Convexity properties of the condition number II [†]

Carlos Beltrán [‡] Jean-Pierre Dedieu [§]

Gregorio Malajovich [¶] Mike Shub [‖]

October 30, 2009

This paper is dedicated to Steve Smale, on his 80th birthday.

Abstract

The condition metric, studied in our previous paper [1], induces a Lipchitz-Riemann structure on the space of all $n \times m$ full rank matrices. After investigating geodesics in such a nonsmooth structure, we show that the inverse of the smallest singular value of a matrix is a log-convex function (Theorem 1).

[†]Mathematics Subject Classification (MSC2010): 53C23 (Primary), 65F35, 15A12 (Secondary).

[‡]C. Beltrán, Departamento de Matemáticas, Estad. y Comput. Universidad de Cantabria, Santander, España (carlos.beltran@unican.es). CB was supported by MTM2007-62799.

[§]J.-P. Dedieu, Institut de Mathématiques, Université Paul Sabatier, 31062 Toulouse cedex 09, France (jean-pierre.dedieu@math.univ-toulouse.fr). J.-P. Dedieu was supported by the ANR Gecko and by the Fields Institute at Toronto.

[¶]G. Malajovich, Departamento de Matemática Aplicada, Universidade Federal de Rio de Janeiro, Caixa Postal 68530, CEP 21945-970, Rio de Janeiro, RJ, Brazil (gregorio@ufrj.br). He was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil), by FAPERJ (Fundação Carlos Chagas de Amparo à Pesquisa do Estado do Rio de Janeiro) and by the Brazil-France agreement of cooperation in Mathematics.

[‖]M. Shub, Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4 (shub.michael@gmail.com). M. Shub was upported by an NSERC Discovery Grant.

We also show that a similar result holds for the solution variety of linear systems (Theorem 23).

Some of our intermediate results, such as Theorem 7, on the Hessian of a function with symmetries on a manifold, and Theorem 21 on piecewise self-convex functions, are of independent interest.

1 Introduction

Let two integers $1 \le n \le m$ be given and let us consider the space of matrices $\mathbb{K}^{n\times m}$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, equipped with the Frobenius inner product

$$\langle M, N\rangle_F = \text{trace } (N^* M) = \sum_{i,j} m_{ij}\overline{n_{ij}}.$$

We denote by

$$\sigma_1(A) \ge \ldots \ge \sigma_{n-1}(A) \ge \sigma_n(A) \ge 0$$

the singular values of a matrix $A \in \mathbb{K}^{n\times m}$, by $\mathbb{GL}_{n,m}$ the space of matrices $A \in \mathbb{K}^{n\times m}$ with maximal rank, that is rank $A = n$ or, equivalently, $\sigma_n(A) > 0$, and by $\mathcal{N}$ the set of singular (or rank deficient) matrices:

$$\mathcal{N} = \mathbb{K}^{n\times m} \setminus \mathbb{GL}_{n,m} = \left\{A \in \mathbb{K}^{n\times m} \ : \ \sigma_n(A) = 0\right\}.$$

The distance of a matrix $A \in \mathbb{K}^{n\times m}$ from $\mathcal{N}$ is given by its smallest singular value:

$$d_F(A, \mathcal{N}) = \min_{S\in\mathcal{N}} \|A - S\|_F = \sigma_n(A).$$

Consider now the problem of connecting two matrices with the shortest possible path staying, as much as possible, away from the set of singular matrices. We realize this objective by considering an absolutely continuous path $A(t)$, $a \le t \le b$, with given endpoints (say $A(a) = A$ and $A(b) = B$) which minimizes its **condition length** defined by

$$L_\kappa = \int_a^b \left\| \frac{dA(t)}{dt} \right\|_F \sigma_n(A(t))^{-1} dt.$$

We call **minimizing condition path** an absolutely continuous path which minimizes this integral in the set of absolutely continuous paths with the same end-points. We define a **minimizing condition geodesic** as a minimizing condition path parametrized by the condition arc length, that is when

$$\left\| \frac{dA(t)}{dt} \right\|_F \sigma_n(A(t))^{-1} = 1 \text{ a. e.}$$

A **condition geodesic** is an absolutely continuous path which is locally a minimizing condition geodesic. This concept of geodesic is related to the Riemannian structure defined on $\mathbb{GL}_{n,m}$ by:

$$\langle M, N \rangle_{\kappa,A} = \sigma_n(A)^{-2} \mathrm{Re} \langle M, N \rangle_F .$$

We call it the **condition Riemann structure** on $\mathbb{GL}_{n,m}$.

Our objective is to investigate the properties of the smallest singular value $\sigma_n(A(t))$ along a condition geodesic. Our main result says:

Theorem 1. *For any condition geodesic $t \to A(t)$ in $\mathbb{GL}_{n,m}$, the map $t \to \log(\sigma_n^{-1}(A(t)))$ is convex.*

This theorem extends our main result in [1]. In that paper, the same theorem is proven for those condition geodesic arcs contained in the open subset

$$\mathbb{GL}_{n,m}^{>} = \{A \in \mathbb{GL}_{n,m} \ : \ \sigma_{n-1}(A) > \sigma_n(A)\}$$

that is when the smallest singular value $\sigma_n(A)$ is simple. The reason for this restriction is easy to explain. The smallest singular value $\sigma_n(A)$ is smooth in $\mathbb{GL}_{n,m}^{>}$, and, in that case, we can use the toolbox of Riemannian geometry. But it is only locally Lipschitz in $\mathbb{GL}_{n,m}$; for this reason we call the condition structure in $\mathbb{GL}_{n,m}$ a **Lipschitz-Riemannian structure**.

This concept is introduced in general in the next section. Using nonsmooth analysis techniques, we prove that any condition geodesic is C^1 with a locally Lipschitz derivative (Theorem 2). Such techniques are already present in Boito-Dedieu [3].

Our improvement depends also on a more systematic use of the symmetries. Theorem 7 gives a simplified computation of the Hessian when there is a Lie group of symmetries. This theorem may be of independent interest. It is so natural we would not be surprised if it is already known, but we have not found it anywhere. We were led to this theorem sometime after a conversation with John Lott on Hessians and Riemannian submersions while he was visiting the University of Toronto.

Another component to our improvement, which also may be of independent interest, is Theorem 21: piecing together convexity results on restrictions of the Lipschitz-Riemann structure to a union of submanifolds of varying dimensions, where the structure is smooth, to obtain a global result. Our impression is that a finer understanding of the structure of the space of geodesics on $\mathbb{GL}_{n,m}$ would give a more natural proof of our main theorem. But this finer understanding is not available to us now.

Let us now say a word about our motivations. The (today) classical papers [18], [19], and [20] by Shub and Smale relate complexity bounds for homotopy methods to solve Bézout's Theorem to the condition number of the encountered problems along the considered homotopy path. Ill-conditioned problems slow the algorithm and increase its complexity. For this reason it is natural to consider paths which avoid ill-posed problems, and, at the same time, are as short as possible. The condition metric has been designed to construct such paths. It has been introduced by Shub in [17], then studied by Beltrán and Shub in [2] in spaces of polynomial equations. The case of linear maps (and related spaces) appears in Beltrán-Dedieu-Malajovich-Shub [1] and Boito-Dedieu [3].

In the linear case, it is rather a remarkable fact the inverse of the distance to singular matrices $\sigma_n^{-1}(A(t))$ is log-convex along the condition geodesics. So, in particular, the maximum of $\log\left(\sigma_n(A(t))^{-1}\right)$ along such paths is necessarily obtained at its endpoints and the condition geodesics stay away from singular matrices. We don't know if such a property is still valid in the polynomial case discussed above.

In our last Theorem (Theorem 23) we prove a version of our main result in the context of the solution variety

$$\mathcal{W} = \{(A, x) \in \mathbb{GL}_{n,n+1} \times \mathbb{P}(\mathbb{K}^n) : Ax = 0\} \ .$$

Acknowledgements. We have benefited greatly from conversations with our colleagues Charles Pugh and Vitaly Kapovitch about Lipschitz-Riemann structures, especially those conformally equivalent to smooth structures by locally Lipschitz scaling maps.

Some of this work was accomplished when we met not only in our intitutions, but also at the Institut de Matemàtica de la Universitat de Barcelona, and at the Thematic Program in the Foundations of Computational Mathematics (FoCM) at the Fields Institute. We thank these institutions.

2 Geodesics in Lipschitz-Riemann structures, and self-convexity

2.1 Lipschitz-Riemann structures

We obtain a **Lipschitz-Riemann structure** on a C^2 manifold $\mathcal{M}$ when, for each $x \in \mathcal{M}$, a scalar product $\langle \cdot, \cdot \rangle_x$ on $T_x\mathcal{M}$ is given, and is a locally Lipschitz function of x.

The **length** of an absolutely continuous path $x(t) \in \mathcal{M}$, $a \leq t \leq b$, is defined as the integral

$$L(x, a, b) = \int_a^b \|\dot{x}(t)\|_{x(t)} dt,$$

where $\dot{x}(t)$ denotes the derivative with respect to t. Its **arc length** is given by the map

$$t \in [a, b] \to L(x, a, t) \in [0, L(x, a, b)].$$

The **distance** $d(a, b)$ between two points $a, b \in \mathcal{M}$ is the infimum of all the lengths of the paths containing a and b in their image. We call **minimizing path** an absolutely continuous path such that $L(x, a, b) = d(a, b)$. We define a **minimizing geodesic** as a minimizing path parametrized by arc length, that is when

$$\|\dot{x}(t)\|_{x(t)} = 1 \text{ a. e.}$$

A path in $\mathcal{M}$ parametrized by arc length is a **geodesic** when it is locally a minimizing geodesic.

The main result of this section is the following:

Theorem 2. *Any geodesic belongs to the class C^{1+Lip} that is C^1 with a locally Lipschitz derivative.*

This theorem is proved in sections 2.5 and 2.6, it extends a similar result by Charles Pugh [15] who proves the existence of locally minimizing C^{1+Lip} geodesics. His argument is based on a smooth approximation of the Lipschitz structure where the classical toolbox of Riemannian geometry applies, followed by a *passage à la limite*.

Using different techniques we prove here this regularity assumption for all geodesics.

According to Rademacher's Theorem, $C^{1+Lip} = W^{2,\infty}$ the Sobolev space of maps f with $f'' \in L^\infty$.

2.2 Existence of geodesics in a Lipschitz-Riemann structure

Existence of minimizing geodesics with given endpoints may be deduced from the Hopf-Rinow Theorem (see Gromov [11] Theorem 1.10). A metric space (X, d) is a **path metric space** if the distance between each pair of points equals the infimum of the lengths of curves joining the points.

Theorem 3. *If* (X, d) *is a complete, locally compact path metric space, then*

- *Each bounded, closed subset is compact,*
- *Each pair of points can be joined by a minimizing geodesic.*

Two examples of such spaces are given by Boito-Dedieu [3] for linear maps (X is one of the connected components of $\mathbb{GL}_{n,m}$ equipped with the condition structure), and by Shub [17] when X is the solution variety associated with the homogeneous polynomial system solving problem equipped with the corresponding condition structure.

2.3 Lipschitz-Riemann structures in $\mathbb{R}^k$

An important example of Lipschitz-Riemann structure is given by an open set $\Omega \subset \mathbb{R}^k$ equipped with the scalar product

$$\langle u, v \rangle_x = v^T H(x) u$$

where H is a locally Lipschitz map from Ω into the set of positive definite $n \times n$ matrices. The length and distance for this structure are denoted by L_H and d_H while L and d correspond to the Euclidean structure.

A minimizing geodesic $x(t) \in \Omega$, $a \leq t \leq b$, minimizes the integral

$$\int_a^b \sqrt{\dot{y}(t)^T H(y(t)) \dot{y}(t)} dt$$

in the set of absolutely continuous paths $y(t)$ with endpoints $y(a) = x(a)$, and $y(b) = x(b)$. This is an instance of the Bolza problem.

For a smooth integrand L, a local solution $x(t)$ of the Bolza problem

$$\inf \int_a^b L(y(t), \dot{y}(t)) dt,$$

where the infimum is taken in the set of a.c. paths with given endpoints, satisfies the Euler-Lagrange differential equation

$$-\frac{d}{dt}\frac{\partial L}{\partial \dot{x}}(x(t),\dot{x}(t)) + \frac{\partial L}{\partial x}(x(t),\dot{x}(t)) = 0 \quad a.e.$$

In our context, $L(x,\dot{x}) = \sqrt{\dot{x}^T H(x)\dot{x}}$ is smooth in the variable $\dot{x}$ (if we avoid $\dot{x} = 0$, which will be the case), and locally Lipschitz in the variable x. For this reason we shall use a generalized version of the Euler-Lagrange equation based on generalized gradients.

