_1|_{\mathcal{Y}}(A) + (1-\lambda)|\mathbf{u}_2|_{\mathcal{Y}}(A) \leq \frac{1}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} |\boldsymbol{\mu}_\gamma|_{\mathcal{Y}}(A).$$

Comme nous avons toujours l'inégalité (triangulaire) opposée, nous obtenons (3.7). Nous déduisons également de l'équation (3.7) que :

$$\|\mathbf{u}_1\|_{\mathcal{Y}} = \|\mathbf{u}_2\|_{\mathcal{Y}} = 1. \quad (3.8)$$

Notons que le même argument peut être utilisé sur les deux termes de la norme  $\mathcal{Y}$  pour obtenir, pour tout borélien  $A$  de  $\mathcal{X}$  :

$$\frac{1}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} |\boldsymbol{\mu}_\gamma|_{\text{TV}^2}(A) = \lambda|\mathbf{u}_1|_{\text{TV}^2}(A) + (1-\lambda)|\mathbf{u}_2|_{\text{TV}^2}(A) \quad (3.9)$$

$$\frac{1}{\|\boldsymbol{\mu}_\gamma\|_{\mathcal{Y}}} |\text{div } \boldsymbol{\mu}_\gamma|_{\text{TV}}(A) = \lambda|\text{div } \mathbf{u}_1|_{\text{TV}}(A) + (1-\lambda)|\text{div } \mathbf{u}_2|_{\text{TV}}(A). \quad (3.10)$$

Soit maintenant  $A \subset \mathcal{X}$  un borélien tel que  $A \cap \Gamma = \emptyset$ , alors d'après (3.7) on en déduit :

$$0 = \lambda|\mathbf{u}_1|_{\mathcal{Y}}(A) + (1-\lambda)|\mathbf{u}_2|_{\mathcal{Y}}(A) \implies \mathbf{u}_1(A) = \mathbf{u}_2(A) = 0. \quad (3.11)$$

En prenant  $A = \Gamma^c$  dans l'équation (3.11), il vient  $\text{spt } \mathbf{u}_i \subset \Gamma$ . Maintenant, en utilisant les théorèmes 3.6 et 3.7, il existe deux mesures de Borel finies positives  $\rho_i$ <sup>10</sup> sur  $\mathfrak{S}$  telles que :

10. Comme  $\|\mathbf{u}_i\|_{\mathcal{Y}} = \int_{\mathfrak{S}} \underbrace{\|\mathbf{R}\|_{\mathcal{Y}}}_{=1} d\rho_i(\mathbf{R}) = \rho_i(\mathfrak{S})$  pour  $\|\mathbf{u}_i\|_{\mathcal{Y}} = 1$ , il est clair que  $\rho_i(\mathfrak{S}) = 1$ .

$$\mathbf{u}_i = \int_{\mathfrak{S}} \mathbf{R} d\rho_i(\mathbf{R}).$$

On note  $\rho \stackrel{\text{def.}}{=} \lambda\rho_1 + (1 - \lambda)\rho_2$ , ce qui entraîne :

$$\lambda\mathbf{u}_1 + (1 - \lambda)\mathbf{u}_2 = \int_{\mathfrak{S}} \mathbf{R} d\rho(\mathbf{R}).$$

Tout d'abord, notons que chaque courbe de la décomposition est incluse dans le support de la charge  $T$ , dit autrement  $\text{spt } \mathbf{R} \subseteq \text{spt } \mu_\gamma$   $\rho$ -p.p. comme l'avait relevé Smirnov dans [82, Remarque 5]. De surcroît, en utilisant l'équation (3.9), on peut noter que :

$$\begin{aligned} \frac{1}{\|\mu_\gamma\|_{\mathcal{V}}} \|\mu_\gamma\|_{\text{TV}^2} &= \|\lambda\mathbf{u}_1 + (1 - \lambda)\mathbf{u}_2\|_{\text{TV}^2} \\ &= \lambda\|\mathbf{u}_1\|_{\text{TV}^2} + (1 - \lambda)\|\mathbf{u}_2\|_{\text{TV}^2} \\ &= \lambda \int_{\mathfrak{S}} \|\mathbf{R}\|_{\text{TV}^2} d\rho_1(\mathbf{R}) + (1 - \lambda) \int_{\mathfrak{S}} \|\mathbf{R}\|_{\text{TV}^2} d\rho_2(\mathbf{R}) \\ &= \int_{\mathfrak{S}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}). \end{aligned}$$

Par conséquent,

$$\frac{1}{\|\mu_\gamma\|_{\mathcal{V}}} \|\mu_\gamma\|_{\text{TV}^2} = \int_{\mathfrak{S}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}). \quad (3.12)$$

Ainsi,  $\mathbf{R}$  doit être de longueur maximale, c'est-à-dire  $\text{spt } \mathbf{R} = \text{spt } \mu_\gamma$ . Supposons au contraire que  $\text{spt } \mathbf{R} \subsetneq \text{spt } \mu_\gamma$  sur un ensemble de mesure  $\rho$  non-nulle (positive). La courbe associée à  $\mathbf{R}$  est nécessairement ouverte, alors :

$$\|\mathbf{R}\|_{\text{TV}^2} = \frac{\mathcal{H}_1(\text{spt } \mathbf{R})}{\mathcal{H}_1(\text{spt } \mathbf{R}) + 2} < \frac{\mathcal{H}_1(\text{spt } \mu_\gamma)}{\mathcal{H}_1(\text{spt } \mu_\gamma) + 2} \leq \frac{\|\mu_\gamma\|_{\text{TV}^2}}{\|\mu_\gamma\|_{\mathcal{V}}}.$$

Donc  $\|\mathbf{R}\|_{\text{TV}^2} < \frac{\|\mu_\gamma\|_{\text{TV}^2}}{\|\mu_\gamma\|_{\mathcal{V}}}$ , on en déduit :

$$\begin{aligned} \int_{\mathfrak{G}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}) &< \frac{\|\boldsymbol{\mu}_\gamma\|_{\text{TV}^2}}{\|\boldsymbol{\mu}_\gamma\|_\gamma} \int_{\mathfrak{G}} d\rho \\ &< \frac{\|\boldsymbol{\mu}_\gamma\|_{\text{TV}^2}}{\|\boldsymbol{\mu}_\gamma\|_\gamma} \underbrace{\rho(\mathfrak{G})}_{=1} = \int_{\mathfrak{G}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}), \end{aligned}$$

ce qui implique

$$\int_{\mathfrak{G}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}) < \int_{\mathfrak{G}} \|\mathbf{R}\|_{\text{TV}^2} d\rho(\mathbf{R}),$$

et on obtient donc une contradiction. On note  $\gamma_{\mathbf{R}}$  la paramétrisation de la courbe à vitesse unitaire associée à  $\mathbf{R}$ , résumons maintenant les propriétés que nous avons recueillies :

- $\text{spt } \mathbf{R} = \text{spt } \boldsymbol{\mu}_\gamma, \rho\text{-p.p.}$  ;
- $\gamma_{\mathbf{R}}$  est Lipschitz simple.

Soit  $x \in \text{spt } \mathbf{R}$ . Alors il existe  $t \in ]0, 1[$  tel que  $\gamma_{\mathbf{R}}(t) = x$ . Comme les supports de  $\mathbf{R}$  et de  $\boldsymbol{\mu}_\gamma$  sont les mêmes, observons qu'il existe  $\tilde{t} \in ]0, 1[$  tel que  $\gamma(\tilde{t}) = x$ , donc  $\gamma_{\mathbf{R}}(t) = \gamma(\tilde{t})$ . Les deux courbes étant simples, il est clair que  $t = \gamma_{\mathbf{R}}^{-1}(\gamma(\tilde{t}))$ , ce qui implique que  $\gamma_{\mathbf{R}}$  est une reparamétrisation de  $\gamma$ . Par conséquent, comme  $\|\mathbf{R}\|_\gamma = \|\boldsymbol{\mu}_\gamma / \|\boldsymbol{\mu}_\gamma\|_\gamma\|_\gamma = 1$ , on a  $\mathbf{R} = \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma} \rho\text{-p.p.}$  ce qui nous permet finalement d'écrire :

$$\mathbf{u}_i = \int_{\mathfrak{G}} \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma} d\rho_i = \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma} \underbrace{\rho_i(\mathfrak{G})}_{=1} = \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma}.$$

Il s'ensuit donc que  $\lambda = 0$  ou  $\lambda = 1$ , et  $\mathbf{u}_1 = \mathbf{u}_2 = \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma}$  ; ce qui conclut cette partie.

Nous pouvons maintenant l'inclusion inverse, à savoir :

$$\text{Ext}(\mathcal{B}_\gamma^1) \subset \left\{ \frac{\boldsymbol{\mu}_\gamma}{\|\boldsymbol{\mu}_\gamma\|_\gamma}, \gamma \text{ est une courbe 1-rectifiable simple Lipschitz} \right\}.$$

Précisons d'emblée la propriété qui suit :

**Lemme 3.1.** *Soit  $X$  un espace de Banach muni de la norme  $\|\cdot\|_X$ . Alors, pour tout  $x \in \text{Ext}(\mathcal{B}_X^1)$ ,  $\|x\|_X = 1$ .*

*Démonstration.* Soit  $x \in \text{Ext}(\mathcal{B}_X^1)$ , supposons par l'absurde que  $\|x\|_X < 1$ , on en déduit que  $\overline{\mathcal{B}_r} \subset \mathcal{B}_X^1$  où  $r \stackrel{\text{def.}}{=} 1 - \|x\|_X$ . Par conséquent, pour tout  $\xi \in X$  tel que  $\|\xi\|_X = 1$ , on a  $y \stackrel{\text{def.}}{=} x + r\xi$  et  $z \stackrel{\text{def.}}{=} x - r\xi$  qui appartiennent à la boule unité  $\mathcal{B}_X^1$ . En outre,  $x = \frac{y+z}{2}$  donc  $x$  n'est pas un point extrémal, ce qui est absurde.  $\square$

Soit  $T \in \text{Ext}(\mathcal{B}_{\mathcal{V}}^1)$  un point extrémal de la boule unité de la norme de  $\mathcal{V}$ , on a alors  $\|T\|_{\mathcal{V}} = 1$  par le lemme (3.1) et il existe<sup>11</sup> une mesure (de probabilité) de Borel finie  $\rho$  telle que :

$$T = \int_{\mathfrak{S}} R \, d\rho(R),$$

par la décomposition de Smirnov, théorème 3.7 [82, 62, 6]. Comme  $T$  est un point extrémal, par [76, Proposition 1.4]  $\rho$  est une mesure atomique, portée par un singleton de  $\mathfrak{S}$  et la preuve est achevée : il existe une courbe de Lipschitz simple 1-rectifiable  $\gamma$  tel que  $\frac{\mu_\gamma}{\|\mu_\gamma\|_{\mathcal{V}}} \in \mathfrak{S}$  et  $\rho = \delta_{\mu_\gamma/\|\mu_\gamma\|_{\mathcal{V}}}$  par conséquent  $T = \mu_\gamma/\|\mu_\gamma\|_{\mathcal{V}}$ , ce qui termine ainsi la preuve.  $\square$

**Attention —** La preuve retour est facilitée grâce au résultat très général de [76]. Donnons quelques éléments de la démonstration pour éclairer la construction de ce résultat. Si  $\rho$  est une mesure atomique, alors elle est supportée par un singleton de  $\mathfrak{S}$  [61, Théorème 2] et la preuve serait achevée : il existerait une courbe Lipschitz simple  $\gamma$  tel que  $\mu_\gamma/\|\mu_\gamma\|_{\mathcal{V}} \in \mathfrak{S}$  et  $\rho = \delta_{\mu_\gamma/\|\mu_\gamma\|_{\mathcal{V}}}$  d'où  $T = \mu_\gamma/\|\mu_\gamma\|_{\mathcal{V}}$ . Par contradiction, supposons que  $\rho$  est une mesure non-atomique. Alors, il existe un borélien  $A \subset \mathfrak{S}$  tel que  $0 < \rho(A) < 1$  avec  $\rho(A)$  arbitraire [77, Proposition A.1] et :

$$\rho = |\rho|(A) \left( \frac{1}{|\rho|(A)} \rho \llcorner A \right) + |\rho|(A^c) \left( \frac{1}{|\rho|(A^c)} \rho \llcorner A^c \right).$$

En écrivant :

$$u_1 = \int_{\mathfrak{S}} \frac{1}{|\rho|(A)} \rho \llcorner A, \quad u_2 = \int_{\mathfrak{S}} \frac{1}{|\rho|(A^c)} \rho \llcorner A^c,$$

on obtient deux points tels que  $\|u_1\|_{\mathcal{V}} \leq 1$  et  $\|u_2\|_{\mathcal{V}} \leq 1$ , alors qu'il est clair que la combinaison est convexe avec  $\lambda \stackrel{\text{def.}}{=} |\rho|(A)$  et  $1 - \lambda = |\rho|(A^c)$ . Toute la difficulté réside alors dans la démonstration de  $u_1 \neq u_2$  : la preuve de [76, Proposition 1.4] y parvient en utilisant le fait que (à un voisinage près) l'ensemble  $A$  est localement convexe.

11. Pour être précis, ce résultat est permis par la décomposition  $T = P + Q$  du théorème 3.7 et  $d = 2$ . Par conséquent,  $P$  et  $Q$  se décomposent sur  $\mathfrak{S}$  et leurs mesures de Borel respectives peuvent être sommées pour obtenir une unique mesure de Borel  $\rho$  concentrée sur  $\mathfrak{S}$ .

$u_1$  est par définition une combinaison convexe d'éléments de  $A$ , par conséquent  $u_1$  appartient à  $A$ . De même  $u_2 \in A^c$ ; alors que  $A$  et  $A^c$  sont disjoints, ce qui entraîne  $u_1 \neq u_2$ . Ainsi,

$$T = \lambda u_1 + (1 - \lambda)u_2$$

donc une combinaison convexe alors que  $T$  est un point extrémal : contradiction.

**Remarque.** On peut considérer les théorèmes A et C de Smirnov dans [82] comme un raffinement de l'intégrale de Choquet définie en théorème 3.2. En effet, l'article de Smirnov aboutit aux mêmes conclusions (avec quelques extensions) que l'application du théorème de Choquet couplée à notre résultat sur les points extrémaux, à savoir que toute charge se décompose en une combinaison convexe de l'ensemble des points extrémaux  $\mathfrak{S}$ .

Nous avons donc réussi à établir un résultat sur les points extrémaux, ce qui ouvre un horizon prometteur concernant la mise en œuvre numérique. En effet, les algorithmes les plus performants dans la littérature sans-grille consistent en une reconstruction itérative d'une combinaison linéaire finie de points extrémaux du régularisateur.

**Résumé** — Les points extrémaux de la boule unité du régulariseur  $\|\cdot\|_{\mathcal{V}}$  sont précisément  $\mu_{\gamma} / \|\mu_{\gamma}\|_{\mathcal{V}}$  les mesures portées par des courbes  $\gamma$ . Un minimiseur de CROC est une combinaison linéaire finie de mesures portées par une courbe.

### 3.5 Conclusion partielle

Dans ce chapitre, nous avons introduit un nouveau terme de régularisation, qui utilise la divergence pour implémenter l'*a priori* structurel de solutions en courbe. Après avoir évoqué différentes pistes de recherche pour reconstruire des courbes, nous avons discuté de l'espace des mesures vectorielles à divergence finie  $\mathcal{V}$  et prouvé quelques propriétés algébriques (espace de Banach, convergence des sous-suites pour une suite bornée) ainsi que fonctionnelles (éléments qui y vivent ou en sont exclus). La mesure portée par une courbe  $\mu_{\gamma}$  est un exemple d'élément de  $\mathcal{V}$ . Puis, nous avons défini une fonctionnelle appelée CROC sur cet espace, qui possède au moins une solution et dont nous pouvons définir les certificats duaux. Finalement, nous avons prouvé que les points extrémaux de la boule unité de la norme  $\|\cdot\|_{\mathcal{V}}$  sont précisément les mesures portées par une courbe simple Lipschitz.

## Chapitre 4

# Mise en œuvre numérique de la reconstruction sans-grille de courbes.

**Points clefs – la reconstruction sans-grille de courbes doit être mise en œuvre :**

1. L'espace des charges  $\mathcal{V}$  a été introduit, c'est un espace de **mesures vectorielles** telles que leur divergence soit une mesure scalaire. Il contient les **mesures portées par les courbes**.
2. La **fonctionnelle CROC** définie sur l'espace des charges a également été présentée, elle est régularisée par la norme  $\mathcal{V}$ .
3. Les **points extrémaux** de la boule unité de la norme de  $\mathcal{V}$  sont les mesures portées par les courbes : il existe donc un minimiseur de CROC qui est précisément une combinaison linéaire de courbes.

**Contributions – une démarche et un algorithme sans-grille pour reconstruire numériquement des courbes :**

4. Nous mettrons en œuvre nos résultats théoriques avec un algorithme dérivé d'un algorithme **Frank-Wolfe**.
5. Nous discuterons des problèmes d'implémentation pratique que cette mise en œuvre engendre telle que la **discrétisation**, le choix de l'**opérateur**, *etc.*
6. Si ce premier travail ouvre la voie à la reconstruction de courbes sans-grille, **quelques questions restent en suspens** : ces points sont étudiés et plusieurs pistes de travail sont dégagées.

**Chapitre 4 – Reconstruction sans-grille numérique de courbes :**


---

4.1	Introduction . . . . .	80
4.2	L'algorithme du <i>Charge (Sliding) Frank-Wolfe</i> pour la reconstruction de courbes . . . . .	81
4.2.1	Rappels . . . . .	81
4.2.2	Un algorithme dans le cadre général . . . . .	82
4.3	Illustration numérique pour le problème de super-résolution . . . . .	84
4.3.1	Discrétisation des courbes . . . . .	84
4.3.2	Définition d'un opérateur d'observation . . . . .	85
4.3.3	Déroulement de l'algorithme : cas polygonal . . . . .	86
4.3.4	Discussion sur le cas Bézier . . . . .	92
4.3.5	Des pistes d'amélioration . . . . .	95
4.4	Conclusion . . . . .	96

---

**4.1 Introduction**

Nous avons vu dans le chapitre précédent une nouvelle approche permettant la reconstruction sans-grille de courbes, en exploitant une nouvelle régularisation. En définissant un nouvel espace noté  $\mathcal{V}$  de mesures vectorielles à divergence finie, nous avons pu caractériser le rôle régularisant de la norme  $\|\cdot\|_{\mathcal{V}}$ . En particulier, nous avons déterminé les points extrémaux de sa boule unité, qui sont précisément les mesures portées par des courbes  $\mu_{\gamma}$ . On rappelle la fonctionnelle CROC, pour *Curves Represented On Charges* :

$$\operatorname{argmin}_{\mathbf{m} \in \mathcal{V}} T_{\alpha}(\mathbf{m}) \stackrel{\text{def.}}{=} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{m}\|_{\mathcal{H}}^2 + \alpha \|\mathbf{m}\|_{\mathcal{V}}. \quad (\mathcal{Q}_{\alpha}(\mathbf{y}))$$

Par le théorème 3.3 dit de la représentation et le résultat sur les points extrémaux, nous savons qu'un *minimum* de  $(\mathcal{Q}_{\alpha}(\mathbf{y}))$  est une combinaison linéaire **finie** de mesures portées par des courbes. Il est maintenant nécessaire de proposer une mise en œuvre numérique pour reconstruire ces courbes de manière sans-grille. La construction de cet algorithme est en fait simplifiée par l'approche choisie : nous avons travaillé sur la caractérisation géométrique des minimiseurs de  $(\mathcal{Q}_{\alpha}(\mathbf{y}))$  en étudiant les points extrémaux, et nous allons exploiter une adaptation de l'algorithme du gradient conditionnel (vu au chapitre 1), car il reconstruit itérativement des points extrémaux du régulariseur. Dans ce chapitre, nous proposons d'abord d'écrire l'algorithme du *Charge Sliding Frank-Wolfe* dans un cadre général, avant de le préciser dans le cadre de la super-résolution. Nous aborderons les difficultés qu'il pose à cet endroit, et nous discuterons des possibles améliorations pour cette méthode. Dans la suite de ce chapitre, nous travaillerons avec l'ensemble  $\mathcal{X}$  défini au chapitre 3.

