Convexity properties of the condition number. *

Carlos Beltrán † Jean-Pierre Dedieu ‡
Gregorio Malajovich § Mike Shub

April 15, 2008

Abstract

In the space of $n \times m$ matrices of rank n, $n \leq m$, consider the "condition metric", obtained by multiplying the usual Frobenius Hermitian product by the inverse of the square of the smallest singular value. We prove that this last quantity is logarithmically convex along geodesics in that space. Let $\mathcal{N}$ be a complete submanifold of $\mathbb{R}^j$ and let $\mathbb{R}^j$ be endowed with the analogous "condition metric", obtained by multiplying the usual metric by the square of the inverse of the distance to $\mathcal{N}$. We prove that the distance to $\mathcal{N}$ is logarithmically convex along geodesics in that space.

---

*Mathematics Subject Classification (MSC2000): 65F35 (Primary), 15A12 (Secondary).

†C. Beltran, M. Shub, Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4 (`beltranc@math.toronto.edu`), (`shub@math.toronto.edu`). CB was supported by MTM2004-01167 and by a Spanish postdoctoral grant. CB and MS were supported by an NSERC Discovery Grant.

‡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.

§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`). GM is partially supported by CNPq grants 303565/2007-1 and 470031/2007-7, 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.

1 Introduction

In this paper we investigate the convexity properties of the condition number in certain spaces of matrices. We also study more general situations.

Let two integers $1 \leq n \leq m$ be given and let us denote by $\mathbb{GL}_{n,m}$ the space of matrices $A \in \mathbb{K}^{n\times m}$ with maximal rank : rank $A = n$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. The singular values of such matrices are denoted in decreasing order:

$$\sigma_1(A) \geq \ldots \geq \sigma_{n-1}(A) \geq \sigma_n(A) > 0.$$

The smallest singular value $\sigma_n(A)$ is a locally Lipschitz map in $\mathbb{GL}_{n,m}$. It is smooth on the open subset

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

This set is equipped with a structure induced by the Hermitian (inner) product of $\mathbb{K}^{n\times m}$,

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

which is invariant by linear isometries. In this paper we will also consider the following Riemannian structure:

$$\langle M, N\rangle_A = \sigma_n(A)^{-2}\text{Re}\,\langle M, N\rangle_F$$

where $M, N \in \mathbb{K}^{n\times m}$ and $A \in \mathbb{GL}_{n,m}^{>}$. We call this metric the *condition metric* (the number $\sigma_1(A)/\sigma_n(A)$ is the classical condition number of a rectangular matrix). One of our objectives is to study the behaviour of the condition number along the geodesics for the condition metric.

The interest of considering this metric comes from recent papers by Shub [5] and Beltrán-Shub [1] where these authors follow geodesics in the condition metric in certain incidence varieties to improve classical complexity bounds for solving systems of polynomial equations.

In this paper we investigate more deeply the linear case.

A minimizing geodesic in the condition metric $A(t)$, $a \leq t \leq b$, with given endpoints $A(a)$ and $A(b)$ minimizes the integral

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

in the set of absolutely continuous curves with the same endpoints. Thus, along such a curve, the (non-normalized) condition number $\sigma_n(A(t))^{-1}$ cannot be too big. In fact, we have obtained a much more precise result: the maximum of $\log\left(\sigma_n(A(t))^{-1}\right)$ along the geodesic is necessarily obtained at its endpoints; in other words this function is convex. Our first main theorem is:

Theorem 1. *σ_n^{-1} is logarithmically convex i.e. for any geodesic curve $\gamma(t)$ in $\mathbb{GL}_{n,m}^{>}$ for the condition metric the map $t \to \log\left(\sigma_n^{-1}(\gamma(t))\right)$ is convex.*

See Corollary 4 for a proof, and corollaries 5 and 6 for extension of this result to the sphere and projective space.

Problem 1. *Extend the metric in $\mathbb{GL}_{n,m}^{>}$ to $\mathbb{GL}_{n,m}$ by the same formula. Note that it is now only Lipschitz. Is Theorem 1 still true for $\mathbb{GL}_{n,m}$?*

In relation with this problem we might replace the nonsmooth $\sigma_n(A)^{-2}$ by $g(A) = \sigma_1(A)^{-2} + \cdots + \sigma_n(A)^{-2}$, which is a smooth function in $\mathbb{GL}_{n,m}$. In that case we would consider the Riemannian metric given by

$$\langle M, N\rangle_A = g(A)\langle M, N\rangle_F.$$

We will prove that, for this new metric, $g(A)$ is not geodesically convex. Hence, the metric defined by $\sigma_n(A)^{-2}$ seems special with respect to log-convexity.

In our second main theorem (theorem 3) we consider the homogeneous version of theorem 1 (see section 5 for precise statements). There is a natural analogue to Problem 1 in the context of Theorem 3.