2.4 Generalized gradients and the problem of Bolza

Let $f : \Omega \subset \mathbb{R}^k \to \mathbb{R}$ be a locally Lipschitz function defined on an open set. Its **one-sided directional derivative** at $x \in \Omega$ in the direction $d \in \mathbb{R}^k$ is defined as

$$f'(x,d) = \lim_{t \to 0_+} \frac{f(x+td) - f(x)}{t}.$$

The **generalized directional derivative in Clarke's sense** of f at $x \in \Omega$ in the direction d is defined as

$$f^o(x,d) = \limsup_{\substack{y \to x \\ t \to 0_+}} \frac{f(y+td) - f(y)}{t}$$

and the **generalized gradient** of f at x is the nonempty compact subset of $\mathbb{R}^k$ given by

$$\partial f(x) = \left\{ s \in \mathbb{R}^k \ : \ \langle s, d \rangle \leq f^o(x,d) \text{ for all } d \in \mathbb{R}^k \right\}.$$

When $f \in C^1(\Omega)$ the generalized gradient is just the usual one: $\partial f(x) = \left\{ \nabla f(x) \right\}$. The generalized directional derivative is related to the gradient via the equality

$$f^o(x,d) = \max_{s \in \partial f(x)} \langle s, d \rangle .$$

We say that f is **regular at** x when the two directional derivatives exist and are equal:

$$f^o(x,d) = f'(x,d) \text{ for any } d \in \mathbb{R}^k.$$

When f is defined on a C^1 manifold $\mathcal{M}$, we say that f is **regular at** $m \in \mathcal{M}$ when its composition with a local chart at m gives a regular map in the usual meaning.

Good references for this topic is Clarke [6] or Schirotzek [16].

For the problem of Bolza described above the counterpart of the Euler-Lagrange equation is given by the following result (see [6] Theorem 4.3.3, and [5]).

Theorem 4. *Let x solve the Bolza problem in the case in which $L(x, \dot{x})$ is a locally Lipschitz map and suppose that $\dot{x}$ is essentially bounded. Then there is an absolutely continuous map p such that*

$$\dot{p}(t) \in \partial_x L(x(t), \dot{x}(t)) \text{ and } p(t) \in \partial_{\dot{x}} L(X(t), \dot{x}(t)) \quad a.e.$$

2.5 Proof of Theorem 2 in $\mathbb{R}^k$

By definition, a geodesic is a locally minimizing geodesic. Thus, it suffices to establish the theorem in this case.

A minimizing geodesic $x(t) \in \Omega$, $a \leq t \leq b$, is parametrized by arc length so that

$$\dot{x}(t)^T H(x(t))\dot{x}(t) = 1 \quad a.e.,$$

Thus, $\dot{x}(t)$ is $\neq 0$ and essentially bounded:

$$\dot{x} \in L^{\infty}\left([a, b], \mathbb{R}^k\right).$$

Moreover, x minimizes the integral

$$\int_a^b \sqrt{\dot{y}(t)^T H(y(t))\dot{y}(t)} dt$$

in the set of absolutely continuous paths with endpoints $y(a) = x(a)$, and $y(b) = x(b)$. Thus, according to Theorem 4, there is an absolutely continuous arc p such that

$$\dot{p}(t) \in \partial_x \sqrt{\dot{x}(t)^T H(x(t))\dot{x}(t)},$$
$$p(t) \in \partial_{\dot{x}} \sqrt{\dot{x}(t)^T H(x(t))\dot{x}(t)}$$

for almost all $t \in [a, b]$. Since our integrand is smooth in the $\dot{x}$ variable we may write this second equation

$$p(t) = \frac{H(x(t))\dot{x}(t)}{\sqrt{\dot{x}(t)^T H(x(t))\dot{x}(t)}} = H(x(t))\dot{x}(t).$$

Thus, $\dot{x}(t) = H(x(t))^{-1}p(t)$ is absolutely continuous and $x(t)$ possesses a.e. a second derivative $\ddot{x}(t) \in L^1([a,b], \mathbb{R}^k)$.

We now have to show that this second derivative is essentially bounded. This comes from the first equation. Since $\sqrt{\cdot}$ is a smooth function we get

$$\partial_x \sqrt{\dot{x}^T H(x)\dot{x}} \subset \frac{\dot{x}^T \partial H(x)\dot{x}}{2\sqrt{\dot{x}^T H(x)\dot{x}}} = \frac{1}{2}\dot{x}^T \partial H(x)\dot{x},$$

with

$$\dot{x}^T \partial H(x)\dot{x} = \sum_{i,j} \dot{x}_i \dot{x}_j \partial h_{ij}(x).$$

The first equation implies

$$\dot{p}(t) \in \frac{1}{2}\dot{x}(t)^T \partial H(x(t))\dot{x}(t).$$

From the hypothesis, the functions $h_{ij}(x)$ are locally Lipschitz. Their generalized gradients are compact convex sets in $\mathbb{R}^k$. The union of all these sets along the path $x(t)$ gives us a bounded set. Since the curve $\dot{x}(t)$ is continuous, we deduce from these considerations, and from the second equation that $\dot{p}(t)$ is bounded. Thus $p(t)$ is Lipschitz, and $\dot{x}(t) = H(x(t))^{-1}p(t)$ is also Lipschitz. The second derivative $\ddot{x}(t)$ is thus bounded by the Lipschitz constant of $\dot{x}(t)$, and we are done.

Remark 5. *The previous lines give the following properties for a geodesic x in Ω: $x \in C^{1+Lip}$, $\dot{x}^T H(x)\dot{x} = 1$, and*

$$\frac{d}{dt}(H(x)\dot{x}) \in \frac{1}{2}\dot{x}^T \partial H(x)\dot{x} = \frac{1}{2}\sum_{i,j} \dot{x}_i \dot{x}_j \partial h_{ij}(x).$$

The initial value problem, and even the boundary value problem associated with this second order differential inclusion, may have many solutions. Examples are given in [3]. Moreover, solutions are not necessarily locally minimizing geodesics and geodesics are not necessarily unique. We will give examples in a future paper.

2.6 Proof of Theorem 2 in $\mathcal{M}$

Since this theorem is of local nature, it suffices to prove it locally. Take a local chart, transfer the Lipschitz-Riemann structure of $\mathcal{M}$ to an open set $\Omega \subset \mathbb{R}^k$ where the theorem is already proved and we are done.

2.7 Conformal Lipschitz-Riemann structure

The example of a L-R structure which motivates this paper is given by the condition structure on $\mathbb{GL}_{n,m}$. It is obtained in multiplying the Frobenius scalar product by the locally Lipschitz function σ_n^{-2}. Let us put it in a more general setting.

Let $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ be a C^2 Riemannian manifold of real dimension k and let $\alpha : \mathcal{M} \to \mathbb{R}$ be a locally Lipschitz function with positive values. Let $\mathcal{M}_\kappa$ be the manifold $\mathcal{M}$ with the new metric

$$\langle\cdot,\cdot\rangle_{\kappa,x} = \alpha(x)\langle\cdot,\cdot\rangle_x, \ \|\cdot\|_{\kappa,x} = \langle\cdot,\cdot\rangle_{\kappa,x}^{1/2},$$

called **condition Lipschitz-Riemann structure** or simply **condition structure**. We denote by L (respectively L_κ) the length of a curve γ in the $\mathcal{M}$-structure (respectively in the $\mathcal{M}_\kappa$-structure). We will speak of **length** or **condition length**, and also of **distance** or **condition distance**, **geodesics** or **condition geodesics** and so on.

We say that α is **self-convex** when $\log\circ\alpha\circ\gamma$ is a convex function for any condition geodesic γ. Examples of self-convex maps are given in [1] where this concept is introduced for the first time.

Using this definition Theorem 1 above reads

$$\alpha(A) = \sigma_n(A)^{-2} \text{ is self-convex in } \mathbb{GL}_{n,m}.$$

3 Self-convexity in the smooth case and the computation of Hessians

3.1 Self-convexity in the smooth case

When $\alpha : \mathcal{M} \to \mathbb{R}$ is C^2, self-convexity of α is equivalent to the second covariant derivative of $\log(\alpha)$ being positive semi-definite in the α-condition Riemann structure (see [22] Chap. 3, Theorem 6.2). Note that the second covariant derivative of a map $\mathcal{M} \to \mathbb{R}$ is different in $\mathcal{M}$ and in $\mathcal{M}_\kappa$. We denote them respectively by D^2, and by D^2_κ, or by Hess and Hess_κ (Hess like Hessian). Thus, self-convexity of α is equivalent to $D^2_\kappa(\log(\alpha))$ being positive semi-definite. In Proposition 2 of [1] we proved that this is equivalent to

$$2\alpha(x)D^2\alpha(x)(\dot{x},\dot{x}) + \|D\alpha(x)\|_x^2\|\dot{x}\|_x^2 - 4(D\alpha(x)\dot{x})^2 \geq 0 \qquad (3.1)$$

for any $x \in \mathcal{M}$ and for any vector $\dot{x} \in T_x\mathcal{M}$, the tangent space at x.

3.2 Self-convexity in a product space

This proposition has an immediate corollary which can be useful. Suppose $\mathcal{N}$ is another smooth Riemannian manifold. Give $\mathcal{M}\times\mathcal{N}$ the product structure. Let $\pi : \mathcal{M}\times\mathcal{N} \to \mathcal{M}$ be the projection on the first factor and $\hat{\alpha} : \mathcal{M}\times\mathcal{N} \to \mathbb{R}$ be the composition $\hat{\alpha} = \alpha \circ \pi$.

Proposition 6. *If α is C^2 and self-convex in $\mathcal{M}$ then $\hat{\alpha}$ is self-convex in $\mathcal{M} \times \mathcal{N}$.*

Proof. Let $(x, y) \in \mathcal{M} \times \mathcal{N}$ and assume normal (geodesic) coordinates in a neighborhood of $x \in \mathcal{M}$ (with respect to the α-structure). Also, assume normal coordinates around $y \in \mathcal{N}$ with respect to the inner product $\alpha(x)\langle \cdot, \cdot \rangle_{\mathcal{N}}$.

We claim that this defines a system of normal coordinates in $\mathcal{M} \times \mathcal{N}$. By construction, $g_{ij}(x, y) = \delta_{ij}$. Also, it is easy to see that for all indexes i, j, k,

$$\Gamma^k_{ij}(x, y) = 0 \ .$$

Indeed, if indices (i, j, k) correspond to the same component $\mathcal{M}$ or $\mathcal{N}$ this follows from the choice of normal coordinates in each component. Otherwise, say that i, j correspond to coordinates in $\mathcal{M}$ and k to coordinates in $\mathcal{N}$. Then $g_{ik} \equiv g_{jk} \equiv 0$ and furthermore,

$$\frac{\partial}{\partial u^k} g_{ij}(x, y) = 0.$$

Thus $\Gamma_{ikj}(x, y) = 0$ for all indexes i, j, k. This implies that $\Gamma^k_{ij}(x, y) = 0$ as well. Thus we have a normal system of coordinates around $(x, y) \in \mathcal{M} \times \mathcal{N}$.

In that system of coordinates,

$$\text{Hess } \hat{\alpha}(\text{x}, \text{y}) = \begin{bmatrix} \text{Hess } \alpha(\text{x}) & 0 \\ 0 & 0 \end{bmatrix}$$

It follows that

$$2\hat{\alpha}(x, y) D^2\hat{\alpha}(x, y) \left(\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}, \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \right) + \|D\hat{\alpha}(x, y)\|^2_{(x,y)} (\|\dot{x}\|^2_x + \|\dot{y}\|^2_y)$$

$$-4 \left(D\hat{\alpha}(x, y) \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \right)^2 \geq$$

$$\geq 2\alpha(x) D^2\alpha(x)(\dot{x}, \dot{x}) + \|D\alpha(x)\|^2_x \|\dot{x}\|^2_x - 4(D\alpha(x)\dot{x})^2 \geq 0$$

(the last inequality by Equation (3.1)). Applying again Proposition 2 of [1], we conclude that $\hat{\alpha}$ is self-convex. □

We have raised the question in the introduction of whether self-convexity of the condition number holds for the condition Riemann structure on the solution variety considered in [17]. The theorems proven in this paper apply to the case of linear system, but with the use of Proposition 6 they give us some information on polynomial systems almost for free. The proof follows from Propositions 13, 6 and Theorem 21.

Let $\mathbf{d} = (d_1, \ldots, d_n)$. Consider the vector space

$$\mathcal{P}_{\mathbf{d},0} = \{(f_1, \ldots, f_n) : f_i \in \mathbb{C}[x_1, \ldots, x_n] \text{ with } \deg f_i = d_i \text{ and } f_i(0) = 0\}.$$

This vector space splits as $\mathcal{P}_{\mathbf{d},0} = \mathcal{L}_0 \oplus (\text{H.O.T.})_0$ where $\mathcal{L}_0$ are linear and $(\text{H.O.T.})_0$ are higher order polynomials vanishing at 0.

Put any inner product in $\mathcal{P}_{\mathbf{d},0}$ for which $\mathcal{L}_0$ and $\mathcal{P}_{\mathbf{d},0}$ are orthogonal, with $\langle A, B \rangle = \text{trace } (B^* A)$ for $A, B \in \mathcal{L}_0$.

