

From Generalized Arithmetic Means to Bregman Divergences and Back

Henryk Gzyl

Center for Finance, IESA, Caracas, Venezuela

Received: 2 Feb. 2022, Revised: 22 Apr. 2022, Accepted: 7 May 2022

Published online: 1 Jul. 2022

Abstract: Here we examine some connections between the notions of generalized arithmetic means, geodesics, Lagrange-Hamilton dynamics and Bregman divergences. The key ingredient for the relationship is the case in which a Riemannian metric has a square root that is the Jacobian of a diffeomorphism. In such case the geodesics of the of the Riemannian metric turn out be the pullback of straight lines by the diffeomorphism. This is interesting when the Riemann metric is the Hessian of a convex function because in this case we obtain comparison results between the Bregman divergence determined by the convex function and the geodesic distance determined by its square root.

Keywords: Generalized arithmetic means, Riemannian distances, Hamilton-Lagrange equivalence, Bregman divergence

1 Introduction and Preliminaries

In this paper we establish a relationship between concepts arising in different fields. We shall see that the concept of generalized arithmetic can be given a geometric and a variational interpretation as well as a multidimensional extension. The geometric interpretation will be as normal coordinates in a Riemannian metric that can be factored as the square of the Jacobian of a diffeomorphism. The variational interpretation relates the generalized arithmetic mean to the best predictor in a geodesic distance given by the aforementioned metric.

Generalized arithmetic mean is a natural extension of the concept of certain equivalent introduced by Bernoulli to deal with the Saint Petersburg paradox in the XVIII-th century. It can be traced back to the work of Bonferroni [1] and de Finetti [2]. Historically speaking, generalized arithmetic means were introduced by Bernoulli under the name of certainty equivalent. They capture the price of a random cash flow (the payoff of a bet) for an investor with utility function u . The first variational characterization of generalized arithmetic mean seems to be due to Berger and Casella, [3], and is worth recalling because it is the germ of the ideas in this work.

Consider a strictly monotone function $u : \mathcal{I} \rightarrow \mathbb{R}$, where $\mathcal{I} \subset \mathbb{R}$ is an open interval. If $\{x_1, \dots, x_n\}$ is a collection of points in \mathcal{I} , their generalized arithmetic

mean is defined by:

$$c = u^{-1} \left(\frac{1}{n} \sum_{i=1}^n u(x_i) \right). \quad (1)$$

That is, c is the point at which the distance $\left(\sum (u(x_i) - u(c))^2 \right)^{1/2}$ is minimal. This is what we extend to a many dimensional setup, and we provide a geometrical interpretation to the change of variables and to the distance as a geodesic distance.

1.1 Organization of the paper

In Section 2 we consider the manifold $\mathcal{M} = \mathcal{I}^n$ and verify that the distance given by

$$d_u^2(\mathbf{x}, \mathbf{y}) = \left(\sum (u(x_i) - u(y_i))^2 \right)^{1/2} \quad (2)$$

is actually a geodesic distance between \mathbf{x} and \mathbf{y} in $c\mathcal{M}$. This simple case is nice because it leads to a diagonal Riemannian metric in \mathcal{M} and the determination of the geodesic distance in that metric is trivial. In Section 3 we replace u by a general diffeomorphism \mathbf{U} defined on \mathcal{M} and use it to define a Riemannian metric on \mathcal{M} . The geodesic distance in this metric is a multidimensional

* Corresponding author e-mail: henryk.gzyl@iesa.edu.ve

version of the generalized arithmetic mean. It is here and Section 5 where the main results are contained.

In Section 4 we consider the changes of variables and the integration of the geodesic equations within the framework of Hamilton-Lagrange formalism just to provide a point of contact with classical mechanics. The idea is to regard the integration of the equations of the geodesics from another point of view.

In Section 5 we relate the results in Sections 2 and 3 to the work of [4]. The relationship goes in two directions. On the one hand, for every $\mathbf{x} \in \mathcal{M}$ we can reinterpret $\mathbf{U}(\mathbf{y}) - \mathbf{U}(\mathbf{x})$ as normal coordinates in a neighborhood of \mathbf{x} . On the other hand, we show that the arithmetic means introduced in Sections 2 and 3 coincides with the notion of intrinsic mean in [4].

It is in Section 6 where we establish conditions for a two way connection between Hessian matrices of convex functions and diffeomorphisms in \mathbb{R}^n , while in Section 7 we compare the geodesic distance to the Bregman divergence defined by the convex function. In Section 8 we collect some examples and address some collateral themes, and in the Final remarks section we say a few words on how this work intersects an older line of work on Hessian structures on \mathbb{R}^n .

A word about notations.

In the remainder of this paper we use $\mathcal{M} = (\mathcal{I})^n$, where as said, \mathcal{I} is an interval (bounded or unbounded). We use this setup because in many cases we want to consider convex functions of the type $\sum u(x_i)$ with $u(x)$ convex. As the manifold that we are considering is a simple subset of \mathbb{R}^n naturally provided with a global chart, the standard constructs of differential geometry in this case are very simple. For example, the tangent bundle and the cotangent bundles are trivially identifiable with $\mathcal{M} \times \mathbb{R}^n$. We use the standard $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ to denote the usual Euclidean scalar product between the vectors $\mathbf{v}_1, \mathbf{v}_2$. And to finish, we use the standard $h'(x)$ to define the derivative of the function $h(x) : \mathcal{I} \rightarrow \mathbb{R}$, and \dot{x} and \ddot{x} to define the first two derivatives of $x(t) : (a, b) \rightarrow \mathbb{R}$ where t is thought of as "time."

2 Geodesics and one-dimensional changes of scale

Let $u : \mathcal{I} \rightarrow \mathbb{R}$ be a twice continuously differentiable, strictly increasing function and $u'(x)$ its derivative. At each $\mathbf{x} \in \mathcal{I}$ define the Riemannian metric (on the tangent space to \mathcal{M} at \mathbf{x}) by

$$g_{i,j}(\mathbf{x}) = g(x_i) \delta_{i,j} = u'(x_i)^2 \delta_{i,j}. \quad (3)$$

That is, not only is the coordinate system orthogonal (that is, the metric is diagonal), but it is separable as well, that is, $g_{i,i}$ is a function of x_i only. The equation of the geodesic that minimizes the distance between two points $\mathbf{x}, \mathbf{y} \in \mathcal{M}$

is obtained minimizing

$$\int_0^1 \left(\sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j \right)^{1/2} dt \quad (4)$$

over the class of continuous functions $\mathbf{x} : [0, 1] \rightarrow \mathcal{M}$, twice continuously differentiable on $(0, 1)$. It is a standard result, that in this case, the equations of the geodesics are (see [5], for example):

$$\begin{aligned} & \frac{d}{dt} \left(\frac{\partial}{\partial \dot{x}_k} \frac{1}{2} \sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j \right) \\ &= \frac{\partial}{\partial x_k} \left(\frac{1}{2} \sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j \right), \quad k = 1, \dots, n. \end{aligned} \quad (5)$$

