Hessian of the Riemannian Squared Distance

Xavier Pennec*
Université Côte d'Azur and Inria Sophia-Antipolis Méditerranée
Asclepios team, Inria Sophia Antipolis
2004 Route des Lucioles, BP93
F-06902 Sophia-Antipolis Cedex, France

March 23, 2017

Abstract

This supplementary material details the notions of Riemannian geometry that are underlying the paper *Barycentric Subspace Analysis on Manifolds*. In particular, it investigates the Hessian of the Riemannian square distance whose definiteness controls the local regularity of the barycentric subspaces. This is exemplified on the sphere and the hyperbolic space.

1 Riemannian manifolds

A Riemannian manifold is a differential manifold provided with a smooth collection of scalar products $\langle\, . \mid . \,\rangle_x$ on each tangent space $T_x\mathcal{M}$ at point x of the manifold, called the Riemannian metric. In a chart, the metric is expressed by a symmetric positive definite matrix $G(x) = [g_{ij}(x)]$ where each element is given by the dot product of the tangent vector to the coordinate curves: $g_{ij}(x) = \langle\, \partial_i \mid \partial_j \,\rangle_x$. This matrix is called the *local representation of the Riemannian metric* in the chart x and the dot products of two vectors v and w in $T_x\mathcal{M}$ is now $\langle\, v \mid w \,\rangle_x = v^{\text{T}}\, G(x)\, w = g_{ij}(x) v^i w^j$ using the Einstein summation convention which implicitly sum over the indices that appear both in upper position (components of [contravariant] vectors) and lower position (components of covariant vectors (co-vectors)).

1.1 Riemannian distance and geodesics

If we consider a curve $\gamma(t)$ on the manifold, we can compute at each point its instantaneous speed vector $\dot\gamma(t)$ (this operation only involves the differential structure) and its norm $\|\dot\gamma(t)\|_{\gamma(t)}$ to obtain the instantaneous speed (the Riemannian metric is needed for this operation). To compute the length of the curve, this value is integrated along the curve:

$$\mathcal{L}_a^b(\gamma) = \int_a^b \|\dot\gamma(t)\|_{\gamma(t)}\, dt = \int_a^b \Big(\langle\, \dot\gamma(t) \mid \dot\gamma(t) \,\rangle_{\gamma(t)} \Big)^{\frac{1}{2}}\, dt$$

*xavier.pennec@inria.fr

The distance between two points of a connected Riemannian manifold is the minimum length among the curves joining these points. The curves realizing this minimum are called geodesics. Finding the curves realizing the minimum length is a difficult problem as any time-reparameterization is authorized. Thus one rather defines the metric geodesics as the critical points of the energy functional $\mathcal{E}(\gamma) = \frac{1}{2}\int_0^1 \|\dot\gamma(t)\|^2\, dt$. It turns out that they also optimize the length functional but they are moreover parameterized proportionally to arc-length.

Let $[g^{ij}] = [g_{ij}]^{(-1)}$ be the inverse of the metric matrix (in a given coordinate system) and $\Gamma^i_{jk} = \frac{1}{2} g^{im}\left(\partial_k g_{mj} + \partial_j g_{mk} - \partial_m g_{jk}\right)$ the Christoffel symbols. The calculus of variations shows the geodesics are the curves satisfying the following second order differential system:

$$\ddot\gamma^i + \Gamma^i_{jk}\dot\gamma^j\dot\gamma^k = 0.$$

The fundamental theorem of Riemannian geometry states that on any Riemannian manifold there is a unique (torsion-free) connection which is compatible with the metric, called the Levi-Civita (or metric) connection. For that choice of connection, shortest paths (geodesics) are auto-parallel curves ("straight lines"). This connection is determined in a local coordinate system through the Christoffel symbols: $\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k$. With these conventions, the covariant derivative of the coordinates v^i of a vector field is $v^i_{;j} = (\nabla_j v)^i = \partial_j v^i + \Gamma^i_{jk} v^k$.

In the following, we only consider the Levi-Civita connection and we assume that the manifold is geodesically complete, i.e. that the definition domain of all geodesics can be extended to $\mathbb{R}$. This means that the manifold has no boundary nor any singular point that we can reach in a finite time. As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold (i.e. whose length is the distance between the two points).

1.2 Normal coordinate systems

Let x be a point of the manifold that we consider as a local reference and v a vector of the tangent space $T_x\mathcal{M}$ at that point. From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma_{(x,v)}(t)$ starting from that point with this tangent vector. This allows to wrap the tangent space onto the manifold, or equivalently to develop the manifold in the tangent space along the geodesics (think of rolling a sphere along its tangent plane at a given point). The mapping $\exp_x(v) = \gamma_{(x,v)}(1)$ of each vector $v \in T_x\mathcal{M}$ to the point of the manifold that is reached after a unit time by the geodesic $\gamma_{(x,v)}(t)$ is called the *exponential map* at point x. Straight lines going through 0 in the tangent space are transformed into geodesics going through point x on the manifold and distances along these lines are conserved.