Set $\mathbf{f} = (f_1, \ldots, f_n) \in \mathcal{P}_{\mathbf{d},0}$ and define $\alpha(\mathbf{f}) = \|D\mathbf{f}(0)^{-1}\|_2^2$. Then α is self-convex in $\mathcal{P}_{\mathbf{d},0}$.

3.3 Computation of the Hessian

When analyzing the convexity properties of $\sigma_n(A)$, we first note that this function has a big symmetry group, as

$$\sigma_n(A) = \sigma_n(UAV^*)$$

for unitary matrices $U \in \mathbb{U}_n$, and $V \in \mathbb{U}_m$ (resp. orthogonal matrices $U \in \mathbb{O}_n$, and $V \in \mathbb{O}_m$). Let us consider this situation in a general framework.

Let G be a Lie group of isometries acting on $\mathcal{M}$ and leaving α invariant, i.e.

$$\alpha \circ g = \alpha \text{ for all } g \in G.$$

Then we say that G is a **group of symmetries of** α.

Theorem 7. *With the notations above, let $w = b + k \in T_p\mathcal{M}$ where $k \in T_pG(p)$, $b \perp T_pG(p)$, and where $G(p)$ is the orbit of p under the action of G. Let the vector field K be the infinitesimal generator associated with some element a in the Lie Algebra $\mathfrak{g}$ of G, where $k = \frac{d}{dt}(exp(ta)p)\mid_{t=0}$. Namely,*

$$K(q) = \frac{d}{dt}(exp(ta)q)\mid_{t=0}, \quad q \in \mathcal{M}.$$

Let $\phi_t(q) = \phi(t, q)$ be the flow of grad α*, defined for $t \in (-\varepsilon, \varepsilon)$ and q close enough to p. Let B be a smooth vector field in $\mathcal{M}$ such that $B(\phi_t(p)) = D\phi_t(p)b$. Then, the following equality holds:*

$$\text{Hess } \alpha(p)(w, w) =$$

$$\text{Hess } \alpha(p)(b, b) + \frac{1}{2}\langle \text{grad } (\|K\|^2)(p), \text{grad } \alpha(p)\rangle + \text{grad } \alpha(\langle B, K\rangle).$$

Let us recall the intrinsic definition of Hessian,

$$\text{Hess } \alpha(p)(v, w) = X(Y(\alpha)) - (\nabla_X Y)(\alpha),$$

where X, Y are vector fields, $X(p) = v$, $Y(p) = w$, and ∇ is the Levi-Civita connection. The proof of Theorem 7 is a consequence of the two following lemmas:

Lemma 8. *For any vector field X on M, we have*

$$2\text{Hess } \alpha(X, K) = \text{grad } \alpha(\langle X, K\rangle) - \langle[\text{grad } \alpha, X], K\rangle.$$

Moreover,

$$\text{Hess } \alpha(k, k) = \frac{1}{2}\langle \text{grad } (\|K\|^2), \text{grad } \alpha\rangle(p), \tag{3.2}$$

Proof. We recall that for vector fields X, Y, Z,

$$2\langle \nabla_X Y, Z\rangle = X(\langle Y, Z\rangle) + Y(\langle X, Z\rangle) - Z(\langle X, Y\rangle) + \tag{3.3}$$
$$\langle [X, Y], Z\rangle + \langle [Z, X], Y\rangle - \langle [Y, Z], X\rangle.$$

Note that $K(p) = k$ and $K(q) \in T_q G(q)$ for $q \in \mathcal{M}$. As α is G-invariant,

$$K(\alpha) = \langle K, \text{grad } \alpha\rangle = 0. \tag{3.4}$$

Moreover, the one-parameter group generated by K consists of global isometries, thus K is a Killing vector field, which implies that for any pair of vector fields X, Y,

$$\langle \nabla_Y K, X\rangle + \langle \nabla_X K, Y\rangle = 0, \text{ or equivalently using (3.3)}$$

$$K(\langle Y, X\rangle) + \langle [Y, K], X\rangle + \langle [X, K], Y\rangle = 0. \tag{3.5}$$

We can now compute

$$2\text{Hess } \alpha(X,K) = 2X(K(\alpha)) - 2(\nabla_X K)(\alpha) = -2\langle \nabla_X K, \text{grad } \alpha\rangle =$$

$$-X(\langle K, \text{grad } \alpha\rangle) - K(\langle X, \text{grad } \alpha\rangle) + \text{grad } \alpha(\langle X, K\rangle)$$
$$-\langle [X,K], \text{grad } \alpha\rangle - \langle [\text{grad } \alpha, X], K\rangle + \langle [K, \text{grad } \alpha], X\rangle.$$

From (3.5) we know that

$$-K(\langle X, \text{grad } \alpha\rangle) - \langle [X,K], \text{grad } \alpha\rangle + \langle [K, \text{grad } \alpha], X\rangle = 0.$$

Using $\langle \text{grad } \alpha, K\rangle = 0$, we conclude

$$2\text{Hess } \alpha(X,K) = \text{grad } \alpha(\langle X, K\rangle) - \langle [\text{grad } \alpha, X], K\rangle,$$

which proves the first assertion.

When $X = K$, the second term above vanishes:

$$\begin{aligned} \langle [K, \text{grad } \alpha], K\rangle &= \langle \nabla_K \text{grad } \alpha, K\rangle - \langle \nabla_{\text{grad } \alpha} K, K\rangle \\ &= \langle \nabla_K \text{grad } \alpha, K\rangle + \langle \text{grad } \alpha, \nabla_K K\rangle \\ &= K(\langle K, \text{grad } \alpha\rangle) \\ &= 0. \end{aligned}$$

Equation (3.2) follows.

□

Lemma 9.

$$2\text{Hess } \alpha(p)(k,b) = \text{grad } \alpha(\langle B, K\rangle).$$

Proof. The case $k = 0$ is easy. The case grad $\alpha(p) = 0$ is also easy, because the orbit $G(p)$ is a critical manifold of α and therefore Hess $\alpha(p)(k,\cdot) \equiv 0$. Thus, we can assume that $k, \text{grad } \alpha(p) \neq 0$. Let N_0 be a codimension 2 submanifold of $\mathcal{M}$ with p in its interior. Assume that $b \in T_p\mathrm{N}_0$, k is orthogonal to $T_p\mathrm{N}_0$, and grad $\alpha(p) \notin T_p N_0$.

Let $\mathrm{N} = \cup\phi_t(\mathrm{N}_0)$ with ϕ_t the flow associated with grad α and where the union is taken in a small interval around $t = 0$. N is a codimension 1 submanifold. For small ε, the integral curve of grad α is thus contained in N, and for $q = \phi_t(p)$, we have $B(q) = D\phi_t(p)b \in T_q\mathrm{N}$. Both grad α and B are tangent to N by construction. By Frobenius Theorem, $[B, \text{grad } \alpha]$ is again tangent to N. In particular, $[\text{grad } \alpha, B](p) \in T_p\text{N}$, and hence $\langle [\text{grad } \alpha, B], K\rangle(p) = 0$. From Lemma 8,

$$2\text{Hess } \alpha(B, K) = \text{grad } \alpha(\langle B, K\rangle) - \langle [\text{grad } \alpha, B], K\rangle = \text{grad } \alpha(\langle B, K\rangle)$$

at p as wanted. □

Proof of Theorem 7. The Hessian is a symmetric bilinear form. Thus,

$$\text{Hess } \alpha(p)(v, v) = \text{Hess } \alpha(p)(b, b) + \text{Hess } \alpha(p)(k, k) + 2\text{Hess } \alpha(p)(b, k).$$

Theorem 7 follows from lemmas 8 and 9. □

Corollary 10. *Assume that for every $p \in \mathcal{M}$:*

- $\text{Hess}_\kappa \log(\alpha)$ *is positive semi-definite in* $(T_pG(p))^\perp$,
- *For $b \in T_p\mathcal{M}$, $b \perp T_pG(p)$, we have that $D\phi_t(p)b \perp T_{\phi_t(p)}G(\phi_t(p))$. Here, $\phi_t(q) = \phi(t, q)$ is the flow of* grad α, *defined for $t \in (-\varepsilon, \varepsilon)$ and q close enough to p.*
- *For every $a \in \mathfrak{g}$, the associated vector field $K(q) = \frac{d}{dt}(exp(ta)q)\mid_{t=0}$, $q \in M$, satisfies*

$$\alpha D(\|K\|^2)(\text{grad } \alpha) + \|K\|^2\|\text{grad } \alpha\|^2 \geq 0.$$

Then, α is self-convex in $\mathcal{M}$.

Proof. α is self-convex if and only if $\text{Hess}_\kappa \log(\mu^2)$ is positive semi-definite. Now, let $v = b + k \in \mathcal{M}$. According to Theorem 7,

$$\text{Hess}_\kappa \log(\alpha)(p)(v, v) = \text{Hess}_\kappa \log(\alpha)(p)(b, b)+$$

$$\frac{1}{2}\langle \text{grad}_\kappa((\|K\|_\kappa)^2)(p), \text{grad}_\kappa \log(\alpha)(p)\rangle_\kappa + \text{grad}_\kappa \alpha(\langle B, K\rangle_\kappa),$$

where K is as defined in Theorem 7 and B is a vector field such that $B(\phi_t(p)) = D\phi_t(p)b$. Note that $\text{grad}_\kappa \alpha(\langle B, K\rangle_\kappa)$ depends only on the value of B, and K along the integral curve $\phi_t(p)$. Moreover,

$$\langle B, K\rangle_\kappa(\phi_t(p)) = \alpha(\phi_t(p))\langle B(\phi_t(p)), K(\phi_t(p))\rangle =$$

$$\alpha(\phi_t(p))\langle D\phi_t(p)(b), K(\phi_t(p))\rangle = 0,$$

from the second item in the hypotheses of our corollary. Thus, we have

$$\mathrm{Hess}_\kappa \log(\alpha)(p)(v,v) =$$

$$\mathrm{Hess}_\kappa \log(\alpha)(p)(b,b) + \frac{1}{2}\langle \mathrm{grad}_\kappa((\|K\|_\kappa)^2)(p), \mathrm{grad}_\kappa \log(\alpha)(p)\rangle_\kappa.$$

This quantity has to be non-negative for every v or, equivalently,

- $\mathrm{Hess}_\kappa \log\alpha$ has to be positive semi-definite in $(T_pG(p))^\perp$, and
- $\langle \mathrm{grad}_\kappa((\|K\|_\kappa)^2)(p), \mathrm{grad}_\kappa \log(\alpha)(p)\rangle_\kappa \geq 0$ for every vector field K, $K(q) = \frac{d}{dt}(exp(ta)q)\|_{t=0}$ where $a \in \mathfrak{g}$.

The second of these two items can be re-written using the original Riemannian structure $\langle\cdot,\cdot\rangle$. Let us define $\mu = \sqrt{\alpha}$. Note that

$$(\|K\|_\kappa)^2 = \alpha\|K\|^2,$$

$$\mathrm{grad}_\kappa((\|K\|_\kappa)^2) = \frac{1}{\alpha}\mathrm{grad}\ (\alpha\|K\|^2) = \frac{1}{\mu}\left(\mu\ \mathrm{grad}\ (\|K\|^2) + 2\|K\|^2\mathrm{grad}\ \mu\right),$$

$$\mathrm{grad}_\kappa \log(\alpha) = \frac{1}{\alpha}\mathrm{grad}\ (\log\alpha) = \frac{2\mathrm{grad}\ \mu}{\mu^3}.$$

Thus,

$$\begin{aligned}
\langle \mathrm{grad}_\kappa((\|K\|_\kappa)^2), \mathrm{grad}_\kappa \log(\alpha)\rangle_\kappa &= \frac{1}{\alpha}\left(2\mu\langle \mathrm{grad}\ (\|K\|^2), \mathrm{grad}\ \mu\rangle + 4\|K\|^2\|\mathrm{grad}\ \mu\|^2\right) \\
&= \frac{2}{\alpha}\left(\mu D(\|K\|^2)(\mathrm{grad}\ \mu) + 2\|K\|^2\|\mathrm{grad}\ \mu\|^2\right) \\
&= \frac{1}{\alpha}\left(\mu D(\|K\|^2)(\mathrm{grad}\ \alpha) + \frac{1}{\alpha}\|K\|^2\|\mathrm{grad}\ \alpha\|^2\right).
\end{aligned}$$

The corollary follows. □

4 Self-convexity in spaces of matrices

Let $u \leq n$ and $(k) = (k_1, \dots, k_u) \in \mathbb{N}^u$ such that $k_1 + \cdots + k_u = n$. We define $\mathcal{P}_{(k)}$ as the set of matrices $A \in \mathbb{GL}_{n,m}$ with u distinct singular values

$$\sigma_1(A) > \cdots > \sigma_u(A) > 0,$$

$\sigma_i(A)$ having the multiplicity k_i. Such a matrix has a singular value decomposition $A = UDV^*$ with $U \in \mathbb{U}_n$, $V \in \mathbb{U}_m$ and $D \in \mathbb{GL}_{n,m}$ with

$$D = \operatorname{diag}\left(\overbrace{\sigma_1, \dots, \sigma_1}^{k_1}, \dots, \overbrace{\sigma_u, \dots, \sigma_u}^{k_u}\right) = \operatorname{diag}\left(\sigma_1 I_{k_1}, \dots, \sigma_u I_{k_u}\right).$$

We also let

$$\mathcal{D}_{(k)} = \left\{ D \in \mathcal{P}_{(k)} \ : \ D = \operatorname{diag}\left(\sigma_1 I_{k_1}, \dots, \sigma_u I_{k_u}\right),\ \sigma_1 > \cdots > \sigma_u \right\}.$$

Proposition 11. *$\mathcal{P}_{(k)}$ is a real smooth embedded submanifold of $\mathbb{GL}_{n,m}$. Its real codimension is*

- $k_1^2 + \cdots + k_u^2 - u$ *if* $\mathbb{K} = \mathbb{C}$.
- $\frac{1}{2}(k_1^2 + \cdots + k_u^2 - k_1 - \cdots - k_u) + n - u$ *if* $\mathbb{K} = \mathbb{R}$.