**Attention** —  $\mathcal{X}$  est un borné de  $\mathbb{R}^2$  : si l'on voulait traiter le cadre  $\mathbb{R}^d$  pour  $d > 2$ , nous n'aurions plus le résultat sur les points extrémaux. Nous pourrions à la rigueur prouver que  $\text{Ext}(\mathcal{B}_\gamma^1) = \mathfrak{S}_\ell$ , l'ensemble des courbes de longueur arbitraire  $\ell > 0$  [82], mais alors d'une part, les courbes ne seraient plus simples (elles peuvent repasser sur elles-mêmes), et d'autre part, il serait nécessaire de fixer la longueur  $\ell$  *a priori* : elle ne pourrait pas être estimée dans l'optimisation.

## 4.2 L'algorithme du *Charge (Sliding) Frank-Wolfe* pour la reconstruction de courbes

### 4.2.1 Rappels

Nous rappelons la définition suivante, à savoir l'espace des mesures portées par des courbes :

**Définition 4.1** (Ensemble des courbes). *On note  $\mathfrak{S}$  l'espace des mesures portées par des courbes simples, qu'elles soient ouvertes ou fermées, muni de la topologie faible-\** :

$$\mathfrak{S} \stackrel{\text{def.}}{=} \left\{ \frac{\mu_\gamma}{\|\mu_\gamma\|_{\mathcal{V}}}, \gamma \text{ est une courbe orientée simple Lipschitz} \right\}.$$

*C'est un espace métrique (non-complet) pour la topologie faible-\**.

**Attention** — Nous insistons sur le fait que les mesures portées par des courbes, éléments de  $\mathfrak{S}$ , n'encodent **aucune variation d'amplitude** le long du support, et qu'elles sont **normalisées**. Cela affecte évidemment les termes de la norme  $\mathcal{V}$  : désormais, et pour éviter toute ambiguïté, nous désignons par  $\nu_\gamma \stackrel{\text{def.}}{=} \mu_\gamma / \|\mu_\gamma\|_{\mathcal{V}}$  un élément de  $\mathfrak{S}$ , mesure portée par une courbe  $\gamma$  et normalisée.

Comme nous l'avons vu dans le chapitre 1, l'algorithme de Frank-Wolfe effectue la minimisation d'une fonction différentiable convexe  $f$  sur un sous-ensemble  $E$  convexe faiblement compact d'un espace de Banach. Il repose sur la minimisation itérative d'une version linéarisée de la fonction objectif, bénéficiant du fait qu'il n'utilise que les dérivées directionnelles de  $f$  et qu'il ne nécessite aucune structure hilbertienne sur  $E$ , contrairement aux algorithmes proximaux plus traditionnels. Il a suscité un grand intérêt dans les communautés de l'optimisation et de l'apprentissage, grâce à sa capacité à produire des itérées qui sont une combinaison de quelques atomes seulement, spécifiques au régularisateur choisi. Comme d'autres implémentations sans-grille, cet algorithme n'est pas directement applicable à CROC :  $T_\alpha$  n'est pas différentiable, et l'ensemble d'optimisation  $\mathcal{V}$  n'est pas borné. Il est donc nécessaire de procéder à un *epigraphical lift* [53, 35] pour obtenir une fonctionnelle différentiable qui partage les mêmes mesures *minimum* que  $T_\alpha$ . L'algorithme que nous proposons pour reconstruire des courbes est donné en algorithme

3 : de la même manière qu'au chapitre 1, la fonctionnelle  $T_\alpha$  doit être entendue comme la version *epigraphical lift* où l'on omet par abus de langage la composante sur la norme  $\mathcal{V}$ .

#### 4.2.2 Un algorithme dans le cadre général

Précisons quelques notations : on note  $\nu_\gamma^{[k]} \stackrel{\text{def.}}{=} (\nu_{\gamma_1}^{[k]}, \dots, \nu_{\gamma_{N^{[k]}}}^{[k]})$  le vecteur itéré des atomes reconstruits à la  $k$ -ième itération, et  $\Phi_{\nu_\gamma^{[k]}}(a) \stackrel{\text{def.}}{=} \sum_{i=0}^k a_i \Phi \nu_{\gamma_i}^{[k]}$  où  $a \in \mathbb{R}^{N^{[k]}}$  est le vecteur des poids estimés, voir algorithme 3. On note  $\mathcal{S}(\mathcal{X})$  l'ensemble des courbes engendrant  $\mathfrak{S}$ , à savoir :

$$\mathcal{S}(\mathcal{X}) \stackrel{\text{def.}}{=} \{\gamma, \gamma \text{ est une courbe simple Lipschitz à support dans } \mathcal{X} \text{ tel que } \nu_\gamma \in \mathfrak{S}\}.$$

La longueur de  $\gamma$  est donnée par  $\ell(\gamma) \stackrel{\text{def.}}{=} \mathcal{H}_1(\gamma([0, 1]))$  et son équivalent de la norme  $\mathcal{V}$  s'écrit  $\ell_{\text{div}}(\gamma)$  :

$$\ell_{\text{div}}(\gamma) = \begin{cases} \ell(\gamma) + 2 & \text{si } \gamma \text{ ouverte.} \\ \ell(\gamma) & \text{sinon.} \end{cases}$$

Notre algorithme bénéficie de l'amélioration apportée par la technique de *sliding*, qui consiste à améliorer l'algorithme classique de Frank-Wolfe par une étape de *sliding* en ligne 8–9. Nous avons en effet vu dans le cadre des mesures ponctuelles que la modification de ces deux lignes par n'importe quelle méthode qui décroît l'énergie  $T_\alpha$  permet de conserver les propriétés de convergence (que nous établissons plus bas pour les courbes). Cette étape non-convexe permet de surcroît une convergence en temps fini pour la reconstruction de mesures ponctuelles : ceci étant un argument similaire pour les mesures surfaciques et les mesures portées par des courbes reste à trouver, bien qu'observé en pratique. La propriété suivante, dérivée des propriétés de l'algorithme de Frank-Wolfe [31, 35], donne quelques garanties de convergence de l'algorithme :

**Proposition 4.1.** *Soit  $(m^{[k]})_{k \geq 0}$  une suite produite par l'algorithme 3. Elle admet alors un point d'accumulation  $m^*$  pour la topologie faible-\*, qui est solution de  $(\mathcal{Q}_\alpha(y))$ . En outre, il existe  $C > 0$  tel que pour tout minimiseur  $m^*$  de  $(\mathcal{Q}_\alpha(y))$  on ait :*

$$\forall k \in \mathbb{N}^*, \quad T_\alpha(m^{[k]}) - T_\alpha(m^*) \leq \frac{C}{k}.$$

Les étapes des lignes 7 à 9 sont une adaptation assez naturelle aux courbes, comme cela l'était pour [35, 28]. L'étape 4 renferme quant à elle plus d'originalité, et constitue en réalité l'étape la plus difficile dans la mise en œuvre de l'algorithme. Pour être plus précis, le problème

$$\gamma_* \in \operatorname{argmax}_{\gamma \in \mathcal{S}(\mathcal{X})} \frac{1}{\ell_{\text{div}}(\gamma)} \int_0^1 \eta^{[k]}(\gamma(t)) \cdot \dot{\gamma}(t) dt$$

**Algorithme 3** : Algorithme du Charge Sliding Frank-Wolfe

**Entrées** : Acquisition  $y \in \mathcal{H}$ , nombre d'itérations  $K$ , paramètre de régularisation  $\alpha > 0$ .

1 Initialisation :  $m^{[0]} = 0, N^{[k]} = 0$ .

2 **pour**  $k, 0 \leq k \leq K$  **faire**

3 Pour  $m^{[k]} = \sum_{i=1}^{N^{[k]}} a_i^{[k]} \mathbf{v}_{\gamma_i}^{[k]}$  tel que  $a_i^{[k]} \in \mathbb{R}, \mathbf{v}_{\gamma_i}^{[k]} \in \mathfrak{S}$ , soit

$$\boldsymbol{\eta}^{[k]}(x) \stackrel{\text{def.}}{=} \frac{1}{\alpha} \Phi^*(\Phi m^{[k]} - y).$$

4 Trouver  $\gamma_*^{[k]} \in \mathcal{S}(\mathcal{X})$  tel que :

$$\gamma_* \in \operatorname{argmax}_{\gamma \in \mathcal{S}(\mathcal{X})} \frac{1}{\ell_{\text{div}}(\gamma)} \int_0^1 \boldsymbol{\eta}^{[k]}(\gamma(t)) \cdot \dot{\gamma}(t) dt.$$

**si**  $|\boldsymbol{\eta}^{[k]}(\gamma_*)| \leq 1$  **alors**

5 |  $m^{[k]}$  est une solution de CROC. Stop.

6 **sinon**

7 Soit  $m^{[k+1/2]} = \sum_{i=1}^{N^{[k]}} a_i^{[k+1/2]} \mathbf{v}_{\gamma_i} + a_{N^{[k]}+1}^{[k+1/2]} \mathbf{v}_{\gamma_*^{[k+1/2]}}$  tel que :

$$a_i^{[k+1/2]} \in \operatorname{argmin}_{a \in \mathbb{R}^{N^{[k]}+1}} T_\alpha \left( \sum_{i=1}^{N^{[k]}} a_i \mathbf{v}_{\gamma_i}^{[k]} + a_{N^{[k]}+1} \mathbf{v}_{\gamma_*^{[k+1/2]}} \right)$$

8 Calculer  $m^{[k+1]} = \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathbf{v}_{\gamma_i}$ , solution de l'optimisation initialisée par  $m^{[k+1/2]}$  :

$$\left( a^{[k+1]}, \mathbf{v}_{\gamma}^{[k+1]} \right) \in \operatorname{argmin}_{(a, \gamma) \in \mathbb{R}^{N^{[k]}+1} \times \mathcal{S}(\mathcal{X})^{N^{[k]}+1}} T_\alpha \left( \sum_{i=1}^{N^{[k]}+1} a_i \mathbf{v}_{\gamma_i}^{[k]} \right).$$

9 Prendre  $m^{[k+1]} = \sum_{i=1}^{N^{[k]}+1} a_i^{[k+1]} \mathbf{v}_{\gamma_i}^{[k]}$ . Retirer les atomes d'amplitudes trop faibles.

10 **fin**

11 **fin**

**Résultat** : Mesure discrète  $m^{[k]}$  où  $k$  est l'itération d'arrêt.

n'est que la traduction du problème sur les charges  $\operatorname{argmax}_{m \in \mathcal{Y}} \langle m, \boldsymbol{\eta}^{[k]} \rangle_{\mathcal{M}^2}$  dans l'espace des courbes. La difficulté est alors plus évidente :  $\mathcal{S}(\mathcal{X})$  n'est pas *a priori* un espace vectoriel<sup>1</sup> ce qui limite les possibilités pour la méthode d'optimisation, et la minimisation de  $\langle m, \boldsymbol{\eta}^{[k]} \rangle_{\mathcal{M}^2}$  est un problème encore moins trivial, eu égard à la difficulté

1. Sauf peut-être à définir une addition qui raccorderait les paramétrisations, mais il est alors clair que cette addition ne serait pas symétrique

d'optimiser sur un espace de mesures.

**Attention** — Le lecteur pourrait se demander pourquoi de telles difficultés n'apparaissent pas pour les mesures ponctuelles ou surfaciques. Nous l'invitons alors à considérer les deux observations suivantes :

- le problème pour les mesures 0-rectifiables  $\operatorname{argmax} \langle m, \eta^{[k]} \rangle_{\mathcal{M}}$  se réduit simplement à  $\operatorname{argmax}_{x \in \mathcal{X}} \eta^{[k]}(x)$ . Ce problème est assez simple à résoudre, tout du moins avec une méthode *grid search* [29] : on parcourt tous les pixels de la grille pour trouver celui qui a la plus forte luminosité.
- le problème pour les mesures 2-rectifiables, s'il peut certes s'écrire comme un problème sur les ensembles de Cheeger [28], peut se réduire au problème plus classique :

$$\operatorname{argmin}_{u \in L^2(\mathcal{X})} \int_{\mathcal{X}} \eta^{[k]}(x) u(x) dx, \quad \text{tel que } |Du|(\mathcal{X}) \leq 1.$$

Le cadre des mesures 1-rectifiables ne bénéficie ni de l'épaisseur d'une mesure 2-rectifiable, ni de la simplicité d'écriture des mesures 0-rectifiables, et nous détaillons dans les paragraphes suivants comment aborder cette question.

Afin d'illustrer le fonctionnement de cet algorithme et de soulever d'autres difficultés qui affleurent avec le problème de super-résolution, nous discutons maintenant de l'implémentation pratique pour une observation convolée et bruitée.

**Résumé** — Nous avons introduit l'algorithme du *Charge Sliding Frank-Wolfe*. Pure adaptation des algorithmes du gradient conditionnel, il permet de reconstruire itérativement une charge comme une combinaison linéaire de mesures portées par des courbes.

### 4.3 Illustration numérique pour le problème de super-résolution

#### 4.3.1 Discrétisation des courbes

Il est nécessaire en pratique de discrétiser les courbes que nous allons optimiser. Nous pouvons relever au moins deux types de discrétisation pour nos courbes Lipschitz et simples.

**Définition 4.2** (Courbes polygonales). *On appelle courbe polygonale la courbe paramétrée, pour<sup>2</sup>  $n \geq 2$  par  $x \in \mathbb{R}^{n \times 2}$ , tel que  $[x_1, x_2], \dots, [x_{n-1}, x_n]$  soit simple.*

Le vecteur  $x$  est donc la liste des sommets par lesquels passe la courbe. L'avantage qu'il procure est assez clair : la discrétisation est en quelque sorte un échantillonnage

2.  $n \geq 3$  évidemment pour une boucle.

du tracé de la courbe. Ceci se fait de toute évidence au détriment de la régularité : le caractère polygonal implique d'avoir un nombre suffisant de points pour que la courbe apparaisse régulière. Si l'on souhaite disposer de courbes plus lisses, avec moins de points de régularisation, on peut utiliser des courbes déterminées par des points de contrôle, comme c'est le cas de la classe suivante :

**Définition 4.3** (Courbes de Bézier). *On appelle courbe de Bézier la courbe contrôlée par les points  $(\mathbf{P}_0, \dots, \mathbf{P}_p)$  et définie par la représentation paramétrique  $\mathbf{P}(t) = \sum_{i=0}^p B_i^p(t) \mathbf{P}_i$ , pour tout  $t \in [0, 1]$  et  $B_i^p$  les polynômes de Bernstein.*

Soit  $\gamma$  une courbe simple, on note  $\gamma_n$  la courbe de Bézier à  $n$  points de contrôle et de degré  $n$  choisie pour interpoler la courbe  $\gamma$  en  $n - 1$  points différents. La réalisation de cette interpolation est un peu plus délicate puisqu'il ne suffit pas d'évaluer la courbe en  $n$  points pour avoir les points de contrôle : nous serons amenés à en discuter plus en détail au moment de la mise en œuvre pour le problème de super-résolution.

**Remarque.** *Il convient de noter qu'à  $n$  fixé cette courbe  $\gamma_n$  est unique. En effet, soient deux ensembles de points de contrôle  $\mathbf{P}$  et  $\mathbf{Q}$  produisant la même courbe de degré  $n$  alors :*

$$\begin{aligned} \sum_{i=0}^p B_i^p(t) \mathbf{P}_i &= \sum_{i=0}^p B_i^p(t) \mathbf{Q}_i, \quad \text{pour tout } t \in [0, 1] \\ \iff \sum_{i=0}^p B_i^p(t) (\mathbf{P}_i - \mathbf{Q}_i) &= 0. \end{aligned}$$

*Les polynômes de Bernstein forment une famille libre, donc  $\mathbf{P}_i = \mathbf{Q}_i$ .*

Notons que la courbe polygonale peut être comprise comme une discrétisation type *spline* (avec un polynôme) de degré 1. Une discrétisation avec des polynômes d'ordre supérieur pourrait être pertinente, mais les inconvénients (dont nous discuterons plus loin) seraient sensiblement les mêmes que la discrétisation polygonale. Nous discuterons donc la suite principalement des choix de courbes polygonales ou de courbes de Bézier.

Un résultat faisant le lien entre solutions des problèmes discrets (CROC sur les courbes discrétisées) et solution du problème continu (CROC sur  $\mathcal{V}$ ) serait le bienvenu : pour établir cette relation, nous avons dessein d'utiliser dans de prochains travaux la notion de  $\Gamma$ -convergence.

### 4.3.2 Définition d'un opérateur d'observation

L'application choisie pour illustrer le principe de l'algorithme est un problème de super-résolution similaire à ceux rencontrés dans les précédents chapitres, dans le contexte d'une convolution gaussienne. Soit  $g \in \mathcal{X}^N$  l'image observée. Toute la difficulté provient de la nature *vectorielle* des mesures portées par des courbes  $\nu_\gamma$  : il n'est pas évident de donner un sens aux composantes  $(\nu_\gamma^1, \nu_\gamma^2) = \nu_\gamma$ . Le choix de l'opérateur  $\Phi : \mathcal{V} \rightarrow \mathcal{H}$  doit donc tenir compte de cette dimension vectorielle. De surcroît, le

certificat sera nécessairement un élément de  $\mathcal{C}_0(\mathcal{X})^2$  (voir chapitre 3), dit autrement  $\Phi^* : \mathcal{H} \rightarrow \mathcal{C}_0(\mathcal{X})^2$ . Il paraît donc naturel de préserver ce caractère vectoriel dans  $\mathcal{H}$ .

**Attention** — La divergence, et par extension le caractère vectoriel, est essentielle pour assurer l'extrémalité des mesures portées par des courbes, en permettant par exemple d'évacuer la mesure de Dirac de l'ensemble d'optimisation.

Il faut donc trouver une quantité *vectorielle* de  $\mathcal{H}$  que l'on pourrait lier à la mesure  $\nu_\gamma$  par l'opérateur  $\Phi$ . Un choix naturel serait de prendre le gradient de l'image  $g$ , noté  $\mathbf{y}$ . En pratique, nous utilisons une astuce permise par la propriété de dérivation de la convolution. En effet, considérons formellement l'image source discrète  $I$ , tel que l'on ait  $g = h * I$ , alors  $\mathbf{y} = \nabla g = h * \nabla I = \nabla h * I$ . Ainsi, nous évaluerons  $\Phi \nu_\gamma$  comme la convolution du support avec le gradient du noyau gaussien. L'opérateur vectoriel  $\Phi$  aura donc pour noyau  $\varphi(x)$  avec  $x \in \mathcal{X}$  et  $N \in \mathbb{N}^*$  le nombre de points dans une image de  $\mathcal{H} = \mathbb{R}^N$  :

$$\forall x \in \mathcal{X}, \quad \varphi(x) \stackrel{\text{def.}}{=} \frac{1}{2\pi\sigma} \begin{pmatrix} -x_1 e^{-(\frac{i-1}{N-1}-x_1)^2/2\sigma^2} \\ -x_2 e^{-(\frac{i-1}{N-1}-x_2)^2/2\sigma^2} \end{pmatrix}_{1 \leq i \leq N}.$$

Il est clair que le gradient  $\mathbf{y}$  doit passer par un post-traitement (lissage, super-résolution), car c'est elle qui est utilisée en pratique dans l'algorithme, alors que cette quantité est calculée avec un gradient discret : le bruit dans  $g$  se répercute d'autant plus sur  $\mathbf{y}$ .