The exponential map is defined in the whole tangent space $T_x\mathcal{M}$ (since the manifold is geodesically complete) but it is generally one-to-one only locally around 0 in the tangent space (i.e. around x in the manifold). In the sequel, we denote by $\overrightarrow{xy} = \log_x(y)$ the inverse of the exponential map: this is the smallest vector (in norm) such that $y = \exp_x(\overrightarrow{xy})$. It is natural to search for the maximal domain where the exponential map is a diffeomorphism. If we follow a geodesic $\gamma_{(x,v)}(t) = \exp_x(t\,v)$ from $t = 0$ to infinity, it is either always minimizing all along or it is minimizing up to a time $t_0 < \infty$ and not any more after (thanks to the geodesic completeness). In this last case, the point $\gamma_{(x,v)}(t_0)$ is called a *cut point* and the corresponding tangent vector $t_0\, v$ a *tangential cut point*. The set of tangential cut points at x is called the *tangential cut locus* $C(x) \in T_x\mathcal{M}$, and the set of cut points of the geodesics starting from x is the *cut locus* $\mathcal{C}(x) = \exp_x(C(x)) \in \mathcal{M}$. This is the closure of the set of points where several minimizing geodesics starting from x meet. On the sphere $\mathcal{S}_2(1)$ for instance, the cut locus of a point x is its antipodal point and the tangential cut locus is the circle of radius π.

The maximal bijective domain of the exponential chart is the domain $D(x)$ containing 0 and delimited by the tangential cut locus ($\partial D(x) = C(x)$). This domain is connected and star-shaped with respect to the origin of $T_x\mathcal{M}$. Its image by the exponential map covers all the manifold except the cut locus, which has a null measure. Moreover, the segment $[0, \overrightarrow{xy}]$ is mapped to the unique minimizing geodesic from x to y: geodesics starting from x are straight lines, and the distance from the reference point are conserved. This chart is somehow the "most linear" chart of the manifold with respect to the reference point x.

When the tangent space is provided with an orthonormal basis, this is called *an normal coordinate systems at* x. A set of normal coordinate systems at each point of the manifold realize an atlas which allows to work very easily on the manifold. The implementation of the exponential and logarithmic maps (from now on exp and log) is indeed the basis of programming on Riemannian manifolds, and we can express using them practically all the geometric operations needed for statistics [Pennec, 2006] or image processing [Pennec et al., 2006].

The size of the maximal definition domain is quantified by the *injectivity radius* $\mathrm{inj}(\mathcal{M}, x) = \mathrm{dist}(x, \mathcal{C}(x))$, which is the maximal radius of centered balls in $T_x\mathcal{M}$ on which the exponential map is one-to-one. The injectivity radius of the manifold $\mathrm{inj}(\mathcal{M})$ is the infimum of the injectivity over the manifold. It may be zero, in which case the manifold somehow tends towards a singularity (think e.g. to the surface $z = 1/\sqrt{x^2 + y^2}$ as a sub-manifold of $\mathbb{R}^3$).

In a Euclidean space, normal coordinate systems are realized by orthonormal coordinates system translated at each point: we have in this case $\overrightarrow{xy} = \log_x(y) = y - x$ and $\exp_x(\overrightarrow{v}) = x + \overrightarrow{v}$. This example is more than a simple coincidence. In fact, most of the usual operations using additions and subtractions may be reinterpreted in a Riemannian framework using the notion of *bipoint*, an antecedent of vector introduced during the 19th Century. Indeed, vectors are defined as equivalent classes of bipoints in a Euclidean space. This is possible because we have a canonical way (the translation) to compare what happens at two different points. In a Riemannian manifold, we can still compare things locally (by parallel transportation), but not any more globally. This means that each "vector" has to remember at which point of the manifold it is attached, which comes back to a bipoint.