The tangent space to $\mathcal{P}_{(k)}$ at a matrix

$$D = \begin{pmatrix} \sigma_1 I_{k_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_u I_{k_u} & 0 & \cdots & 0 \end{pmatrix}$$

is the set of matrices

$$\begin{pmatrix} \lambda_1 I_{k_1} + A_1 & * & * & * & \cdots & * \\ * & \ddots & * & * & \cdots & * \\ * & * & \lambda_u I_{k_u} + A_u & * & \cdots & * \end{pmatrix}$$

where $A_1, \dots, A_u$ are skew-symmetric matrices of respective sizes $k_1, \dots, k_u$, $\lambda_1, \dots, \lambda_u \in \mathbb{R}$, and the other entries are complex numbers (real, if $\mathbb{K} = \mathbb{R}$). Moreover, for any $i = 1, \dots, u$, $\sigma_i : \mathcal{P}_{(k)} \to \mathbb{R}$ is a smooth function.

Proof. To prove that $\mathcal{P}_{(k)}$ is a real smooth embedded submanifold of $\mathbb{GL}_{n,m}$ we use Lemma 25 (see the appendix). We take $G = \mathbb{U}_n \times \mathbb{U}_m$, $\mathcal{M} = \mathbb{GL}_{n,m}$, and $\mathcal{D} = \mathcal{D}_{(k)}$. The group action of G on $\mathcal{M}$ is given by

$$(U, V, X) \in G \times \mathbb{GL}_{n,m} \to UXV^* \in \mathbb{GL}_{n,m}.$$

Under this action, the image of $\mathcal{D}_{(k)}$ is $\mathcal{P}_{(k)}$, and the equivalence relation $(U, V, D)\mathcal{R}(U', V', D')$ if and only if $UDV^* = U'D'V'^*$ is equivalent to

$$D' = D, \quad U' = UM, \quad V' = VM_W,$$

where M and M_W are unitary block-diagonal matrices

$$M = \text{diag }(U_1, \ldots, U_u), \quad M_W = \text{diag }(U_1, \ldots, U_u, W),$$

with $U_i \in \mathbb{U}_{k_i}$, and $W \in \mathbb{U}_{m-n}$. Note that the set $\mathcal{I}_{(k)}$ of such pairs (M, M_W) is the isotropy group of any $X \in \mathbb{GL}_{n,m}$. Also, the relation $\mathcal{R}$ is invariant under left $\mathbb{U}_n \times U_m$ action, namely:

$$(U, V, D)\mathcal{R}(U', V', D') \Leftrightarrow (QU, RV, D)\mathcal{R}(QU', RV', D')$$

for any $(Q, R) \in \mathbb{U}_n \times U_m$.

It is easy to see that the graph of this equivalence relation, that is the set of pairs $((U, V, D), (UM, VM_W, D))$, with U, V, D, M, and W as before, is a closed submanifold in $(G \times \mathcal{D}_{(k)}) \times (G \times \mathcal{D}_{(k)})$. Thus the quotient space $(G \times \mathcal{D}_{(k)})/\mathcal{R}$ is equipped with a unique manifold structure making π (the canonical surjection) a submersion.

Let us define

$$i : (G \times \mathcal{D}_{(k)})/\mathcal{R} \to \mathbb{GL}_{n,m}, \quad i(\pi(U, V, D)) = UDV^*.$$

We have to check that this map is an immersion. For any $(\dot{U}, \dot{V}, \dot{D})$ in the tangent space $T_{(U,V,D)}G \times \mathcal{D}_{(k)}$ we have

$$D(i \circ \pi)(U, V, D)(\dot{U}, \dot{V}, \dot{D}) = \dot{U}DV^* + U\dot{D}V^* + UD\dot{V}^* = U(AD + \dot{D} - DB)V^*$$

with $\dot{U} = UA$, $\dot{V} = VB$, A and B skew-symmetric matrices of respective size n and m. When $AD + \dot{D} - DB = 0$, we obtain, via an easy computation,

$$\dot{D} = 0, \quad A = \text{diag }(A_1, \ldots, A_u), \quad B = \text{diag }(A_1, \ldots, A_u, C),$$

where A_i and C are skew-symmetric matrices of respective sizes k_i and $m-n$. Thus $(\dot{U}, \dot{V}, \dot{D}) = (UA, VB, 0)$ is tangent to the fiber of π in $G \times \mathcal{D}_{(k)}$ above $\pi(U, V, D)$ so that $D\pi(U, V, D)(\dot{U}, \dot{V}, \dot{D}) = 0$. In other words

$$Di(\pi(U, V, D))(D\pi(U, V, D)(\dot{U}, \dot{V}, \dot{D})) = 0 \Longrightarrow D\pi(U, V, D)(\dot{U}, \dot{V}, \dot{D}) = 0$$

that is $Di(\pi(U, V, D))$ is injective.

The last point to check to apply Lemma 25 is the continuity of the inverse of i. Suppose that $X_p \to X$ with $X_p, X \in \mathrm{Im}\ i = \mathcal{P}_{(k)}$. We can write them $X_p = U_p D_p V_p^*$ and $X = UDV^*$. Let (U_{p_q}, V_{p_q}) be a subsequence which converges to $(\tilde{U}, \tilde{V})$ (G is compact). Since $X_{p_q} \to X$ we have $D_{p_q} \to \tilde{U}^* X \tilde{V} = \tilde{D}$, and $\tilde{U}\tilde{D}\tilde{V}^* = UDV^*$. Now we consider the sequence $\tilde{U}^* X_p \tilde{V}$. It is a convergent sequence, hence it has a unique limit $\tilde{D}$ and $(\tilde{U}, \tilde{V}, \tilde{D})\mathcal{R}(U, V, D)$. Thus, $\pi(\tilde{U}^* U_p, \tilde{V}^* V_p, D_p)$ converges to $\pi(I, I, D)$. By left $\mathbb{U}_n \times U_m$ action, we conclude that $\pi(U_p, V_p, D_p)$ converges to $\pi(U, V, D)$ as required.

Thus, the hypothesis of Lemma 25 is satisfied and $\mathcal{P}_{(k)}$ is a real smooth embedded submanifold of $\mathbb{GL}_{n,m}$.

The computation of its dimension is easy: it is given by the difference of the dimension of $G \times \mathcal{D}_{(k)}$ and the dimension of the fiber above any point in the quotient space, that is

$$\dim \mathbb{U}_n + \dim \mathbb{U}_m + u - \dim \mathbb{U}_{k_1} - \ldots - \dim \mathbb{U}_{k_u} - \dim \mathbb{U}_{m-n}.$$

The tangent space $T_D \mathcal{P}_{(k)}$, $D = \mathrm{diag}\ (\sigma_1 I_{k_1}, \ldots, \sigma_u I_{k_u})$, is the image of the tangent space $T_{(I_n, I_m, D)} G \times \mathcal{D}_{(k)}$ by the derivative $D(i \circ \pi)(I_n, I_m, D)$. It is the set of matrices $AD + \dot{D} - DB$ with $\dot{D} = \mathrm{diag}\ (\lambda_1 I_{k_1}, \ldots, \lambda_u I_{k_u})$, A and B skew symmetric of sizes n and m. They all have the type described in Proposition 11 and this space of matrices has the right dimension.

Let us prove the smoothness of the map $X \in \mathcal{P}_{(k)} \to \sigma_i(X) \in \mathbb{R}$. Since the map $(U, V, D) \in G \times \mathcal{D}_{(k)} \to \sigma_i(D)$ is smooth, and constant in the equivalence classes, the map $\pi(U, V, D) \in (G \times \mathcal{D}_{(k)})/\mathcal{R} \to \sigma_i(D) = \sigma_i(UDV^*)$ is also smooth. Thus the map $X = UDV^* \in \mathcal{P}_{(k)} \to \sigma_i(X)$ is smooth as the composition of the previous map by i^{-1}. □

Before proving the self-convexity of $\alpha = \sigma_u^{-2}$ in $\mathcal{P}_{(k)}$ we first consider the case of diagonal matrices.

Lemma 12. *Let $\mathcal{P}_{(k)}$ be equipped with the condition metric structure*

$$\langle \cdot, \cdot \rangle_\kappa = \sigma_u^{-2} \mathrm{Re} \langle \cdot, \cdot \rangle_F .$$

1. *If $\Sigma_1, \Sigma_2 \in \mathcal{D}_{(k)}$, then any minimizing condition geodesic in $\mathcal{P}_{(k)}$ joining Σ_1 and Σ_2 lies in $\mathcal{D}_{(k)}$,*
2. *The set $\mathcal{D}_{(k)}$ is geodesically complete (or totally geodesic) in $\mathcal{P}_{(k)}$ for this metric, namely, every geodesic in $\mathcal{D}_{(k)}$ for the induced structure is also a geodesic in $\mathcal{P}_{(k)}$, or equivalently:*
3. *If $\Sigma \in \mathcal{D}_{(k)}$ and $\dot{\Sigma} \in T_\Sigma \mathcal{D}_{(k)}$, then the unique geodesic in $\mathcal{P}_{(k)}$ through Σ with tangent vector $\dot{\Sigma}$ at Σ, remains in $\mathcal{D}_{(k)}$.*

Moreover, $\alpha = \sigma_u^{-2}$ is log-convex in $\mathcal{D}_{(k)}$.

Proof. According to Proposition 11, $\mathcal{P}_{(k)}$ is a smooth Riemannian manifold for the condition structure.

Let $\gamma(t)$, $0 \leq t \leq T$, be a minimizing condition geodesic with endpoints Σ_1 and $\Sigma_2 \in \mathcal{D}_{(k)}$. Let $\gamma(t) = U_t \Sigma_t V_t^*$ be a singular value decomposition of $\gamma(t)$, chosen such a way that U_t and V_t are smooth functions of t. Let $\sigma_u(t)$ be the smallest singular value of $\gamma(t)$. It suffices to see that $L_\kappa(\Sigma) \leq L_\kappa(\gamma)$ that is

$$\int_0^T \|\dot{\Sigma}_t\|_F \sigma_u(t)^{-1} dt \leq \int_0^T \|\dot{\gamma}_t\|_F \sigma_u(t)^{-1} dt.$$

Since

$$\dot{\gamma}_t = \dot{U}_t \Sigma_t V_t^* + U_t \dot{\Sigma}_t V_t^* + U_t \Sigma_t \dot{V}_t^*,$$

with $\dot{U}_t = U_t A_t$, $\dot{V}_t = U_t B_t$, A_t and B_t skew-symmetric, we see that

$$\|\dot{\gamma}_t\|_F^2 = \|A_t \Sigma_t + \dot{\Sigma}_t - \Sigma_t B_t\|_F^2 = \|\dot{\Sigma}_t\|_F^2 + \|A_t \Sigma_t - \Sigma_t B_t\|_F^2 \geq \|\dot{\Sigma}_t\|_F^2$$

because the diagonal terms in $\dot{\Sigma}_t$ are real numbers and those of $A_t \Sigma_t - \Sigma_t B_t$ are purely imaginary.

The second assertion is an easy consequence of the first one, The third assertion is another classical characterization of totally geodesic submanifolds, see [14] Chapter 4, Theorem 5.