### 4.3.3 Déroutement de l'algorithme : cas polygonal

Dans la suite, nous allons considérer un exemple synthétique où l'on cherche à reconstruire des courbes à partir d'une image « classique » observée  $g$ , dégradée par une convolution d'écart-type  $\sigma = 2 \times 10^{-2}$  et par du bruit blanc additif gaussien d'écart-type  $\sigma_b = 4 \times 10^{-3}$ . Soit deux charges, plus précisément  $\mu_1$  une courbe ouverte et  $\mu_2$  une courbe fermée, appartenant toutes les deux à  $\mathfrak{S}$ . Considérons maintenant la charge source  $T = 4\mu_1 + \mu_2$ , le facteur quatre étant choisi pour des questions de visualisation. L'image  $g$  et sa source  $T$  sont présentées en figure 4.1. Le gradient de l'image  $\mathbf{y} \stackrel{\text{def.}}{=} \nabla g$  qui jouera le rôle d'observation est quant à lui portée en figure 4.2, il est calculé en pratique avec les différences finies.

La reconstruction sera d'abord produite avec des courbes polygonales, ce qui nous permettra d'observer le fonctionnement du *Charge Sliding Frank-Wolfe* dans un cas simple et de retirer les principaux avantages et défauts de cette discrétisation. Déroutons l'algorithme à la première étape, auquel cas notre mesure itérée  $m^{[k]}$  est nulle.

Au sein d'une itération, la première étape de notre algorithme consiste à estimer le support de l'atome ajouté à l'itérée. Comme nous l'avons indiqué, cette étape d'estimation est propre à chaque approche sans-grille : l'équivalent pour la norme TV classique

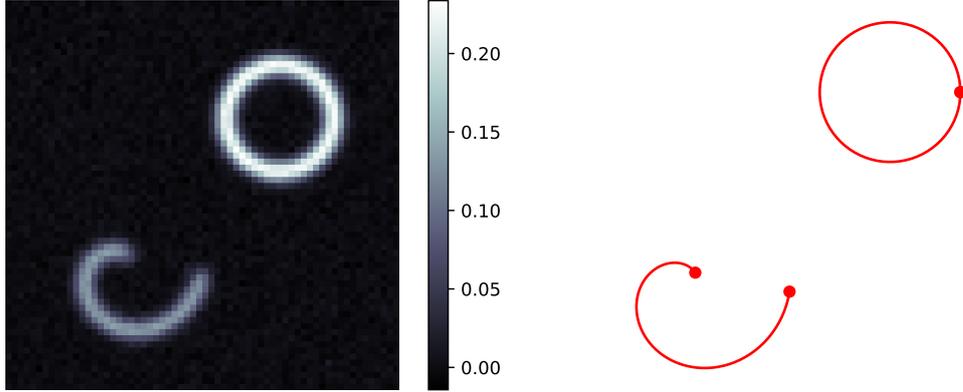


FIGURE 4.1 – À gauche : l'image observée  $g$ , générée à partir d'une charge composée d'une spirale  $\mu_1$  et d'une boucle  $\mu_2$ , qui ont des intensités différentes. À droite : les supports des deux courbes. Noter la courbure lisse de  $\mu_1$  et  $\mu_2$  que nous cherchons à retrouver.

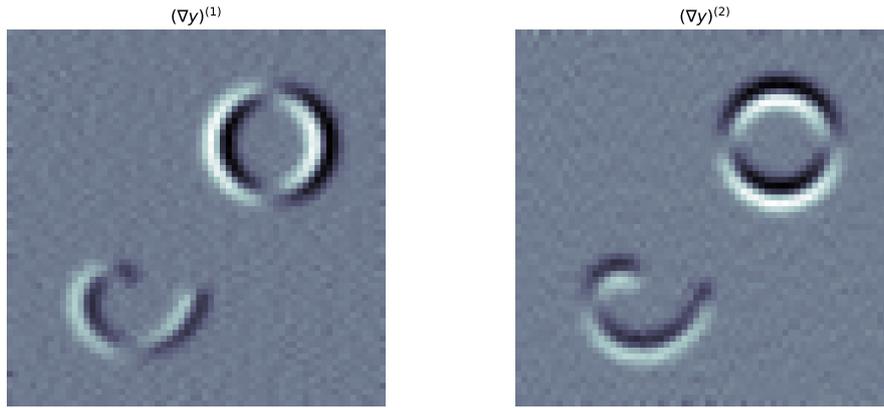


FIGURE 4.2 – Les deux composantes du gradient  $\mathbf{y} = \nabla g$  constituent la quantité pertinente pour le terme d'attache aux données de  $(\mathcal{Q}_\alpha(\mathbf{y}))$ . Notez que le bruit sur l'image a un effet d'autant plus grand sur le gradient, de sorte qu'une stratégie de débruitage du gradient doit être adoptée, en particulier pour les images expérimentales.

serait une simple recherche de la position de la plus grande valeur de  $|\boldsymbol{\eta}^{[k]}|$ , comme cela est défini en ligne 3 de l'algorithme 3. Dans notre cas, il est nécessaire de résoudre

$$\operatorname{argmax}_{m \in \mathcal{V}} \langle \boldsymbol{\eta}^{[k]}, m \rangle_{\mathcal{M}^2}$$

équivalent à

$$\operatorname{argmax}_{\gamma \in \mathcal{S}(\mathcal{X})} \frac{1}{\ell_{\operatorname{div}}(\gamma)} \int_0^1 \boldsymbol{\eta}^{[k]}(\gamma(t)) \cdot \dot{\gamma}(t) dt. \quad (4.1)$$

Cette quantité peut s'interpréter comme la longueur de la courbe  $\gamma$  pondérée par la « métrique »  $\eta^{[k]}$ . Si le noyau de  $\Phi$  est le gradient d'une gaussienne 2D, l'opérateur adjoint sera en pratique une divergence.

Le certificat dual est composé de dérivées secondes, ce qui le rapproche du Laplacien de l'image  $g$ , une quantité qui est connue pour « exploser » au voisinage du support de la courbe. Les points d'amplitude *maximum* dans  $|\eta^{[k]}|$  donnent donc une première approximation de la courbe. Les difficultés sont alors de deux ordres :

- contrairement au cas ponctuel où il y avait un seul point *maximum*, il faut désormais retrouver un ensemble de pixels pour former  $\gamma_*$  le support du nouvel atome estimé. Nous l'avons vu, le problème (4.1) est loin d'être évident, il faut *a minima* lui trouver un substitut;
- les courbes peuvent être soit ouvertes, soit fermées (contrairement au cas 2-rectifiable où la frontière de l'ensemble simple est nécessairement une courbe fermée). Il faut donc à ce titre prévoir une heuristique pour décider si l'atome ajouté est une courbe ouverte ou une courbe fermée.

En première approximation, nous pouvons déjà réaliser un seuillage pour identifier les pixels de plus forte amplitude du support. Nous récupérons ainsi une liste de points qui donne grossièrement le support de l'estimée, mais cette liste n'est pas triée.

**Attention** — Un seuillage trop fort évacuerait les *endpoints* de la liste de positions estimées. *A contrario*, un seuillage trop faible estime trop de points dans le support, ce qui donne une « épaisseur » à la courbe.

En ce qui concerne l'heuristique permettant de déterminer si  $\gamma_*$  est ouverte ou fermée, notons que les courbes en jeu sont simples. Par conséquent, nous exploitons le fait qu'une boucle sépare  $\mathcal{X}$  en deux composantes connexes, contrairement à une courbe ouverte dont le complémentaire du support n'a qu'un seul ensemble connexe. Il convient de distinguer dès maintenant les cas ouvert et fermé, car la manière de gérer les extrémités de la courbe va différer : une courbe fermée ne requiert pas d'attention particulière sur la détermination de ces *endpoints*. La figure 4.3 présente le certificat ainsi que le seuillage permettant de donner une initialisation du support de la courbe. Il est nécessaire à ce stade :

- d'identifier les extrémités (*endpoints*), points de départ et d'arrivée;
- d'ordonner la liste de points.

Nous pourrions utiliser ce premier résultat pour initialiser  $\gamma_*$ . Ceci étant, le seuillage que nous utilisons en pratique donne une épaisseur d'au moins deux pixels à la courbe : par conséquent la paramétrisation n'emprunte pas le chemin de la ligne moyenne, et oscille au contraire autour du tracé de la courbe vérité-terrain. Pour réduire le nombre de points dans la courbe numérique, nous proposons de réaliser une approximation lisse avec des fonctions de *python* telles que `splprep`. Ceci permet d'une part, de réduire le

nombre de points discrétisant le support (et par extension de réduire le temps de calcul des optimisations) et d'autre part, de contenir le risque d'oscillations que nous aurons l'occasion d'examiner plus en détail.

Une fois toutes ces étapes mises en œuvre, nous obtenons une courbe polygonale simple notée  $\gamma_*$ , et une initialisation du support pour l'atome ajouté est obtenue avec succès.

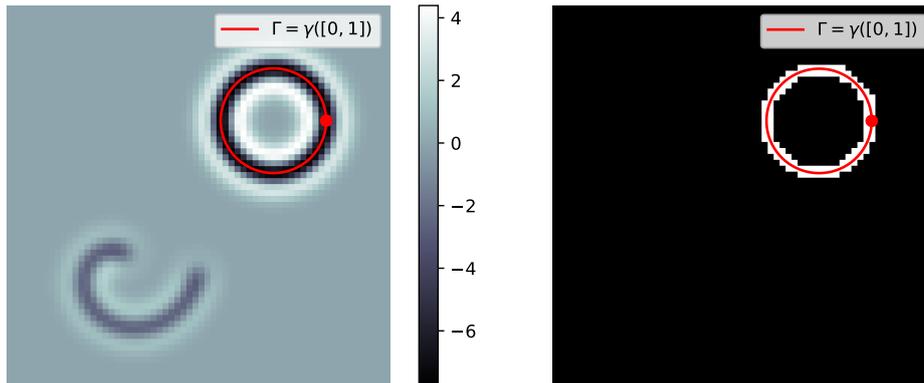


FIGURE 4.3 – À gauche : amplitude du vecteur  $\Phi^* \mathbf{y}$ , le support de la courbe estimée  $\gamma_*$  se trouve dans les pixels (quasi) optimaux. À droite : seuillage permettant d'obtenir une estimation approximative du support. La courbe vérité-terrain  $\gamma$ , de support  $\Gamma = \gamma([0, 1])$ , est tracée en rouge.

Ce support  $\gamma_*$  du nouvel atome estimé, nous pouvons ajouter la mesure portée par la courbe  $\gamma_*$  à l'itérée. L'optimisation qui suit est la résolution du problème convexe présenté en ligne 7 : c'est encore une fois une adaptation du BLASSO, dans la veine des travaux portant sur les mesures 0-rectifiables ou 2-rectifiables [35, 28]. L'amplitude des atomes, à supports fixés,  $\mathbf{y}$  est estimé. Cette optimisation est produite avec n'importe quel solveur convexe, tel qu'une implémentation d'ISTA/FISTA [4] (*Fast Iterative Shrinkage-Thresholding Algorithm*).

L'étape non-convexe en lignes 8 et 9 est pour sa part bien plus contraignante : le *sliding* qui est ici à l'œuvre ajuste les atomes à la fois en amplitude et en position. La discrétisation des courbes avec l'approximation polygonale revient à effectuer l'optimisation de « chaînes », plus précisément, on réalise une descente de gradient à la fois sur les amplitudes, qui est commune à chaque chaîne/courbe, et sur les positions. Ceci est réalisé en pratique par un optimiseur non-convexe, tel que (L-)BFGS ou dans notre cas le solveur ADAM [63] qui s'est montré empiriquement plus rapide et plus précis que les autres optimiseurs proposés dans *PyTorch*.

**Attention** — Les paramètres d'ADAM tels que le taux d'apprentissage ( $lr$  pour *learning rate*) doivent être choisis avec précaution. Ce dernier doit être pris assez faible (bien inférieur à 0,1), sous peine d'obtenir des solutions fortement divergentes.

La sortie de l'étape convexe est portée en figure 4.4 à gauche, elle partage alors le même support que le support « brut » estimé dans l'étape linéaire ou ligne 4. Elle est à comparer avec l'estimé issue de l'étape non-convexe sur la droite, où l'on peut voir l'effet de la descente de gradient qui régularise la courbure.

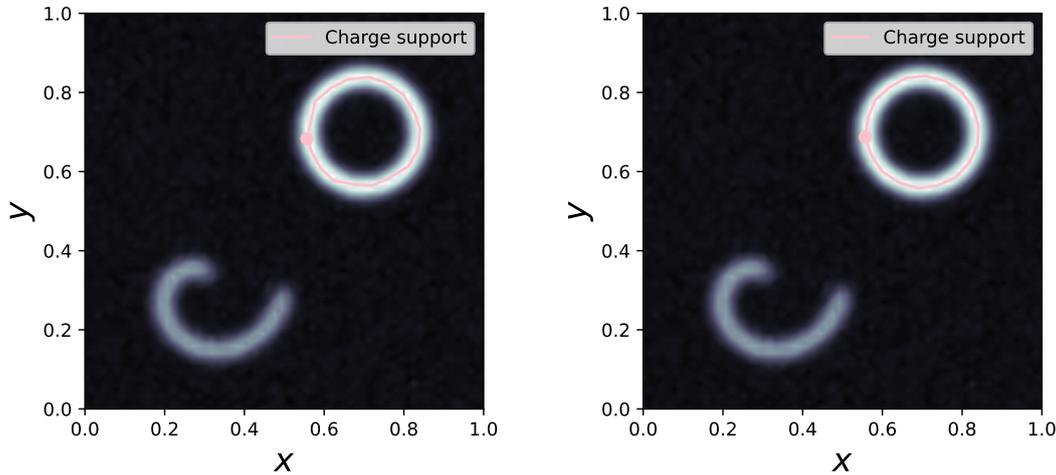


FIGURE 4.4 – À gauche : estimation brute obtenue à l'issue de l'étape linéaire. À droite : estimation à la fin de la première itération. Il est à espérer que l'étape de *sliding* lisse la courbure et corrige l'estimation approximative du support.

Il est intéressant de constater que les extrémités des boucles (c'est-à-dire les courbes fermées)  $\gamma(0)$  et  $\gamma(1)$ , égales par définition, tendent en pratique à s'éloigner l'une de l'autre au fil des itérations de l'étape non-convexe. Ceci est peut-être lié aux faces et, plus généralement, à la structure géométrique de la boule unité de la norme de  $\mathcal{V}$  : à notre connaissance, aucun travail n'a encore pu étudier ce curieux changement de topologie puisque l'espace  $\mathcal{V}$  est un concept assez nouveau pour la communauté sans-grille. Notre observation est à ce stade purement empirique, et nous avons simplement mis en œuvre une stratégie visant à (re)fusionner les extrémités de courbe en réalité fermées. Notre algorithme a alors reconstruit une courbe, et nous itérons ainsi l'algorithme pour obtenir une reconstruction de la charge source  $T$ , voir la figure 4.5.

Finalement, la reconstruction est portée dans la figure 4.6. Les supports des courbes sont correctement reconstruits<sup>3</sup>, avec bien sûr quelques petites différences dues au bruit.

Les résultats proposés peuvent sembler un peu simples. Cependant, nous soulignons que nous pouvons exploiter cette méthode avec d'autres types d'opérateur d'acquisition rencontrés en problèmes inverses, tel que Fourier ou Laplace, permettant de traiter des acquisitions plus difficiles à interpréter visuellement. Cet algorithme pourrait être appliqué à des données expérimentales, mais cela nécessite un post-traitement précis du

3. 20 secondes sur un *Intel Xeon E5-2687W v3* pour une image de  $64 \times 64$ . Une implémentation CUDA est mise en place pour les images plus grandes, garantissant toujours la reconstruction d'une image de  $512 \times 512$  en moins de 4 minutes sur une *NVIDIA Quadro K2200*. Voir <https://gitlab.inria.fr/blaville/amg>.

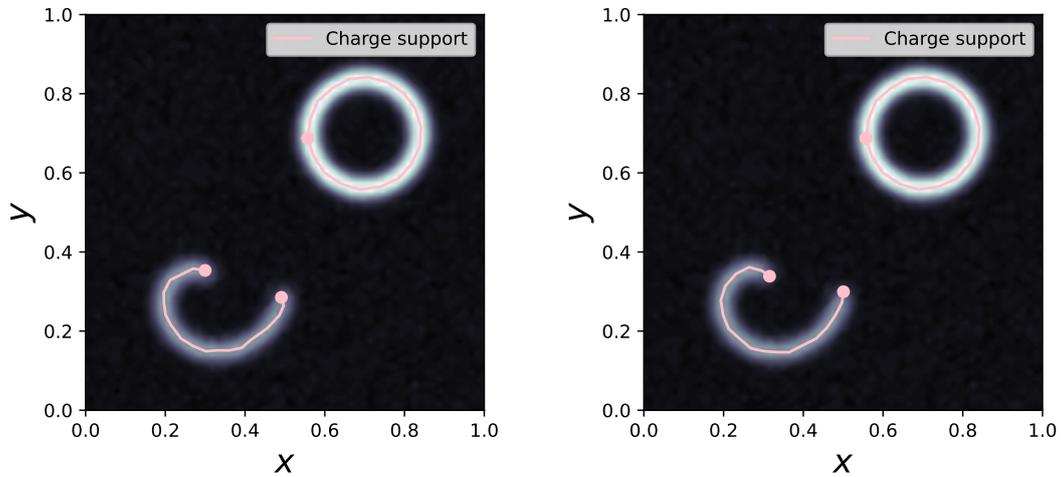


FIGURE 4.5 – À gauche : estimation brute à partir de l'étape de l'oracle. À droite : l'estimation à la fin de la première itération. Les estimations correspondent à la structure des courbes mais présentent quelques « cassures » : en raison du bruit, la reconstruction ne semble pas aussi lisse que la vérité de terrain. L'étape non convexe permet de mieux reconstruire l'enroulement de la spirale.