2 Hessian of the squared distance

2.1 Computing the differential of the Riemannian log

On $\mathcal{M}/C(y)$, the Riemannian gradient $\nabla^a = g^{ab}\partial_b$ of the squared distance $d_y^2(x) = \mathrm{dist}^2(x, y)$ with respect to the fixed point y is well defined and is equal to $\nabla d_y^2(x) = -2\log_x(y)$. The Hessian operator (or double covariant derivative) $\nabla^2 f(x)$ from $T_x\mathcal{M}$ to $T_x\mathcal{M}$ is the covariant derivative of the gradient, defined by the identity $\nabla^2 f(v) = \nabla_v(\nabla f)$. In a normal coordinate system at point x, the Christoffel symbols vanish at x, so that the Hessian operator of the squared distance can be expressed with the standard differential

D_x with respect to the point x:

$$\nabla^2 d_y^2(x) = -2(D_x \log_x(y)).$$

The points x and $y = \exp_x(v)$ are called conjugate if $D\exp_x(v)$ is singular. It is known that the cut point (if it exists) occurs at or before the first conjugate point along any geodesic [Lee, 1997]. Thus, $D\exp_x(v)$ has full rank inside the tangential cut-locus of x. This is in essence why there is a well posed inverse function $\overrightarrow{xy} = \log_x(y)$, called the Riemannian log, which is continuous and differentiable everywhere except at the cut locus of x. Moreover, its differential can be computed easily: since $\exp_x(\log_x(y)) = y$, we have $D\exp_x|_{\overrightarrow{xy}} D\log_x(y) = \text{Id}$, so that

$$D\log_x(y) = \left(D\exp_x|_{\overrightarrow{xy}}\right)^{-1} \tag{1}$$

is well defined and of full rank on $\mathcal{M}/C(x)$.

We can also see the Riemannian log $\log_x(y) = \overrightarrow{xy}$ as a function of the foot-point x, and differentiating $\exp_x(\log_x(y)) = y$ with respect to it gives: $D_x \exp_x|_{\overrightarrow{xy}} + D\exp_x|_{\overrightarrow{xy}} . D_x \log_x(y) = 0$. Once again, we obtain a well defined and full rank differential for $x \in \mathcal{M}/C(y)$:

$$D_x \log_x(y) = -\left(D\exp_x|_{\overrightarrow{xy}}\right)^{-1} D_x \exp_x|_{\overrightarrow{xy}} . \tag{2}$$

The Hessian of the squared distance can thus be written:

$$\frac{1}{2}\nabla^2 d_y^2(x) = -D_x \log_x(x_i) = \left(D\exp_x|_{\overrightarrow{xy}}\right)^{-1} D_x \exp_x|_{\overrightarrow{xy}} .$$

If we notice that $J_0(t) = D\exp_x|_{t\overrightarrow{xy}}$ (respectively $J_1(t) = D_x \exp_x|_{t\overrightarrow{xy}}$) are actually matrix Jacobi field solutions of the Jacobi equation $\ddot{J}(t) + R(t)J(t) = 0$ with $J_0(0) = 0$ and $\dot{J}_0(0) = \text{Id}_n$ (respectively $J_1(0) = \text{Id}_n$ and $\dot{J}_1(0) = 0$), we see that the above formulation of the Hessian operator is equivalent to the one of Villani [2011][Equation 4.2]: $\frac{1}{2}\nabla^2 d_y^2(x) = J_0(1)^{(-1)} J_1(1)$.

2.2 Taylor expansion of the Riemannian log

In order to better figure out what the dependence of the Hessian of the squared Riemannian distance on curvature, we compute here the Taylor expansion of the Riemannian log function. Following Brewin [2009], we consider a normal coordinate system centered at x and $x_v = \exp_x(v)$ a variation of the point x. We denote by $R_{ihjk}(x)$ the coefficients of the curvature tensor at x and by ϵ a conformal gauge scale that encodes the size of the path in terms of $\|v\|_x$ and $\|\overrightarrow{xy}\|_x$ normalized by the curvature (see Brewin [2009] for details).

In a normal coordinate system centered at x, we have the following Taylor expansion of the metric tensor coefficients:

$$\begin{aligned} g_{ab}(v) = & g_{ab} - \frac{1}{3}R_{cabd}v^c v^d - \frac{1}{6}\nabla_e R_{cabd}v^e v^c v^d \\ & + \left(-\frac{1}{20}\nabla_e \nabla_f R_{cabd} + \frac{2}{45}R^g_{cad}R^h_{ebf}\delta_{gh}\right) v^c v^d v^e v^f + O(\epsilon^5). \end{aligned} \tag{3}$$

A geodesic joining point z to point $z+\delta z$ has tangent vector:

$$\begin{aligned}[\log_z(z+\Delta z)]^a &= \Delta z^a + \frac{1}{3} z^b \Delta z^c \Delta z^d R^a_{cbd} + \frac{1}{12} z^b z^c \Delta z^d \Delta z^e \nabla_d R^a_{bce} \\ &+ \frac{1}{6} z^b z^c \Delta z^d \Delta z^e \nabla_b R^a_{dce} + \frac{1}{24} z^b z^c \Delta z^d \Delta z^e \nabla^a R_{bdce} \\ &+ \frac{1}{12} z^b \Delta z^c \Delta z^d \Delta z^e \nabla_c R^a_{dbe} + O(\epsilon^4).\end{aligned}$$