Since $\sigma_n^{-1}(A)$ is equal to the inverse of the distance from A to the set of singular matrices (i.e. with non-maximal rank) a natural question is to ask whether our main result remains valid for the inverse of the distance from certain sets.

In our third main theorem we prove this property for the distance function to a complete C^2 submanifold without boundary $\mathcal{N} \subset \mathbb{R}^j$. Let us denote by

$$\rho(x) = d(x, \mathcal{N}) = \min_{y\in\mathcal{N}} \|x - y\| \text{ and } g(x) = \frac{1}{\rho(x)^2}.$$

Let $\mathcal{U}$ be the largest open set in $\mathbb{R}^j$ such that, for any $x \in \mathcal{U}$, there is a unique closest point in $\mathcal{N}$ to x. When $\mathcal{U}$ is equipped with the new metric $g(x)\langle .,.\rangle$ (called: condition metric) we have:

Theorem 2. *The function $g : \mathcal{U} \setminus \mathcal{N} \to \mathbb{R}$ is logarithmically convex i.e. for any geodesic curve $\gamma(t)$ in $\mathcal{U} \setminus \mathcal{N}$ for the condition metric the map $t \to \log g(\gamma(t))$ is convex.*

Notice that our first main theorem cannot be deduced from the second one because the set of matrices with non-maximal rank is not a submanifold.

Finally, Theorem 2 can be extended to the projective case.

Corollary 1. *Let $\mathcal{N}$ be a $\mathcal{C}^2$ complete submanifold without boundary of $\mathbb{P}(\mathbb{R}^j)$. Let $g(x) = d_{\mathbb{P}}(x, \mathcal{N})^{-2}$, where $d_{\mathbb{P}} = \sin d_R$ and d_R is the Riemannian distance in projective space. Let $\mathcal{U}$ be the largest open subset of $\mathbb{P}(\mathbb{R}^j)$ such that for $x \in \mathcal{U}$ there is a unique closest point from $\mathcal{N}$ to x. Then, $g : \mathcal{U} \setminus \mathcal{N} \to \mathbb{R}$ is self-convex (see Definition 3 below).*

2 Self-convexity

2.1 The definition of self-convexity

Let us first start to recall some basic definitions about convexity on Riemannian manifolds. A good reference on this subject is Udrişte [6].

Definition 1. *A function $f : C \subset \mathbb{R}^n \to \mathbb{R}$ defined on the convex set C is convex when $f((1-\theta)x + \theta y) \le (1-\theta)f(x) + \theta f(y)$ whenever x and $y \in C$ and $0 \le \theta \le 1$. When f has positive values we say that f is log-convex when $\log \circ f$ is convex.*

Log-convexity implies convexity and it is equivalent to the convexity of f^α for every $\alpha > 0$.

Let $\mathcal{M}$ be a Riemannian manifold. Let x and y be two points in $\mathcal{M}$; we denote by $\gamma_{xy} : [0,1] \to \mathcal{M}$ a geodesic arc in $\mathcal{M}$ joining x and y: $\gamma_{xy}(0) = x$ and $\gamma_{xy}(1) = y$. Such a geodesic arc is not necessarily unique.

Definition 2. *We say that a function $g : \mathcal{M} \to \mathbb{R}$ is convex (one also says: geodesically convex) whenever*

$$g(\gamma_{xy}(\theta)) \le (1-\theta)g(x) + \theta g(y)$$

for every $x, y \in \mathcal{M}$, for every geodesic arc γ_{xy} joigning x and y and $0 \le \theta \le 1$. When g has positive values we say that g is log-convex when $\log \circ g$ is convex.

The convexity of g in $\mathcal{M}$ is equivalent to the convexity of $g \circ \gamma_{xy}$ on $[0, 1]$ for every $x, y \in \mathcal{M}$ and arc γ_{xy} or also to the convexity of $g \circ \gamma$ on $[a, b]$ for every geodesic $\gamma : [a, b] \to \mathcal{M}$ ([6] Chap. 3, Th. 2.2).

Definition 3. *Let $\mathcal{M} = (\mathcal{M}, \langle \cdot, \cdot \rangle)$ be Riemannian and $g : \mathcal{M} \to \mathbb{R}$ a function of class C^2 with positive values. Let $\mathcal{M}' = (\mathcal{M}, \langle \cdot, \cdot \rangle')$ be the manifold $\mathcal{M}$ with the new metric*

$$\langle \cdot, \cdot \rangle'_x = g(x) \langle \cdot, \cdot \rangle_x.$$

We say that g is self-convex when it is log-convex on $\mathcal{M}'$.

For example, with $\mathcal{M} = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ x_n > 0\}$ equipped with the usual metric, $g(x) = x_n^{-2}$ is self-convex. The space $\mathcal{M}'$ is the Poincaré model of hyperbolic space.