Finally, for log-convexity of $\alpha(X) = \sigma_u(X)^{-2}$, using [1] Proposition 3, it suffices to see that for $\Sigma \in \mathcal{D}_{(k)}$ and $\dot{\Sigma} \in T_\Sigma \mathcal{D}_{(k)}$,

$$2\|\dot{\Sigma}\|^2 \|D\sigma_u(\Sigma)\|^2 \geq D^2 \sigma_u^2(\Sigma)(\dot{\Sigma}, \dot{\Sigma}), \tag{4.1}$$

where the second derivative is computed in the Frobenius metric structure. Now,

$$D^2(\sigma_u^2)(\Sigma)(\dot{\Sigma}, \dot{\Sigma}) = 2\sigma_u(\dot{\Sigma})^2, \quad \|D\sigma_u(\Sigma)\|^2 = \frac{1}{k_u^2},$$

and equation (4.1) follows. □

Proposition 13. *The map $\alpha = \sigma_u^{-2}$ is self-convex in $\mathcal{P}_{(k)}$.*

Proof. By unitary invariance, we may choose as initial point a matrix $\Sigma \in \mathcal{D}_{(k)}$ with ordered distinct diagonal entries $\sigma_1 > \ldots > \sigma_u > 0$. We use Corollary 10, with the group G being $\mathbb{U}_n \times \mathbb{U}_m$ and the action

$$\begin{array}{ccc} \mathbb{U}_n \times \mathbb{U}_m \times \mathcal{P}_{(k)} & \longrightarrow & \mathcal{P}_{(k)} \\ ((U,V),A) & \mapsto & UAV^*. \end{array}$$

The Lie algebra of G is the set $\mathcal{A}_n \times \mathcal{A}_m$ where $\mathcal{A}_k$ is the set of $k \times k$ skew-symmetric matrices. We now check the three conditions of Corollary 10.

- $\mathrm{Hess}_\kappa \log(\alpha)(\Sigma)$ is positive semi-definite in $(T_pG(p))^\perp$: Let

  $$\dot{\Sigma} \in (T_pG(p))^\perp = T_\Sigma \mathcal{D}_{(k)},$$

  and let γ be a condition geodesic in $\mathcal{D}_{(k)}$ such that $\gamma(0) = \Sigma$, $\dot{\gamma}(0) = \dot{\Sigma}$. We have to check that

  $$\frac{d^2}{dt^2} \log \alpha(\gamma(t)) \mid_{t=0} \geq 0.$$

  This is true as α is log-convex in $\mathcal{D}_{(k)}$ from Lemma 12.

- We have to check that for small enough t, and for

  $$b \in T_\Sigma \{U\Sigma V^* : U \in \mathbb{U}_n, V \in \mathbb{U}_m\}^\perp \subseteq T_p\mathcal{P}_{(k)},$$

  $D\phi_t(D)b$ is perpendicular to

  $$T_{\phi_t(\Sigma)}\{U\phi_t(\Sigma)V^* : U \in \mathbb{U}_n, V \in \mathbb{U}_m\},$$

  where ϕ_t is the flow of $\mathrm{grad}_\kappa \alpha$. First, we note that for any $L \in \mathcal{D}_{(k)}$, we have

  $$T_L \{ULV^* : U \in \mathbb{U}_n, V \in \mathbb{U}_m\}^\perp =$$
  $$\{B_1 L + L B_2^* : (B_1, B_2) \in \mathcal{A}_n \times \mathcal{A}_m\}^\perp .$$

  Let us denote by S this last set. We claim that $S = \mathcal{D}_{(k)}$. Indeed, $\mathcal{D}_{(k)} \subseteq S$, because the diagonal of any matrix of the form $B_1 L + L B_2^*$ is purely imaginary and hence orthogonal to $\mathcal{D}_{(k)}$. The other inclusion is easily checked by a dimensional argument: The dimension of $\mathcal{D}_{(k)}$ is u and the dimension of S is

$$\dim(\mathcal{P}_{(k)}) - \dim\left\{B_1 L + L B_2^* : (B_1, B_2) \in \mathcal{A}_n \times \mathcal{A}_m\right\},$$

that is $\dim(\mathcal{P}_{(k)})$ minus the dimension of the orbit of L under the action of $\mathbb{U}_n \times \mathbb{U}_m$. We have computed these two quantities in Proposition 11, and we immediately conclude that $\dim(S) = u$, for both $\mathbb{K} = \mathbb{C}$ and $\mathbb{K} = \mathbb{R}$. Thus, for all $L \in \mathcal{D}_{(k)}$,

$$T_L \left\{ULV^* : U \in \mathbb{U}_n, V \in \mathbb{U}_m\right\}^{\perp} = \mathcal{D}_{(k)}.$$

In our case, ϕ_t can be computed exactly. Indeed,

$$\text{grad}_\kappa \alpha = \frac{1}{\alpha} \text{grad } \alpha = -\frac{2}{k\sigma_u} E,$$

where

$$E = \text{diag } (0, \ldots, 0, \overbrace{1, \ldots, 1}^{k}).$$

Thus, grad α preserves the diagonal form, and $\phi_t(\Sigma) \in \mathcal{D}_{(k)}$ is a diagonal matrix, for every t while defined. Thus, $D\phi_t(\Sigma)(\dot{\Sigma})$ is again a diagonal matrix, for every diagonal matrix $\dot{\Sigma}$. This proves that the second condition of Corollary 10 applies to our case.

- For $(B_1, B_2) \in \mathcal{A}_n \times \mathcal{A}_m$, the vector field K on $\mathbb{GL}_{n,m}$ generated by (B_1, B_2) is

$$K(A) = \frac{d}{dt}\left(e^{tB_1} A e^{tB_2^*}\right)|_{t=0} = B_1 A + A B_2^*.$$

Note that

1. K^* as a linear operator on $\mathbb{GL}_{n,m}$ satisfies $K^*(A) = B_1^* A + A B_2$.
2. $\|K(\Sigma)\|^2 = \|B_1 \Sigma + \Sigma B_2^*\|^2$,
3. For $w \in T_\Sigma \mathcal{P}_{(k)}$, $D(\|K\|^2)(\Sigma)w = 2\text{Re}\langle K^* K(\Sigma), w\rangle = 2\text{Re}\langle B_1 B_1^* \Sigma + \Sigma B_2 B_2^* - 2 B_1 \Sigma B_2^*, w\rangle$.
4. grad $\alpha(\Sigma) = -\frac{2}{k\sigma_u^3} E$ where $E = \text{diag } (0, \ldots, 0, \overbrace{1, \ldots, 1}^{k})$.

Thus,

$$\alpha(\Sigma)D(\|K\|^2)(\Sigma)(\text{grad }\alpha(\Sigma)) + \|K(\Sigma)\|^2\|\text{grad }\alpha(\Sigma)\|^2 =$$

$$\frac{4}{k_u\sigma_u^6}\left(-\sigma_u \text{Re}\langle B_1B_1^*\Sigma + \Sigma B_2B_2^* - 2B_1\Sigma B_2^*, D\rangle + \|B_1\Sigma - \Sigma B_2\|^2\right).$$

Hence, it suffices to see that $J \geq 0$ where

$$J = \|B_1\Sigma + \Sigma B_2\|^2 - \sigma_u \text{Re}\langle B_1B_1^*\Sigma + \Sigma B_2B_2^* - 2B_1\Sigma B_2^*, D\rangle.$$

Expanding this expression and writing $\Sigma' = \Sigma^* - \sigma_u D^*$, we have

$$J = \text{Re}\left(\text{trace }(B_1B_1^*\Sigma\Sigma') + \text{trace }(\Sigma'\Sigma B_2B_2^*) - 2\text{trace }(B_1\Sigma B_2^*\Sigma')\right),$$

which by Lemma 14 below is a non-negative quantity. The proposition follows.

□

Lemma 14. *Let $\Sigma = diag\ (\sigma_1 I_{k_1}, \ldots, \sigma_{u-1}I_{k_{u-1}}, \sigma_u I_{k_u}) \in \mathbb{GL}_{n,m}$ and $\Sigma' = diag\ (\sigma_1 I_{k_1}, \ldots, \sigma_{u-1}I_{k_{u-1}}, 0I_{k_u}) \in \mathbb{GL}_{m,n}$. Then, for any skew-symmetric matrices B, C of respective sizes n, m, we have:*

$$\text{Re}\left(\text{trace }(BB^*\Sigma\Sigma') + \text{trace }(\Sigma'\Sigma CC^*) - 2\text{ trace }(B\Sigma C^*\Sigma')\right) \geq 0.$$

Proof. We denote

$$J = \text{Re}\left(\text{trace }(BB^*\Sigma\Sigma') + \text{trace }(\Sigma'\Sigma CC^*) - 2\text{ trace }(B\Sigma C^*\Sigma')\right).$$

Write

$$\Sigma = \begin{pmatrix} L & 0 & 0 \\ 0 & \sigma_u I_k & 0 \end{pmatrix}, \quad \Sigma' = \begin{pmatrix} L & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and let us write B, C by blocks,

$$B = \begin{pmatrix} B_1 & B_2 \\ -B_2^* & B_4 \end{pmatrix}, \quad C = \begin{pmatrix} C_1 & C_2 & C_3 \\ -C_2^* & C_4 & C_5 \\ -C_3^* & -C_5^* & C_6 \end{pmatrix}$$

where B_1, C_1 are of the size of L and B_4, C_4 are of the size of I_k. Then,

$$\begin{aligned} \text{trace }(BB^*\Sigma\Sigma') &= \text{trace }((B_1B_1^* + B_2B_2^*)L^2), \\ \text{trace }(\Sigma'\Sigma CC^*) &= \text{trace }(L^2(C_1C_1^* + C_2C_2^* + C_3C_3^*)), \\ \text{trace }(B\Sigma C^*\Sigma') &= \text{trace }(B_1LC_1^*L + \sigma_u B_2C_2^*L). \end{aligned}$$

Thus,

$$J \geq \operatorname{Re}\left(\operatorname{trace}\left((B_1B_1^* + C_1C_1^*)L^2 - 2B_1LC_1^*L\right)\right) +$$
$$\operatorname{Re}\left(\operatorname{trace}\left((B_2B_2^* + C_2C_2^* + C_3C_3^*)L^2 - 2\sigma_u B_2C_2^*L\right)\right).$$

We will prove that these two terms are non-negative. For the first one, note that

$$\operatorname{Re}\left(\operatorname{trace}\left((B_1B_1^* + C_1C_1^*)L^2 - 2B_1LC_1^*L\right)\right) = \|B_1L - LC_1\|^2 \geq 0.$$

For the second one, we check that for every l, $1 \leq l \leq n-k$ the l-th diagonal entry of the matrix $(B_2B_2^* + C_2C_2^* + C_3C_3^*)L^2 - 2\sigma_u B_2C_2^*L$ has a positive real part. Indeed, if we denote by $v \in \mathbb{K}^k$ the l-th row of B_2 by $w \in \mathbb{K}^k$ the l-th row of C_2 and by x the l-th row of C_3, we have

$$\operatorname{Re}\left((B_2B_2^* + C_2C_2^* + C_3C_3^*)L^2 - 2\sigma_u B_2C_2^*L\right)_{l,l} =$$

$$\sigma_l^2\left((\|v\|^2 + \|w\|^2 + \|x\|^2) - 2\frac{\sigma_u}{\sigma_l}\operatorname{Re}\langle v, w\rangle\right) \geq \sigma_l^2\|v - w\|^2 \geq 0$$

as $\sigma_u < \sigma_l$. This finishes the proof of Lemma 14 and hence of Proposition 13. □

5 Puting pieces together

Before stating the main result of this section we have to introduce the following machinery:

5.1 Second symmetric derivatives

In the case of Lipschitz-Riemann structures, the mappings we want to consider are not necessarily C^2 and, to study their convexity properties, an approach based on Hessians is insufficient. We will use instead the second symmetric upper derivative.

Let $U \subseteq \mathbb{R}^k$ be an open set and $\phi : U \to \mathbb{R}$ be any function. The **second symmetric upper derivative** of ϕ at $x \in U$ in the direction $v \in \mathbb{R}^k$ is

$$\overline{\mathcal{SD}^2}\phi(x; v) = \limsup_{h \mapsto 0} \frac{\phi(x + hv) + \phi(x - hv) - 2\phi(x)}{h^2}$$

which is allowed to be $\pm\infty$. If $U \subseteq \mathbb{R}$ is an interval, we simply write $\overline{\mathcal{SD}^2}\phi(x)$.

It is well-known that a continuous function ϕ on an interval is convex if and only if $\overline{\mathcal{SD}}^2\phi(x) \geq 0$ for all x (see for example [21] Theorem 5.29). There is a stronger result due to Burkill [4] Theorem 1.1 (see also [21] Corollary 5.31) which uses a weaker hypothesis:

Theorem 15 (Burkill)**.** *Let $\phi :]a,b[\to \mathbb{R}$ be a continuous function such that $\overline{\mathcal{SD}}^2\phi(x) \geq 0$ for almost all $x \in]a,b[$, and assume that $\overline{\mathcal{SD}}^2\phi(x) > -\infty$ for $x \in]a,b[$. Then, ϕ is a convex function.*

Theorem 15 will allow us to assemble the pieces where convexity is proven in Proposition 13 to prove our main results (theorems 1 and 23). We proceed a little more generally as the result may be of interest in other circumstances. Let $\mathcal{M}$ be a k-dimensional C^2 manifold (not necessarily having a Riemannian structure).