In this special diagonal metric, if we put $g'(x) = dg(x)/dx$, the equations of the geodesic are

$$g(x) \ddot{x}_k + \frac{1}{2} g'(x_i) (\dot{x}_k)^2, \quad k = 1, \dots, n.$$

Substituting $g(x_k) = (u'(x_k))^2$ we obtain

$$u'(x_k) \ddot{x}_k + u''(x_k) \dot{x}_k^2 = 0, \quad k = 1, \dots, n$$

This can be simply integrated. At the first step, note that we obtain that $u'(x_k(t)) \dot{x}_k = C_k$, where C_k is some unknown constant. This equation can be trivially integrated to obtain $u(x_k(t)) = u(x_k(0)) + tC_k$. To determine C_k make use of the fact that $x_k(0) = x_k$ and $x_k(1) = y_k$ to obtain $C_k = u(y_k) - u(x_k)$. To close the circle, note that the geodesic distance between \mathbf{x} and \mathbf{y} is, according to (4)

$$\begin{aligned} & \int_0^1 \left(\sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j \right)^{1/2} dt \\ &= \int_0^1 \left(\sum_i C_i^2 \right)^{1/2} dt = \left(\sum_i (u(y_k) - u(x_k))^2 \right)^{1/2} \end{aligned}$$

which equals $d_u(\mathbf{x}, \mathbf{y})$.

3 Geodesics induced by a diffeomorphism

Here we extend the setup of the previous section. Let $\mathbf{U} : \mathcal{M} \rightarrow \mathbb{R}^n$ be a twice continuously differentiable diffeomorphism between \mathcal{M} and $\mathbf{U}(\mathcal{M})$. On the tangent bundle $\mathbf{U}(\mathcal{M}) \times \mathbb{R}^n$ we consider the Euclidean metric. Its pullback to $\mathcal{M} \times \mathbb{R}^n$ at $\mathbf{x} \in \mathcal{M}$ is given by

$$g_{i,j}(\mathbf{x}) = \sum_{k=1}^n U_i^k U_j^k, \quad \text{for } 1 \leq i, j \leq n. \quad (6)$$

Here, $U^k(\mathbf{x})$ denotes the k -th component of \mathbf{U} and U_i^k denotes $\partial U^k / \partial x_j$, that is, the Jacobian of \mathbf{U} . We denote by V_j^k the inverse of U_j^k , that is, the Jacobian of the inverse of \mathbf{U} . Again, the geodesics minimize the distance

given by (4) with $g_{i,j}$ given by (6). If $t \rightarrow \mathbf{x}(t)$ is a geodesic from $\mathbf{x}(0) = \mathbf{x}$ to $\mathbf{x}(1) = \mathbf{y}$, then it must satisfy equation (5).

If in (5) we substitute (6), after some simple but boring arithmetics we obtain

$$\sum_{i=1}^n U_k^i \left(\sum_{n=1}^n U_n^i \dot{x}_n + \sum_{k,l} U_{n,l}^i \dot{x}_n \dot{x}_l \right) = 0, \quad k = 1, \dots, n.$$

Now, multiply both sides by V_j^k and sum over $k = 1, \dots, n$ to obtain

$$\sum_{n=1}^n U_n^j \dot{x}_n + \sum_{k,l} U_{n,l}^j \dot{x}_n \dot{x}_l = 0, \quad j = 1, \dots, n.$$

This can be written as

$$\frac{d}{dt} \left(\sum_{n=1}^n U_n^j \dot{x}_n \right) = 0, \quad j = 1, \dots, n.$$

Which implies that

$$\frac{d}{dt} U^j(\mathbf{x}(t)) = \sum_{n=1}^n U^j(\mathbf{x}(t)) \dot{x}_j(t) = C_j, \quad j = 1, \dots, n$$

for some constants $C_j, j = 1, \dots, n$. Again, this implies that $U^j(\mathbf{x}(t)) = U^j(\mathbf{x}(0)) + tC_j$ and that $C_j = U^j(\mathbf{y}) - U^j(\mathbf{x})$ for $j = 1, \dots, n$. As above, inserting this into the definition of geodesic distance between \mathbf{x} and \mathbf{y} we obtain

$$\int_0^1 \left(\sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j \right)^{1/2} dt = \int_0^1 \left(\sum_i C_i^2 \right)^{1/2} dt \tag{7}$$

$$= \left(\sum_i (U^i(\mathbf{y}) - U^i(\mathbf{x}))^2 \right)^{1/2} = d_U(\mathbf{x}, \mathbf{y}).$$

Notice that $d_U(\mathbf{x}, \mathbf{y})$ is positive, symmetric, satisfies the triangle inequality, and since \mathbf{U} is a diffeomorphism, $d_U(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$, thus the notation is consistent. The comparison with the first case is clear: There $U^k(\mathbf{x}) = u(x_k)$ for $k = 1, \dots, n$.

4 A detour into classical mechanics

Here we address the issue of integrating the geodesic equations from the point of view of classical mechanics. This is to explain the trivial integration of the geodesic equations within the context of an elegant framework. We refer the reader to [6] for the essential notions. Cutting some edges, the general approach to Newton's equations, consists of specifying some manifold (\mathcal{M} in our case), on whose tangent bundle $\mathcal{M} \times \mathbb{R}^n$ a Lagrangian function $L(\mathbf{x}, \dot{\mathbf{x}})$ is defined. A trajectory of the system between points $\mathbf{x}(1)$ and $\mathbf{x}(2)$ at times t_1, t_2 respectively, is the curve that minimizes the (action) integral:

$$\int_{t_1}^{t_2} L(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) dt$$

over the class of all (twice continuously) differentiable curves joining the said points at the given times. It can be shown, see [6], that such trajectory satisfies (the Euler-Lagrange) extension of the Newton's equation of motion:

$$\frac{d}{dt} \left(\frac{\partial}{\partial \dot{x}_k} L(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) \right) = \frac{\partial}{\partial x_k} L(\mathbf{x}(t), \dot{\mathbf{x}}(t), t). \tag{8}$$

If we put

$$L(\mathbf{x}(t), \dot{\mathbf{x}}(t)) = \frac{1}{2} \sum_{i,j} g_{i,j}(\mathbf{x}) \dot{x}_i \dot{x}_j,$$

We recognize the equations of the geodesics as the Euler-Lagrange equations of some dynamical system. Notice that if we consider the curves $\mathbf{y}(t) = \mathbf{U}(\mathbf{x}(t))$ in $\mathbf{U}(\mathcal{M})$, then the Lagrangian in these coordinates becomes

$$L(\mathbf{y}, \dot{\mathbf{y}}, t) = \frac{1}{2} \sum_{k=1}^n (\dot{y}_k)^2$$

and the Euler-Lagrange (Newton) equations become $\ddot{y}_k = 0$, namely the equations of straight lines. We already saw that in Section 3, hidden in the notation. Another way of understanding that change of variables is to begin with the Hamiltonian associated with the Lagrangian. This is defined by introducing the momentum variables p_k by

$$p_k = \frac{\partial}{\partial \dot{x}_k} L(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) = \sum_{j=1}^n g_{j,k}(\mathbf{x}) \dot{x}_j.$$