2.2 Convexity and the second derivative

When $C \subset \mathbb{R}^n$ is convex and open and when f is C^2, the convexity of f is equivalent to $D^2 f(x) \geq 0$ (here ≥ 0 means positive semidefinite) for every $x \in C$ while log-convexity is equivalent to

$$f(x) D^2 f(x) - Df(x) \otimes Df(x) \geq 0$$

for every $x \in C$.

When g is a function of class C^2 in the Riemannian manifold $\mathcal{M}$, we define its second derivative $D^2 g(x)$ as the second covariant derivative. It is a symmetric bilinear form on $T_x\mathcal{M}$. Note ([6, Chapter 1]) that if $x \in \mathcal{M}$ and $\dot{x} \in T_x\mathcal{M}$, and if $\gamma = \gamma(t)$ is a geodesic in $\mathcal{M}$, $\gamma(0) = x, \dot{\gamma}(0) = \dot{x}$, then

$$D^2 g(x)(\dot{x}, \dot{x}) = (g \circ \gamma)''(0).$$

This second derivative depends on the Levi-Civita connection on $\mathcal{M}$. Since $\mathcal{M}$ is equipped with two different metrics: $\langle ., . \rangle_x$ and $\langle ., . \rangle'_x$ we have to distinguish between the corresponding second derivatives: they are denoted by $D^2 g(x)$ and $D^2 g(x)'$ respectively. No such distinction is necessary for the first derivative $Dg(x)$.

Convexity on Riemannian manifold is characterized by (see [6] Chap. 3, Th. 6.2):

Proposition 1. *A function $g : \mathcal{M} \to \mathbb{R}$ of class C^2 is convex if and only if $D^2 g(x)$ is positive semidefinite for every $x \in \mathcal{M}$.*

2.3 Characterization of self-convexity.

Proposition 2. *For a function $g : \mathcal{M} \to \mathbb{R}$ of class C^2 with positive values self-convexity is equivalent to*

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

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

Proof. Let $x \in \mathcal{M}$ be given. Let $\varphi : \mathbb{R}^m \to \mathcal{M}$ be a coordinate system such that $\varphi(0) = x$ and with first fundamental form $h_{ij}(0) = \delta_{ij}$ (Kronecker's delta) and Christoffel's symbols $\Gamma^i_{jk}(x) = 0$. Those coordinates are called "normal" or "geodesic". Note that this implies

$$\frac{\partial h_{ij}}{\partial z_k}(0) = 0$$

for all i, j, k. Let $\varphi' : \mathbb{R}^n \longrightarrow \mathcal{M}'$ be the coordinate system defined by $\varphi'(z) = \varphi(z)$ for all $z \in \mathbb{R}^m$. We denote by h'_{ij} and $(\Gamma^i_{jk})'$ respectively the first fundamental form and the Christoffel symbols for φ'. Let us compute them. Note that

$$h'_{ij}(z) = h_{ij}(z)g(\varphi(z)),$$

$$\frac{\partial h'_{ij}}{\partial z_k}(0) = D(h'_{ij})(0)(e_k) = D((g \circ \varphi)h_{ij})(0)(e_k) =$$

$$h_{ij}(0)D(g \circ \varphi)(0)(e_k) + g(x)Dh_{ij}(0)(e_k) = \delta_{ij}\frac{\partial(g \circ \varphi)}{\partial z_k}(0).$$

Moreover,

$$(\Gamma^i_{jk})' = \frac{1}{2g(x)}\left(\frac{\partial h'_{ij}}{\partial z_k}(0) + \frac{\partial h'_{ik}}{\partial z_j}(0) - \frac{\partial h'_{jk}}{\partial z_i}(0)\right) =$$

$$\frac{1}{2g(x)}\left(\delta_{ij}\frac{\partial(g \circ \varphi)}{\partial z_k}(0) + \delta_{ik}\frac{\partial(g \circ \varphi)}{\partial z_j}(0) - \delta_{jk}\frac{\partial(g \circ \varphi)}{\partial z_i}(0)\right).$$

That is,

$$\begin{cases} (\Gamma^i_{ik})' = (\Gamma^i_{ki})' = \frac{1}{2g(x)}\frac{\partial(g\circ\varphi)}{\partial z_k}(0) & \text{for all } i, k, \\ (\Gamma^i_{jj})' = \frac{-1}{2g(x)}\frac{\partial(g\circ\varphi)}{\partial z_i}(0), & j \neq i, \\ (\Gamma^i_{jk})' = 0 & \text{otherwise.} \end{cases}$$

The second derivative of the composition of two maps is given by the identity (see [6] Chap. 1-3)

$$D^2(\phi \circ g)(x) = \phi'(g(x))D^2g(x) + \phi''(g(x))Dg(x) \otimes Dg(x)$$

which gives in our context

$$D^2(\log \circ g)(x)' = \frac{1}{g(x)} D^2 g(x)' - \frac{1}{g(x)^2} Dg(x) \otimes Dg(x).$$