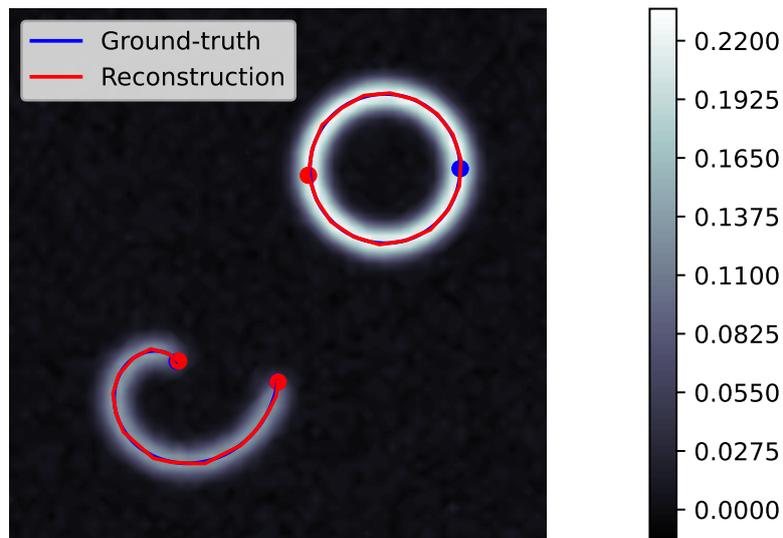


FIGURE 4.6 – La reconstruction finale rend bien compte de la courbure des courbes de la source.

gradient tel que le débruitage ou une super-résolution, car l'algorithme a besoin d'une première estimation (au pire médiocre, mais tout du moins utilisable) pour estimer un point de  $\mathcal{V}$  qui se trouve dans le bassin d'attraction du minimiseur global.

Pour finir, une limite théorique de l'erreur d'estimation du support n'est pas triviale à estimer : elle a été étudiée pour les mesures ponctuelles [43] et un résultat similaire

pour les ensembles de niveaux a récemment été annoncé<sup>4</sup>. Par analogie avec ces travaux, il est peut-être possible de définir un difféomorphisme entre les courbes reconstruites et celles de la vérité-terrain, la norme de ce difféomorphisme quantifiant et bornant l'écart entre les deux courbes. Cet équivalent pour la reconstruction de courbes sans-grille reste ceci étant à explorer.

**Résumé** — L'algorithme du *Charge Sliding Frank-Wolfe* reconstruit des courbes aussi bien ouvertes que fermées. Les spécificités des mesures 1-rectifiables constituent une subtilité supplémentaire, et la difficulté du problème d'optimisation nécessite un soin particulier dans le choix des paramètres de discrétisation, de régularisation, d'optimisation, etc.

L'algorithme a été ici mis en œuvre avec des courbes polygonales : si elles permettent une implémentation simplifiée (les points de contrôle sont exactement les points par lesquels la courbe chemine), elles présentent un désavantage au niveau du nombre de points nécessaires pour représenter une courbe avec une régularité importante. Il est nécessaire de prendre suffisamment de points de discrétisation pour avoir une certaine régularité visuelle. En revanche, il faut en mettre avec parcimonie, sous peine de voir des oscillations apparaître dans la reconstruction. On observe ainsi un phénomène similaire au chapitre 2 : les points discrétisant la courbe ont tendance à s'écartier les uns des autres, comme s'ils respectaient une distance de séparation minimale.

Ce phénomène est assez pénible pour la mise en œuvre numérique, puisqu'il nous contraint à choisir judicieusement le nombre de points de discrétisation. Il serait plus intéressant de disposer d'un vrai contrôle sur la finesse de discrétisation, et c'est ce que nous proposons dans le prochain paragraphe avec les courbes de Bézier.

#### 4.3.4 Discussion sur le cas Bézier

Les courbes de Bézier, telles que présentées plus haut en définition 4.3 présentent l'avantage d'avoir des points de contrôle qui **gèrent** le tracé de la courbe. À partir de  $p \in \mathbb{N}^*$  points de contrôle  $(\mathbf{P}_0, \dots, \mathbf{P}_p)$ , on peut gérer un nombre  $n \in \mathbb{N}^*$  arbitraire de points  $(t_i)_{1 \leq i \leq n}$  discrétisant le tracé  $(\gamma(t_i))_{1 \leq i \leq n}$ . Ceci se réalise en pratique avec l'algorithme récursif de Casteljau<sup>5</sup>. Soit  $t \in [0, 1]$  un point de la paramétrisation, le calcul de  $\gamma(t)$  pour  $\gamma$  la courbe de Bézier contrôlée par  $(\mathbf{P}_0, \dots, \mathbf{P}_p)$  est donné dans l'algorithme 4.

Il suffit ensuite d'appliquer cet algorithme sur  $n$  points de discrétisation  $(t_i)_{1 \leq i \leq n}$  pour obtenir le tracé de la courbe. Si l'on rajoute une couche d'abstraction à notre mise en œuvre, on dispose alors d'un contrôle sur la discrétisation du tracé, indépendamment

4. D'après les travaux de M. Romain Petit, présenté au CANUM 2022.

5. Paul de Casteljau est un mathématicien français, à l'origine de l'algorithme éponyme qu'il a développé dans les années 50 pour le compte du constructeur automobile Citroën. Historiquement, cet algorithme a précédé le cadre théorique permettant de définir les courbes de Bézier, pour sa part remontant aux années 80.

**Algorithme 4** : Algorithme de Casteljau.

---

**Entrées** : Points de contrôle  $\mathbf{P}_0, \dots, \mathbf{P}_p$ .

- 1 **pour**  $0 \leq i \leq p$  **faire**
- 2 |  $\beta[i] = \mathbf{P}_i$
- 3 **fin**
- 4 **pour**  $1 \leq j \leq p$  **faire**
- 5 | **pour**  $0 \leq i \leq p - j$  **faire**
- 6 | |  $\beta[i] = (1 - t) \times \beta[i] + t \times \beta[i + 1]$
- 7 | **fin**
- 8 **fin**

**Résultat** :  $\beta[0]$ , qui est la position estimée de  $\gamma(t)$ .

---

du nombre de points de contrôle. De surcroît, les courbes de Bézier permettent de représenter des courbes lisses avec peu de points de contrôle, contrairement aux courbes polygonales. *A contrario*, il est plus difficile pour une courbe de Bézier de modéliser des courbes affines par morceaux, les courbes polygonales étant plus adaptées à cette fin.

En reprenant les opérateurs et les observations utilisés plus haut pour le cas polygonal, nous allons illustrer la reconstruction avec une paramétrisation Bézier. Précisons au passage une considération numérique utile pour les étapes d'estimation du support : pour trouver les points de contrôle  $\mathbf{P}$  d'une courbe  $\gamma$  à partir de sa discrétisation polygonale, il est nécessaire de minimiser la fonction  $\mathbf{P} \mapsto \|\Phi \nu_\gamma - \Phi(\mathbf{F}(\mathbf{P}))\|_2^2$  avec  $\mathbf{F}$  l'algorithme de Casteljau décrit en algorithme 4. La spirale  $\mu_1$  présentée plus haut est reproduite avec une discrétisation Bézier en figure 4.7, à gauche dans sa version floutée, bruitée; et à droite avec ses points de contrôle. En pratique, la vérité-terrain sera constituée de la spirale  $\mu_1$  avec une discrétisation **polygonale**.

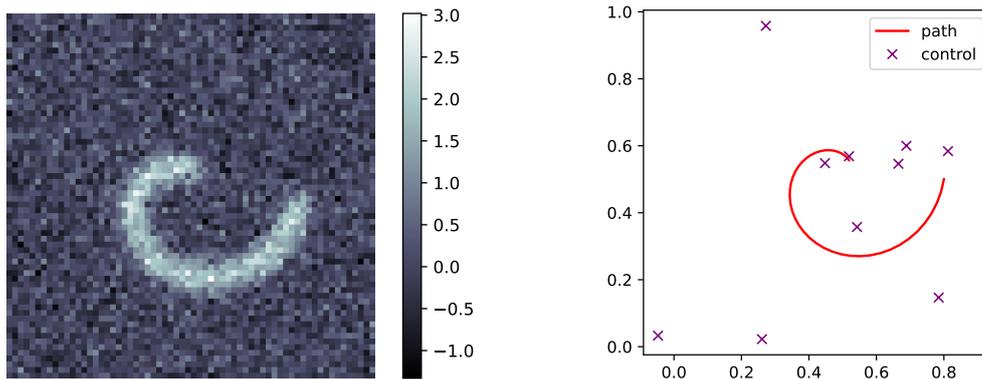


FIGURE 4.7 – À gauche : Une image  $g$  issue de l'observation de  $T$  composée d'une spirale  $\mu_1$ . À droite : le support que nous souhaitons reconstruire, par exemple ici paramétré par une courbe de Bézier contrôlée par les croix.

Il est intéressant de noter que les points de contrôle ne s'agglutinent pas aux mêmes endroits, et que l'ajout de point de contrôle ne provoque pas de phénomène d'oscillations.

L'algorithme 3 est alors adapté pour cette version en courbe de Bézier. L'optimisation ne se fait donc plus sur la liste des positions échantillonnant la courbe, mais sur les points de contrôle. En pratique, la mise en œuvre reste sensiblement similaire, dans le sens où la connaissance des points de contrôle permet de générer les points d'une courbe<sup>6</sup>, et nous retrouvons alors le cadre polygonal détaillé plus haut ; à ceci près que l'on travaille ici avec les points de contrôle. Les résultats obtenus par notre algorithme sont présentés dans la figure 4.8.

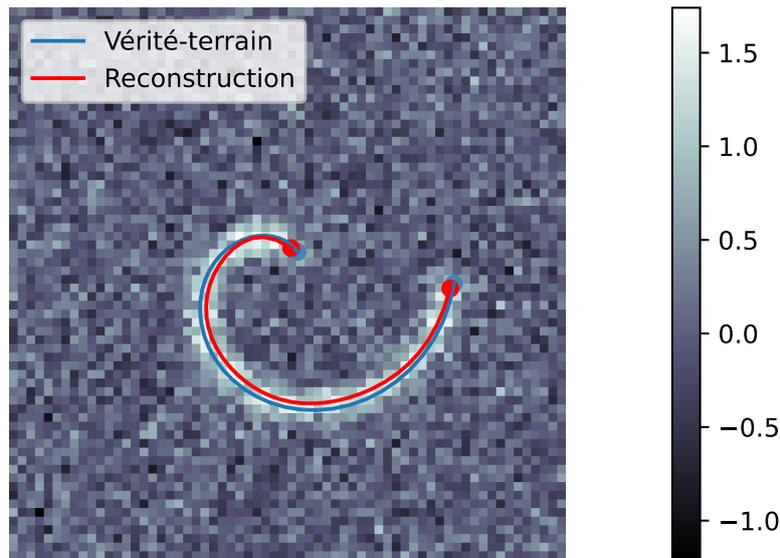


FIGURE 4.8 – Reconstruction de la source par l'algorithme *Charge Sliding Frank-Wolfe* avec une discrétisation en courbes de Bézier.

La reconstruction n'est paramétrée que par dix points de contrôle, contre cent points pour la discrétisation polygonale de la vérité-terrain<sup>7</sup>. Ainsi, contrairement aux courbes polygonales, nous pouvons décrire correctement une courbe régulière avec peu de points d'optimisation, ce qui est un véritable atout pour les applications dans lesquelles il sera nécessaire d'estimer un nombre conséquent de courbes, le temps de calcul en étant d'autant plus réduit.

**Attention** — La courbe reconstruite est plus courte que la vérité-terrain, ce qui est une conséquence assez logique de la régularisation par la norme  $\|\cdot\|_{\mathcal{C}^1}$ , et par extension  $\|\cdot\|_{\text{TV}^2}$  qui pénalise la longueur de la courbe (voir chapitre 3).

Il convient également de noter que l'ajout de points de contrôle ne dégrade pas l'estimation : il n'y a pas d'oscillations artificielles introduites par des points de discrétisation trop proches. Les points de contrôle peuvent se répartir librement dans  $\mathbb{R}^2$ , et

6. on prendra typiquement  $n = 100$  points.

7. Compte tenu du bruit il n'est pas nécessaire d'en rajouter plus.

l'optimisation ne semble pas souffrir d'une distance de séparation minimale. En contrepartie, l'algorithme tend empiriquement à sous-estimer plus fortement la longueur de la courbe que dans le cas polygonal. Toutes les caractéristiques des deux discrétisations sont finalement rappelées dans la table 4.1.

Discrétisation	Polygonale	Bézier
Tracé de $\gamma$	Direct	Algorithme de Casteljau
Régularité	Beaucoup de points	Peu de points
Oscillations	Si trop de points	Non
Prise en compte des discontinuités	Aisée	Difficile
Sous-estimation de la longueur	Très légère	Légère

TABLE 4.1 – Avantages et inconvénients des deux discrétisations.

**Résumé** — La paramétrisation Bézier est une bonne alternative à la paramétrisation polygonale. Elle permet de modéliser efficacement des courbes régulières avec peu de points de contrôle, et s'affranchit des considérations sur le choix des points de discrétisation qui provoquent des oscillations.

#### 4.3.5 Des pistes d'amélioration

Nous recueillons dans la suite les pistes d'amélioration dégagées dans ce chapitre :

- la modélisation *vectorielle* implique le calcul d'un gradient de l'image observée, qui n'est pas aisé eu égard aux phénomènes venant perturber ou dégrader l'observation. L'utilisation d'un terme *scalaire* pour le terme d'attache aux données, dans le sens où l'image n'aurait qu'une composante en chaque point, permettrait de simplifier cette démarche. Ceci étant, cette modification n'est pas aisée dans la mesure où le calcul du  $\Phi^*$  ne semble plus aussi trivial : quelle application ferait le lien entre une acquisition « scalaire » et un certificat vectoriel ?
- l'estimation du support reste assez grossière puisqu'elle ne résout pas de problème d'optimisation et repose sur un seuillage qui produit des résultats nécessairement incomplets. Il faudrait peut-être chercher à résoudre un problème sur l'ensemble des fonctions Lipschitz (qui est lui un espace vectoriel) plutôt que  $\mathcal{S}(\mathcal{X})$ , mais nous n'avons quoi qu'il en soit pas encore trouvé de démarche satisfaisante dans la littérature pour l'optimisation de la circulation du certificat  $\eta$  sur  $\gamma$  ;
- une modélisation avec des courbes *splines* pourrait être pertinente, bien qu'elle présente probablement les mêmes avantages et inconvénients qu'une discrétisation polygonale, qui n'est jamais qu'un *spline* d'ordre un ;

- les courbes que nous avons présentées dans ce chapitre ne se croisaient pas. C’est pourtant une situation que l’on rencontre souvent en pratique dans les données réelles, voir chapitres 1 et 2. Des discussions avec des chercheurs de la communauté sans-grille ont pu soulever quelques réserves sur ce problème : il semblerait que la reconstruction soit nécessairement composée de mesures singulières, ou dit autrement qu’elles soient des points extrémaux aux supports **disjoints**. Ce phénomène est attesté pour les mesures de Dirac dynamiques, les mesures ponctuelles dont la trajectoire au cours du temps est une courbe : les chemins reconstruits s’évitent nécessairement [12, 44]. Pour résoudre ce problème, nous préparons un article qui propose dans le cadre dynamique de réaliser un *lift* du problème dans l’espace des roto-translations [18], et de formuler une nouvelle fonctionnelle régularisée par la métrique de Reeds-Shepp [71, 40] afin de reconstruire des courbes dotées d’une information d’orientation. Ce *lift* permet ainsi de décroiser des courbes en relevant la dimension du problème : les premiers résultats sont concluants, mais il n’est pas tout à fait évident que ce qui fonctionne dans le cadre dynamique (et qui incorpore donc une information temporelle) s’étende naturellement à  $\mathcal{V}$ . Les liens entre ces problèmes restent encore à explorer et à clarifier.

#### 4.4 Conclusion

Grâce aux résultats théoriques sans-grille établis au chapitre 3, nous avons pu étudier la mise en œuvre numérique de la reconstruction sans-grille de courbes. Nous avons discuté des problèmes liés au cadre vectoriel pour la modélisation de l’opérateur d’acquisition, ainsi que les questions liées à la discrétisation et la paramétrisation des courbes numériques. Puis, nous avons proposé un nouvel algorithme, qui repose sur l’algorithme du gradient conditionnel, pour reconstruire itérativement une combinaison linéaire d’atomes, précisément ici des courbes. Les résultats ont été illustrés pour des paramétrisations polygonales ainsi que par courbes de Bézier, puis mis en perspective. Dans de prochains travaux, il serait pertinent d’effectuer des expériences avec des données réelles en microscopie de fluctuation, par exemple avec une adaptation de  $\Lambda$  du chapitre 2 ; tout en creusant certaines propriétés numériques de notre algorithme : une équivalence du taux de convergence en temps fini du *Sliding Frank-Wolfe* valide dans le cas des mesures ponctuelles, par exemple, serait une propriété assez utile.

## Chapitre 5

# Conclusion et perspectives

Ce manuscrit de thèse s'est proposé de détailler nos contributions en termes de construction mathématique et de mise en œuvre numérique, dans l'optique de résoudre des problèmes rencontrés en imagerie biomédicale et en particulier dans le cadre de la microscopie de fluorescence.

Plus précisément, nous avons présenté l'état de l'art des méthodes sans-grille, que nous avons mises en œuvre dans une bibliothèque qui se veut accessible et que nous avons illustrées sur des images SMLM. Nous avons ensuite construit une méthode sans-grille en microscopie de fluctuations, à travers la formulation d'un opérateur et d'une énergie exploitant une hypothèse d'indépendance statistique des fluctuations. Devant l'impossibilité de reconstruire des structures biologiques supportées par des courbes dans ce cadre, et souhaitant combler le fossé théorique entre reconstruction de mesures ponctuelles et surfaciques, nous avons proposé une nouvelle approche permettant la reconstruction de mesures portées par des courbes. L'introduction d'un espace de mesures, la formulation d'une nouvelle énergie appelée CROC ainsi que la preuve d'un théorème de caractérisation géométrique des solutions de CROC comme mesures portées par des courbes ont permis la construction de cette nouvelle approche. Finalement, nous avons proposé un algorithme mettant en œuvre cette méthode sans-grille reconstruisant des courbes, et nous avons dégagé les pistes d'amélioration qui se devraient d'être explorées.

Pour conclure ce manuscrit, nous détaillons précisément dans la suite nos contributions, qui ont elles-mêmes dégagé de nouvelles perspectives que nous nous efforcerons de détaillons dans la suite.

### 5.1 Conclusions et contributions

#### Reconstruction de mesures ponctuelles à partir d'une image

Le premier chapitre fournit un état de l'art sur les méthodes variationnelles sans-grille pour la reconstruction de sources ponctuelles. Tous ces résultats présentent un réel intérêt pratique, qui est démontré par l'application de ces méthodes sans-grille

sur des problèmes concrets tels que le problème de super-résolution. Nous avons ainsi montré l'utilité de l'approche sans-grille à travers un exemple dans le cadre de la microscopie de fluorescence SMLM : la reconstruction obtenue est de qualité comparable aux productions de l'apprentissage profond, en ne nécessitant pas de jeux de données et avec la seule connaissance de l'opérateur d'observation. Ces travaux ont été publiés dans [68].

### Reconstruction de mesures ponctuelles en microscopie de fluctuations

Nous avons formulé une modélisation temporelle du cadre d'acquisition en microscopie de fluorescence par fluctuation. La source est une mesure dynamique, combinaison linéaire de mesures de Dirac aux positions fixes mais aux amplitudes fluctuantes au cours du temps. À l'aide de l'hypothèse d'indépendance des fluctuations de luminosité des fluorophores, il est possible de construire un opérateur qui fait le lien entre matrice de covariance spatiale et une mesure de Radon sans dépendance temporelle : cette mesure partage les mêmes positions que la mesure source. Les résultats de la méthode sans-grille sur des données synthétiques et expérimentales illustrent la capacité de reconstruction des détails fins par cette approche, permettant par exemple de déduire un enlacement de filaments oblitéré par le flou de la diffraction. Les résultats sont néanmoins en pointillé, ce qui n'est que la conséquence logique de la régularisation promouvant la géométrie ponctuelle. Eu égard à cette limite théorique, il n'est pas possible en pratique de reconstruire des solutions exhibant une géométrie de courbe, qui constitue une part pourtant non négligeable des structures à imager dans ce type de microscopie de fluorescence. Ces résultats ont été publiés dans [65].