Under the assumption of invertibility of the Jacobian $U_{j,i}^i$, solving for $\dot{\mathbf{x}}$ and substituting in the expression for L we obtain:

$$H(\mathbf{x}, \mathbf{p}) = \frac{1}{2} \sum_{k,j} g^{k,j}(\mathbf{x}) p_k p_j.$$

where we use the conventional $g^{k,j}$ to denote the inverse of $g_{k,j}$. The analogue of the Euler-Lagrange equations are the Hamilton equations for (\mathbf{x}, \mathbf{p}) is:

$$\dot{x}_k = \frac{\partial H}{\partial p_k} \quad \dot{p}_k = - \frac{\partial H}{\partial x_k}. \tag{9}$$

To complete explaining the change of variables we need to show that there exists a mapping $(\mathbf{x}, \mathbf{p}) \rightarrow (\mathbf{y}, \boldsymbol{\pi})$ from $\mathcal{M} \times \mathbb{R}^n$ to $\mathbf{U}(\mathcal{M}) \times \mathbb{R}^n$, under which the Hamiltonian becomes

$$\tilde{H} = \frac{1}{2} \sum_{k=1}^n (\pi_k)^2,$$

and that the change of variables preserves the form of the Hamiltonian equations of motion (9). The mapping is:

$$y^i = U^i(\mathbf{x}) \quad \text{and} \quad \pi_i = \sum_{k=1}^n p_j V_i^j; \quad i = 1, \dots, n$$

The condition for this mapping to preserve the form of the equations is that the following holds for the given Poisson “brackets”

$$[y^i, y^j] = [\pi_i, \pi_j] = 0, \text{ and } [y^i, \pi_j] = \delta_{i,j}, \text{ } i, j = 1, \dots, n$$

which, for any pair of continuously differentiable functions f, g on $\mathcal{M} \times \mathbb{R}^n$ is given by:

$$[f, g] = \sum_{k=1}^n \frac{\partial f}{\partial x_k} \frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k} \frac{\partial g}{\partial x_k}.$$

As we said, we were to cut a lot of corners. See [6] for full details. That the system is trivially integrable, means that (9), in the new coordinates becomes

$$\dot{y}^i = \frac{\partial \tilde{H}}{\partial \pi_i} = \pi_i \text{ and } \dot{\pi}_i = -\frac{\partial \tilde{H}}{\partial y^i} = 0,$$

which imply that $\pi_i = \text{constant} = C_i$ and $y^i = y^i(0) + tC_i$. That is, the constants of integration that determine the solutions to the equations of the geodesics are the constant values of the momenta. For trajectories go from a point \mathbf{x} to a point \mathbf{y} in a unit of time, the momenta are related to the distance between the two points by $y^i(1) - y^i(0) = C^i$.

5 Hessians of convex functions and diffeomorphisms in \mathbb{R}^n

In the two (sub)sections that come up next, we do two things. First we extend the results in [7] that were summed up in Section 2, to a more general convex function, after that we examine in which case a diffeomorphism \mathbf{U} determines a convex function whose Hessian coincides with the Riemann metric given by (6).

5.1 From a convex function to diffeomorphisms

Let us begin by stating some necessary properties of the convex function.

Assumptions 51 Let us now consider an open, convex subset \mathcal{M} of \mathbb{R}^n and a strictly convex function $\Phi : \mathcal{M} \rightarrow \mathbb{R}^n$ satisfying

1 Φ is at least twice continuously differentiable in all variables.

Denote its Hessian matrix by Φ'' , that is, $(\Phi'')_{i,j} = \partial^2 \Phi / \partial x_i \partial x_j$. The strict convexity of Φ implies that the Hessian Φ'' is positive definite. Let us denote by $k(\mathbf{x})$ its smallest eigenvalue and let us suppose that:

2 There exists a $a > 0$ such that $a \leq k(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{M}$, or equivalently

$$0 < a \langle \xi, \xi \rangle \leq k(\mathbf{x}) \langle \xi, \xi \rangle \leq \langle \xi, \Phi'' \xi \rangle, \quad \forall \xi \in \mathbb{R}^n. \quad (10)$$

And we now have:

Theorem 1. With the notations introduced above, let us suppose that there exists a continuously differentiable square root S of Φ'' , that is, $\Phi'' = S^t S$ such that:

a $S(\mathbf{x})$ is invertible at each $\mathbf{x} \in \mathcal{M}$.

b For every $i = 1, \dots, n$ we have $\partial S_{i,j} / \partial x_k = \partial S_{i,k} / \partial x_j$, for all $1 \leq j, k \leq n$.

Fix some $\mathbf{x}_0 \in \mathcal{M}$ and define

$$U^i(\mathbf{x}) = \int S_{i,j}(\gamma(s)) \dot{\gamma}(s) ds$$

where γ denotes any continuously differentiable trajectory between \mathbf{x}_0 and \mathbf{x} . Then, the mapping $\mathbf{U} : \mathcal{M} \rightarrow \mathbb{R}^n$ with components $U^i(\mathbf{x})$ is well defined (up to a constant), and it is a global diffeomorphism satisfying

$$\Phi''_{i,j} = \sum_{m=1}^n \frac{\partial U^m}{\partial x_i} \frac{\partial U^m}{\partial x_j}$$

Before proving the theorem we need the following result.

Lemma 1. With the notations introduced above, let T denote the inverse S^{-1} of S . Then $\|T\| \leq 1/a$.

Proof. In item (2) of Assumption 51 replace ξ by $\mathbf{V}\xi$ to obtain

$$\begin{aligned} a \langle \mathbf{V}\xi, \mathbf{V}\xi \rangle &\leq k(\mathbf{x}) \langle \mathbf{V}\xi, \mathbf{V}\xi \rangle \leq \langle \mathbf{V}\xi, \Phi'' \mathbf{V}\xi \rangle \\ &= \langle \xi, \mathbf{V}^t \Phi'' \mathbf{V}\xi \rangle = \langle \xi, \mathbf{V}^t U^t U \mathbf{V}\xi \rangle = \langle \xi, \xi \rangle. \end{aligned}$$

In other words

$$\|\mathbf{V}^t \mathbf{V}\| \leq 1/k(\mathbf{x}) \leq 1/a \Rightarrow \|\mathbf{V}\| < K = (\sqrt{a^{-1}})$$

since $\|\mathbf{V}\|$ is given by the square root of the largest eigenvalue of $\mathbf{V}^t \mathbf{V}$.

Let us now complete the proof of Theorem 1.

Proof.

Since \mathcal{M} is convex, it is simply connected and assumption (b) in the statement implies (via Stokes theorem) that \mathbf{U} is a well defined mapping on \mathcal{M} , and that its Jacobian S is non vanishing (by assumption), and thus \mathbf{U} is a local diffeomorphism. Invoking the previous lemma and Hadamard’s theorem, see, for example Theorem 2 in [8], we conclude that \mathbf{U} is a diffeomorphism satisfying

$$\Phi''_{i,j} = \sum_{m=1}^n \frac{\partial U^m}{\partial x_i} \frac{\partial U^m}{\partial x_j}$$

which concludes the proof.

5.2 From diffeomorphisms to convex functions

A convex function has a positive definite Hessian. The next result imposes a condition on the diffeomorphism U so that the metric that it defines, as explained in Section 3, can be the Hessian of a convex function. The result that we obtain is a variation on the theme of Section 2 of [9].