Our objective is now to give a necessary and sufficient condition for $D^2(\log \circ g)(x)'$ to be ≥ 0. Let us denote

$$G = g \circ \varphi.$$

In our system of local coordinates the components of $D^2 g(x)$ are (see [6] Chap. 1-3)

$$G_{jk} = \frac{\partial^2 G}{\partial x_j \partial x_k} - \sum_i \Gamma^i_{jk} \frac{\partial G}{\partial x_i} = \frac{\partial^2 G}{\partial x_j \partial x_k}$$

while the components of $D^2 g(x)'$ are

$$G'_{jk} = \frac{\partial^2 G}{\partial x_j \partial x_k} - \sum_i \left(\Gamma^i_{jk}\right)' \frac{\partial G}{\partial x_i}.$$

If we replace the Christoffel symbols in this last sum by the values previously computed we obtain, when $j = k$

$$\sum_i \left(\Gamma^i_{jj}\right)' \frac{\partial G}{\partial x_i} = \left(\Gamma^j_{jj}\right)' \frac{\partial G}{\partial x_j} + \sum_{i \neq j} \left(\Gamma^i_{jj}\right)' \frac{\partial G}{\partial x_i} =$$

$$\frac{1}{2g}\left(\frac{\partial G}{\partial x_j}\right)^2 - \frac{1}{2g}\sum_{i \neq j}\left(\frac{\partial G}{\partial x_i}\right)^2 = \frac{1}{g}\left(\frac{\partial G}{\partial x_j}\right)^2 - \frac{1}{2g}\sum_i \left(\frac{\partial G}{\partial x_i}\right)^2$$

while when $j \neq k$

$$\sum_i \left(\Gamma^i_{jk}\right)' \frac{\partial G}{\partial x_i} = \left(\Gamma^j_{jk}\right)' \frac{\partial G}{\partial x_j} + \left(\Gamma^k_{jk}\right)' \frac{\partial G}{\partial x_k} =$$

$$\frac{1}{2g}\frac{\partial G}{\partial x_k}\frac{\partial G}{\partial x_j} + \frac{1}{2g}\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k} = \frac{1}{g}\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k}.$$

Both cases are subsumed in the identity

$$\sum_i \left(\Gamma^i_{jk}\right)' \frac{\partial G}{\partial x_i} = \frac{1}{g}\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k} - \frac{\delta_{jk}}{2g}\sum_i \left(\frac{\partial G}{\partial x_i}\right)^2$$

with δ_{jk} the Kronecker symbol. Putting together all these identities gives the following expression for the components of $D^2(\log\circ g)(x)'$:

$$D^2(\log\circ g)(x)'_{jk} = \frac{1}{g}\left(\frac{\partial^2 G}{\partial x_j \partial x_k} - \frac{1}{g}\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k} + \frac{\delta_{jk}}{2g}\sum_i \left(\frac{\partial G}{\partial x_i}\right)^2\right) - \frac{1}{g^2}\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k} =$$

$$\frac{1}{2g^2}\left(2g\frac{\partial^2 G}{\partial x_j \partial x_k} + \delta_{jk}\sum_i \left(\frac{\partial G}{\partial x_i}\right)^2 - 4\frac{\partial G}{\partial x_j}\frac{\partial G}{\partial x_k}\right).$$

Thus, $D^2(\log\circ g)(x)' \geq 0$ if and only if $2g(x)D^2g(x) + \|Dg(x)\|_x^2 \langle .,.\rangle_x - 4Dg(x)\otimes Dg(x) \geq 0$ that is when

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

for any $x \in \mathcal{M}$ and for any vector $\dot{x} \in T_x\mathcal{M}$. This finishes the proof. □

Proposition 3. *The following condition is equivalent for a function $g = 1/\rho^2 : \mathcal{M} \longrightarrow \mathbb{R}$ to be self-convex on $\mathcal{M}$: For every $x \in \mathcal{M}$ and $\dot{x} \in T_x\mathcal{M}$,*

$$\|\dot{x}\|^2\|D\rho(x)\|^2 - (D\rho(x)\dot{x})^2 - \rho(x)D^2\rho(x)(\dot{x},\dot{x}) \geq 0,$$

or what is the same

$$2\|\dot{x}\|^2\|D\rho(x)\|^2 \geq D^2(\rho^2)(x)(\dot{x},\dot{x}).$$

Proof. Note that

$$Dg(x)\dot{x} = \frac{-2}{\rho(x)^3}D\rho(x)\dot{x},$$

$$D^2g(x)(\dot{x},\dot{x}) = \frac{6}{\rho(x)^4}(D\rho(x)\dot{x})^2 - \frac{2}{\rho(x)^3}D^2\rho(x)(\dot{x},\dot{x}).$$