Definition 16. *Let $\alpha : \mathcal{M} \to \mathbb{R}$. We say that $\overline{\mathcal{SD}}^2\alpha$* **is bounded from** $-\infty$ *(denoted $\overline{\mathcal{SD}}^2\alpha > -\infty$) if for every $x \in \mathcal{M}$ there is an open neighborhood $U_x \subseteq \mathcal{M}$ and a coordinate chart $\varphi_x : U_x \to \mathbb{R}^k$, $\varphi_x(x) = 0$ such that*

$$\overline{\mathcal{SD}}^2(\alpha \circ \varphi_x^{-1})(0; v) > -\infty$$

for every $v \in \mathbb{R}^k$.

The following lemma is a consequence of Definition 16.

Lemma 17. *Let $\mathcal{M}$ be a C^2 manifold of dimension k, let $\alpha : \mathcal{M} \to]0,\infty[$ be a locally Lipschitz mapping, and let $\phi :]0,\infty[\to \mathbb{R}$ be a C^2 function. If $\overline{\mathcal{SD}}^2\alpha > -\infty$ then $\overline{\mathcal{SD}}^2(\phi \circ \alpha) > -\infty$. In particular, $\overline{\mathcal{SD}}^2\alpha > -\infty$ if and only if $\overline{\mathcal{SD}}^2(\log \circ \alpha) > -\infty$.*

Proof. Assume that $\overline{\mathcal{SD}}^2\alpha > -\infty$. Let $x \in \mathcal{M}$ and let $\varphi_x : U_x \to \mathbb{R}^k$ be a coordinate chart such that $\varphi_x(x) = 0$ and $\overline{\mathcal{SD}}^2(\alpha \circ \varphi_x^{-1})(0; v) > -\infty$ for each $v \in \mathbb{R}^k$. There is a sequence $h_p \to 0$ such that

$$\lim_{p\to\infty} \frac{\alpha(\varphi_x^{-1}(h_p v)) + \alpha(\varphi_x^{-1}(-h_p v)) - 2\alpha(x)}{h_p^2} = C > -\infty.$$

Let us define $H_p = \alpha(\varphi_x^{-1}(h_p v)) - \alpha(x)$, and similarly $K_p = \alpha(\varphi_x^{-1}(-h_p v)) - \alpha(x)$. By Taylor's formula we get

$$\phi(\alpha(\varphi_x^{-1}(h_p v))) = \phi(\alpha(x)) + \phi'(\alpha(x))H_p + \phi''(\alpha(x))\frac{H_p^2}{2} + o(H_p^2),$$

and similarly

$$\phi(\alpha(\varphi_x^{-1}(-h_p v))) = \phi(\alpha(x)) + \phi'(\alpha(x))K_p + \phi''(\alpha(x))\frac{K_p^2}{2} + o(K_p^2),$$

so that

$$\frac{\phi(\alpha(\varphi_x^{-1}(h_p v))) + \phi(\alpha(\varphi_x^{-1}(-h_p v))) - 2\phi(\alpha(x)}{h_p^2} =$$

$$\phi'(\alpha(x))\frac{H_p + K_p}{h_p^2} + \phi''(\alpha(x))\frac{H_p^2 + K_p^2}{2h_p^2} + \frac{o(H_p^2) + o(K_p^2)}{h_p^2}.$$

Notice that $\lim_{p\to\infty} \frac{H_p+K_p}{h_p^2} = C$. Since $h \to \alpha(\varphi_x^{-1}(hv))$ is Lipschitz in a neighborhood of 0 we have, for a suitable constant $D > 0$, $H_p^2 \leq Dh_p^2$ and $K_p^2 \leq Dh_p^2$. Thus, taking the lim sup as $p \to \infty$ gives $\overline{\mathcal{SD}}^2\phi(\alpha \circ \varphi_x^{-1})(0, v) \geq C + D > -\infty$ and we are done. □

5.2 Projecting geodesics on submanifolds : the Euclidean case

The following technical lemma, interesting by itself, is a consequence of Lebesgue's Density Theorem.

Lemma 18. *For any locally integrable function f defined in $\mathbb{R}$ with values in $\mathbb{R}^n$, let F denote an antiderivative of f, and let $x \in \mathbb{R}$ be such that $F'(x) = f(x)$. Then*

$$\lim_{\varepsilon\to 0} \frac{2}{\varepsilon^2}\int_x^{x+\varepsilon} (y - x)f(y)dy = f(x).$$

Proof. Notice that, by Lebesgue's differentiation theorem, an antiderivative F of f exists a. e. and it is absolutly continuous. Suppose that $F(x) = 0$. Let us define

$$h(y) = \begin{cases} F(y)/(y - x) & \text{if } y \neq x, \\ f(x) & \text{if } y = x, \end{cases}$$

so that h is a continuous function and $F(y) = (y - x)h(y)$ for any y. Integrating by parts gives

$$\int_x^{x+\varepsilon} (y - x)f(y)dy = \varepsilon F(x + \varepsilon) - \int_x^{x+\varepsilon} F(y)dy$$

so that

$$\frac{2}{\varepsilon^2}\int_x^{x+\varepsilon}(y-x)f(y)dy = 2\frac{F(x+\varepsilon)-F(0)}{\varepsilon} - \frac{2}{\varepsilon^2}\int_x^{x+\varepsilon}(y-x)h(y)dy.$$

Since h is continuous, by the Mean Value Theorem, there exists $\zeta \in [x, x+\varepsilon]$ such that

$$\frac{2}{\varepsilon^2}\int_x^{x+\varepsilon}(y-x)h(y)dy = \frac{2h(\zeta)}{\varepsilon^2}\int_x^{x+\varepsilon}(y-x)dy = h(\zeta) \to h(x) = f(x)$$

as $\varepsilon \longrightarrow 0$. On the other hand

$$\lim_{\varepsilon\to 0} 2\frac{F(x+\varepsilon)-F(x)}{\varepsilon} = 2f(x).$$

Thus

$$\lim_{\varepsilon\to 0}\frac{2}{\varepsilon^2}\int_x^{x+\varepsilon}(y-x)f(y)dy = 2f(x) - f(x) = f(x)$$

and we are done. □

Our aim is now to see how close are a geodesic in a Lipschitz-Riemannian manifold and a geodesic in a submanifold when they have the same tangent at a given point. Let us start to study a simple case.

Let us consider the Lipschitz-Riemann structure defined on an open set $\Omega \subset \mathbb{R}^k$ containing 0 by the scalar product $\langle u, v\rangle_x = v^T H(x)u$ (see section 2.3).

1. The matrix $H(0)$ is supposed to have the following block structure

$$H(0) = \begin{pmatrix} H_p(0) & 0 \\ 0 & H_{k-p}(0) \end{pmatrix}.$$

We also suppose that (see section 2.4)

2. The entries $h_{ij}(x)$ of $H(x)$ are regular at $x = 0$,

The set $\Omega_p = \Omega \cap (\mathbb{R}^p \times \{0\})$ is a submanifold in Ω. We suppose that

3. H_p is C^2 in Ω_p,

so that Ω_p is in fact a smooth C^2 Riemannian manifold for the induced $H-$structure. Let us now condider a vector $a \in \mathbb{R}^p \times \{0\}$ and three parametrized curves denoted by x, x_p, and y defined in a neighborhood of 0 in $\mathbb{R}$, and such that:

4. $x(0) = x_p(0) = y(0) = 0$,
5. $\dot{x}(0) = \dot{x}_p(0) = \dot{y}(0) = a$,
6. x is a geodesic in $\mathbb{R}^k$,
7. x_p is its orthogonal projection onto $\mathbb{R}^p \times \{0\}$,
8. y is a geodesic in $\mathbb{R}^p \times \{0\}$ for the induced structure.

According to Theorem 2, x has regularity C^{1+Lip} so that its second derivative exists a.e. We suppose here that

9. The second derivative $\ddot{x}(t)$ is defined at $t = 0$, and

$$\frac{d}{dt}(H(x(t))\dot{x}(t)) \in \frac{1}{2}\sum_{i,j} \dot{x}_i(t)\dot{x}_j(t)\partial h_{ij}(x(t)).$$

In this context we have:

Lemma 19. *Under the hypotheses* 1 *to* 9 *above, the curves* x_p *and* y *have a contact of order* 2 *at* 0*:* $x_p(s) = y(s) + o(s^2)$.

Proof. The geodesic $x(t) \in \mathbb{R}^k$ satisfies the following differential system (Remark 5):

$$\begin{cases} x \in C^{1+Lip}, \\ \dot{x}^T(t)H(x(t))\dot{x}(t) = 1, \\ \frac{d}{dt}(H(x(t))\dot{x}(t)) \in \frac{1}{2}\sum_{i,j} \dot{x}_i(t)\dot{x}_j(t)\partial h_{ij}(x(t)), \text{ a.e.}. \end{cases}$$

Similarly, we obtain for y (taking into account hypothesis 3):

$$\begin{cases} y \in C^2, \\ \dot{y}^T(s)H_p(y(s))\dot{y}(s) = 1, \\ \frac{d}{ds}(H_p(y(s))\dot{y}(s)) = \frac{1}{2}\sum_{i,j} \dot{y}_i(s)\dot{y}_j(s)\text{grad } h_{p,ij}(y(s)). \end{cases}$$

The first equation computed at $t = 0$ gives (thanks to hypothesis 9)

$$H(x(0))\ddot{x}(0) + \frac{d}{dt}_{|t=0}(H(x(t)))\dot{x}(0) \in \frac{1}{2}\sum_{i,j=1}^{k} \dot{x}_i(0)\dot{x}_j(0)\partial h_{ij}(x(0)).$$

When we project it onto $\mathbb{R}^p$ we get, with $x(t) = \begin{pmatrix} x_p(t) \\ x_{k-p}(t) \end{pmatrix}$,

$$H_p(0)\ddot{x}_p(0) + \frac{d}{dt}_{|t=0}(H_p(x_p(t)))\dot{x}_p(0) \in \frac{1}{2}\sum_{i,j=1}^{p} \dot{x}_{p,i}(0)\dot{x}_{p,j}(0)\Pi_{\mathbb{R}^p}\partial h_{ij}(x(0)).$$

Since the functions $h_{ij}(x)$ for $i, j = 1 \dots p$ are regular (hypothesis 2), from Clarke [6] Proposition 2.3.15, we obtain

$$\Pi_{\mathbb{R}^p}\partial h_{ij}(x(0)) = \partial h_{p,ij}(x_p(0)) = \text{grad } h_{p,ij}(x_p(0))$$

so that

$$H_p(0)\ddot{x}_p(0) + \frac{d}{dt}(H_p(x_p(0)))\dot{x}_p(0) = \frac{1}{2}\sum_{i,j=1}^{p} \dot{x}_{p,i}(0)\dot{x}_{p,j}(0)\text{grad } h_{p,ij}(x_p(0)).$$

Taking at $t = 0$ the differential equation giving y and noting that $y(0) = x_p(0)$, $\dot{y}(0) = \dot{x}_p(0)$ gives

$$\ddot{y}(0) = \ddot{x}_p(0).$$

We want to prove that $x_p(s) = y(s) + o(s^2)$. According to Taylor's formula with integral remainder, we have

$$x_p(s) - y(s) = x_p(0) - y(0) + s(\dot{x}_p(0) - \dot{y}(0)) + \int_0^s (\ddot{x}_p(\sigma) - \ddot{y}(\sigma))\sigma d\sigma,$$

so that,

$$2\frac{x_p(s) - y(s)}{s^2} = \frac{2}{s^2}\int_0^s (\ddot{x}_p(\sigma) - \ddot{y}(\sigma))\sigma d\sigma.$$

From Lemma 18 and hypothesis 9, the limit of this expression exists at $s = 0$, and it is equal to $\ddot{x}_p(0) - \ddot{y}(0) = 0$. This achieves the proof. □

5.3 Projecting geodesics on submanifolds : the Riemannian case

Our aim, in this section, is to prove another version of Lemma 19 in a different geometric context. Let $\mathcal{M}$ be a C^3 Riemannian manifold with distance d, of dimension k, and let $\mathcal{N}$ be a submanifold of dimension p.

Let us first define the **projection onto** $\mathcal{N}$. To each $q \in \mathcal{N}$ and to a vector $u \neq 0$ normal to $\mathcal{N}$ at q we associate the geodesic $\gamma_{q,u}$ in $\mathcal{M}$ such that $\gamma_{q,u}(0) = q$ and $\dot{\gamma}_{q,u}(0) = u$. Let $n \in \mathcal{N}$ be given, and let U be an open neighborhood of n such that, for each $m \in U$ there exists a unique geodesic arc $\gamma_{q,u}(t)$, t in an open interval containing 0, contained in U and containing m. Thus U is the union of such geodesic arcs and two of them have always a void intersection. This picture defines a map $K : U \to \mathcal{N}$ by $K(m) = q$ if $m = \gamma_{q,u}(t)$. The map K is the **projection map onto** $\mathcal{N}$. It has the following classical properties:

1. It is defined in the neighborhood U of $n \in \mathcal{N}$,
2. For each $m \in U$, $K(m)$ is the unique point in $\mathcal{M}$ such that
$$\inf_{q \in \mathcal{N}} d(m, q) = d(m, K(m))$$
3. K is C^2.