Using $z = v$ and $z + \Delta z = \overrightarrow{xy}$ (i.e. $\Delta z = \overrightarrow{xy} - v$) in a normal coordinate system centered at x, and keeping only the first order terms in v, we obtain the first terms of the series development of the log:

$$\left[\log_{x+v}(y)\right]^a = \overrightarrow{xy}^a - v^a + \frac{1}{3} R^a_{cbd} v^b \overrightarrow{xy}^c \overrightarrow{xy}^d + \frac{1}{12} \nabla_c R^a_{dbe} v^b \overrightarrow{xy}^c \overrightarrow{xy}^d \overrightarrow{xy}^e + O(\epsilon^4). \quad (4)$$

Thus, the differential of the log with respect to the foot point is:

$$-\left[D_x \log_x(y)\right]^a_b = \delta^a_b - \frac{1}{3} R^a_{cbd} \overrightarrow{xy}^c \overrightarrow{xy}^d - \frac{1}{12} \nabla_c R^a_{dbe} \overrightarrow{xy}^c \overrightarrow{xy}^d \overrightarrow{xy}^e + O(\epsilon^3) \quad (5)$$

Since we are in a normal coordinate system, the zeroth order term is the identity matrix, like in the Euclidean space, and the first order term vanishes. The Riemannian curvature tensor appear in the second order term and its covariant derivative in the third order term. The important point here is to see that the curvature is the leading term that makes this matrix departing from the identity (i.e. the Euclidean case) and which may lead to the non invertibility of the differential.

3 Example on spheres

We consider the unit sphere in dimension $n \geq 2$ embedded in $\mathbb{R}^{n+1}$ and we represent points of $\mathcal{M} = \mathcal{S}_n$ as unit vectors in $\mathbb{R}^{n+1}$. The tangent space at x is naturally represented by the linear space of vectors orthogonal to x: $T_x\mathcal{S}_n = \{v \in \mathbb{R}^{n+1}, v^{\mathrm{T}} x = 0\}$. The natural Riemannian metric on the unit sphere is inherited from the Euclidean metric of the embedding space $\mathbb{R}^{n+1}$. With these conventions, the Riemannian distance is the arc-length $d(x,y) = \arccos(x^{\mathrm{T}} y) = \theta \in [0, \pi]$. Denoting $f(\theta) = 1/\mathrm{sinc}(\theta) = \theta / \sin(\theta)$, the spherical exp and log maps are:

$$\begin{aligned}\exp_x(v) &= \cos(\|v\|)x + \mathrm{sinc}(\|v\|)v & (6)\\ \log_x(y) &= f(\theta)\,(y - \cos(\theta)x) \quad \text{with} \quad \theta = \arccos(x^{\mathrm{T}} y). & (7)\end{aligned}$$

Notice that $f(\theta)$ is a smooth function from $]-\pi;\pi[$ to $\mathbb{R}$ that is always greater than one and is locally quadratic at zero: $f(\theta) = 1 + \theta^2/6 + O(\theta^4)$.

3.1 Hessian of the squared distance on the sphere

To compute the gradient and Hessian of functions on the sphere, we first need a chart in a neighborhood of a point $x \in \mathcal{S}_n$. We consider the unit vector $x_v = \exp_x(v)$ which is a variation of x parametrized by the tangent vector $v \in T_x\mathcal{S}_n$ (i.e. verifying $x^{\mathrm{T}} v = 0$). In order to extend this mapping to the embedding space to simplify computations, we consider that v is the orthogonal projection of an unconstrained vector $w \in \mathbb{R}^{n+1}$ onto the tangent space at x: $v = (\,\mathrm{Id} - xx^{\mathrm{T}})w$. Using the above formula for the exponential map, we get at first order $x_v = x - v + O(\|v\|^2)$ in the tangent space or $x_w = x + (\,\mathrm{Id} - xx^{\mathrm{T}})w + O(\|w\|^2)$ in the embedding space.