Hence, the necessary and sufficient condition of Proposition 2 reads

$$\frac{4\|\dot{x}\|^2\|D\rho(x)\|^2}{\rho(x)^6} - \frac{16}{\rho(x)^6}(D\rho(x)\dot{x})^2 + \frac{12}{\rho(x)^6}(D\rho(x)\dot{x})^2 - \frac{4}{\rho(x)^5}D^2\rho(x)(\dot{x},\dot{x}) \geq 0,$$

and the proposition follows. □

Corollary 2. *Each of the following conditions is sufficient for a function $g = 1/\rho^2 : \mathcal{M} \longrightarrow \mathbb{R}$ to be self-convex at $x \in \mathcal{M}$: For every $\dot{x} \in T_x\mathcal{M}$,*

$$D^2\rho(x)(\dot{x}, \dot{x}) \leq 0,$$

$$\|D^2(\rho^2)(x)\| \leq 2\|D\rho(x)\|^2.$$

Proposition 4. *Let $\mathcal{M}$ be a $\mathcal{C}^2$ Riemannian manifold with metric $\langle\cdot,\cdot\rangle_x$ and $g : \mathcal{M} \to \mathbb{R}$ of class $\mathcal{C}^2$. Let $\mathcal{M}'$ be $\mathcal{M}$ with the new metric $\langle\cdot,\cdot\rangle'_x = g(x)\langle\cdot,\cdot\rangle_x$. Then, $g(x)$ is convex along geodesics in $\mathcal{M}'$ if and only if*

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

for any $x \in \mathcal{M}$ and any vector $\dot{x} \in T_x\mathcal{M}$.

Proof. We follow the lines of the proof of Proposition 2 with ϕ the identity map. □

3 Some general formulas for matrices

For a given matrix $B \in \mathbb{GL}^{>}_{n,m}$, we denote by $\sigma_1(B) \geq \ldots \geq \sigma_{n-1}(B) > \sigma_n(B) > 0$ its singular values.

Proposition 5. *Let $A = (\Sigma, 0) \in \mathbb{GL}^{>}_{n,m}$, where $\Sigma = diag\ (\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n > 0) \in \mathbb{K}^{n\times n}$ so that $\sigma_k(A) = \sigma_k$. Then, $\sigma_n : \mathbb{GL}^{>}_{n,m} \to \mathbb{R}$ is a smooth map and, for every $U \in \mathbb{K}^{n\times m}$,*

$$\begin{cases} D\sigma_n(A)U = \mathrm{Re}(u_{nn}), \\ D^2\sigma_n^2(A)(U,U) = 2\sum_{j=1}^m |u_{nj}|^2 - 2\sum_{k=1}^{n-1} \frac{|u_{kn}\sigma_n + \overline{u_{nk}}\sigma_k|^2}{\sigma_k^2 - \sigma_n^2}. \end{cases}$$

Proof. Since $\sigma_n^2 = \sigma_n^2(A)$ is an eigenvalue of AA^* with multiplicity 1, the implicit function theorem proves the existence of smooth functions $\sigma_n^2(B) \in \mathbb{R}$ and $u(B) \in \mathbb{K}^n$, defined in an open neighborhood of A and satisfying

$$\begin{cases} BB^*u(B) = \sigma_n^2(B)u(B), \\ \|u(B)\|^2 = 1, \\ u(A) = e_n = (0, \ldots, 0, 1)^T \in \mathbb{K}^n, \\ \sigma_n^2(A) = \sigma_n^2. \end{cases}$$

Differentiating these equations at B gives, for any $U \in \mathbb{K}^{n\times m}$,

$$\begin{cases} (UB^* + BU^*)u(B) + BB^*\dot{u}(B) = (\sigma_n^2)'\, u(B) + \sigma_n^2(B)\dot{u}(B), \\ u(B)^*\dot{u}(B) = 0 \end{cases}$$

with $\dot{u}(B) = Du(B)U$ and $(\sigma_n^2)' = D\sigma_n^2(B)U$. Pre-multiplying the first equation by $u^*(B)$ gives

$$u^*(B)(UB^*+BU^*)u(B)+u^*(B)BB^*\dot{u}(B) = \left(\sigma_n^2\right)' u^*(B)u(B)+\sigma_n^2(B)u^*(B)\dot{u}(B)$$

so that

$$D\sigma_n^2(B)U = \left(\sigma_n^2\right)' = 2\mathrm{Re}(u^*(B)UB^*u(B))$$

and

$$D\sigma_n(B)U = \frac{\mathrm{Re}(u^*(B)UB^*u(B))}{\sigma_n(B)}.$$

The derivative of the eigenvector is now easy to compute:

$$Du(B)U = \dot{u}(B) = (\sigma_n^2(B)I_n - BB^*)^\dagger(UB^* + BU^* - \left(\sigma_n^2\right)' I_n)u(B)$$

where $(\sigma_n^2(B)I_n - BB^*)^\dagger$ denotes the generalized inverse (or Moore-Penrose inverse) of $\sigma_n^2(B)I_n - BB^*$.