Theorem 2. Let \mathcal{M} be an open convex subset of \mathbb{R}^n and $U : \mathcal{M} \rightarrow \mathbb{R}^n$. be a diffeomorphism satisfying

$$\sum_{m=1}^n \frac{\partial^2 U^m}{\partial x_k \partial x_i} \frac{\partial U^m}{\partial x_j} = \sum_{m=1}^n \frac{\partial^2 U^m}{\partial x_i \partial x_j} \frac{\partial U^m}{\partial x_k} \quad (11)$$

Then there exists a strictly convex function $\Phi : \mathcal{M} \rightarrow \mathbb{R}^n$ such that

$$\frac{\partial^2 \Phi}{\partial x_i \partial x_j} = \sum_{m=1}^n \frac{\partial U^m}{\partial x_i} \frac{\partial U^m}{\partial x_j}$$

Proof. As before, $g_{i,j} = \sum_{m=1}^n \frac{\partial U^m}{\partial x_i} \frac{\partial U^m}{\partial x_j}$. Fix an $1 \leq i \leq n$ and an $\mathbf{x}(0) \in \mathcal{M}$ and consider the following curve joining $\mathbf{x}(0)$ to \mathbf{x} defined piecewise by increasing one coordinate at a time: At the k -th step, move along the k -th coordinate axis from $x_k(0)$ to x_k , that is, along the line:

$$(x_1, x_2, \dots, x_{k-1}, \xi_k, x_{k+1}(0), \dots, x_n(0)) \text{ with } x_k(0) \leq \xi_k \leq x_k.$$

Call this trajectory γ . The import of condition(11) is to make the integral defined below to be independent of the trajectory.

$$A^i(\mathbf{x}) = A^i(\mathbf{x}_0) + \sum_{k=1}^{x_k} \int_{x_k(0)}^{x_k} g_{i,k}(\gamma) d\xi_k.$$

where $A^i(\mathbf{x}_0)$ are constants of integration. Now, since $\partial A^i / \partial x_j = g_{i,j} = g_{j,i} = \partial A^j / \partial x_i$, then the following line integral is also independent of the trajectory. Thus integrating along the same piecewise trajectory we put:

$$\Phi(\mathbf{x}) = \Phi(\mathbf{x}(0)) + \sum_{k=1}^{x_k} \int_{x_k(0)}^{x_k} A^k(\gamma) d\xi_k.$$

We choose the special trajectory so that the verification that the Hessian of Φ is g is trivial. Thus we prove the claim.

6 Generalized means, intrinsic means and exponential barycenters

Throughout this section we suppose that $U : \mathcal{M} \rightarrow \mathbb{R}^n$ is as in Section 3, namely a twice continuously differentiable diffeomorphism between \mathcal{M} and its image $U(\mathcal{M})$. As said above, here we relate the generalized arithmetic means to the notion of intrinsic mean and barycentric mean presented in [4].

As \mathcal{M} is a rather simple manifold with tangent at every point \mathbf{x} denoted by \mathcal{M}_x . We might think of it as the set

of all displacements starting at \mathbf{x} . On \mathcal{M}_x we consider the metric given by (6) If $\mathbf{x}(t) : (a, b) \rightarrow \mathcal{M}$ is a continuously differentiable curve (a, b) is some open interval containing $[0, 1]$, its length between $\mathbf{x}(0) = \mathbf{x}_0$ and $\mathbf{x}(1) = \mathbf{x}_1$ is given by

$$L[\mathbf{x}_0, \mathbf{x}_1] = \int_0^1 \sqrt{\left\langle \frac{d\mathbf{U}(\mathbf{x}(t))}{dt}, \frac{d\mathbf{U}(\mathbf{x}(t))}{dt} \right\rangle} dt. \quad (12)$$

In Section3 we showed that the geodesic between \mathbf{x}_0 and \mathbf{x}_1 is given by

$$\mathbf{x}(t) = U^{-1} \left(U(\mathbf{x}_0) + t\xi \right), \text{ where } \xi = U(\mathbf{x}_1) - U(\mathbf{x}_0). \quad (13)$$

If we want to fix the initial point, call it \mathbf{x} now, and think of ξ as a possible initial velocity we have to be careful. The problem shows up if \mathcal{M} is bounded. In this case if we want to let ξ vary, it will have to be in a ball about 0 such that $U(\mathbf{x}) + \xi \in U(\mathcal{M})$. Such an ball is easy to determine in our setup because both U and U^{-1} are diffeomorphisms. So, given \mathbf{x} let $B(r)$ be a ball about 0 in \mathcal{M}_x such that $U(\mathbf{x}) + \xi \in U(\mathcal{M})$ for $\xi \in B(r)$. In this case, the exponential mapping $\exp_x : B(0) \subset \mathcal{M}_x \rightarrow \mathcal{M}$ is given by

$$\exp_x(\xi) = U^{-1} \left(U(x) + \xi \right). \quad (14)$$

Note that given \mathbf{y} in $U(\mathcal{M})$, it comes from $\xi = \exp_x^{-1}(\mathbf{y}) = U(\mathbf{y}) - U(\mathbf{x})$. And the characterization of $m \in \mathcal{M}$ as intrinsic mean value of a \mathcal{M} valued random variable \mathbf{X} given in [4] or of m as barycenter of \mathbf{X} is as follows.

Definition 1. The point $m \in \mathcal{M}$ is the (intrinsic) mean value of \mathbf{X} if there is a normal neighborhood W of m such that $P(\mathbf{X} \in W) = 1$ and

$$E_P[\exp_m^{-1}(\mathbf{X})] = E_P[U(\mathbf{X}) - U(m)] = 0.$$

Clearly, $\xi = U(\mathbf{X}) - U(m)$ is the initial velocity of a trajectory that takes m to \mathbf{X} along the geodesic (13) in a unit of time.

7 A comparison result: Divergence versus geodesic distance

Bregman divergences are a common measure of discrepancy. They are used to compare how different are two objects that can be described by points in a convex subset of some many-dimensional space. The definition goes as follows:

Definition 2. Let $\mathcal{M} \subset \mathbb{R}^n$. be an open convex set $\Phi : \mathcal{M} \rightarrow \mathbb{R}$ be a strictly convex, continuously differentiable function and put

$$\delta_{\Phi}^2(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) - \Phi(\mathbf{y}) - \langle (\mathbf{x} - \mathbf{y}), \nabla \Phi(\mathbf{y}) \rangle. \quad (15)$$

Comment: Since the right hand is non-negative, and vanishes if and only if $\mathbf{y} = \mathbf{x}$, the notation chosen for the left hand side is consistent, even though in general, it is not a true distance. This is why it is called a discrimination function between \mathbf{x} and \mathbf{y} .

This notion was introduced in [10], and has been used in a variety of applications. For a short list, consult with [7]. Actually, the function Φ considered there is rather simple:

$$\Phi(\mathbf{x}) = \sum_{i=1}^n \phi(x_i)$$

with $\phi : \mathcal{I} \rightarrow \mathbb{R}$ being a convex function defined on the interval \mathcal{I} and $\mathcal{M} = \mathcal{I}^n$. This is a typical example in many applications. The thrust in [7] was to compute the geodesic distance defined by the Riemannian metric defined by the Hessian of Φ , and to compare it to the pseudo distance δ_Φ defined in (15).