It is worth verifying first that the gradient of the squared distance $\theta^2 = d_y^2(x) = \arccos^2(x^{\mathrm{T}}y)$ is indeed $\nabla d_y^2(x) = -2\log_x(y)$. We considering the variation $x_w = \exp_x((\mathrm{Id}-xx^{\mathrm{T}})w) = x + (\,\mathrm{Id} - xx^{\mathrm{T}})w + O(\|w\|^2)$. Because $D_x \arccos(y^{\mathrm{T}}x) = -y^{\mathrm{T}}/\sqrt{1-(y^{\mathrm{T}}x)^2}$, we get:

$$D_w \arccos^2(x_w^{\mathrm{T}}y) = \frac{-2\theta}{\sin\theta} y^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}}) = -2f(\theta)y^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}}),$$

and the gradient is as expected:

$$\nabla d_y^2(x) = -2f(\theta)(\,\mathrm{Id} - xx^{\mathrm{T}})y = -2\log_x(y). \tag{8}$$

To obtain the Hessian, we now compute the Taylor expansion of $\log_{x_w}(y)$. First, we have

$$f(\theta_w) = f(\theta) - \frac{f'(\theta)}{\sin\theta} y^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}})w + O(\|w\|^2),$$

with $f'(\theta) = (1 - f(\theta)\cos\theta)/\sin\theta$. Thus, the first order Taylor expansion of $\log_{x_w}(y)$ is:

$$\log_{x_w}(y) = \left(f(\theta) - \frac{f'(\theta)}{\sin\theta} y^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}})w\right)(\,\mathrm{Id} - xx^{\mathrm{T}} - (\,\mathrm{Id} - xx^{\mathrm{T}})wx^{\mathrm{T}} - xw^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}}))\,y + O(\|w\|^2)$$

so that

$$-2D_w \log_{x_w}(y) = \frac{f'(\theta)}{\sin\theta}(\,\mathrm{Id} - xx^{\mathrm{T}})yy^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}}) - f(\theta)\left(x^{\mathrm{T}}y\,\mathrm{Id} + xy^{\mathrm{T}}\right)(\,\mathrm{Id} - xx^{\mathrm{T}})$$

Now, since we have computed the derivative in the embedding space, we have obtained the Hessian with respect to the flat connection of the embedding space, which exhibits a non-zero normal component. In order to obtain the Hessian with respect to the connection of the sphere, we need to project back on $T_x\mathcal{S}_n$ (i.e. multiply by $(\,\mathrm{Id} - xx^{\mathrm{T}})$ on the left) and we obtain:

$$\begin{aligned}\frac{1}{2}H_x(y) &= \left(\frac{1 - f(\theta)\cos\theta}{\sin^2\theta}\right)(\,\mathrm{Id} - xx^{\mathrm{T}})\,yy^{\mathrm{T}}(\,\mathrm{Id} - xx^{\mathrm{T}}) + f(\theta)\cos\theta(\,\mathrm{Id} - xx^{\mathrm{T}})\\ &= (\,\mathrm{Id} - xx^{\mathrm{T}})\left((1 - f(\theta)\cos\theta)\frac{yy^{\mathrm{T}}}{\sin^2\theta} + f(\theta)\cos\theta\,\mathrm{Id}\right)(\,\mathrm{Id} - xx^{\mathrm{T}}),\end{aligned}$$

To simplify this expression, we note that $\|(\mathrm{Id}-xx^{\mathrm{T}})y\|^2 = \sin\theta$, so that $u = \frac{(\,\mathrm{Id}-xx^{\mathrm{T}})y}{\sin\theta} = \frac{\log_x(y)}{\theta}$ is a unit vector of the tangent space at x (for $y \neq x$ so that $\theta > 0$). Using this unit vector and the intrinsic parameters $\log_x(y)$ and $\theta = \|\log_x(y)\|$, we can rewrite the Hessian:

$$\frac{1}{2}H_x(y) = f(\theta)\cos\theta(\,\mathrm{Id} - xx^{\mathrm{T}}) + \left(\tfrac{1-f(\theta)\cos\theta}{\theta^2}\right)\log_x(y)\log_x(y)^{\mathrm{T}} \tag{9}$$

$$= uu^{\mathrm{T}} + f(\theta)\cos\theta(\,\mathrm{Id} - xx^{\mathrm{T}} - uu^{\mathrm{T}}) \tag{10}$$

The eigenvectors and eigenvalues of this matrix are now very easy to determine. By construction, x is an eigenvector with eigenvalue $\mu_0 = 0$. Then the vector u (or equivalently $\log_x(y) = f(\theta)(\,\mathrm{Id} - xx^{\mathrm{T}})y = \theta u$) is an eigenvector with eigenvalue $\mu_1 = 1$. Lastly, every vector u which is orthogonal to these two vectors (i.e. orthogonal to the plane spanned by 0, x and y) has eigenvalue $\mu_2 = f(\theta)\cos\theta = \theta\cot\theta$. This last eigenvalue is positive for $\theta \in [0, \pi/2[$, vanishes for $\theta = \pi/2$ and becomes negative for $\theta \in]\pi/2\pi[$. We retrieve here the results of [Buss and Fillmore, 2001, lemma 2] expressed in a more general coordinate system.