### Un nouveau régulariseur : la divergence pour les mesures portées par des courbes

Nous avons introduit un nouveau terme de régularisation, qui utilise la divergence pour incorporer l'*a priori* structurel de solutions en courbe. Après avoir évoqué différentes pistes de recherche pour reconstruire des courbes, nous avons discuté de l'espace  $\mathcal{V}$  des mesures vectorielles à divergence finie et prouvé quelques propriétés algébriques (espace de Banach, convergence de sous-suites) ainsi que fonctionnelles (éléments qui y vivent ou qui en sont exclus). La mesure  $\mu_\gamma$  portée par une courbe  $\gamma$  est un exemple d'élément de  $\mathcal{V}$ . Puis, nous avons défini une fonctionnelle appelée CROC sur cet espace, qui possède au moins une solution et dont nous pouvons définir les certificats duaux que nous avons explicités. Finalement, nous avons prouvé le résultat le plus essentiel de cette thèse, à savoir que les points extrémaux de la boule unité de la norme  $\|\cdot\|_\gamma$  sont précisément les mesures portées par une courbe simple Lipschitz. Ces travaux ont été publiés dans [67].

### **Une mise en œuvre numérique de la reconstruction des mesures portées par des courbes**

Nous avons discuté des problèmes liés au cadre vectoriel pour la modélisation de l'opérateur d'acquisition, ainsi que les questions liées à la discrétisation et la paramétrisation des courbes numériques. Puis, nous avons proposé un nouvel algorithme, qui repose sur l'algorithme du gradient conditionnel, pour reconstruire itérativement une combinaison linéaire d'atomes, précisément ici des courbes. Les résultats ont été illustrés pour différentes paramétrisations et mis en perspective. Ces résultats ont été publiés dans [66], et donneront lieu à une version étendue dans *Journal of Mathematical Imaging and Vision* de l'éditeur Springer.

### **5.2 *The Off-the-grid Cookbook* : des perspectives en sans-grille**

Nous clôturons cette thèse en discutant maintenant des perspectives ouvertes au fil des chapitres, que nous avons rassemblées dans la suite :

#### **Une bibliothèque à destination des biologistes**

La préparation de cette thèse a présidé à la création d'un module appelé `offgrid` en *python* se proposant de rassembler les méthodes les plus pertinentes en méthode sans-grille. À travers un travail sur l'efficacité des opérations (par exemple de convolution ou d'auto-dérivation) et sur le choix des implémentations pratiques, notre bibliothèque s'efforce de mettre à disposition la théorie sans-grille au plus grand-nombre.

Dans cette optique, nous avons commencé le développement d'une version GUI, sur la plateforme *Napari*, pour permettre à un opérateur de manipuler toutes ces méthodes simplement depuis une interface graphique. Ce travail est encore parcellaire, puisqu'il a été mis en œuvre uniquement sur la reconstruction de mesures de Dirac statiques. Il serait ainsi pertinent d'incorporer tous les algorithmes de `offgrid` dans ce paquet, tels que les méthodes sans-grille pour la microscopie de fluctuations, pour les mesures portées des courbes, *etc.*

#### **Reconstruction de courbes en covariance**

La définition de l'opérateur de covariance  $\Lambda$  présentée au chapitre 2 pourrait être étendue aux mesures portées par des courbes. Plus précisément, il faudrait adapter la reconstruction de courbes au cadre de la microscopie de fluctuations. Il est alors nécessaire de relever plusieurs autres défis pour y parvenir pleinement, les points qui suivent en sont une illustration.

#### **Variations d'amplitudes et décroissement**

Par construction, les amplitudes le long d'une mesure portée par une courbe de  $\mathcal{V}$  sont constantes. Cependant, ces amplitudes varient en pratique, avec en particulier

un aspect « guirlandes de Noël » en microscopie de fluctuations. La luminosité de la structure biologique courbe peut ainsi être nulle sur certaines portions, et le modèle n'en tient actuellement pas compte. Dans la même veine, les filaments, vaisseaux sanguins, *etc.* imagés en dimension deux sont régulièrement enchevêtrés. Cependant, des phénomènes (pour une mesure source qui est courbe, surfacique, *etc.*) constatés en pratique semblent indiquer qu'il n'est pas possible de reconstruire itérativement des mesures aux supports s'intersectant. La reconstruction sur  $\mathcal{V}$  pourrait à ce titre bénéficier des idées développées dans le cadre dynamique, que nous abordons ci-dessous.

### Mesures dynamiques en localisation par ultrasons

L'état de l'art en problème sans-grille dynamique permet la reconstruction de mesures ponctuelles se déplaçant au cours du temps, grâce à une minimisation par l'énergie de Benamou-Brenier faisant intervenir  $\mathcal{M}(\Gamma)$  (à un plongement près). Un futur article de notre cru, aujourd'hui en cours de rédaction, propose d'effectuer un *lift* du problème dans l'espace des roto-translations  $\mathbb{R}^2 \times \mathbb{S}^1$ , d'étendre l'énergie de Benamou-Brenier à n'importe quelle métrique *a minima* Riemannienne et de formuler une nouvelle fonctionnelle régularisée par la métrique de Reeds-Shepp qui reconstruirait des courbes dotées d'une information d'orientation. Ce *lift* permet ainsi de décroiser des courbes en relevant la dimension du problème. La méthode devrait être testée dans un cadre de microscopie de localisation par ultrasons (ULM), où la dimension dynamique intervient dans le suivi de chaque bulle.

À ce titre, il serait également pertinent de s'interroger sur la dimension aveugle du problème inverse : la connaissance de la PSF est partielle en imagerie ULM, et le problème pourrait bénéficier d'une formulation du terme d'attaches aux données relaxées. L'idée serait de travailler non pas avec des moindres carrées, mais avec la distance de Wasserstein  $\mathcal{W}^2$ .

### Des propriétés supplémentaires sur $\mathcal{V}$

Nous avons prouvé quelques propriétés sur  $\mathcal{V}$  (espace de Banach, convergence des suites bornées, structure de la boule unité de la norme, *etc.*), mais il reste encore beaucoup de questions en suspens à ce jour. Il serait par exemple naturel de se demander si les mesures portées par des courbes  $\mu_\gamma \in \mathfrak{G}$  sont denses dans  $\mathcal{V}$ , au moins pour la topologie faible-\*. Dans ce même esprit, une caractérisation plus fine des faces de la boule unité de  $\|\cdot\|_{\mathcal{V}}$  pourrait offrir un nouvel éclairage sur certains phénomènes observés numériquement, tels que les courbes fermées ayant tendance à s'ouvrir au cours de la reconstruction. En outre, les ensembles  $\mathcal{V}$  et  $\mathcal{M}(\Gamma)$  semblent reliés : les travaux d'Aneva Doliciane Tsafack au sein de l'équipe Morpheme vont dans ce sens, en établissant une « égalité » entre  $\langle \mu_\gamma, \varphi \rangle$  et  $\langle 1 \cdot \delta_\Gamma, \psi \rangle$  pour  $\delta_\Gamma \in \mathcal{M}(\Gamma)$ , faisant ainsi le lien entre mesures vectorielles et scalaires. La reconstruction sans-grille de courbes pourrait alors bénéficier des résultats sur  $\mathcal{M}(\Gamma)$  où l'on sait incorporer une variation d'amplitude le long des courbes.

D'autres interrogations restent à ce jour sans réponse, et risquent de le rester : l'identification du prédual de  $\mathcal{V}$  par exemple n'a pas encore été élucidée, et par analogie avec BV, risque fort de le rester.



# Annexe 1 : précisions sur le chapitre 1

## Définitions et notions utiles

### Détails d'analyse fonctionnelle

**Définition .1** (Banach). Soit  $V$  un espace vectoriel normé, on dit que  $V$  est un espace de Banach s'il est complet pour la distance induite par sa norme.

**Définition .2** (Hilbert). Soit  $V$  un espace de Banach, on dit que c'est un espace de Hilbert si sa norme est induite par un produit scalaire ou hermitien.

**Définition .3** (Mesures singulières). Deux mesures de Radon  $\mu$  et  $\nu$  de  $\mathcal{M}(\mathcal{X})$  sont dites singulières s'il existe deux sous-ensembles disjoints  $A, B$  de la tribu  $\sigma$  de  $\mathcal{X}$  dont l'union est  $\mathcal{X}$ ; de telle sorte que  $\mu$  est nul sur tous les sous-ensembles mesurables de  $B$  tandis que  $\nu$  est nul sur tous les sous-ensembles mesurables de  $A$ .

**Proposition .1** (Décomposition de Jordan). Soit  $\mu \in \mathcal{M}(\mathcal{X})$ , la décomposition de Jordan indique qu'il existe deux mesures de Radon positives  $\mu^+, \mu^- \in \mathcal{M}^+(\mathcal{X})$ , associée à  $\mu$ , ces deux premières mesures sont singulières et on a  $\mu = \mu^+ - \mu^-$ .

**Définition .4** (Topologie faible-\*). La convergence faible-\* peut être vue comme la convergence localement en moyenne. On dit qu'une suite de mesures de Radon  $(m_n)_{n \geq 0}$  converge faiblement-\* vers  $m \in \mathcal{M}(\mathcal{X})$  si et seulement si pour tout  $f \in \mathcal{C}_0(\mathcal{X})$  :

$$\int_{\mathcal{X}} f \, dm_n \xrightarrow{n \rightarrow +\infty} \int_{\mathcal{X}} f \, dm.$$

On la note  $m_n \xrightarrow{*} m$ , cette convergence est aussi appelée la convergence vague.

**Définition .5** (Réflexivité). Un espace vectoriel  $E$  est dit réflexif si le bidual  $E^{**}$  s'identifie à  $E$ .

**Remarque.** Puisque  $\mathcal{C}_0(\mathcal{X})$  n'est pas un espace réflexif lorsqu'il est muni de sa norme supremum, le dual de  $\mathcal{M}(\mathcal{X})$  pour la topologie induite par sa norme-TV est un espace compliqué, strictement plus grand [60] que  $\mathcal{C}_0(\mathcal{X})$ . Néanmoins, si  $\mathcal{M}(\mathcal{X})$  est équipé de sa topologie faible-\*, alors  $\mathcal{M}(\mathcal{X})$  est un espace localement convexe dont le dual est précisément  $\mathcal{C}_0(\mathcal{X})$  [43].

Nous précisons également la notion de métrisabilité, utile pour la définition des quantités en transport optimal :

**Définition .6.** *Un espace topologique  $(E, \mathcal{T})$  est dit métrisable s'il existe une distance  $d : E \times E \rightarrow [0, +\infty[$  telle que la topologie induite par celle-ci  $d$  s'identifie à  $\mathcal{T}$ .*

$(\mathcal{M}(\mathcal{X}), *)$  n'est pas un espace à bases dénombrables de voisinages, et par conséquent n'est pas un espace métrisable. Pour s'en convaincre :

**Lemme .1.** *Si  $E$  est un espace de Banach, la topologie faible-\* n'est pas métrisable sur  $E^*$ , sauf si  $E$  est de dimension finie.*

Tous les sous-ensembles bornés de  $\mathcal{M}(\mathcal{X})$  sont néanmoins métrisable pour la topologie faible-\*. Cette propriété est capitale pour la bonne définition des métriques classiques en transport optimal telles que la distance de Wasserstein, ainsi que pour la preuve de  $\Gamma$ -convergence du LASSO vers le BLASSO [75, 41]. Pour résumer toutes les propriétés de ces différentes topologies, nous fournissons le tableau 1 :

Propriétés	Topologie TV	Topologie faible-*
Complétude	Oui	Sur les bornés
Séparabilité	Non	Yes
Réflexivité	Non	Yes
Métrisable	Oui	Sur les bornés
Espace polonais	Non	Sur les bornés

TABLE 1 – Propriétés algébriques de  $\mathcal{M}(\mathcal{X})$  pour ses deux principales topologies.

## Preuve de la dualité de Fenchel

**Proposition .2.** *Soit le problème :*

$$\operatorname{argmax}_{\|\phi^* p\|_{\infty, \mathcal{X}} \leq 1} \langle y, p \rangle_{\mathcal{H}} - \frac{\lambda}{2} \|p\|_{\mathcal{H}}^2.$$

*C'est le problème dual du BLASSO ( $\mathcal{P}_\lambda(y)$ ).*

*Démonstration.* Nous allons utiliser les résultats de [45, Remarque 4.2], avec une petite subtilité qu'il convient de noter : l'espace de Banach  $V$  doit être réflexif, ce qui n'est clairement pas le cas pour  $V = \mathcal{M}(\mathcal{X})$ . Ceci étant, l'hypothèse de réflexivité n'est nécessaire que pour la preuve d'existence. Comme nous avons déjà prouvé l'existence de solutions, cette contrainte n'est pas nécessaire dans notre cas. Si l'on revient à la remarque 4.2, elle précise, pour  $\Lambda : V \rightarrow Y$  linéaire,  $F : V \rightarrow \mathbb{R}$  convexe, que le problème primal

$$\inf_{u \in V} F(u) + G(\Lambda u)$$

a un problème dual qui s'écrit :

$$\sup_{p^* \in Y^*} -F^*(\Lambda^* p^*) - G^*(-p^*). \quad (1)$$

Si  $u$  et  $p^*$  sont respectivement solutions des problèmes primal et dual, les conditions d'extrémalité s'écrivent :

$$\begin{cases} \Lambda^* p^* \in \partial F(u) \\ -p^* \in \partial G(\Lambda u). \end{cases}$$

Précisons d'emblée les fonctions en jeu, à savoir  $V \stackrel{\text{def.}}{=} \mathcal{M}(\mathcal{X})$ ,  $Y \stackrel{\text{def.}}{=} \mathcal{H}$ ,  $F(m) \stackrel{\text{def.}}{=} |m|(\mathcal{X})$  et  $G(p) \stackrel{\text{def.}}{=} \frac{1}{2} \|y - p\|_{\mathcal{H}}^2$ . Il est clair que l'adjoint de  $G$  s'écrit  $G^*(p^*) = \langle y, p^* \rangle_{\mathcal{H}} + \frac{1}{2} \|p^*\|_{\mathcal{H}}^2$  pour  $p^* \in \mathcal{H}$ ; soit maintenant  $\psi \in V^* \stackrel{\text{def.}}{=} \mathcal{C}_0(\mathcal{X})$  :

$$\begin{aligned} F^*(\psi) &= \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathcal{M}} - |m|(\mathcal{X}) \\ &\geq \langle \psi, m \rangle_{\mathcal{M}} - |m|(\mathcal{X}), \quad \forall m \in \mathcal{M}(\mathcal{X}). \end{aligned}$$

Soit  $x \in \mathcal{X}$ , et  $m = \lambda \delta_x$  avec  $\lambda > 0$ . Alors, on a :

$$\sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathcal{M}} - |m|(\mathcal{X}) \geq \lambda(\psi(x) - 1).$$

Lorsque  $\lambda \rightarrow +\infty$  il vient  $F^*(\psi) \geq +\infty$  si  $\psi(x) > 1$ . Un résultat similaire pour  $\psi(x) < 1$  est obtenu avec la mesure  $m = -\lambda \delta_x$ . On obtient finalement  $F^*(\psi) = +\infty$  si  $\|\psi\|_{\infty, \mathcal{X}} > 1$ . Supposons que  $\|\psi\|_{\infty, \mathcal{X}} \leq 1$ , nous pouvons déjà noter que  $F^*(\psi) = \sup_{m \in \mathcal{M}(\mathcal{X})} \langle \psi, m \rangle_{\mathcal{M}} - |m|(\mathcal{X}) \geq 0$  (cas  $m = 0$ ). De plus,

$$\begin{aligned} \langle \psi, m \rangle_{\mathcal{M}} - |m|(\mathcal{X}) &\leq \|\psi\|_{\infty, \mathcal{X}} |m|(\mathcal{X}) - |m|(\mathcal{X}) \\ &\leq |m|(\mathcal{X})(\|\psi\|_{\infty, \mathcal{X}} - 1) \\ &\leq 0 \quad \text{puisque } \|\psi\|_{\infty, \mathcal{X}} \leq 1. \end{aligned}$$

En introduisant le sup de part et d'autre de la dernière inégalité, on obtient finalement  $F^*(\psi) = 0$  si  $\|\psi\|_{\infty, \mathcal{X}} \leq 1$  ce qui donne ainsi la condition sur la norme *supremum*.

Finalement, par l'équation (1), nous obtenons le problème dual :

$$\sup_{\|\Phi^* p^*\|_{\infty, \mathcal{X}} \leq 1} \langle y, p^* \rangle_{\mathcal{H}} - \frac{1}{2} \|p^*\|_{\mathcal{H}}^2,$$

et pour  $m^*$  et  $p^*$  respectivement solutions des problèmes primal et dual, les conditions d'extrémalité s'écrivent :

$$\begin{cases} \Phi^* p^* \in \partial |m^*|(\mathcal{X}) \\ -p^* = \Phi m^* - y. \end{cases}$$

Ceci conclut la preuve. □

### Précisions sur les algorithmes par flot de gradient

Tous les résultats suivants sont prouvés pour un domaine  $\mathcal{X}$  sans bord, *p.ex.* le  $d$ -tore  $\mathbb{T}^d$ . Le cas décrit dans les sections précédentes –  $\mathcal{X}$  est un compact quelconque de  $\mathbb{R}^d$  – est inclus dans ce nouveau cadre, puisque tout compact  $\mathcal{X}$  peut être périodisé pour donner un domaine sans bord. Le noyau de l'opérateur direct  $\varphi : \mathcal{X} \rightarrow \mathcal{H}$  doit également être Fréchet-différentiable. Le terme des moindres carrés dans le BLASSO est désigné par le terme de données plus général  $R : \mathcal{H} \rightarrow \mathbb{R}^+$ , la fonctionnelle  $T_\lambda$  du BLASSO sera maintenant restreinte à  $\mathcal{M}^+(\mathcal{X})$  et notée  $J$ ; sa différentielle de Fréchet au point  $\nu \in \mathcal{M}^+(\mathcal{X})$  est notée  $J'_\nu$  :

$$J(\nu) = \|y - \Phi \nu\|_{\mathcal{H}}^2 + \lambda |\nu|(\mathcal{X}), \quad (2)$$

$$J'_\nu(x) = \langle \varphi(x), \nabla R \rangle_{\mathcal{H}} + \lambda \quad \text{pour tout } x \in \mathcal{X}. \quad (3)$$

Dans ce qui suit, nous décrivons le cadre de travail pour les mesures positives de  $\mathcal{M}^+(\mathcal{X})$ , mais il peut être étendu sans difficulté [21] aux mesures signées de  $\mathcal{M}(\mathcal{X})$  en appliquant la méthode à la partie positive puis négative de la mesure signée (voir la décomposition de Jordan évoquée plus haut, en annexe 5.2). La figure 1 résume les principales quantités et relations introduites dans cette section, le lecteur est invité à s'y référer chaque fois qu'il aura besoin d'une vue d'ensemble sur le problème de transport optimal.