In this section we generalize a comparison result previously obtained in [7] for the separable case. Before we state the result, we shall introduce some notations to declutter the typography and carry out a few elementary calculations that form the basis of the proof of the result.

We shall use ∂_i , $\partial_{i,j}^2$ to denote partial (and repeated partial) derivatives with respect to x_i (respectively x_i and x_j). We shall use Einstein summation convention whenever it does not lead to confusion. That is, for example $\partial_i U^m \partial_j U^m$ stands for $\sum_{m=1}^n \partial_i U^m \partial_j U^m$. Next we present three instances of the same computation: Once for functions, once for vector fields and once for matrix valued functions. Just to refer to them when the time comes up.

Again let $\mathcal{M} \subset \mathbb{R}^n$ be an open, convex, connected set and $\Phi : \mathcal{M} \rightarrow \mathbb{R}$ be a twice continuously differentiable function. Let $\gamma : [0, 1] \rightarrow \mathcal{M}$ be a continuously differentiable curve. We shall denote by γ_k the k -th component of γ . Let $\mathbf{y}, \mathbf{x} \in \mathbb{R}^n$ and suppose that $\gamma(0) = \mathbf{x}$ and $\gamma(1) = \mathbf{y}$. Below, anytime that $\gamma(t)$ appears as the argument, say of a function defined on \mathcal{M} , we shorten it to t .

Starting from

$$\Phi(\mathbf{y}) = \Phi(\mathbf{x}) + \int_0^1 \partial_i \Phi(s) \dot{\gamma}_i(t) dt.$$

As above $\dot{\gamma}$ stands for the time derivative of γ . Apply this same computation to the function $\partial_i \Phi(\mathbf{x})$ that appears under the integral sign to obtain.

$$\Phi(\mathbf{y}) = \Phi(\mathbf{x}) + \int_0^1 \left(\partial_i \Phi(0) + \int_0^t \partial_{k,i}^2 \Phi(s) \dot{\gamma}_k(s) dt \right) \dot{\gamma}_i(t) dt.$$

Now integrate the first term to obtain $(\mathbf{y} - \mathbf{x})_i \partial_i \Phi(\mathbf{x})$. To complete, exchange the integration over t with that over s , notice that the integral over t becomes $(\mathbf{y} - \gamma(s))_i$ and the whole identity becomes

$$\begin{aligned} \Phi(\mathbf{y}) &= \Phi(\mathbf{x}) \\ &+ (\mathbf{y} - \mathbf{x})_i \partial_i \Phi(\mathbf{x}) + \int_0^1 (\mathbf{y} - \gamma(s))_i \partial_{k,i}^2 \Phi(s) \dot{\gamma}_k(s) ds. \end{aligned} \quad (16)$$

A similar identity holds componentwise for vector valued functions $\mathbf{A} : \mathcal{M} \rightarrow \mathbb{R}^n$. In this case the analogue of (16) becomes

$$\begin{aligned} A_m(\mathbf{y}) &= A_m(\mathbf{x}) + (\mathbf{y} - \mathbf{x})_i \partial_i A_m(\mathbf{x}) \\ &+ \int_0^1 (\mathbf{y} - \gamma(u))_i \partial_{k,i}^2 A_m(u) \dot{\gamma}_k(u) du. \end{aligned} \quad (17)$$

Let us now rewrite (16) as

$$\begin{aligned} \delta_\Phi^2(\mathbf{y}, \mathbf{x}) &= \Phi(\mathbf{y}) - \Phi(\mathbf{x}) + (\mathbf{y} - \mathbf{x})_i \partial_i \Phi(\mathbf{x}) \\ &= \int_0^1 (\mathbf{y} - \gamma(s))_i \partial_{k,i}^2 \Phi(s) \dot{\gamma}_k(s) ds \end{aligned} \quad (18)$$

and proceed with the right hand side as follows. Suppose that as in Section 3 that there exists a diffeomorphism \mathbf{U} such that $\partial_{i,k}^2 \Phi = \partial_i U^m \partial_k U^m$. Now, for a fixed m consider only $(\mathbf{y} - \gamma(s))_i \partial_i U^m$. According to (17) in which \mathbf{x} is replaced by $\gamma(s)$, this can be rewritten as

$$(U^m(\mathbf{y}) - U^m(s)) - \int_s^1 (\mathbf{y} - \gamma(u))_i \partial_{i,j}^2 U^m(u) \dot{\gamma}_j(u) du.$$

and reinserted back in (19) to obtain

$$\begin{aligned} \delta_\Phi^2(\mathbf{y}, \mathbf{x}) &= \int_0^1 \left((U^m(\mathbf{y}) - U^m(s)) \right. \\ &\left. - \int_s^1 (\mathbf{y} - \gamma(u))_i \partial_{i,j}^2 U^m(u) \dot{\gamma}_j(u) du \right) \partial_k U^m \dot{\gamma}_k(s) ds. \end{aligned}$$

Therefore, the first term under the outer integral becomes

$$\begin{aligned} &\int_0^1 (U^m(\mathbf{y}) - \gamma(s)) \partial_k U^m \dot{\gamma}_k(s) ds \\ &= \frac{1}{2} \|U(\mathbf{y}) - U(\mathbf{x})\|^2 = \frac{1}{2} d_U^2(\mathbf{x}, \mathbf{y}) \end{aligned} \quad (19)$$

To rewrite the second term, notice that $\partial_k U^m \dot{\gamma}_k(s) ds = dU^m(\gamma(s))$. Now, we shall consider a specific trajectory: $\gamma(t) = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$. With this choice we have

$$(\mathbf{y} - \gamma(u))_i \partial_{i,j}^2 U^m(u) \dot{\gamma}_j(u) = (1-u)(\mathbf{y} - \mathbf{x})_i U_{i,j}^m(u) (\mathbf{y} - \mathbf{x})_j$$

and therefore, the second integral becomes

$$\int_0^1 \left(\int_s^1 (1-u) ((\mathbf{y} - \mathbf{x})_i \partial_{i,j}^2 U^m(u) (\mathbf{y} - \mathbf{x})_j) du \right) dU^m. \quad (20)$$

We now gather these results in the main result of this section.

Theorem 3. *With the notations introduced above, suppose that the convex function $\Phi : \mathcal{M} \rightarrow \mathbb{R}$ is at least three times continuously differentiable, that its Hessian can be factored as $\partial_{i,k}^2 \Phi = \partial_i U^m \partial_k U^m$, and that the diffeomorphism \mathbf{U} is at least twice continuously*

differentiable and such that the sign of $K(\xi) \equiv \partial_{i,j}^2 U^m(\xi) \partial_k U^m(\xi)$ is constant over $\xi \in \mathcal{M}$, then

$$\delta_{\Phi}^2(\mathbf{y}, \mathbf{x}) \leq \frac{1}{2} d_U^2(\mathbf{x}, \mathbf{y}) \text{ whenever } K(\xi) \geq 0, \tag{21}$$

$$\delta_{\Phi}^2(\mathbf{y}, \mathbf{x}) \geq \frac{1}{2} d_U^2(\mathbf{x}, \mathbf{y}) \text{ whenever } K(\xi) \leq 0. \tag{22}$$

The proof is contained in the computations carried out above. The conclusion is obtained after substituting (19) and (20) into (16).