The second derivative of σ_n^2 at B is given by

$$D^2\sigma_n^2(B)(U,U) = 2\mathrm{Re}(\dot{u}(B)^*UB^*u(B)+u^*(B)UU^*u(B)+u(B)^*UB^*\dot{u}(B)) =$$
$$2\mathrm{Re}(u^*(B)UU^*u(B) + u(B)^*(UB^* + BU^*)\dot{u}(B)) = 2\mathrm{Re}(u^*(B)UU^*u(B)+$$
$$u(B)^*(UB^* + BU^*)(\sigma_n^2(B)I_n - BB^*)^\dagger(UB^* + BU^* - \left(\sigma_n^2\right)' I_n)u(B)).$$

Using $u(A) = e_n$ and $\sigma_n(A) = \sigma_n$ we get

$$\begin{cases} D\sigma_n^2(A)U = 2\mathrm{Re}(UA^*)_{nn} = 2\sigma_n\mathrm{Re}(u_{nn}), \\ D\sigma_n(A)U = \mathrm{Re}(u_{nn}), \end{cases}$$

and the second derivative is given by

$$D^2\sigma_n^2(A)(U,U) =$$
$$2\mathrm{Re}\left((UU^*)_{nn} + \sum_{k=1}^{n-1}(UA^* + AU^*)_{nk}(\sigma_n^2 - \sigma_k^2)^{-1}(UA^* + AU^* - \left(\sigma_n^2\right)' I_n)_{kn}\right) =$$
$$2\mathrm{Re}\left((UU^*)_{nn} + \sum_{k=1}^{n-1}\frac{\left|(UA^* + AU^*)_{kn}\right|^2}{\sigma_n^2 - \sigma_k^2}\right) = 2\sum_{j=1}^{m}\left|u_{nj}\right|^2 - 2\sum_{k=1}^{n-1}\frac{\left|u_{kn}\sigma_n + \overline{u_{nk}}\sigma_k\right|^2}{\sigma_k^2 - \sigma_n^2}.$$

□

Corollary 3. *Let* $A = (\Sigma, 0) \in \mathbb{GL}^{>}_{n,m}$*, where* $\Sigma = diag\,(\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n > 0) \in \mathbb{K}^{n\times n}$*. Let us define* $\rho(A) = \sigma_n(A)/\left\|A\right\|_F$*. Then, for any* $U \in \mathbb{K}^{n\times m}$ *such that* $\mathrm{Re}\left\langle A, U\right\rangle_F = 0$*, we have*

$$\begin{cases} D\rho(A)U = \frac{\mathrm{Re}(u_{nn})}{\|A\|_F}, \\ D^2\rho^2(A)(U,U) = \frac{2}{\|A\|_F^2}\left(\sum_{j=1}^m |u_{nj}|^2 - \sum_{k=1}^{n-1}\frac{|u_{kn}\sigma_n+\overline{u_{nk}}\sigma_k|^2}{\sigma_k^2-\sigma_n^2} - \frac{\|U\|_F^2}{\|A\|_F^2}\sigma_n^2\right). \end{cases}$$

Proof. Note that

$$D\rho(A)U = \frac{D\sigma_n(A)U\|A\|_F - \sigma_n(A)\frac{2\mathrm{Re}\langle A,U\rangle_F}{2\|A\|_F}}{\|A\|_F^2} = \frac{D\sigma_n(A)U}{\|A\|_F},$$

and the first assertion of the corollary follows from Proposition 5. For the second one, note that $h = \frac{h_1}{h_2}$ (for real valued $\mathcal{C}^2$ functions h, h_1, h_2 with $h_2(0) \neq 0$) implies

$$D^2h = \frac{h_2^2D^2h_1 - h_1h_2D^2h_2 - 2h_2Dh_1Dh_2 + 2h_1(Dh_2)^2}{h_2^3}.$$

Now, $\rho^2(A) = \sigma_n^2(A)/\|A\|_F^2$, $D(\|A\|_F^2)U = 2\mathrm{Re}\langle A, U\rangle_F = 0$, $D^2(\|A\|_F^2)(U,U) = 2\|U\|_F^2$, and $D^2\sigma_n^2(A)(U,U)$ is known from Proposition 5. The formula for $D^2\rho^2(A)$ follows after some elementary calculations. □

4 The affine linear case

We consider here the Riemannian manifold $\mathcal{M} = \mathbb{GL}^{>}_{n,m}$ equipped with the usual Frobenius Hermitian product. Let $g : \mathbb{GL}^{>}_{n,m} \to \mathbb{R}$ be defined as $g(A) = 1/\sigma_n^2(A)$.