L'optimisation parcimonieuse par transport optimal sur les mesures [20, 21] repose sur l'approximation de la mesure source  $m_{a_0, x_0}$  par un « système de  $N \in \mathbb{N}^*$  particules », c'est-à-dire un élément de l'espace  $\Omega^N \stackrel{\text{def.}}{=} (\mathbb{R}^+ \times \mathcal{X})^N$ . Il s'agit alors d'estimer la source par une optimisation utilisant le gradient de la fonction objective :

$$F_N((r_1, x_1), \dots, (r_N, x_N)) \stackrel{\text{def.}}{=} \left\| y - \frac{1}{N} \sum_{i=1}^N r_i^2 \varphi(x_i) \right\|_{\mathcal{H}}^2 + \frac{\lambda}{N} r_i^2, \quad (4)$$

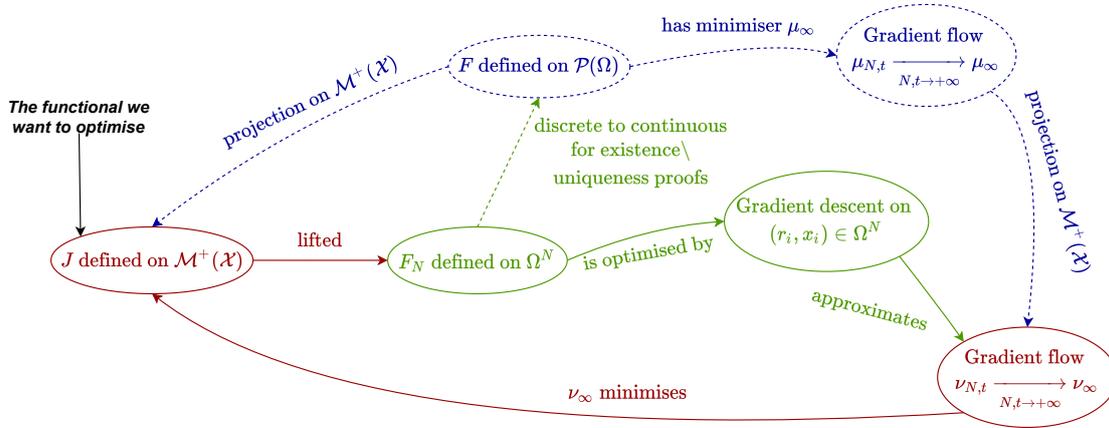


FIGURE 1 – Résumé des quantités importantes mentionnées dans [20, 21] : le rouge renvoie aux quantités de  $\mathcal{M}^+(\mathcal{X})$ , le vert à  $\Omega^N \stackrel{\text{def.}}{=} (\mathbb{R}^+ \times \mathcal{X})$  et le bleu à l'espace de Wasserstein  $\mathcal{P}_2(\Omega)$  et aux résultats théoriques. Les lignes en pointillé correspondent à la partie théorique, et les lignes continues à la partie numérique.

où  $(r_i, x_i)$  appartiennent à l'espace augmenté  $\Omega \stackrel{\text{def.}}{=} \mathbb{R}^+ \times \mathcal{X}$ , muni d'une métrique. On espère donc que la descente de gradient sur  $F_N$  converge vers les amplitudes et les positions de la mesure source, et ce, en dépit de la non-convexité de la fonctionnelle (4). L'auteur de [21] propose la définition d'une métrique appropriée pour le gradient de  $F_N$ , qui permet la séparation des variables dans chaque itération de la descente de gradient. Soit  $\alpha, \beta$  deux paramètres tels que  $\alpha > 0$  et  $\beta > 0$  et pour tout  $(r, \theta) \in \Omega$ , nous définissons le produit scalaire riemannien de  $\Omega$  appelé *métrique du cône* équipant  $\Omega$  est défini par  $\forall (\delta r_1, \delta r_2) \in \mathbb{R}_+^2, \forall (\delta \theta_1, \delta \theta_2) \in \mathcal{X}^2$  :

$$\langle (\delta r_1, \delta \theta_1), (\delta r_2, \delta \theta_2) \rangle_{(r,\theta)} \stackrel{\text{def.}}{=} \frac{\delta r_1 \delta r_2}{\alpha} + r^2 \frac{\langle \delta \theta_1, \delta \theta_2 \rangle_\theta}{\beta}.$$

Nous désignons par  $\langle \cdot, \cdot \rangle_\theta$  la métrique sur la variété  $\mathcal{X}$  au point  $\theta$ . Le gradient de la fonctionnelle  $F_N$  pour tout  $i \in \llbracket 1, N \rrbracket$  pour la métrique du cône s'écrit [21, 29] :

$$\begin{cases} \nabla_{r_i} F_N = 2\alpha r_i J'_v(x_i) = -2\alpha r_i \lambda (\eta_\lambda - 1) \\ \nabla_{x_i} F_N = \beta \lambda \nabla J'_v(x_i) = -\beta \lambda \nabla \eta_\lambda \end{cases} \quad \text{avec } v \stackrel{\text{def.}}{=} \sum_{i=1}^N r_i^2 \delta_{x_i}, \eta_\lambda \stackrel{\text{def.}}{=} -J'_v / \lambda. \quad (5)$$

Nous présentons maintenant les résultats théoriques sur la descente de gradient particulière, qui correspondent aux lignes pointillées bleues de la figure 1. Le lecteur est invité à se référer à cette figure chaque fois qu'il en ressent le besoin, et ce afin de garder une bonne vue d'ensemble du problème.

### Résultats théoriques

L'idée principale des articles fondateurs [20, 21] peut se résumer à l'observation suivante : la minimisation de la fonction (4) est un cas particulier d'un problème plus général, formulé en termes de mesures appartenant à l'espace augmenté  $\Omega$ . L'espace est plus précisément  $\mathcal{P}_2(\Omega)$  sous-ensemble de  $\mathcal{M}(\Omega)$ , à savoir l'espace des probabilités de second moment fini, muni de la métrique de 2-Wasserstein c'est-à-dire la distance de transport optimal : voir l'annexe 4 pour plus de détails. Ainsi, l'augmentation (*lift*) de l'inconnue  $m \in \mathcal{M}^+(\mathcal{X})$  en  $\mu \in \mathcal{P}_2(\Omega)$  permet d'escamoter l'asymétrie dans les mesures discrètes entre la position  $x \in \mathcal{X}$  et l'amplitude  $a \in \mathbb{R}^+$  en transformant  $a\delta_x$  en  $\delta_{(a,x)}$ . L'énergie augmentée s'écrit désormais pour le paramètre  $\lambda > 0$  :

$$\forall \mu \in \mathcal{P}_2(\Omega), \quad F(\mu) \stackrel{\text{def.}}{=} \|y - \tilde{\Phi}\mu\|_{\mathcal{H}}^2 + \lambda \tilde{V}(\mu), \quad (6)$$

où  $\tilde{\Phi}\mu \stackrel{\text{def.}}{=} \int_{\Omega} \phi(a, x) d\mu(a, x)$  pour  $\phi(a, x) \stackrel{\text{def.}}{=} a\varphi(x)$  et  $\tilde{V}$  est la norme TV sur la composante spatiale de la mesure  $\mu$ . La fonctionnelle est non convexe, sa différentielle de Fréchet est notée  $F'$  et pour  $u \in \Omega$  :

$$F'(\mu)(u) \stackrel{\text{def.}}{=} \langle \tilde{R}'(\mu), \phi(u) \rangle_{\mathcal{H}} + \lambda,$$

avec  $\tilde{R}' \stackrel{\text{def.}}{=} \|y - \int_{\Omega} \nabla \phi(a, x) d\mu(a, x)\|_{\mathcal{H}}^2$ . Ainsi, une mesure discrète (c'est-à-dire somme pondérée de mesures de Dirac)  $\mu_N \stackrel{\text{def.}}{=} \frac{1}{N} \sum_i \delta_{a_i, x_i}$  de  $\mathcal{P}_2(\Omega)$  peut également être comprise comme un élément de  $\Omega^N$ , en prenant le point de vue de ses composantes  $(a_i, x_i)$ . Elle permet aux auteurs de [20, 21] de donner une caractérisation précise des conditions de reconstruction de la source, à travers les mesures et les outils du transport optimal tels que le flot de gradient (voir ci-dessous).

On peut alors effectuer une descente de gradient sur les amplitudes et les positions  $(a_i, x_i) \in (\mathbb{R}^+ \times \mathcal{X})^N$  de la mesure  $\mu_N$ , afin d'exploiter la différentiabilité du noyau  $\varphi$ . Notez que la mesure  $\mu_N$  est surparamétrée, c'est-à-dire que le nombre de  $\delta$ -pics qui la composent est plus important que le nombre de  $\delta$ -pics de la mesure vérité-terrain : les particules, à savoir les  $\delta$ -pics de l'espace  $\Omega$ , couvrent donc le domaine  $\mathcal{X}$  pour la composante spatiale; voir la figure 3 à titre d'exemple, où  $\mu_N$  est représenté par des points rouges.

Avant de donner les principaux résultats, nous devons clarifier la notion généralisée de descente de gradient aux mesures appelée *gradient flow* [1, 79] provenant de la théorie du transport optimal, et qui constitue l'ingrédient principal de la descente de gradient particulière. Soit  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  la fonction objectif, qui dispose d'une certaine régularité, un flot de gradient décrit l'évolution d'une courbe  $x(t)$  telle que son point de départ à  $t = 0$  est  $x_0 \in \mathbb{R}^d$ , évoluant à tout moment  $t$  vers la direction qui décroît le plus la fonction  $F$  [79].

$$\begin{cases} x'(t) = -\nabla F(x(t)) & \text{pour } t > 0 \\ x(0) = x_0. \end{cases}$$

L'intérêt du flot de gradient réside dans son extension aux espaces  $X$  qui ne possèdent pas de structure différentiable. Dans le cas différentiable, on peut considérer la discrétisation du flot de gradient comme la suite définie pour un pas  $\tau > 0, k \in \mathbb{N}^*$  par :

$$x_{k+1}^\tau \in \operatorname{argmin}_{x \in \mathcal{X}} F(x) + \frac{|x - x_k^\tau|^2}{2\tau}.$$

Il s'agit du schéma d'Euler implicite pour l'équation<sup>1</sup>  $(x^\tau)' = -\nabla F(x^\tau)$ . Le flot de gradient est alors la limite (sous certaines hypothèses) de la suite  $(x_k^\tau)_{k \geq 0}$  pour  $\tau \rightarrow 0$  pour un point de départ  $x_0 \in X$ . La notion de flot de gradient peut être étendue aux espaces métriques : en effet, pour un espace métrique  $(X, d)$  et une application  $F : X \rightarrow \mathbb{R}$  semi-continue inférieurement, on peut définir la discrétisation du flot de gradient par la suite :

$$x_{k+1} \in \operatorname{argmin}_{x \in \mathcal{X}} F(x) + \frac{d(x, x_k)^2}{2\tau}. \quad (7)$$

Dans le cas de l'espace métrique des mesures de probabilité, la limite  $\tau \rightarrow 0$  du schéma existe et converge vers l'unique flot de gradient commençant en  $x_0$ , un élément de l'espace métrique. Un cas typique se présente avec l'espace des probabilités de seconds moments finis  $\mathcal{P}_2(\Omega)$ , muni de la 2-métrique de Wasserstein (ou la distance de transport optimal, voir l'annexe 4) : un flot de gradient dans cet espace  $\mathcal{P}_2(\Omega)$  est une courbe  $t \mapsto \mu_t$  appelée *flot de gradient de Wasserstein* commençant à  $\mu_0 \in \mathcal{P}_2(\Omega)$ , pour tout  $t$  dans  $\mathbb{R}^+$  on a  $\mu_t \in \mathcal{P}_2(\Omega)$ , obéissant à l'équation aux dérivées partielles au sens des distributions<sup>2</sup> :

$$\partial_t \mu_t = -\operatorname{div}(\mu_t \nabla F'(\mu_t)). \quad (8)$$

Cette équation garantit la conservation de la masse, à savoir qu'à chaque instant  $t > 0$ , on a  $|\mu_t|(\Omega) = |\mu_0|(\Omega)$ . Ainsi, malgré l'absence d'une structure de différentiabilité sur  $\mathcal{P}_2(\Omega)$  qui interdit l'application directe d'un algorithme classique utilisant le gradient, on peut effectuer une optimisation sur l'espace grâce à un flot de gradient qui permet de minimiser  $F$  en discrétisant (8).

---

1. ou le schéma plus faible  $(x^\tau)' \in \partial F(x^\tau)$  si  $F$  est convexe et non lisse  
 2.  $\operatorname{div}(m) = \sum_{i=1}^d \frac{\partial m}{\partial x_i}$  pour tout  $m \in \mathcal{M}(\mathcal{X})$ . Les dérivées doivent être comprises au sens des distributions.

Un cas pertinent de flot de gradient dans  $\mathcal{P}_2(\Omega)$  est celui du flot débutant en  $\mu_{N,0} \stackrel{\text{def.}}{=} 1/N \sum_{i=1}^N \delta_{(a_i^0, x_i^0)}$ , défini de manière unique par l'équation (8), qui s'écrit pour tout  $t \in \mathbb{R}^+$  :  $\mu_{N,t} \stackrel{\text{def.}}{=} 1/N \sum_{i=1}^N \delta_{(a_i(t), x_i(t))}$  où  $a_i : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  et  $x_i : \mathbb{R}^+ \rightarrow \mathcal{X}$  sont des applications continues. Cette trajectoire  $(\mu_{N,t})_{t \geq 0}$  est un flot de gradient de Wasserstein, et utilise  $N$  mesures de Dirac sur  $\Omega$  pour optimiser la fonction objective  $F$  dans (6). Lorsque le nombre de particules  $N$  tend vers l'infini et si  $\mu_{N,0}$  converge vers un certain  $\mu_0 \in \mathcal{P}_2(\mathcal{X})$ , le flot de gradient  $(\mu_{N,t})_{t \geq 0}$  converge pour  $N \rightarrow \infty$  vers l'unique flot de gradient de Wasserstein de  $F$  débutant en  $\mu_0$ , décrit par la densité dépendante du temps  $(\mu_t)_{t \geq 0}$  évaluée dans  $\mathcal{P}_2(\mathcal{X})$  et obéissant à cette dernière équation aux dérivées partielles (8).

Pour ces flots de gradient non convexes, les auteurs de [20] fournissent un résultat concret : sous certaines hypothèses, le flot de gradient  $(\mu_{N,t})_{t \geq 0}$  converge vers un *minimum* global pour la limite de sur-paramétrisation c'est-à-dire pour  $N \rightarrow +\infty$ . Elle repose sur deux hypothèses importantes qui empêchent l'optimisation d'être bloquée en des points non optimaux :

- l'homogénéité<sup>3</sup> de  $\phi$  afin de sélectionner la bonne amplitude pour chaque caractéristique, ou tout du moins la 1-homogène partielle (c'est-à-dire  $\phi$  est borné dans [20]);
- la diversité dans l'initialisation des paramètres, afin d'explorer toutes les combinaisons de caractéristiques. Des particules trop peu nombreuses ou trop proches n'atteindront pas tous les  $\delta$ -pics de la source et ne produiront que des *minimum* locaux.

Nous pouvons alors introduire le résultat fondamental pour la limite du nombre de particules [20], les limites de champ moyen des flots de gradient  $(\mu_{N,t})_{t \geq 0}$  malgré leur absence de convexité :

**Théorème 1** (Convergence globale – informelle). *Si l'initialisation  $\mu_{N,0}$  est telle que le support de  $\mu_0 \stackrel{\text{def.}}{=} \lim_{N \rightarrow +\infty} \mu_{N,0}$  sépare<sup>4</sup>  $\{-\infty\} \times \mathcal{X}$  de  $\{+\infty\} \times \mathcal{X}$  alors le flot de gradient  $\mu_t$  converge faiblement-\* (voir l'annexe 5.2) dans  $\mathcal{P}_2(\Omega)$  vers un minimum global de  $F$  et nous avons également :*

$$\lim_{N,t \rightarrow \infty} F(\mu_{N,t}) = \min_{m \in \mathcal{M}^+(\mathcal{X})} J(m).$$

Les limites en  $N$  et  $t$  sont interchangeable, le lecteur intéressé pourra jeter un œil à [20] pour des énoncés précis et des hypothèses exactes (conditions aux limites, régularité de type « Sard » *p.ex.*  $\phi$  est  $d$ -fois continûment différentiable, *etc.*)

Maintenant que nous disposons d'un résultat de convergence, nous pouvons dès lors étudier la mise en œuvre numérique. Ce problème d'optimisation est traité grâce

3. Une fonction  $f$  entre des espaces vectoriels est positivement  $p$ -homogène si, pour  $\lambda > 0$  et l'argument  $x$ , on a  $f(\lambda x) = \lambda^d f(x)$ .

4. Le support d'une mesure  $m$  est le complément du plus grand ouvert sur lequel  $m$  s'évanouit. Dans un espace ambiant  $\mathcal{X}$ , on dit qu'un ensemble  $C$  sépare les ensembles  $A$  et  $B$  si tout chemin continu dans  $\mathcal{X}$  avec des extrémités dans  $A$  et  $B$  intersecte  $C$ .

à l'algorithme du *Conic Particle Gradient Descent* [21] que l'on note aussi CPGD : le cadre proposé implique un flot de gradient légèrement différent  $(v_t)_{t \geq 0}$  défini par une projection de  $(\mu_t)_{t \geq 0}$  sur  $\mathcal{M}^+(\mathcal{X})$ . Ce nouveau flot de gradient  $(v_t)_{t \geq 0}$  est défini pour une métrique spécifique dans  $\mathcal{M}^+(\mathcal{X})$ , qui est maintenant un compromis entre les métriques de Wasserstein et de Fisher-Rao<sup>5</sup> [21], il est alors appelé un *flot de gradient de Wasserstein-Fisher-Rao*. Ainsi le flot de gradient de Wasserstein-Fisher-Rao commençant en  $v_{N,0} \stackrel{\text{def.}}{=} \sum_{i=1}^N a_i^0 \delta_{x_i^0}$  dans  $\mathcal{M}^+(\mathcal{X})$  s'écrit  $t \mapsto v_{N,t} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N r_i(t)^2 \delta_{x_i(t)}$  dans  $\mathcal{M}^+(\mathcal{X})$ , plutôt que le flot de Wasserstein  $t \mapsto \mu_{N,t} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N \delta_{r_i(t), x_i(t)}$  à partir de  $\mu_{N,0}$  dans  $\mathcal{P}_2(\Omega)$ . L'équation aux dérivées partielles d'un flot de Wasserstein-Fisher-Rao s'écrit :

$$\partial_t v_t = -4\alpha v_t T_\lambda(v_t) + \beta \operatorname{div}(v_t \nabla J'_v(v_t)), \quad (9)$$

pour les deux paramètres  $\alpha, \beta > 0$  de la métrique du cône,  $\alpha$  règle le poids de la métrique de Fisher-Rao tandis que  $\beta$  contrôle celui de la métrique de Wasserstein. Toutes les propriétés formulées sur la convergence peuvent être appliquées à  $\mu_t$  ou  $v_t$ , nous avons en effet le même théorème :

**Théorème .2** (Convergence globale – informelle). *Si  $v_0$  a un support complet (son support est l'ensemble  $\mathcal{X}$ ) et que  $(v_t)_{t \geq 0}$  converge pour  $t \rightarrow +\infty$ , alors la limite est un minimum global de  $J$ . Si  $v_{N,0} \xrightarrow[N \rightarrow +\infty]{} v_0$  au sens faible-\* alors :*

$$\lim_{N,t \rightarrow \infty} J(v_{N,t}) = \min_{m \in \mathcal{M}^+(\mathcal{X})} J(m).$$

Nous examinons maintenant les résultats numériques de la descente de gradient particulière. Nous invitons le lecteur à jeter un coup d'œil à la figure 1, plus précisément aux ellipses rouges et vertes, pour se familiariser avec la partie numérique.