In the most common case in applications, when $\Phi(\mathbf{x}) = \sum_{k=1}^n \phi(x_k)$, that is in the separable case, then $\phi''(x) = (u'(x))^2 > 0$, then $u'(x) > 0$ and the condition upon the sign of $u''(x)$ is equivalent to a condition upon the sign of $\phi'''(x)$. In this case, the result of Theorem 3 was obtained in [7].

8 Examples

8.1 Extended generalized arithmetic means

Here is a direct extension of the notion of generalized arithmetic mean.

Theorem 4. Let \mathcal{M} be a convex open subset of \mathbb{R}^d and $\mathbf{U} : \mathcal{M} \rightarrow \mathbb{R}^d$ be a diffeomorphism such that $\mathbf{U}(\mathcal{M})$ is a convex open subset of \mathbb{R}^d . Consider a set $\{\mathbf{x}(1), \dots, \mathbf{x}(M)\}$ of points in $c\mathcal{M}$. There is a unique point \mathbf{c} in \mathcal{M} which minimizes the d_U distance to the set $\{\mathbf{x}(1), \dots, \mathbf{x}(M)\}$. It is given by

$$\mathbf{c} = \mathbf{U}^{-1} \left(\frac{1}{M} \sum_{m=1}^M \mathbf{U}(\mathbf{x}(m)) \right). \tag{23}$$

The proof is easy and follows the pattern of the simple one-dimensional case: To find $\xi \in \mathbf{U}(\mathcal{M})$ such that minimizes $\sum_{m=1}^M \|\mathbf{U}(\mathbf{x}(m)) - \xi\|^2$ is easy. It is given by $\xi = \frac{1}{M} \sum_{m=1}^M \mathbf{U}(\mathbf{x}(m))$. Since $\xi \in \mathbf{U}(\mathcal{M})$, let $\mathbf{c} = \mathbf{U}(\xi)$ and we are through.

8.2 Means defined by a flow

With the notations of Section 3 and the previous example in mind, let $\mathbf{U}(t, \mathbf{x})$ be the flow associated to the geodesics determined by the diffeomorphism \mathbf{U} , and the solution to the geodesic equation such that $\mathbf{x}(0) = \mathbf{x}$ and $\dot{\mathbf{x}}(0) = \xi$. That is

$$\mathbf{U}(t, \mathbf{x}) = \mathbf{U}^{-1}(\mathbf{U}(\mathbf{x}) + t\xi). \tag{24}$$

To verify that for any real t, s $\mathbf{U}(t+s, \mathbf{x}) = \mathbf{U}(s, \mathbf{U}(t, \mathbf{x}))$ is routine -as long as the solution to the geodesic equations is defined for all times. With the aid of (24) one can construct a family of transition kernels $\{P_t(\mathbf{x}, A) : t \geq 0, A \in \mathcal{B}(\mathbb{R}^n)\}$, where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel-subsets of \mathbb{R}^n , as follows:

$$P_t(\mathbf{x}, A) = I_A(\mathbf{U}(t, \mathbf{x})) \tag{25}$$

where I_A stands for the usual indicator function of the set A . To verify that $\{P_t\}$ is indeed a (semi)group is trivial using (24). Skipping a considerable amount of detail, it is intuitive that this semi group defines a Markov process $\{\mathbf{X}_t : t \geq 0\}$ having \mathbb{R}^n as state space, and $\{P_t\}$ as transition semigroup. Furthermore, this process is such that for any Borel, measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we have

$$E[f(\mathbf{X}_{t+s}) | \mathbf{X}_t] = f(\mathbf{U}(s, \mathbf{X}_t)).$$

In particular, if $X_i(t)$ denotes the i -th coordinate of \mathbf{X}_t for $i = 1, \dots, n$ we have $E[X_i(t+s) | \mathbf{X}_t] = U_i(s, \mathbf{X}_t)$, or in vector notation:

$$E[\mathbf{X}(t+s) | \mathbf{X}_t] = \mathbf{U}(s, \mathbf{X}_t),$$

that is, the current position is the best predictor of the future values of the position.

8.3 Harmonic means in \mathbb{R}^n

As second example, consider the inversion with respect to the unit sphere in $\mathcal{D} = \mathbb{R}^n \setminus \{0\}$. The mapping $\mathbf{U}(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|^2$ is an involution of \mathcal{D} . Let $\mathcal{M} \subset \mathcal{D}$ such that $\mathbf{U}(c\mathcal{M})$ is convex. If $\{\mathbf{x}(1), \dots, \mathbf{x}(M)\}$ is a finite set of points in \mathcal{M} , then

$$\mathbf{c} = \mathbf{U} \left(\frac{1}{M} \sum_{m=1}^M \frac{\mathbf{x}(m)}{\|\mathbf{x}(m)\|^2} \right)$$

is the point closest to $\{\mathbf{x}(1), \dots, \mathbf{x}(M)\}$ is the d_U distance. Thus we have a variational interpretation of the n -dimensional extension of the notion of harmonic mean. The metric defined by \mathbf{U} is:

$$g_{i,j}(\mathbf{x}) = \left(\frac{1}{\|\mathbf{x}\|^4} \right) (\delta_{i,j} + 4n_i(\mathbf{x})n_j(\mathbf{x})), \quad n_i(\mathbf{x}) = x_i/\|\mathbf{x}\|.$$

The geodesic that starts at $\mathbf{x}_0 \in \mathcal{D}$ with initial speed ξ is $\mathbf{x}(t) = \mathbf{U}(\mathbf{U}(\mathbf{x}_0) + t\xi)$. It is defined for all t if $\mathbf{U}(\mathbf{x}_0) + t\xi$ never passes through 0.

8.4 A metric defined by a Jacobian

Consider $\mathcal{M} = (0, \infty)^2$ and $\mathbf{U} : \mathcal{M} \rightarrow (0, 1)^2$ defined by

$$U^1(x_1, x_2) = \frac{x_1}{1+x_1+x_2},$$

$$U^2(x_1, x_2) = \frac{x_2}{1+x_1+x_2}.$$

The inverse $\mathbf{V} =: (0, 1)^2 \rightarrow (0, \infty)^2$ of that mapping is

$$V^1(y_1, y_2) = \frac{y_1}{1-(y_1+y_2)},$$

$$V^2(y_1, y_2) = \frac{y_2}{1-(y_1+y_2)}.$$

The Jacobian of the mapping \mathbf{U} is given by

$$U_j^i(\mathbf{x}) = \frac{1}{(1+x_1+x_2)^2} \begin{bmatrix} x_1+x_2 & -x_2 \\ -x_1 & x_1+x_2 \end{bmatrix}.$$

As proposed in Section 3, this leads to a metric given by

$$g_{i,j}(\mathbf{x}) = \frac{1}{(1+x_1+x_2)^4} \begin{bmatrix} (x_1+x_2)^2+x_1^2 & -(x_1+x_2)^2 \\ -(x_1+x_2)^2 & (x_1+x_2)^2+x_2^2 \end{bmatrix}.$$

To determine the geodesics in this metric might be quite laborious. But knowing the connection established in Section 3, we know that the geodesic through a point $\mathbf{x}(0)$ with initial speed ξ is given by

$$\mathbf{x}(t) = \mathbf{V} \left(\mathbf{U}(\mathbf{x}(0)) + t\xi \right).$$

8.5 An example of a convex function whose Hessian is a product of Jacobians

For this example, let $\mathcal{M} \subset \mathbb{R}^d = \prod_{k=1}^d \mathcal{I}$ be a product of open intervals and let $\phi_k : \mathcal{I} \rightarrow \mathbb{R}$ be twice continuously differentiable, strictly monotone convex functions, and let u_k be strictly increasing and such that $(u_k')^2 = \phi_k''$ as in Section 2. Let \mathbf{A} be a $d \times d$ invertible matrix and think of its rows as the transpose of the d -vectors \mathbf{a}^k .