Corollary 4. *The function* g *is self-convex in* $\mathbb{GL}^{>}_{n,m}$*.*

Proof. From Proposition 3, it suffices to see that

$$2\|U\|_F^2\|D\sigma_n(A)\|_F^2 \geq D^2\sigma_n^2(A)(U,U).$$

Since unitary transformations are isometries in $\mathbb{GL}^{>}_{n,m}$ we may suppose, via a singular value decomposition that $A = (\Sigma, 0) \in \mathbb{GL}^{>}_{n,m}$, where $\Sigma =$

diag $(\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n > 0) \in \mathbb{K}^{n\times n}$. Now, the inequality to verify is obvious from Proposition 5, as $\|D\sigma_n(A)\|_F = 1$ and

$$D^2\sigma_n^2(A)(U,U) = 2\sum_{j=1}^{m}|u_{nj}|^2 - 2\sum_{k=1}^{n-1}\frac{|u_{kn}\sigma_n + \overline{u_{nk}}\sigma_k|^2}{\sigma_k^2 - \sigma_n^2} \leq 2\sum_{j=1}^{m}|u_{nj}|^2 \leq 2\|U\|_F^2.$$

$\square$

Corollary 5. *Let $r > 0$. The function g is self-convex in the sphere $S^r_{\mathbb{GL}^>_{n,m}}$ of radius r in $\mathbb{GL}^>_{n,m}$.*

Proof. Let $S^r_{\mathbb{GL}^>_{n,m}}$ and $\mathbb{GL}^>_{n,m}$ be equipped with the condition metric. Note that for $r, r' > 0$ the mapping $S^r_{\mathbb{GL}^>_{n,m}} \to S^{r'}_{\mathbb{GL}^>_{n,m}}$, $x \mapsto r'x/r$ is an isometry. Hence, $\mathbb{GL}^>_{n,m}$ is isometric to the cylinder $S^r_{\mathbb{GL}^>_{n,m}} \times \mathbb{R}$ and the geodesics of $S^r_{\mathbb{GL}^>_{n,m}}$ are geodesics of $\mathbb{GL}^>_{n,m}$. Thus, the corollary follows from Corollary 4. $\square$

Proposition 6. *$g(A) = \sigma_1(A)^{-1} + \cdots + \sigma_n(A)^{-2}$ is not geodesically convex in $\mathbb{GL}_{n,m}$ for the metric $\langle M, N\rangle_A = g(A)\langle M, N\rangle_F$.*

Proof. For simplicity we consider the case of real square matrices. We have $g(X) = \|X^{-1}\|_F^2$,

$$Dg(X)\dot{X} = -2\langle X^{-1}, X^{-1}\dot{X}X^{-1}\rangle_F = -2\langle X^{-T}X^{-1}X^{-T}, \dot{X}\rangle_F,$$

$$\|Dg(X)\|_F^2 = 4\|X^{-T}X^{-1}X^{-T}\|_F^2,$$

$$D^2g(X)(\dot{X},\dot{X}) = 2\|X^{-1}\dot{X}X^{-1}\|_F^2 + 4\langle X^{-1}, X^{-1}\dot{X}X^{-1}\dot{X}X^{-1}\rangle_F.$$

According to Proposition 4, the geodesic convexity of $g(X)$ in $\mathbb{GL}'_n$ is equivalent to

$$2\|X^{-1}\|_F^2\left(2\|X^{-1}\dot{X}X^{-1}\|_F^2 + 4\langle X^{-1}, X^{-1}\dot{X}X^{-1}\dot{X}X^{-1}\rangle_F\right) +$$

$$4\|\dot{X}\|_F^2\|X^{-T}X^{-1}X^{-T}\|_F^2 - 8\langle X^{-1}, X^{-1}\dot{X}X^{-1}\rangle_F^2 \geq 0$$

This inequality is not satisfied when

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \text{ and } \dot{X} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

$\square$

5 The homogeneous linear case

5.1 The complex projective space.

The matter of this subsection is mainly taken from Gallot-Hulin-Lafontaine [3] sect. 2.A.5.

Let V be a Hermitian space of complex dimension $\dim_{\mathbb{C}} V = d+1$. We denote by $\mathbb{P}(V)$ the corresponding projective space that is the quotient of $V \setminus \{0\}$ by the group $\mathbb{C}^*$ of dilations of V; $\mathbb{P}(V)$ is equipped with its usual smooth manifold structure with complex dimension $\dim \mathbb{P}(V) = d$. We denote by p the canonical surjection.

Let V be considered as a real vector space of dimension $\dim_{\mathbb{R}} V = 2d+2$ equipped with the scalar product $\mathrm{Re}\,\langle .,. \rangle_V$. The sphere S_V is a submanifold in V of real dimension $2d+1$. This sphere being equipped with the induced metric becomes a Riemannian manifold and, as usual, we identify the tangent space at $z \in S_V$ with

$$T_z S_V = \{u \in V \ : \ \mathrm{Re}\,\langle u, z \rangle_V = 0\}.$$

The projective space $\mathbb{P}(V)$ can also be seen as the quotient S_V/S^1 of the unit sphere in V by the unit circle in $\mathbb{C}$ for the action given by $(\lambda, z) \in S^1 \times S_V \to \lambda z \in S_V$. The canonical map is denoted by

$$p_V : S_V \to \mathbb{P}(V).$$

p_V is the restriction of p to S_V.