4 Example on the hyperbolic space $\mathbb{H}^n$

We consider in this section the hyperboloid of equation $x_0^2 + x_1^2 \dots x_n^2 = -1$ (with $x_0 > 0$ and $n \geq 2$) embedded in $\mathbb{R}^{n+1}$. Using the notations $x = (x_0, \hat{x})$ and the indefinite nondegenerate symmetric bilinear form $\langle\, x \mid y \,\rangle_* = x^{\mathrm{T}} J y = \hat{x}^{\mathrm{T}} \hat{y} - x_0 y_0$ with $J = \mathrm{diag}(-1, \mathrm{Id}_n)$, the hyperbolic space can be seen as the sphere $\|x\|_*^2 = -1$ of radius -1 in the $(n+1)$-dimensional Minkowski space:

$$\mathbb{H}^n = \{x \in \mathbb{R}^{n,1} / \|x\|_*^2 = \|\hat{x}\|^2 - x_0^2 = -1\}.$$

A point in $\mathcal{M} = \mathbb{H}^n \subset \mathbb{R}^{n,1}$ can be parametrized by $x = (\sqrt{1 + \|\hat{x}\|^2}, \hat{x})$ for $\hat{x} \in \mathbb{R}^n$. This happen to be in fact a global diffeomorphism that provides a very convenient global chart of the hyperbolic space. We denote $\pi(x) = \hat{x}$ (resp. $\pi^{(-1)}(\hat{x}) = (\sqrt{1 + \|\hat{x}\|^2}, \hat{x})$) the coordinate map from $\mathbb{H}^n$ to $\mathbb{R}^n$ (resp. the parametrization map from $\mathbb{R}^n$ to $\mathbb{H}^n$). The Poincarré ball model is another classical models of the hyperbolic space $\mathbb{H}^n$ which can be obtained by a stereographic projection of the hyperboloid onto the hyperplane $x_0 = 0$ from the south pole $(-1, 0 \dots, 0)$.

A tangent vector $v = (v_0, \hat{v})$ at point $x = (x_0, \hat{x})$ satisfies $\langle\, x \mid v \,\rangle_* = 0$, i.e. $x_0 v_0 = \hat{x}^{\mathrm{T}} \hat{v}$, so that

$$T_x \mathbb{H}^n = \left\{ \left(\frac{\hat{x}^{\mathrm{T}} \hat{v}}{\sqrt{1 + \|\hat{x}\|^2}}, \hat{v} \right), \quad \hat{v} \in \mathbb{R}^n \right\}.$$

The natural Riemannian metric on the hyperbolic space is inherited from the Minkowski metric of the embedding space $\mathbb{R}^{n,1}$: the scalar product of two vectors $u = (\hat{x}^{\mathrm{T}} \hat{u} / \sqrt{1 + \|\hat{x}\|^2}, \hat{u})$ and $v = (\hat{x}^{\mathrm{T}} \hat{v} / \sqrt{1 + \|\hat{x}\|^2}, \hat{v})$ at $x = (\sqrt{1 + \|\hat{x}\|^2}, \hat{x})$ is

$$\langle\, u \mid v \,\rangle_* = u^{\mathrm{T}} J v = -u_0 v_0 + \hat{u}^{\mathrm{T}} \hat{v} = \hat{u}^{\mathrm{T}} \left(-\frac{\hat{x} \hat{x}^{\mathrm{T}}}{1 + \|\hat{x}\|^2} + \mathrm{Id} \right) \hat{v}$$

The metric matrix expressed in the coordinate chart $G = \mathrm{Id} - \frac{\hat{x}\hat{x}^{\mathrm{T}}}{1 + \|\hat{x}\|^2}$ has eigenvalue 1, with multiplicity $n - 1$, and $1/(1 + \|\hat{x}\|^2)$ along the eigenvector x. It is thus positive definite.

With these conventions, geodesics are the trace of 2-planes passing through the origin and the Riemannian distance is the arc-length:

$$d(x, y) = \mathrm{arccosh}(-\langle\, x \mid y \,\rangle_*). \tag{11}$$

The hyperbolic exp and log maps are:

$$\exp_x(v) = \cosh(\|v\|_*) x + \frac{\sinh(\|v\|_*)}{\|v\|_*} v \tag{12}$$

$$\log_x(y) = f_*(\theta) \, (y - \cosh(\theta) x) \quad \text{with} \quad \theta = \mathrm{arccosh}(-\langle\, x \mid y \,\rangle_*), \tag{13}$$

where $f_*(\theta) = \theta / \sinh(\theta)$ is a smooth function from $\mathbb{R}$ to $(0, 1]$ that is always positive and is locally quadratic at zero: $f_*(\theta) = 1 - \theta^2/6 + O(\theta^4)$.