## Résultats numériques

Nous rappelons qu'un flot de gradient  $(v_{N,t})_{t \geq 0}$  débutant en  $\stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N \left(r_i^{(0)}\right)^2 \delta_{x_i^{(0)}}$  peut être vu comme une généralisation (continue par rapport au temps) de la descente de gradient dans l'espace des mesures, permettant de formuler des propositions théoriques précises sur les conditions de reconstruction.

Pour approcher ce flot de gradient, nous utilisons l'algorithme *Conic Particle Gradient Descent* [21] noté CPGD : il consiste à discrétiser l'évolution du flot de gradient  $t \mapsto v_{N,t}$  à travers un schéma numérique sur (9). Cela consiste en une descente de gradient sur les amplitudes  $r$  et les positions  $x$  à travers le gradient de la fonctionnelle  $F_N$  dans l'équation (5), une stratégie qui approxime bien la dynamique du flot de gradient [21].

Ce choix de gradient avec la métrique du cône permet des mises à jour multiplicatives en  $r$  et additives en  $x$ , les deux mises à jour étant indépendantes l'une de l'autre.

5. également appelée métrique de Hellinger.

L'algorithme consiste alors en une descente de gradient avec la définition de  $r'_i(t)$  et  $x'_i(t)$  selon [21, 29] :

$$\begin{cases} r'_i(t) = -2\alpha r_i \lambda (\eta_\lambda(x_i(t)) - 1) \\ x'_i(t) = -\beta \lambda \nabla \eta_\lambda(x_i(t)) \end{cases} \quad (10)$$

grâce au gradient dans l'équation (5), pour la rétraction miroir<sup>6</sup> et  $\eta_\lambda = -J'_v/\lambda$ . La structure du CPGD est présentée dans l'algorithme 5. On notera que les mises à jour multiplicatives en  $r$  utilisent une exponentielle du certificat, et que les mises à jour des quantités  $r, x$  sont bien séparées.

---

**Algorithme 5 :** Algorithme du *Conic Particle Gradient Descent* (CPGD).

---

**Entrées :** Pas de descente de gradient  $\alpha, \beta > 0$  et  $N \geq 1$  le nombre de particules.

- 1 Tirer uniformément l'initialisation, la mesure discrète aux positions et amplitudes  $(r_i^{(0)}, x_i^{(0)})_{i=1}^N \in (\mathbb{R}^+ \times \mathcal{X})^N$  telles que  $a_i^{(0)} = (r_i^{(0)})^2$  :

$$\nu^{(0)} \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^N a_i^{(0)} \delta_{x_i^{(0)}}.$$

- 2 **tant que** la condition d'arrêt n'est pas remplie **faire**

- 3 | Étape de descente miroir, pour tout  $i = 1, \dots, N$  mettre à jour selon :

$$\begin{aligned} r_i^{(k+1)} &= r_i^{(k)} \exp\left(2\alpha \lambda \left(\eta^{(k)}(x_i^{(k)}) - 1\right)\right) \\ x_i^{(k+1)} &= x_i^{(k)} + \beta \lambda \nabla \eta^{(k)}(x_i^{(k)}) \end{aligned}$$

où  $\eta^{(k)} = -\frac{J'(\nu^{(k)})}{\lambda}$ ,  $\nu^{(k)} = \frac{1}{N} \sum_{i=1}^N a_i^{(k)} \delta_{x_i^{(k)}}$  et  $a_i^{(k)} = (r_i^{(k)})^2$ .

- 4 **fin**
- 

Cet algorithme réalise des itérations assez faciles à calculer et peu coûteuses : pour atteindre une précision de  $\varepsilon$  – une distance telle que la distance  $\infty$ -Wasserstein entre la mesure source  $m_{a_0, x_0}$  et la mesure reconstruite  $m^*$  soit inférieure à  $\varepsilon$  – le CPGD a typiquement une complexité en  $\log(\varepsilon^{-1})$ , plutôt qu'en  $\varepsilon^{-1/2}$  pour un programme convexe [21, Théorème 4. 2]. La figure 2 présente une reconstruction à partir de ces dernières mesures de Fourier 1D. La reconstruction est obtenue par deux flots de gradient, le premier sur les mesures positives pour récupérer les  $\delta$ -pics positifs de la vérité terrain et le second sur les mesures négatives pour récupérer la partie négative : la réunion des

---

6. La notion de *rétraction*, compatible avec la structure du cône, est ici fondamentale : dans le contexte riemannien, une rétraction est une application continue qui fait correspondre un vecteur tangent à un point sur la variété. On pourrait l'interpréter comme un moyen d'obliger les valeurs du gradient à rester sur la variété. Voir [21] pour d'autres choix de rétractions compatibles et plus de détails sur ces notions.

deux résultats permet d'obtenir les  $\delta$ -pics reconstruits. La reconstruction sans bruit<sup>7</sup> pour la convolution gaussienne 2D avec les mêmes paramètres que la section sur l'algorithme du Frank-Wolfe est représentée sur la figure 3. On peut voir que les pics sont correctement reconstruits, en notant que certaines particules rouges et bleues non nulles se regroupent autour des trois  $\delta$ -pics.

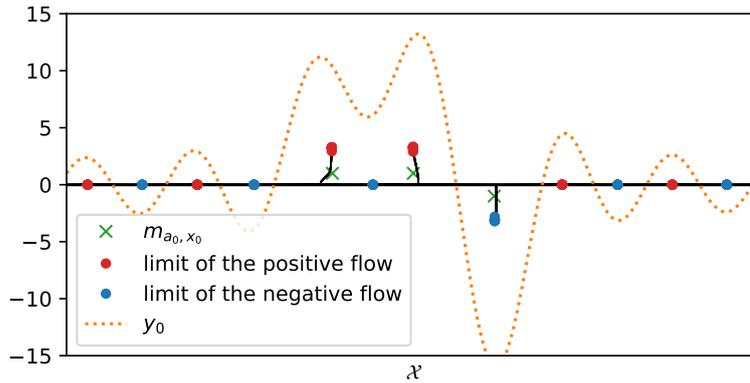


FIGURE 2 – Reconstruction par le *Conic Particle Gradient Descent* pour un opérateur de Fourier 1D dans un environnement sans bruit. La mise en œuvre est une adaptation de [21],  $\alpha = \beta = 1 \times 10^{-3}$  et  $\lambda = 1$  pour 1000 itérations.

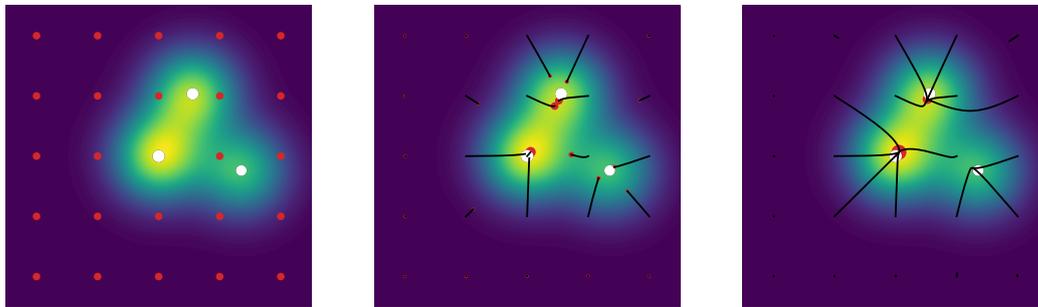


FIGURE 3 – (a) Initialisation  $k = 0$  (b) Milieu du calcul  $k = 150$  (c) Fin du calcul  $k = 1000$ . *Conic Particle Gradient Descent* appliqué pour la déconvolution gaussienne 2D, les points rouges sont la mesure particulaire  $v^{(k)}$  (taille du point proportionnelle à l'amplitude), les trois points blancs sont la mesure source, l'image en arrière-plan est l'acquisition sans bruit  $y_0$  et la ligne noire représentent les trajectoires des particules  $v^{(k)}$  — toutes les trajectoires constituent le flot de gradient  $(v_t)_{t \geq 0}$ . L'implémentation est une adaptation de [21],  $\alpha = \beta = 1 \times 10^{-2}$  et  $\lambda = 1$ .

7. Voir notre dépôt GitHub pour notre implémentation : <https://github.com/XeBasTeX>

### Détails sur l'autodérivation

L'autodérivation, ou dérivation automatique, se propose de calculer le gradient de manière efficace, donc en un nombre d'opérations proche de l'évaluation de la fonction à dériver. Elle a pris son essor dans les dernières années [54], notamment avec l'introduction du calcul sur carte graphique (GPU) et d'interfaces (*wrapper*) tel que CUDA qui permettent de délivrer le plein potentiel de cette méthode. Nous donnons dans la suite quelques précisions sur l'auto-différentiation [74, 48].

**Définition .7** (Différentiable). Soit  $(E, \|\cdot\|_E)$  et  $(F, \|\cdot\|_F)$  deux espaces vectoriels normés. Une fonction  $f : E \rightarrow F$  est dite différentiable au sens de Fréchet en  $x_0 \in X$  s'il existe une forme linéaire  $\mathcal{L} : E \rightarrow F$  telle que :

$$\forall \delta_x \in E, \quad f(x_0 + \delta_x) = f(x_0) + \mathcal{L}(\delta_x) + o(\|\delta_x\|_E).$$

Si elle existe, cette forme linéaire est unique. Elle est appelée la différentielle de  $f$  en  $x_0$  et elle est dénotée  $\mathcal{L} = d_x f(x_0)$ . L'application de  $\mathcal{L}$  à  $\delta_x \in E$  s'écrit :  $\mathcal{L}(\delta_x) = d_x f(x_0) \cdot \delta_x$ .

Il est nécessaire d'imposer à  $X$  d'être un espace de Hilbert pour garantir que la notion suivante soit bien indépendante de la base de  $E$  choisie. Alors, on peut écrire :

**Définition .8** (Gradient). Soit  $(E, \langle \cdot, \cdot \rangle_E)$  un espace de Hilbert,  $x_0 \in E$  et  $f : E \rightarrow \mathbb{R}$  une fonction différentiable en  $x_0$ . D'après le théorème de représentation de Riesz, il existe un unique vecteur  $\nabla_x f(x_0) \in E$  appelé gradient de  $f$  en  $x_0$  tel que :

$$\forall \delta_x \in E, \quad d_x f(x_0) \cdot \delta_x = \langle \nabla_x f(x_0), \delta_x \rangle_E.$$

**Définition .9** (Gradient généralisé). Soit  $(E, \langle \cdot, \cdot \rangle_E)$  et  $(F, \langle \cdot, \cdot \rangle_F)$  deux espaces de Hilbert,  $x_0 \in E$  et  $f : E \rightarrow F$  une fonction différentiable en  $x_0$ . Toujours d'après le théorème de représentation de Riesz, l'adjoint  $(d_x F)^*(x_0) : F^* \rightarrow E^*$  de la différentielle induit l'application continue  $d_x^T f(x_0) : Y \rightarrow X$ , qui est appelée le gradient généralisé de  $F$  en  $x_0$ .

En pratique, l'énergie que l'on souhaite dériver  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  s'écrit comme une composition de  $p \in \mathbb{N}^*$  fonctions élémentaires  $f_i : \mathbb{R}^{N_{i-1}} \rightarrow \mathbb{R}^{N_i}$ , pour  $N_0 = n$  et  $N_p = 1$ . Ainsi :  $f = f_p \circ \dots \circ f_2 \circ f_1$ . Le gradient de  $f$  au point  $x_0 \in \mathbb{R}^n$  se note donc  $\nabla_x f(x_0)$ , et s'écrit précisément comme l'image de 1 par l'application  $d_x^T f(x_0) : \mathbb{R} \rightarrow \mathbb{R}^n$ . La règle de dérivation en chaîne indique que :

$$d_x^T f(x_0) \cdot 1 = d_x^T f_1(x_0) \circ d_x^T f_2(f_1(x_0)) \circ \dots \circ d_x^T f_p(f_{p-1}(\dots f_1(x_0))),$$

en notant  $x_i = f_i \circ \dots \circ f_1(x_0)$  les résultats intermédiaires dans le calcul de  $x_p = f(x_0)$  il vient alors :

$$\begin{aligned} d_x^T f(x_0) \cdot 1 &= d_x^T f_1(x_0) \circ d_x^T f_2(f_1(x_0)) \circ \dots \circ d_x^T f_p(f_{p-1}(\dots f_1(x_0))) \\ &= d_x^T f_1(x_0) \circ d_x^T f_2(x_1) \circ \dots \circ d_x^T f_p(x_{p-1}). \end{aligned}$$

On appelle opérateur *forward*

$$\begin{aligned} f_i : \mathbb{R}^{N_{i-1}} &\longrightarrow \mathbb{R}^{N_i} \\ x &\longmapsto f_i(x), \end{aligned}$$

et *backward*

$$\begin{aligned} d_x^T f_i : \mathbb{R}^{N_{i-1}} \mathbb{R}^{N_i} \times &\longrightarrow \mathbb{R}^{N_{i-1}} \\ (x, e) &\longmapsto d_x^T f_i(x) \cdot e. \end{aligned}$$

Le calcul de  $f(x_0)$  et de  $\delta_x f(x_0)$  est alors réalisé avec une approche *forward-backward*, selon le schéma en figure 4 issu de [48] :

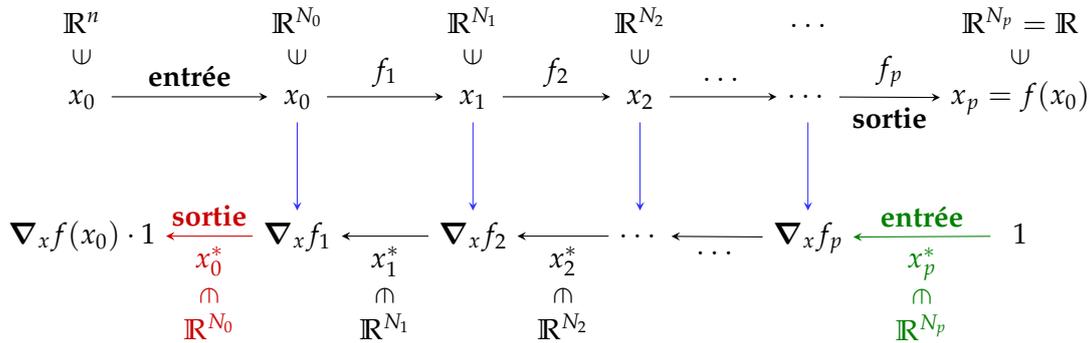


FIGURE 4 – Illustration de la procédure de rétropropagation, d’après un schéma de [48].

L’algorithme de rétropropagation, essentiel pour le calcul de l’auto-dérivation, peut se décomposer en deux étapes qui correspondent aux deux lignes de la figure 4. Ainsi :

1. en prenant l’entrée  $x_0 \in \mathbb{R}^n$ , on calcule et on garde en mémoire les itérations  $x_i \in \mathbb{R}^{N_i}$ . La dernière itération, notée  $x_p \in \mathbb{R}$ , est précisément la valeur de la fonction objective  $F(x_0)$ ;
2. en prenant l’entrée  $x_p^* = 1$ , on calcule successivement les éléments duaux selon :

$$x_i^* = d_x^T F_{i+1}(x_i) \cdot x_{i+1}^*.$$

la dernière itération  $x_0^*$  est précisément égale au gradient  $\nabla_x F(x_0) = d^T(x_0) \cdot 1$ .

**Détails sur le transport optimal**

**Définition .10.** Un espace  $(\mathcal{X}, d_p)$  est appelé espace polonais s'il est séparable, métrisable, et possède une topologie – induite par une distance – qui rend l'espace complet.

Si  $\mathcal{X}$  est un espace de Hilbert séparable,  $(\mathcal{X}, d_p)$  est un espace métrique polonais pour  $d_p$  la distance sur  $\mathbb{R}^d$  restreinte à  $\mathcal{X}$ . Nous pouvons également introduire :

**Définition .11** (Plan de transport). La mesure positive  $\gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{X})$  qui vérifie, pour tout  $A, B \in \mathcal{B}(\mathcal{X})$  où  $\mathcal{B}(\mathcal{X})$  est la tribu borélienne :

$$\gamma(A \times \mathcal{X}) = m_1(A), \quad \gamma(B \times \mathcal{X}) = m_2(B)$$

est appelé le plan de transport [75] entre deux mesures positives  $m_1$  et  $m_2$  de même masse.

On note  $\Gamma(m_1, m_2)$  l'ensemble des plans de transport entre  $m_1$  et  $m_2$ . Les métriques de transport optimal comme la distance de Wasserstein reposent essentiellement sur ces notions, et ne sont définies que sur des espaces polonais : c'est pourquoi nous travaillons avec les mesures dans  $X$  de [41], restriction de  $\mathcal{M}(\mathcal{X})$  avec la topologie faible-\*

**Définition .12** (Distance de Wasserstein). Soit l'espace métrique polonais  $(\mathcal{X}, d_p)$ , et  $p \in [1, +\infty[$ . Pour toutes mesures de probabilité  $\mu$  et  $\nu$  de  $\mathcal{X}$ , la  $p$ -distance de Wasserstein entre  $\mu$  et  $\nu$  est définie par :

$$\mathcal{W}_p(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathcal{X}} d(x, y)^p d\gamma(x, y) \right)^{1/p}.$$

Nous donnons au passage la définition de moment :

**Définition .13.** Si  $r \in \mathbb{N}$ , on appelle moment d'ordre  $r$  d'une mesure  $m \in \mathcal{M}(\mathcal{X})$  la quantité :

$$\int_{\mathcal{X}} x^r dm(x).$$

On dit que  $m$  est de moment  $r$ -fini si la quantité précédent est finie.

## Annexe 2 : précisions sur le chapitre 3

**Définition .14** (Restriction d'une mesure). Soit une mesure  $\rho$ , on note  $\rho \llcorner A$  la restriction de la mesure  $\rho$  à l'ensemble  $A$ , à savoir pour tous les ensembles mesurables  $B$  :  $\rho \llcorner A(B) \stackrel{\text{def.}}{=} \rho(A \cap B)$ .

**Définition .15** (Support d'une mesure). Le support de  $\rho$ , noté  $\text{spt}(\rho)$  est un borélien (élément de la tribu borélienne sur  $\mathcal{X}$ )  $B \subset \mathcal{X}$  tel que

$$\text{spt}(\rho) = \{x \in \mathcal{X} \mid \forall U \text{ voisinage de } x, \rho(U) > 0\}.$$

**Définition .16** (Mesure de Borel, régularité). Une mesure de Borel est une mesure définie sur la tribu borélienne. Une mesure de Borel  $\mu$  est dite régulière si elle est à la fois intérieurement régulière et extérieurement régulière, c'est-à-dire :

- pour tout élément  $A \subset \mathcal{X}$  de la tribu,  $\mu(A) = \sup\{\mu(K) \mid K \text{ compact de } A\}$ ;
- pour tout élément  $A \subset \mathcal{X}$  de la tribu,  $\mu(A) = \inf\{\mu(O) \mid O \text{ ouvert contenant } A\}$ .