See Li-Nirenberg [13] or Beltran-Dedieu-Malajovich-Shub [1].

Let $\alpha : \mathcal{M} \to]0, \infty[$ be a locally Lipschitz map regular at n (see section 2.4). It defines a conformal Lipschitz-Riemann structure on $\mathcal{M}$ associated with the inner product

$$\langle \cdot, \cdot \rangle_{\alpha,m} = \alpha(m) \langle \cdot, \cdot \rangle_m .$$

We call it the $\alpha-$structure. We suppose that α is C^2 when it is restricted to $\mathcal{N}$ so that $\mathcal{N}$ is C^2 and not only Lipschitz for the induced $\alpha-$structure.

Let γ be a geodesic curve in $\mathcal{M}$ for the $\alpha-$structure such that $\gamma(0) = n \in \mathcal{N}$, $\dot{\gamma}(0) \in T_{\gamma(0)}\mathcal{N}$, and such that $\ddot{\gamma}(0)$ exists. We denote by $\gamma_{\mathcal{N}}$ the projection of γ onto $\mathcal{N}$ and by δ the geodesic in $\mathcal{N}$ for the induced $\alpha-$structure with the same initial data. Thus

$$\gamma(0) = \gamma_{\mathcal{N}}(0) = \delta(0) = n, \ \dot{\gamma}(0) = \dot{\gamma}_{\mathcal{N}}(0) = \dot{\delta}(0),$$

and δ is C^2.

Lemma 20. *Under the hypotheses above, the curves $\gamma_{\mathcal{N}}$ and δ have a contact of order 2 at 0: $d(\gamma_{\mathcal{N}}(t), \delta(t)) = o(t^2)$.*

Proof. The proof consists in a transfer from $\mathcal{M}$ to $\mathbb{R}^k$ where we apply Lemma 19.

Since $\mathcal{M}$ is C^3, the normal bundle to $\mathcal{N}$ is C^2 and there exists a C^2 diffeomorphism $\phi : U \to V \subset \mathbb{R}^k$, where V is an open set containing 0, satisfying

1. $\phi(n) = 0$,
2. $\phi(U \cap \mathcal{N}) = V \cap (\mathbb{R}^p \times \{0\})$,
3. For any $q \in \mathcal{N}$ and the vector $u \neq 0$ normal to $\mathcal{N}$ at q, $\phi(\gamma_{q,u})$ is a straight line in $\mathbb{R}^k$ orthogonal to $\mathbb{R}^p \times \{0\}$.

We make ϕ an isometry in defining on $V \subset \mathbb{R}^k$ a Lipschitz-Riemannian structure by

$$\langle D\phi(m)u, D\phi(m)v \rangle_{\phi(m)} = \alpha(m) \langle u, v \rangle_m$$

for any $m \in U$, and $u, v \in T_m\mathcal{M}$. Let us denote $x = \phi(m)$, $a = D\phi(m)u$, $b = D\phi(m)v$, we also write this scalar product

$$\langle a, b \rangle_x = b^T H(x) a$$

where H is a locally Lipschitz map from V into the $k \times k$ positive definite matrices.

Notice that H is regular at 0 because α is regular at n.

Since

$$D\phi(n)(T_n\mathcal{N}) = \mathbb{R}^p \times \{0\} \text{ and } D\phi(n)\left((T_n\mathcal{N})^\perp\right) = \{0\} \times \mathbb{R}^{k-p},$$

$H(0)$ has the block structure

$$H(0) = \begin{pmatrix} H_p(0) & 0 \\ 0 & H_{k-p}(0) \end{pmatrix}.$$

Since α is C^2 when restricted to $\mathcal{N}$ we have the same regularity for the restriction of H to $\mathbb{R}^p \times \{0\}$.

Since ϕ is an isometry the curves $\phi \circ \gamma$ and $\phi \circ \delta$ are geodesics in $\mathbb{R}^k$ and $\mathbb{R}^p \times \{0\}$ respectively, and, from the definition of ϕ, the orthogonal projection (in the Euclidean meaning) of $\phi \circ \gamma$ onto $\mathbb{R}^p \times \{0\}$ is equal to $\phi \circ \gamma_{\mathcal{N}}$.

Finally we notice that the second derivative $\frac{d^2}{dt^2}\phi \circ \gamma(0)$ exists because $\ddot{\gamma}(0)$ exists.

Thus, the hypotheses of Lemma 19 are satisfied so that $\phi \circ \gamma_{\mathcal{N}}$ and $\phi \circ \delta$ have an order 2 contact at $t = 0$. This gives easily an order 2 contact for $\gamma_{\mathcal{N}}$ and δ at $t = 0$ in $\mathcal{M}$ in terms of the $\alpha-$distance but also, since $1/\alpha$ is locally Lipschitz, in terms of the initial Riemannian distance. □

5.4 Arriving to the main theorem

We are now ready to state the main theorem in this section:

Theorem 21 (Piecing together)**.** *$\mathcal{M} = \cup_{i=1}^{\infty}\mathcal{M}_i$ is a C^2 Riemannian manifold, enumerable union of the submanifolds $\mathcal{M}_i$. Let $\alpha : \mathcal{M} \to]0, \infty[$ be a locally Lipschitz mapping. Assume that:*

1. *α is regular,*
2. *For each i, the restriction of α to $\mathcal{M}_i$ is C^2 and self-convex in $\mathcal{M}_i$,*
3. *$\overline{\mathcal{SD}}^2\alpha > -\infty$.*

Then, α is self-convex in $\mathcal{M}$.

Proof. Once again we add to $\mathcal{M}$ the $\alpha-$structure. If this theorem is false, there exists a geodesic γ in $\mathcal{M}$ for the $\alpha-$structure such that

$$\overline{\mathcal{SD}}^2 \log(\alpha(\gamma(t))) < 0$$

on a positive measure set $P \subset \mathbb{R}$ (Theorem 15 and Lemma 17). Since an enumerable union of zero-measure sets is also a zero-measure set, we can suppose that $P \subset \mathcal{M}_i$ for some i, so that $\gamma(t) \in \mathcal{M}_i$ for every $t \in P$. According to the Lebesgue Density Theorem, almost all points $t \in P$ are density points, that is

$$\lim_{\varepsilon \to 0} \frac{\text{meas}\,(P \cap [t-\varepsilon, t+\varepsilon])}{2\varepsilon} = 1.$$

We remove the "non-density points" from P to obtain a new set, also called P, with positive measure and only density points. Since $\gamma \in C^{1+Lip}$ (Theorem 2), the second derivative $\ddot{\gamma}(t)$ exists for almost all t. We also remove from P the points where Hypothesis 9 of Lemma 19 does not hold.

Let $t \in P$ be given. Since it is a density point of P, we have $s \in P$ for "a lot of points" close to t. Since $\gamma(s) \in \mathcal{M}_i$ for such points, and since γ is C^1, we get

$$\dot{\gamma}(t) \in T_{\gamma(t)}\mathcal{M}_i.$$

Take now the geodesic δ in $\mathcal{M}_i$ for the induced $\alpha-$structure such that $\delta(t) = \gamma(t)$ and $\dot{\delta}(t) = \dot{\gamma}(t)$. By self-convexity of α in $\mathcal{M}_i$, and since δ is C^2 we get

$$\overline{\mathcal{SD}}^2 \log \circ\alpha \circ \delta(t) = \frac{d^2}{dt^2} \log \circ\alpha \circ \delta(t) \geq 0.$$

Let us now consider

$$\Delta^2(h) = \frac{\log \circ\alpha \circ \gamma(t+h) + \log \circ\alpha \circ \gamma(t-h) - 2\log \circ\alpha \circ \gamma(t)}{h^2}.$$

It is not difficult to prove that t is a density point of

$$Q = \{s = t + h \in P \ : \ t - h \in P\}.$$

Let us denote by γ_i the projection of γ on $\mathcal{M}_i$ (see section 5.3). For the points $s = t + h \in Q$, one has $\gamma(t+h) = \gamma_i(t+h)$, $\gamma(t-h) = \gamma_i(t-h)$, and $\gamma(t) = \gamma_i(t)$, thus

$$\Delta^2(h) = \frac{\log \circ\alpha \circ \gamma_i(t+h) + \log \circ\alpha \circ \gamma_i(t-h) - 2\log \circ\alpha \circ \gamma_i(t)}{h^2}.$$

By Lemma 20, γ_i and δ have a contact of order 2 at t so that

$$\Delta^2(h) = \frac{\log \circ\alpha \circ \delta(t+h) + \log \circ\alpha \circ \delta(t-h) - 2\log \circ\alpha \circ \delta(t) + o(h^2)}{h^2}.$$

Since δ is C^2, taking the limit as $h \to 0$ gives

$$\lim \Delta^2(h) = \frac{d^2}{dt^2} \log \circ\alpha \circ \delta(t).$$

Since this last expression is nonegative we obtain

$$\overline{\mathcal{SD}}^2 \log(\alpha(\gamma(t))) \geq \lim \Delta^2(h) \geq 0$$

which contradicts our hypothesis $\overline{\mathcal{SD}}^2 \log(\alpha(\gamma(t))) < 0$ on P. $\square$

6 Proof of Theorem 1

Theorem 1 is a consequence of Theorem 21 applied to $\mathcal{M} = \mathbb{GL}_{n,m}$ considered as the union of the submanifolds $\mathcal{P}_{(k)}$ (see section 4) and to the mapping $\alpha(A) = \sigma_n(A)^{-2}$, the inverse of the square of the smallest singular value of $A \in \mathbb{GL}_{n,m}$. According to propositions 11 and 13 we just have to prove that α is a regular map and that $\overline{\mathcal{SD}}^2\alpha > -\infty$. Let us start with this last inequality.

We must prove that for every $A \in \mathbb{GL}_{n,m}$, $B \in \mathbb{K}^{n\times m}$,

$$\overline{\mathcal{SD}}^2\sigma_n^{-2}(A;B) = \limsup_{h\mapsto 0} \frac{\sigma_n^{-2}(A_h) + \sigma_n^{-2}(A_{-h}) - 2\sigma_n^{-2}(A)}{h^2} > -\infty,$$

where $A_h = A + hB$. Now, let $\mathcal{S}_n^+$ be the set of positive definite $n \times n$ matrices. Then,

$$\sigma_n^{-2}(A_h) + \sigma_n^{-2}(A_{-h}) = \lambda_n^{-1}(A_hA_h^*) + \lambda_n^{-1}(A_{-h}A_{-h}^*).$$

where, λ_n denotes the smallest eigenvalue. Since, for any $S \in \mathcal{S}_n^+$,

$$\lambda_n(S) = \inf_{u\in\mathbb{R}^n,\ \|u\|=1} u^TSu,$$

it is a concave function of S, and λ_n^{-1} is convex. Thus,

$$\lambda_n^{-1}(A_hA_h^*)+\lambda_n^{-1}(A_{-h}A_{-h}^*) \geq 2\lambda_n^{-1}\left(\frac{A_hA_h^* + A_{-h}A_{-h}^*}{2}\right) = 2\lambda_n^{-1}(AA^*+h^2BB^*).$$

We conclude that

$$\overline{\mathcal{SD}}^2\sigma_n^{-2}(A;B) \geq \limsup_{h\mapsto 0} \frac{2\lambda_n^{-1}(AA^* + h^2BB^*) - 2\lambda_n^{-1}(AA^*)}{h^2}.$$

This last quantity is bounded in absolute value for λ_n^{-1} is locally Lipschitz, so in particular $\overline{\mathcal{SD}}^2\sigma_n^{-2}(A;B) > -\infty$.

To prove that α is regular it suffices to write it as the composition of C^1 maps and of the convex λ_n^{-1} which is also a regular map (see [6] Prop. 2.3.6). This finishes the proof of our Main Theorem 1.

7 The solution variety

As in [1], we are also interested in the log–convexity of $\sigma_n(A)^{-1}$ in the solution variety:

$$\mathcal{W} = \{(A, x) \in \mathbb{GL}_{n,n+1} \times \mathbb{P}(\mathbb{K}^{n+1}) : Ax = 0\}.$$

Remark 22. *In [1] we have sometimes taken A to lie in the unit sphere of $\mathbb{K}^{n\times m}$ or even the projective space $\mathbb{P}(\mathbb{K}^{n\times m})$. The interested reader can check [1] for the relations between self-convexity in the various settings.*

Theorem 23. *For any condition geodesic $t \to (A(t), x(t))$ in $\mathcal{W}$, the map $t \to \log(\sigma_n^{-1}(A(t)))$ is convex.*

As we have done in the case of $\mathbb{GL}_{n,m}$, we divide the proof in several sections.