4.1 Hessian of the squared distance on the hyperbolic space

We first verify that the gradient of the squared distance $d_y^2(x) = \mathrm{arccosh}^2(-< x, y >_*)$ is indeed $\nabla d_y^2(x) = -2 \log_x(y)$. Let us consider a variation of the base-point along the tangent vector v at x verifying $\langle\, v \mid x \,\rangle_* = 0$:

$$x_v = \exp_x(v) = \cosh(\|v\|_*) x + \frac{\sinh(\|v\|_*)}{\|v\|_*} v = x + v + O(\|v\|_*^2).$$

In order to extend this mapping to the embedding space around the paraboloid, we consider that v is the projection $v = w + \langle\, w \mid x \,\rangle_* \, x$ of an unconstrained vector $w \in \mathbb{R}^{n,1}$ onto the tangent space at $T_x\mathbb{H}^n$. Thus, the variation that we consider in the embedding space is

$$x_w = x + \partial_w x_w + O(\|w\|_Q^2) \qquad \text{with} \qquad \partial_w x_w = w + \langle\, w \mid x \,\rangle_* \, x = (\,\mathrm{Id} + x x^{\mathrm{T}} J) w$$

Now, we are interested in the impact of such a variation on $\theta_w = d_y(x_w) = \operatorname{arccosh}(-\langle\, x_w \mid y \,\rangle_*)$. Since $\operatorname{arccosh}'(t) = \frac{1}{\sqrt{t^2-1}}$, and $\sqrt{\cosh(\theta)^2 - 1} = \sinh(\theta)$ for a positive θ, we have:

$$d/dt \operatorname{arccosh}(t)\big|_{t=\cosh(\theta)} = 1/\sqrt{\cosh(\theta)^2 - 1} = 1/\sinh(\theta).$$

so that

$$\theta_w = \theta - \frac{1}{\sinh(\theta)} \langle\, w + \langle\, w \mid x \,\rangle_* \, x \mid y \,\rangle_* + O(\|v\|_*^2).$$

This means that the directional derivative is

$$\partial_w \theta_w = -\frac{1}{\sinh(\theta)} \langle\, w + \langle\, w \mid x \,\rangle_* \, x \mid y \,\rangle_* = -\frac{1}{\sinh(\theta)} \langle\, w \mid y - \cosh(\theta) x \,\rangle$$

so that $\partial_w \theta_w^2 = -2 f_*(\theta) \langle\, w \mid y - \cosh(\theta) x \,\rangle_*$. Thus, the gradient in the embedding space defined by $< \nabla d_y^2(x), w >_* = \partial_w \theta_w^2$ is as expected:

$$\nabla d_y^2(x) = -2 f_*(\theta)(y - \cosh(\theta) x) = -2 \log_x(y) \tag{14}$$

To obtain the Hessian, we now compute the Taylor expansion of $\log_{x_w}(y)$. First, we compute the variation of $f_*(\theta_w) = \theta_w / \sinh(\theta_w)$:

$$\partial_w f_*(\theta_w) = f_*'(\theta)\, \partial_w \theta_w = -\frac{f_*'(\theta)}{\sinh(\theta)} \langle\, w \mid y - \cosh(\theta) x \,\rangle_* = -\frac{f_*'(\theta)}{\theta} \langle\, w \mid \log_x(y) \,\rangle_*$$

with $f_*'(\theta) = (1 - f_*(\theta)\cosh\theta)/\sinh\theta = (1 - \theta \coth\theta)/\sinh\theta$. The variation of $\cosh\theta_w$ is:

$$\partial_w \cosh\theta_w = \sinh\theta\, \partial_w \theta_w = -\langle\, w \mid y - \cosh(\theta) x \,\rangle_* .$$

Thus, the first order variation of $\log_{x_w}(y)$ is:

$$\begin{aligned}
\partial_w \log_{x_w}(y) &= \partial_w f_*(\theta_w)(y - \cosh\theta x) - f_*(\theta)\,(\partial_w \cosh(\theta_w) x + \cosh(\theta) \partial_w x_w) \\
&= -\frac{f_*'(\theta)\sinh\theta}{\theta^2} \langle\, w \mid \log_x(y) \,\rangle_* \log_x(y) \\
&\quad + f_*(\theta)\,(\langle\, w \mid y - \cosh(\theta) x \,\rangle_* \, x - \cosh(\theta)(w + \langle\, w \mid x \,\rangle_* \, x)) \\
&= -\frac{(1 - \theta\coth\theta)}{\theta^2} \langle\, w \mid \log_x(y) \,\rangle_* \log_x(y) + \langle\, w \mid \log_x(y) \,\rangle_* \, x - \theta\coth(\theta)(w + \langle\, w \mid x \,\rangle_* \, x)
\end{aligned}$$