Thus the mapping $\mathbf{U} : \mathcal{M} \rightarrow \mathbb{R}^d$ whose k -th component is $u_k(\langle \mathbf{a}^k, \mathbf{x} \rangle)$, is a diffeomorphism. To check its invertibility, note that for \mathbf{y} in its range, the solution of $\mathbf{y} = \mathbf{U}(\mathbf{x})$ is given by $x_i = \mathbf{A}_{i,j}^{-1} u_j^{-1}(y_j)$. The Jacobian matrix of \mathbf{U} is

$$\frac{\partial U^k(\mathbf{x})}{\partial x_i} = u_k'(\langle \mathbf{a}^k, \mathbf{x} \rangle) a_i^k.$$

From the above and $\phi_k'' = (u_k')^2$ it is also clear that

$$\frac{\partial^2 \Phi}{\partial x_i \partial x_j} = \sum_{k=1}^d a_i^k a_j^k \frac{\partial^2 \phi_k}{\partial x_i \partial x_j}(\langle \mathbf{a}^k, \mathbf{x} \rangle) = \sum_{k=1}^d \frac{\partial U^k(\mathbf{x})}{\partial x_i} \frac{\partial U^k(\mathbf{x})}{\partial x_j}.$$

8.6 Multidimensional best predictors

Here we extend the situation considered in the second example. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, that is, a set Ω , a σ -algebra of subsets of Ω , and a probability \mathbb{P} . By the customary $E_{\mathbb{P}}$ we shall denote expectation with respect to \mathbb{P} . All the random variables \mathbf{X} in this section will take values in $\mathbf{U}(\mathcal{M})$ and be such that $E_{\mathbb{P}}[\mathbf{U}(\mathbf{X})^2] < \infty$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub- σ -algebra. We have

Theorem 5. Define the \mathbf{U} distance between any two random variables \mathbf{X}, \mathbf{Y} by

$$\delta_U(\mathbf{X}, \mathbf{Y}) = \left(E_{\mathbb{P}}[d_U(\mathbf{X}, \mathbf{Y})^2] \right)^{1/2}.$$

Then, there is a unique, \mathcal{G} -measurable, square integrable random variable $\mathbf{X}_{\mathcal{G}}$ which satisfies

$$\mathbf{X}_{\mathcal{G}} = \operatorname{argmin}\{\delta_U(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} \text{ measurable with respect to } \mathcal{G}\}.$$

It is given by

$$\mathbf{X}_{\mathcal{G}} = \mathbf{U}^{-1} \left(E_{\mathbb{P}}[\mathbf{U}(\mathbf{X}) \mid \mathcal{G}] \right).$$

A particular case of this result is contained in the following result.

Theorem 6. Let $\mathcal{G} = \{\emptyset, \Omega\}$ be the trivial sigma algebra, and Let \mathbf{X} be a random variable taking finitely many values $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ with probabilities $P(\mathbf{X} = \mathbf{x}_i) = p_i$, $i = 1, \dots, M$. The rest of the notations are as before. The best predictor of \mathbf{X} given no information is given by the generalized arithmetic mean. As noted at the outset, it happens to coincide with the notion of certainty equivalent.

$$\langle \mathbf{X} \rangle = \mathbf{U}^{-1} \left(E_{\mathbb{P}}[\mathbf{U}(\mathbf{X})] \right) = \mathbf{U}^{-1} \left(\sum_{k=1}^M \mathbf{U}(\mathbf{x}_i) p_i \right). \quad (26)$$

8.7 Generalized arithmetic means and convex functions defined by the gradient of a convex function

A particular case of the results in sections 3 and 5 is provided by strictly convex (or concave) functions, that is functions whose Hessian is strictly positive (or negative). Suppose $\Psi(\mathbf{x})$ is a strictly positive function on \mathbb{R}^n , then $\nabla \Psi$ is locally invertible. Let us suppose that its inverse is global. If we consider the metric introduced in (6), that is,

$$g_{i,j}(\mathbf{x}) = \Psi_{i,k}(\mathbf{x}) \Psi_{j,k}(\mathbf{x})$$

using Einstein's summation convention, then the generalized mean of the set $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ that it defines is given by

$$(\nabla \Psi)^{-1} \left(\frac{1}{M} \sum_{k=1}^M \nabla \Psi(\mathbf{x}_k) \right).$$

If the condition mentioned in Section 5.2 hold, namely:

$$\text{For each fixed } i = 1, \dots, n \quad \Psi_{i,m,k} \Psi_{j,k} = \Psi_{i,j,k} \Psi_{m,k}$$

then there exists a convex function $\Phi(\mathbf{x})$ on \mathbb{R}^n such that

$$\Phi_{i,j}(\mathbf{x}) = \Psi_{i,k}(\mathbf{x}) \Psi_{j,k}(\mathbf{x})$$

to which the comparison results established in Section 6 apply, whenever the condition mentioned in Theorem 27 holds, namely: The sign of $\Psi_{i,j,m}(\mathbf{x}) \Psi_{k,m}(\mathbf{x})$ is constant over \mathbb{R}^n independently of i, j, k .

A pending question related to the topic of this section is the following: Is there any relationship between the gradient flow determined by Ψ (that is, the solution to $\xi = \pm \nabla(\xi)$) and the geodesic flow of $\Phi_{i,j}$?

8.8 Geodesics and distances determined by the Fenchel-Legendre conjugate

Consider the simplest possible case in which $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ has a strictly positive Hessian and that $\nabla\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a diffeomorphism whose range is \mathbb{R}^n . Recall that:

Definition 3. The Fenchel-Lagrange dual Φ^* is defined by

$$\Phi^*(\xi) = \sup\{\langle \xi, \mathbf{x} \rangle - \Phi(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}.$$

The basic properties of this definition can be seen, for example, in [11]. To begin with, we have:

Lemma 2. With the assumptions made at the outset of the section, a simple computation shows that:

$$\Phi(\xi) = \langle \xi, (\nabla\Phi)^{-1}(\xi) \rangle - \Phi((\nabla\Phi)^{-1}(\xi)).$$

And more importantly

$$\nabla\Phi^*(\xi) = (\nabla\Phi)^{-1}(\xi). \tag{27}$$

And also (using Einstein's summation convention):

$$\partial_{i,k}^2\phi^*(\xi)\partial_{k,j}^2\Phi((\nabla\Phi)^{-1}(\xi)) = \delta_{i,j}. \tag{28}$$

The following result is known. See [12] for example. Its proof drops out of a computation.