Il existe une unique mesure de Borel régulière, invariante par translation, telle que la mesure du cube unité  $[0, 1]^d$  dans  $\mathbb{R}^d$  vaille 1. Cette mesure est appelée la mesure de Lebesgue  $\mathcal{L}^d$ . Malheureusement, elle ne permet pas de mesurer des ensembles «  $p$ -dimensionnels » dans  $\mathbb{R}^d$ , pour  $p < d$ . Felix Hausdorff proposa en 1918 une mesure  $p$ -dimensionnel sur  $\mathbb{R}^d$ , qui donne la même aire pour les sous-variétés de dimension  $p$ , mais qui est définie pour les ensembles mesurables de  $\mathbb{R}^d$ . En particulier pour  $p = d$ , la mesure de Hausdorff coïncide avec la mesure de Lebesgue. Précisons quelques définitions importantes :

**Définition .17** (Préliminaires sur la mesure de Hausdorff [72]). Soit  $S \subset \mathbb{R}^d$ , on appelle diamètre de  $S$  la quantité  $\text{diam}(S) = \sup(\|x - y\| \mid x, y \in S)$ . Définissons maintenant la quantité suivante pour  $S \in \mathbb{R}^d$ ; soit  $\delta > 0$ , on note :

$$\mathcal{H}_\delta^p(S) \stackrel{\text{def.}}{=} \inf \left\{ \sum_{i=1}^{\infty} (\text{diam } U_i)^p : \bigcup_{i=1}^{\infty} U_i \supseteq S, \text{diam } U_i < \delta \right\}.$$

La quantité  $\mathcal{H}_\delta^p(S)$  mesure une surface  $S$  en la recouvrant d'ensembles de formes arbitraires, mais de diamètres contrôlés par  $\delta$ . Il faut maintenant passer à la limite pour pouvoir vraiment définir cette mesure :

**Définition .18** (Mesure de Hausdorff [46]). *On appelle  $p$ -mesure de Hausdorff dans  $\mathbb{R}^d$  la quantité :*

$$\mathcal{H}^d(S) \stackrel{\text{def.}}{=} \sup_{\delta > 0} \mathcal{H}_\delta^d(S) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^d(S).$$

La mesure de Hausdorff d'une variété peut formellement se concevoir comme : dans un premier temps, le recouvrement de la sous-variété par des boules, puis l'addition de toutes les sections efficaces (intersection des boules et de la variété) pour obtenir l'aire de cette variété. On peut finalement préciser la dernière définition

**Définition .19** (Rectifiabilité [72]). *On dit que  $E \subset \mathbb{R}^d$  est  $p$ -rectifiable si c'est une union dénombrable d'image de fonctions lipschitziennes de  $\mathbb{R}^p$  dans  $\mathbb{R}^d$ , à un ensemble négligeable (pour la mesure  $\mathcal{H}^p$ ) près.*

Dans le cas de  $d = 2$  qui va particulièrement nous intéresser au chapitre 3, nous pouvons illustrer cette notion de rectifiabilité : un singleton/un point est un ensemble 0-rectifiable, une surface est un ensemble 2-rectifiable, et une courbe est un ensemble 1-rectifiable.





## Bibliographie

- [1] Luigi AMBROSIO, Nicola GIGLI et Giuseppe SAVARE : *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Springer-Verlag GmbH, 2008.
- [2] Jean-Marc AZAIS, Yohann De CASTRO et Fabrice GAMBOA : Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, mars 2015.
- [3] Laurent BARATCHART, Cristóbal Villalobos GUILLÉN et Douglas P. HARDIN : Inverse potential problems in divergence form for measures in the plane. *ESAIM : Control, Optimisation and Calculus of Variations*, 27:87, 2021.
- [4] Amir BECK et Marc TEBoulLE : A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, avril 2009.
- [5] Arne BEURLING : Sur les intégrales de fourier absolument convergentes et leur application à une transformation fonctionnelle. *Ninth Scandinavian Mathematical Congress*, pages 345–366, 1938.
- [6] Paolo BONICATTO et Nikolay A. GUSEV : On the structure of divergence-free measures on  $\mathbb{R}^2$ . *arXiv :1912.10936 [math]*, novembre 2020. arXiv : 1912.10936.
- [7] Nicholas BOYD, Geoffrey SCHIEBINGER et Benjamin RECHT : The alternating descent conditional gradient method for sparse inverse problems. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, décembre 2015.
- [8] Claire BOYER, Antonin CHAMBOLLE, Yohann De CASTRO, Vincent DUVAL, Frédéric de GOURNAY et Pierre WEISS : On representer theorems and convex regularization. *SIAM Journal on Optimization*, 29(2):1260–1281, janvier 2019.
- [9] Kristian BREDIES et Marcello CARIONI : Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59(1), décembre 2019.

- [10] Kristian BREDIES, Marcello CARIONI, Silvio FANZON et Francisco ROMERO : On the extremal points of the ball of the benamou–brenier energy. *Bulletin of the London Mathematical Society*, juin 2021.
- [11] Kristian BREDIES, Marcello CARIONI, Silvio FANZON et Francisco ROMERO : A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization. *Foundations of Computational Mathematics*, mars 2022.
- [12] Kristian BREDIES et Silvio FANZON : An optimal transport approach for solving dynamic inverse problems in spaces of measures. *ESAIM : Mathematical Modelling and Numerical Analysis*, 54(6):2351–2382, novembre 2020.
- [13] Kristian BREDIES et Hanna Katriina PIKKARAINEN : Inverse problems in spaces of measures. *ESAIM : Control, Optimisation and Calculus of Variations*, 19(1):190–218, mars 2012.
- [14] Richard H. BYRD, Peihuang LU, Jorge NOCEDAL et Ciyou ZHU : A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, septembre 1995.
- [15] Emmanuel CANDÈS, Justin ROMBERG et Terence TAO : Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, février 2006.
- [16] Emmanuel J. CANDÈS et Carlos FERNANDEZ-GRANDA : Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6): 906–956, avril 2013.
- [17] Yohann De CASTRO, F. GAMBOA, Didier HENRION et J.-B. LASSERRE : Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory*, 63(1):621–630, janvier 2017.
- [18] Antonin CHAMBOLLE et Thomas POCK : Total roto-translational variation. *Numerische Mathematik*, 142(3):611–666, mars 2019.
- [19] Da CHEN, Jean-Marie MIREBEAU et Laurent D. COHEN : Global minimum for curvature penalized minimal path method. *In Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association, 2015.
- [20] Lenaïc CHIZAT et Francis BACH : On the global convergence of gradient descent for over-parameterized models using optimal transport. *Advances in Neural Information Processing Systems (NIPS)*, décembre 2018.
- [21] Lénaïc CHIZAT : Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, mars 2021.
- [22] Donald L. COHN : *Measure Theory*. Springer New York, 2013.

- [23] Giovanni E. COMI, Daniel SPECTOR et Giorgio STEFANI : The fractional variation and the precise representative of  $BV^{\alpha,p}$  functions. *arXiv :2109.15263 [math]*, novembre 2021. arXiv : 2109.15263.
- [24] Giovanni E. COMI et Giorgio STEFANI : A distributional approach to fractional Sobolev spaces and fractional variation : Existence of blow-up. *Journal of Functional Analysis*, 277(10):3373–3435, novembre 2019.
- [25] Giovanni E. COMI et Giorgio STEFANI : A distributional approach to fractional sobolev spaces and fractional variation : asymptotics i. *Revista Matemática Complutense*, 36(2):491–569, juin 2022.
- [26] Jean-Baptiste COURBOT et Bruno COLICCHIO : A Fast Homotopy Algorithm for Gridless Sparse Recovery. working paper or preprint, novembre 2020.
- [27] Siân CULLEY, Kalina L. TOSHEVA, Pedro Matos PEREIRA et Ricardo HENRIQUES : SRRF : Universal live-cell super-resolution microscopy. *The International Journal of Biochemistry & Cell Biology*, 101:74–79, août 2018.
- [28] Yohann de CASTRO, Vincent DUVAL et Romain PETIT : Towards off-the-grid algorithms for total variation regularized inverse problems. *In Lecture Notes in Computer Science*, pages 553–564. Springer International Publishing, 2021.
- [29] Yohann de CASTRO, Sébastien GADAT, Clément MARTEAU et Cathy MAUGIS : SuperMix : Sparse regularization for mixtures. *The Annals of Statistics*, 49(3), juin 2021.
- [30] Yohann de CASTRO et Fabrice GAMBOA : Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, novembre 2012.
- [31] Moiseevich Aleksandr DEMIANOV VLADIMIR FEDOROVICH : *Approximate methods in optimization problems*. American Elsevier Pub. Co., 1970.
- [32] Quentin DENOYELLE : *Theoretical and Numerical Analysis of Super-Resolution Without Grid*. Theses, Université Paris sciences et lettres, juillet 2018.
- [33] Quentin DENOYELLE, Thanh an PHAM, Pol del AGUILA PLA, Daniel SAGE et Michael UNSER : Optimal-transport-based metric for smlm. *In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, avril 2021.
- [34] Quentin DENOYELLE, Vincent DUVAL et Gabriel PEYRÉ : Support Recovery for Sparse Super-Resolution of Positive Measures. *Journal of Fourier Analysis and Applications*, septembre 2016.
- [35] Quentin DENOYELLE, Vincent DUVAL, Gabriel PEYRÉ et Emmanuel SOUBIES : The sliding frank–wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001, décembre 2019.

- [36] Thomas DERTINGER, Mike HEILEMANN, Robert VOGEL, Markus SAUER et Shimon WEISS : Superresolution optical fluctuation imaging with organic dyes. *Angewandte Chemie International Edition*, 49(49):9441–9443, octobre 2010.
- [37] J. DIESTEL et J.J. UHL : *Vector Measures*. Mathematical surveys and monographs. American Mathematical Society, 1977.
- [38] Charles DOSSAL, Vincent DUVAL et Clarice POON : Sampling the fourier transform along radial lines. *SIAM Journal on Numerical Analysis*, 55(6):2540–2564, janvier 2017.
- [39] Jérôme DRONIOU : Intégration et espaces de sobolev à valeurs vectorielles. *HAL*, 2001(0), 2001.
- [40] R. DUIJS, S. P. L. MEESTERS, J.-M. MIREBEAU et J. M. PORTEGIES : Optimal paths for variants of the 2d and 3d reeds–shepp car with applications in image analysis. *Journal of Mathematical Imaging and Vision*, 60(6):816–848, février 2018.
- [41] Vincent DUVAL et Gabriel PEYRÉ : Sparse regularization on thin grids i : the lasso. *Inverse Problems*, 33(5):055008, mars 2017.
- [42] Vincent DUVAL et Gabriel PEYRÉ : Sparse spikes super-resolution on thin grids II : the continuous basis pursuit. *Inverse Problems*, 33(9):095008, août 2017.
- [43] Vincent DUVAL et Gabriel PEYRÉ : Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, octobre 2014.
- [44] Vincent DUVAL et Robert TOVEY : Dynamical programming for off-the-grid dynamic inverse problems. working paper or preprint, décembre 2021.
- [45] Ivar EKELAND et Roger TÉMAM : *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics, janvier 1999.
- [46] Herbert FEDERER : *Geometric measure theory*. Springer, Berlin, 1996.
- [47] Carlos FERNANDEZ-GRANDA : Super-resolution of point sources via convex programming, 2016.
- [48] Jean FEYDY : *Analyse de données géométriques, au delà des convolutions*. Theses, Université Paris-Saclay, juillet 2020.
- [49] Marguerite FRANK et Philip WOLFE : An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, mars 1956.
- [50] Arik GIRSAULT, Tomas LUKES, Azat SHARIPOV, Stefan GEISSBUEHLER, Marcel LEUTENEGGER, Wim VANDENBERG, Peter DEDECKER, Johan HOFKENS et Theo LASSER : SOFI simulation tool : A software package for simulating and testing

- super-resolution optical fluctuation imaging. *PLOS ONE*, 11(9):e0161602, septembre 2016.
- [51] M. G. L. GUSTAFSSON : Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. SHORT COMMUNICATION. *Journal of Microscopy*, 198(2):82–87, mai 2000.
- [52] Nils GUSTAFSSON, Siân CULLEY, George ASHDOWN, Dylan M. OWEN, Pedro Matos PEREIRA et Ricardo HENRIQUES : Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications*, 7(1), août 2016.
- [53] Zaid HARCHAOUL, Anatoli JUDITSKY et Arkadi NEMIROVSKI : Conditional gradient algorithms for norm-regularized smooth convex optimization. *Mathematical Programming*, 152(1-2):75–112, avril 2014.
- [54] Laurent HASCOET et Valérie PASCUAL : The tapenade automatic differentiation tool. *ACM Transactions on Mathematical Software*, 39(3):1–43, avril 2013.
- [55] J.F. HAUER, C.J. DEMEURE et L.L. SCHARF : Initial results in prony analysis of power system response signals. *IEEE Transactions on Power Systems*, 5(1):80–89, 1990.
- [56] Stefan W. HELL et Jan WICHMANN : Breaking the diffraction resolution limit by stimulated emission : stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780, juin 1994.
- [57] David HILBERT : Ueber die darstellung definiter formen als summe von formenquadraten. *Mathematische Annalen*, 32(3):342–350, septembre 1888.
- [58] Seamus J HOLDEN, Stephan UPHOFF et Achillefs N KAPANIDIS : DAOSTORM : an algorithm for high- density super-resolution microscopy. *Nature Methods*, 8(4):279–280, mars 2011.
- [59] Adrian JARRET, Julien FAGEOT et Matthieu SIMEONI : A fast and scalable polyatomic frank-wolfe algorithm for the LASSO. *IEEE Signal Processing Letters*, 29:637–641, décembre 2022.
- [60] Hossein JAVANSHIRI et Rasoul NASR-ISFAHANI : The strict topology for the space of radon measures on a locally compact hausdorff space. *Topology and its Applications*, 160(7):887–895, 2013.
- [61] Vladimir KADETS : *A Course in Functional Analysis and Measure Theory*. Springer International Publishing, 2018.
- [62] V. P. KHAVIN et S. K. SMIRNOV : Approximation and extension problems for some classes of vector fields. *St. Petersburg Department of Steklov Institute of Mathematics, Russian Academy of Sciences*, 10(3):507–528, 1998.

- [63] Diederik P. KINGMA et Jimmy BA : Adam : A method for stochastic optimization. *CoRR*, abs/1412.6980, décembre 2014.
- [64] Jean B. LASSERRE : Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, janvier 2001.
- [65] Bastien LAVILLE, Laure BLANC-FÉRAUD et Gilles AUBERT : Off-the-grid covariance-based super-resolution fluctuation microscopy. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, mai 2022.
- [66] Bastien LAVILLE, Laure BLANC-FÉRAUD et Gilles AUBERT : Off-the-grid charge algorithm for curve reconstruction in inverse problems. In *Lecture Notes in Computer Science*, pages 393–405. Springer International Publishing, 2023.
- [67] Bastien LAVILLE, Laure BLANC-FÉRAUD et Gilles AUBERT : Off-the-grid curve reconstruction through divergence regularization : An extreme point result. *SIAM Journal on Imaging Sciences*, 16(2):867–885, mai 2023.
- [68] Bastien LAVILLE, Laure BLANC-FÉRAUD et Gilles AUBERT : Off-The-Grid Variational Sparse Spike Recovery : Methods and Algorithms. *Journal of Imaging*, 7(12):266, décembre 2021.
- [69] Jan LELLMANN, Dirk A. LORENZ, Carola SCHÖNLIEB et Tuomo VALKONEN : Imaging with kantorovich-rubinstein discrepancy. *SIAM Journal on Imaging Sciences*, 7(4):2833–2859, janvier 2014.
- [70] E.S. LEVITIN et B.T. POLYAK : Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):1–50, janvier 1966.
- [71] Jean-Marie MIREBEAU : Fast-marching methods for curvature penalized shortest paths. *Journal of Mathematical Imaging and Vision*, 60(6):784–815, décembre 2017.
- [72] Frank MORGAN : *Geometric measure theory : a beginner's guide*. Academic Press, Boston, 1988.
- [73] Elias NEHME, Lucien E. WEISS, Tomer MICHAELI et Yoav SHECHTMAN : Deep-STORM : super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458, avril 2018.
- [74] Adam PASZKE, Sam GROSS, Soumith CHINTALA, Gregory CHANAN, Edward YANG, Zachary DEVITO, Zeming LIN, Alban DESMAISON, Luca ANTIGA et Adam LERER : Automatic differentiation in pytorch. In *NIPS-W*, 2017.
- [75] Gabriel PEYRÉ et Marco CUTURI : *Computational Optimal Transport*. now Publishers Inc, 2019.

- [76] Robert R. PHELPS, éditeur. *Lectures on Choquet's Theorem*. Springer Berlin Heidelberg, 2001.
- [77] R. Norvaiša R. M. DUDLEY : *Concrete Functional Calculus*. Springer-Verlag GmbH, 2010.
- [78] Daniel SAGE, Hagai KIRSHNER, Thomas PENGO, Nico STUURMAN, Junhong MIN, Suliana MANLEY et Michael UNSER : Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature Methods*, 12(8):717–724, juin 2015.
- [79] Filippo SANTAMBROGIO : Euclidean, Metric, and Wasserstein gradient flows : an overview, 2016.
- [80] Laurent SCHWARZ : Sur l'impossibilité de la multiplication des distributions. *C. R. Acad. Sci. Paris*, 239:847–848, 1954.
- [81] Miroslav ŠILHAVÝ : Divergence measure vectorfields : their structure and the divergence theorem. *Mathematical modelling of bodies with complicated bulk and boundary behavior*, 20:214–237, 2008.
- [82] S. K. SMIRNOV : Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows. *St. Petersburg Department of Steklov Institute of Mathematics, Russian Academy of Sciences*, 5(4):206–238, 1993.
- [83] Oren SOLOMON, Yonina C. ELДАР, Maor MUTZAFI et Mordechai SEGEV : SPARCOM : Sparsity based super-resolution correlation microscopy. *SIAM Journal on Imaging Sciences*, 12(1):392–419, janvier 2019.
- [84] Emmanuel SOUBIES, Laure BLANC-FÉRAUD et Gilles AUBERT : A Continuous Exact l0 penalty (CELO) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3):pp. 1607–1639 (33 pages), juillet 2015.
- [85] Vasiliki STERGIOPOULOU, Luca CALATRONI, Henrique de MORAIS GOULART, Sébastien SCHAUB et Laure BLANC-FÉRAUD : COL0rme : Super-resolution microscopy based on sparse blinking / fluctuating fluorophore localization and intensity estimation. *Biological Imaging*, 2, 2022.
- [86] Vasiliki STERGIOPOULOU, Jose Henrique de MORAIS GOULART, Sebastien SCHAUB, Luca CALATRONI et Laure BLANC-FÉRAUD : Col0rme : Covariance-based l0 super-resolution microscopy with intensity estimation. *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, avril 2021.
- [87] Roger TEMAM : Fonction convexe d'une mesure et applications. *Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi "Séminaire Goulaouic-Schwartz"*, 1982-1983. talk :10.

- [88] Robert TIBSHIRANI : Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society : Series B (Methodological)*, 58(1):267–288, janvier 1996.
- [89] Yann TRAONMILIN et Jean-François AUJOL : The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems*, février 2020.
- [90] Yann TRAONMILIN, Jean-François AUJOL et Arthur LECLAIRE : Projected gradient descent for non-convex sparse spike estimation. *IEEE Signal Processing Letters*, juin 2020.