This vector is a variation in the embedding space: it displays a normal component to the hyperboloid $\langle\, w \mid \log_x(y) \,\rangle_* \, x$ which reflects the extrinsic curvature of the hyperboloid in the Minkowski space (the mean curvature vector is $-x$), and a tangential component which measures the real variation in the tangent space:

$$(\,\mathrm{Id} + x x^{\mathrm{T}} J) \partial_w \log_{x_w}(y) = -\frac{(1 - \theta\coth\theta)}{\theta^2} \langle\, w \mid \log_x(y) \,\rangle_* \log_x(y) - \theta\coth(\theta)(J + x x^{\mathrm{T}}) J w.$$

Thus the intrinsic gradient is:

$$D_x \log_x(y) = -\frac{(1-\theta\coth\theta)}{\theta^2}\log_x(y)\log_x(y)^{\mathrm{T}}J - \theta\coth(\theta)(\,\mathrm{Id} + xx^{\mathrm{T}}J).$$

Finally, the Hessian of the square distance, considered as an operator from $T_x\mathbb{H}^n$ to $T_x\mathbb{H}^n$, is $H_x(y)(w) = -2D_x\log_x(y)w$. Denoting $u = \log_x(y)/\theta$ the unit vector of the tangent space at x pointing towards the point y, we get in matrix form:

$$\frac{1}{2}H_x(y) = uu^{\mathrm{T}}J + \theta\coth\theta(J + xx^{\mathrm{T}} - uu^{\mathrm{T}})J$$

In order to see that the Hessian is symmetric, we have to lower an index (i.e. multiply on the left by J) to obtain the bilinear form:

$$H_x(y)(v,w) = \langle\, v \mid H_x(y)(w)\,\rangle_* = 2v^{\mathrm{T}}J\left(uu^{\mathrm{T}} + \theta\coth\theta(J + xx^{\mathrm{T}} - uu^{\mathrm{T}})\right)Jw.$$

The eigenvectors and eigenvalues of (half) the Hessian operator are now easy to determine. By construction, x is an eigenvector with eigenvalue $\mu_0 = 0$ (restriction to the tangent space). Then, within the tangent space at x, the vector u (or equivalently $\log_x(y) = \theta u$) is an eigenvector with eigenvalue $\mu_1 = 1$. Lastly, every vector v which is orthogonal to these two vectors (i.e. orthogonal to the plane spanned by 0, x and y) has eigenvalue $\mu_2 = \theta\coth\theta$. Since $\theta\coth\theta \geq 1$ (with equality only for $\theta = 0$, we can conclude that the Hessian of the squared distance is always positive definite and does never vanish along the hyperbolic space. This was of course expected since it is well known that the Hessian stay positive definite for negatively curved spaces Bishop and ONeill [1969]. As a consequence, the squared distance is a convex function and has a unique minimum.

References

R. L. Bishop and B. ONeill. Manifolds of negative curvature. *Transactions of the American Mathematical Society*, 145:1–49, 1969. ISSN 0002-9947, 1088-6850. doi: 10.1090/S0002-9947-1969-0251664-4. URL http://www.ams.org/tran/1969-145-00/S0002-9947-1969-0251664-4/.

Leo Brewin. Riemann normal coordinate expansions using cadabra. *Classical and Quantum Gravity*, 26(17):175017, 2009. URL http://iopscience.iop.org/0264-9381/26/17/175017.

Samuel R. Buss and Jay P. Fillmore. Spherical Averages and Applications to Spherical Splines and Interpolation. *ACM Trans. Graph.*, 20(2):95–126, April 2001. ISSN 0730-0301. doi: 10.1145/502122.502124. URL http://doi.acm.org/10.1145/502122.502124.

John M. Lee. *Riemannian Manifolds: An Introduction to Curvature*. Springer, 1997.

Xavier Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. doi: 10.1007/s10851-006-6228-4. URL https://hal.inria.fr/inria-00614994.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. doi: 10.1007/s11263-005-3222-z. URL https://hal.inria.fr/inria-00614990.

Cédric Villani. Regularity of optimal transport and cut locus: From nonsmooth analysis to geometry to smooth analysis. *Discrete and Continuous Dynamical Systems*, 30(2): 559–571, February 2011. ISSN 1078-0947. doi: 10.3934/dcds.2011.30.559. URL http://www.aimsciences.org/journals/displayArticles.jsp?paperID=5986.