Theorem 7. With the notations introduced above, let $\{\mathbf{x}(t) : 0 \leq t \leq 1\}$ be a geodesic in \mathbb{R}^n between the points $\mathbf{x}_1 = \mathbf{x}(0)$ and $\mathbf{x}_2 = \mathbf{x}(1)$ with respect to the metric given by the Hessian matrix of Φ . Let us put $\xi(t) = (\nabla\Phi)(\mathbf{x}(t))$ with $\xi_1 = (\nabla\Phi)(\mathbf{x}_1)$ and $\xi_2 = (\nabla\Phi)(\mathbf{x}_2)$. Then $\xi(t)$ is a geodesic between \mathbf{x}_1 and ξ_2 in the metric given by the Hessian of Φ^* .

Proof. To verify the assertion it suffices invoke (27) and (28) to verify that

$$\partial_{i,j}^2\phi^*(\xi)\dot{\xi}_i(t)\dot{\xi}_j(t) = \partial_{i,j}^2\phi(\mathbf{x})\dot{\mathbf{x}}_i(t)\dot{\mathbf{x}}_j(t).$$

This is left for the reader to carry out.

Let us now verify that if the Hessian of Φ can be factored as in Section 3, then the Hessian of Φ^* can be factored as well.

Theorem 8. Let Φ and Φ^* be as above. Suppose that there is a diffeomorphism U of \mathbb{R}^n such that $\partial_{ij}^2\Phi(\mathbf{x}) = \partial_i U_k(\mathbf{x})\partial_j U_k(\mathbf{x})$.

Then $U^*(\xi) = U((\nabla\Phi)^{-1}(\xi))$ is a factorization of the Hessian of Φ^* .

Proof. Observe that invoking (28) we obtain:

$$\partial_j U_k^*(\xi) = \partial_l U_k((\nabla\Phi)^{-1}(\xi))\partial_n^2 i \Phi^*(\xi).$$

From this and (28) it follows that

$$\partial_i U_k^*(\xi)\partial_j U_k^*(\xi) = \partial_{i,j}\Phi^*(\xi).$$

Thus concludes the proof.

To finish, let us verify that the distances along the geodesics in the Hessians of Φ^* and Φ coincide.

Corollary 1. With the notations introduced above, let $\xi(t)$ and $\mathbf{x}(t)$ be geodesics described in Theorem 27. Then

$$\delta_{\Phi^*}(\xi_1, \xi_2) = \delta_{\Phi}(\mathbf{x}_1, \mathbf{x}_2).$$

Proof. We saw in Section 3 that $\delta_{\Phi^*}^2(\xi_1, \xi_2) = \|U^*(\xi_1) - U^*(\xi_2)\|^2$. From the definition of U^* and since $\xi(t) = \nabla\Phi(\mathbf{x}(t))$ we obtain

$$\begin{aligned} \delta_{\Phi^*}^2(\xi_1, \xi_2) &= \|U^*(\xi_1) - U^*(\xi_2)\|^2 \\ &= \|U(\mathbf{x}_1) - U(\mathbf{x}_2)\|^2 = \delta_{\Phi}^2(\mathbf{x}_1, \mathbf{x}_2). \end{aligned}$$

And thus the assertion is verified.

9 Final remarks

We should mention that Hessian structures on Euclidean spaces is a well studied subject. See for example [9] and [13]. In that line of work, the focus is on the geometry on \mathbb{R}^d resulting from a Hessian metric, that is, a metric of the type $g = Hess\Phi$, or more explicitly, $g_{i,j}(\mathbf{z}) = \partial^2\Phi(\mathbf{x})/\partial x_i\partial x_j$, when $\Phi(\mathbf{x})$ is a twice continuously differentiable convex function defined on an open convex subset of \mathbb{R}^n .

The relationship of that line of work to the work presented here, is that here we are interested in the geometry on \mathbb{R}^n for a metric given by $g = J_U^t J_U$ where J_U is the Jacobian of a diffeomorphism U . This is related to that of a Hessian metric when $Hess\Phi$ can be factored as $Hess\Phi = J_U^t J_U$, which is the subject of Section 5. But as the example in section 8.4 suggests, an open question remains. Given that a Riemannian metric g is given by a positive matrix, does it admit a square root Σ such that Σ is the Jacobian of a diffeomorphism?

Acknowledgements:

I would like to thank Pál Burai for bringing reference [3] up to my attention.

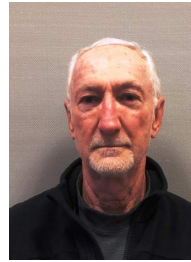
Conflicts of Interests

The authors declare that they have no conflicts of interests.

References

[1] C.E. Bonferroni, La media esponenziale in matematica finanziaria, *Annuario del Regio Istituto di Scienze Economiche e Commerciali di Bari*, AA 23-24, 1-14, (1926).
 [2] B. de Finetti, Sul concetto di media, *Giornale dell'Istituto Italiano degli Attuari*, 2, 369-396, (1931).

- [3] R.L.Berger and G. Casella, Deriving Generalized Means as Least Squares and Maximum Likelihood Estimates, *The American Statistician*, **46**, 279-282, (1992).
- [4] J.M. Oller and J.M. Corcuera, Intrinsic analysis of statistical estimation, *The Annals of Statistics*, **23**, 1562-1581, (1995).
- [5] Hicks, N. *Notes on Differential Geometry*, Van Nostrand, Princeton, (1965).
- [6] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer, New York, (1978).
- [7] H. Gzyl, Prediction in Riemannian metrics derived from divergence functions, *Comm. in Stats.*, <https://doi.org/10.1080/03610926.2020.1752384>, (2020).
- [8] J.D.Miller, Some Global Inverse Function Theorems, *Journal of Mathematical Analysis and its Applications*, **100**, 375-384, (1984).
- [9] H. Kito, On Hessian structures on the Euclidean space and the hyperbolic spaces, *Osaka Journal of Mathematics*, **36**, 51-62, (1994).
- [10] L. Bregman, The relaxation method of finding common points of convex sets and its application to the solution of problems in convex programming, *Comp. Math. Phys., USSR*, **7**, 2100-217, (1967).
- [11] J. M. Borwein and A.S. Lewis, *Convex Analysis and Non linear Optimization*, CMS Books in Mathematics, Springer, New York, (2000).
- [12] F. Nielsen, An elementary introduction to information geometry, <https://arxiv.org/abs/1808.08271>, (2018).
- [13] H. Shima, *The geometry of Hessian structures*, World Scientific Publishers, Singapore (2007).



Henryk Gzyl earned his PhD in Mathematics at the University of California at San Diego. For more than 35 years he taught many subjects at the graduate and undergraduate level in the Physics and Mathematics Departments, mainly at the Sciences Faculty of the

Central University of Venezuela. Currently he teaches subjects related to risk management and portfolio optimization at IESA (Institute for Higher Education in Administration) in Caracas. His interests are mainly in inverse problems, mathematical modeling, stochastic processes.