7.1 The smooth part of $\mathcal{W}$

Let $u \le n$ and $(k) = (k_1, \ldots, k_u) \in \mathbb{N}^u$ such that $k_1 + \cdots + k_u = n$. We define $\mathcal{W}_{(k)} = \{(A, x) \in \mathcal{W} : A \in \mathcal{P}_{(k)}\}$.

Proposition 24. *For any choice of (k), the set $\mathcal{W}_{(k)}$ is a smooth submanifold of $\mathcal{W}$, σ_u is a smooth function and $\alpha = \sigma_u^{-2}$ is self–convex in $\mathcal{W}_{(k)}$.*

Proof. Let us consider the map

$$\begin{array}{llcl} \psi : & \mathcal{P}_{(k)} \times \mathbb{K}^{n+1} \setminus \{0\} & \to & \mathbb{K}^n \\ & (A, x) & \mapsto & Ax \end{array}$$

which is a smooth mapping between two smooth manifolds. Since 0 is a regular value of ψ, its preimage $\psi^{-1}(0)$ is a smooth submanifold of $\mathcal{P}_{(k)} \times \mathbb{K}^{n+1} \setminus \{0\}$. Moreover, σ_u is the composition of the projection onto the first coordinate $\mathcal{W}_{(k)} \to \mathcal{P}_{(k)}$ and the function σ_u which is smooth by Proposition 11. To check that σ_u is self–convex in $\mathcal{W}_{(k)}$ we use Corollary 10 and proceed as in the proof of Proposition 13. Let $G = \mathbb{U}_n \times \mathbb{U}_{n+1}$, and consider the action

$$\begin{array}{ccc} G \times \mathcal{W}_{(k)} & \to & \mathcal{W}_{(k)} \\ ((U, V), (A, x)) & \mapsto & (UAV^*, Vx) \end{array}$$

Let $p = (\Sigma, e_{n+1})$ where $e_{n+1}^T = (0, \ldots, 0, 1)$ and $\Sigma \in \mathcal{D}_{(k)}$ has ordered distinct singular values $\sigma_1 > \cdots > \sigma_u > 0$. As in Propositions 11 and 13, we have

$$T_pG(p) = \{(B_1\Sigma + \Sigma B_2^*, B_2 e_{n+1}) : (B_1, B_2) \in \mathcal{A}_n \times \mathcal{A}_{n+1}\},$$

$$T_pG(p)^\perp = \{(\dot\Sigma, 0) : \dot\Sigma \in \mathcal{P}_{(k)}, \dot\Sigma \text{ is diagonal}, \dot\Sigma e_{n+1} = 0\}.$$

Note that $T_pG(p)^\perp$ is isometric to the set of diagonal $n \times n$ matrices with eigenvalues $\sigma_1 > \ldots > \sigma_u > 0$ of respective multiplicities $k_1, \ldots, k_u$.

Let us check the conditions of Corollary 10. By unitary invariance, we can choose a pair $p = (\Sigma, e_{n+1})$ as above.

1. $\mathrm{Hess}_\kappa \log(\alpha)(\Sigma, e_n)$ is positive semi-definite in $(T_pG(p))^\perp$: Let $(\dot\Sigma, 0) \in T_pG(p)^\perp$. Let γ be a condition geodesic in $T_pG(p)^\perp$ such that $\gamma(0) = (\Sigma, 0)$, $\dot\gamma(0) = (\dot\Sigma, 0)$. We have to check that

$$\frac{d^2}{dt^2} \log \alpha(\gamma(t)) \mid_{t=0} \geq 0.$$

This is true as α is log-convex in the set of diagonal $n \times n$ matrices with eigenvalues $\sigma_1 > \ldots > \sigma_u > 0$ from Proposition 11.

2. We have to check that for small enough t, and for

$$b = (\dot\Sigma, 0) \in T_pG(p)^\perp,$$

$D\phi_t(p)b$ is perpendicular to

$$T_{\phi_t(p)}G(\phi_t(p)),$$

where ϕ_t is the flow of $\mathrm{grad}_\kappa \alpha$ in $\mathcal{W}_{(k)}$. Now, as in the proof of Proposition 13, the operator grad preserves the diagonal form of (Σ, e_n) and hence $D\phi_t(p)b$ is of the form $(\Sigma', 0)$ where Σ' is diagonal with $\Sigma' e_n = 0$. In particular, it is orthogonal to $T_{\phi_t(p)}G(\phi_t(p))$. Thus, the second condition of Corollary 10 applies to our case.

3. For $(B_1, B_2) \in \mathcal{A}_n \times \mathcal{A}_m$, the vector field K on $\mathcal{W}_{(k)}$ generated by (B_1, B_2) is

$$K(A, x) = \frac{d}{dt}\left(e^{tB_1} A e^{tB_2^*}, e^{tB_2} x\right) \mid_{t=0} = (B_1 A + A B_2^*, B_2 x).$$

Note that

$$\|K(A, x)\|^2 = \|B_1 A + A B_2^*\|^2 + \|B_2 x\|^2.$$

Thus,

$$D(\|K\|^2)(A, x)(C, v) = \frac{d}{dt} \mid_{t=0} \left(\|K(A + tC, x + tv)\|^2\right) =$$

$$2\mathrm{Re}\langle B_1 B_1^* A + A B_2 B_2^* + 2 B_1^* A B_2^*, C\rangle + 2\mathrm{Re}\langle B_2^* B_2 x, v\rangle.$$

Moreover,

$$\text{grad } \alpha(\Sigma, e_{n+1}) = \left(-\frac{2}{k_u \sigma_u^2} E, 0\right) \text{ where } E = \text{diag } (0, \ldots, 0, \overbrace{1, \ldots, 1}^{k_u}).$$

Thus,

$$\alpha(\Sigma, e_{n+1}) D(\|K\|^2)(\Sigma, e_{n+1})(\text{grad } \alpha(\Sigma, e_{n+1})) + \|K(\Sigma, e_{n+1})\|^2 \|\text{grad } \alpha(\Sigma, e_{n+1})\|^2 =$$

$$\frac{2}{k_u \sigma_u^6} \left(-\sigma_u \text{Re}\langle B_1 B_1^* \Sigma + \Sigma B_2 B_2^* - 2B_1 \Sigma B_2^*, E\rangle + \|B_1 \Sigma - \Sigma B_2\|^2 + \|B_2 x\|^2\right).$$

This is positive from the proof of Proposition 13.

Hence, all the conditions of Corollary 10 are fulfilled and the Proposition follows. □

7.2 Proof of Theorem 23

Now we can prove Theorem 23 using Theorem 21 and Proposition 24. Note that we have $\mathcal{W} = \cup_{(k)} \mathcal{W}_{(k)}$ and α is smooth and self–convex in each $\mathcal{W}_{(k)}$ by Proposition 24. From Theorem 21 we just need to check that α is regular in $\mathcal{W}$ and that $\overline{\mathcal{SD}^2}\alpha > -\infty$. Since

$$\alpha = \sigma_n^{-2} \circ \pi_1,$$

where π_1 is the projection on the first coordinate, α is a smooth function. Now, consider the chart locally given by π_1^{-1}, and note that $\alpha \circ \pi_1^{-1} = \sigma_n^{-2}$ is regular in $\mathbb{GL}_{n,m}$ from the proof of Theorem 1. By definition, this means that α is regular in $\mathcal{W}_{(k)}$. Using the same argument, $\overline{\mathcal{SD}^2}\sigma_n^{-2} > -\infty$ in $\mathbb{GL}_{n,m}$ also implies that $\overline{\mathcal{SD}^2}\alpha > -\infty$ in $\mathcal{W}$ and we are done. □

8 Appendix

In this appendix we prove the following which gives a sufficient condition for the image of a submanifold under a group action to be a submanifold.

Lemma 25. *Let G be a Lie group acting on a smooth manifold $\mathcal{M}$, and $\mathcal{D}$ a smooth submanifold in $\mathcal{M}$. Define on $G \times \mathcal{D}$ the equivalence relation $(g, d)\mathcal{R}(g', d')$ when $gd = g'd'$. Let us denote by*

$$\pi : G \times \mathcal{D} \to (G \times \mathcal{D})/\mathcal{R}$$

the canonical surjection onto the quotient space, by i the map

$$i : (G \times \mathcal{D})/\mathcal{R} \to \mathcal{M}, \quad i(\pi(g, d)) = gd,$$

and by $\mathcal{P} = i((G\times\mathcal{D})/\mathcal{R})$ the image of i. When the three following conditions are satisfied

1. *The graph of $\mathcal{R}$ is a closed submanifold in $(G \times \mathcal{D}) \times (G \times \mathcal{D})$,*
2. *i is an immersion,*
3. *For every sequence $(x_k) \in (G \times \mathcal{D})/\mathcal{R}$ such that $(i(x_k))$ converges to $y \in \mathcal{P}$ the sequence (x_k) converges,*

then, $\mathcal{P}$ is a submanifold in $\mathcal{M}$.

Proof. Let $\mathcal{X}$ be a manifold and let $\mathcal{R}$ denote an equivalence relation defined on $\mathcal{X}$. A classical necessary and sufficient condition to define on the quotient space $\mathcal{X}/\mathcal{R}$ a unique quotient manifold structure making the canonical surjection $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{R}$ a submersion is the following: the graph $\mathcal{G}$ of the relation is a closed submanifold in $\mathcal{X}\times\mathcal{X}$ and the first projection $pr_1 : \mathcal{G} \to \mathcal{X}$ is a submersion.

In the context of our lemma this condition comes from the first hypothesis and from the definition of the equivalence relation via the group action.

Let $f : \mathcal{Y} \to \mathcal{Z}$ be a smooth map between two manifolds. Its image $f(\mathcal{Y})$ is a submanifold in $\mathcal{Z}$ when f is an immersion and a homeomorphism onto its image.

In the context of our lemma, i is smooth and injective from the construction, it is a homeomorphism by the third hypothesis and an immersion by the second one. □

References

[1] Beltrán C., J.-P. Dedieu, G. Malajovich, and M. Shub, *Convexity properties of the condition number.* http://arxiv.org/abs/0806.0395v2. To appear in SIMAX.

[2] Beltrán C., and M. Shub, *Complexity of Bézout's Theorem VII: Distances Estimates in the Condition Metric.* Foundations of Computational Mathematics, 9 (2009) 179-195.

[3] Boito P., and J.-P. Dedieu, *The condition metric in the space of full rank rectangular matrices.* To appear in SIMAX.

[4] Burkill J. C., *Integrals and trigonometric series*, Proc. London Math. Soc. 3 (1951) 46-57.

[5] Clarke F., *The Erdman condition and Hamiltonian inclusions in optimal control and the calculus of variations.* Can. J. Math. 32 (1980) pp 494-509.

[6] Clarke F. H., *Optimization and Nonsmooth Analysis.* Les Publications CRM (1989) ISBN 2-921120-01-1.

[7] Demmel J. W., *The probability that a Numerical Problem is Difficult.* Mathematics of Computation, 50 (1988) 449-480.

[8] do Carmo M. P., *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser Boston Inc., 1992.

[9] Foote R., *Regularity of the distance function*, Proceedings of the AMS, 92 (1984) pp 153-155.

[10] Gallot S., D. Hulin and J. Lafontaine, *Riemannian Geometry*, Springer, 2004.

[11] Gromov M., *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhuser, 1999.

[12] Jost J., *Riemannian geometry and geometric analysis*, fifth ed., Universitext, Springer-Verlag, Berlin, 2008.

[13] Li Y. and L. Nirenberg, *Regularity of the distance function to the boundary*, Rendiconti Accad. Naz. delle Sc. 123 (2005) pp 257-264.

[14] O,Neil B., *Semi-Riemannian Geometry.* Academic Press, 1983.

[15] Pugh C., *Lipschitz Riemann Structures*. Private communication, 2007.

[16] Schirotzek, W, *Nonsmooth analysis.* Universitext. Springer, Berlin, 2007.

[17] Shub M., *Complexity of Bézout's Theorem VI: Geodesics in the Condition Metric*. Foundations of Computational Mathematics, 9 (2009) 171-178.

[18] Shub, M. and S. Smale *Complexity of Bézout's Theorem I: Geometric Aspects* J. Am. Math. Soc. 6 (1993) 459-501.

[19] Shub, M. and S. Smale *Complexity of Bézout's Theorem II: Volumes and Probabilities* in: Computational Algebraic Geometry, Progress in Mathematics, F. Eyssette and A. Galligo editors, Birkhäuser (1993).

[20] Shub, M. and S. Smale *Complexity of Bézout's Theorem V: Polynomial Time* Theoretical Computer Science, 133 (1994) 141-164.

[21] Thomson B. S., *Symmetric properties of real functions*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 183, Marcel Dekker Inc., New York, 1994.

[22] Udriste, C., *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer (1994) ISBN 0-7923-3002-1.