The horizontal space at $z \in S_V$ related to p_V is defined as the (real) orthogonal complement of $\ker Dp_V(z)$ in $T_z S_V$. This horizontal space is denoted by H_z. Since V is decomposed in the (real) orthogonal sum

$$V = \mathbb{R}z \oplus \mathbb{R}iz \oplus z^{\perp}$$

and since $\ker Dp_V(z) = \mathbb{R}iz$ (the tangent space at z to the circle $S^1 z$) we get

$$H_z = z^{\perp} = \{u \in V \ : \ \langle u, z \rangle = 0\}.$$

There exists on $\mathbb{P}(V)$ a unique Riemannian metric such that p_V is a Riemannian submersion that is, p_V is a smooth submersion and, for any $z \in S_V$, $Dp_V(z)$ is an isometry between H_z and $T_{p(z)}\mathbb{P}(V)$. Thus, for this Riemannian structure, one has:

$$\langle Dp_V(z)u, Dp_V(z)v \rangle_{T_{p(z)}\mathbb{P}(V)} = \mathrm{Re}\,\langle u, v \rangle_V$$

for any $z \in S_V$ and $u, v \in H_z$.

Proposition 7. *Let $z \in S_V$ be given.*

1. *A chart at $p(z) \in \mathbb{P}(V)$ is defined by*
$$\varphi_z : H_z \to \mathbb{P}(V), \quad \varphi_z(u) = p(z+u).$$

2. *Its derivative at 0 is the restriction of $Dp(z)$ at H_z:*
$$D\varphi_z(0) = Dp(z) : H_z \to T_{p(z)}\mathbb{P}(V)$$
which is an isometry.

3. *For any smooth mapping $\psi : \mathbb{P}(V) \to \mathbb{R}$, and for any $v \in H_z$ we have*
$$D\psi(p(z))\,(Dp(z)v) = D(\psi \circ \varphi_z)(0)v$$
and
$$D^2\psi(p(z))(Dp(z)v, Dp(z)v) = D^2(\psi \circ \varphi_z)(0)(v,v).$$

Proof. 1 and 2 are easy. We have $D(\psi \circ \varphi_z)(0) = D\psi(p(z))D(\varphi_z)(0)$ which gives 3 since $D(\varphi_z)(0)v = Dp(z)v$ for any $v \in H_z$. For the second derivative, recall that $D^2\psi(p(z))(Dp(z)v, Dp(z)v) = (\psi \circ \tilde{\gamma})''(0)$, where $\tilde{\gamma}$ is a geodesic curve in $\mathbb{P}(V)$ such that $\tilde{\gamma}(0) = p(z), \tilde{\gamma}'(0) = Dp(z)v$. Now, consider the horizontal p_V-lift γ of $\tilde{\gamma}$ to S_V with base point z. Note that $\gamma(0) = z, \gamma'(0) = v$. Hence,
$$(\psi \circ \tilde{\gamma})''(0) = (\psi \circ p \circ \gamma)''(0) = D^2(\psi \circ p)(z)(v,v) + D\psi(p(z))Dp(z)\gamma''(0).$$
As $\gamma''(0)$ is orthogonal to T_zS_V, we have $Dp(z)\gamma''(0) = 0$. Finally,
$$D^2(\psi\circ p)(z)(v,v) = (\psi\circ p(z+tv))''(0) = (\psi\circ\varphi_z(tv))''(0) = D^2(\psi\circ\varphi_z)(0)(v,v),$$
and the assertion on the second derivative follows. □

The following result will be helpful.

Proposition 8. *Let $\mathcal{M}, \tilde{\mathcal{M}}$ be complete Riemannian manifolds and $\tilde{g} : \tilde{\mathcal{M}} \to [0,\infty)$ be of class $\mathcal{C}^2$. Let $\pi : \mathcal{M} \to \tilde{\mathcal{M}}$ be a Riemannian submersion. Let $\tilde{\mathcal{U}} \subseteq \tilde{\mathcal{M}}$ be an open set and assume that $g = \tilde{g}\circ\pi$ is self-convex in $\mathcal{U} = \pi^{-1}(\tilde{\mathcal{U}})$. Then, $\tilde{g}$ is self-convex in $\tilde{\mathcal{U}}$.*

Proof. Let $\mathcal{M}'$ be $\mathcal{M}$, but endowed with the condition metric given by g, and let $\tilde{\mathcal{M}}'$ be $\tilde{\mathcal{M}}$, but endowed with the condition metric given by $\tilde{g}$. Then, $\pi : \mathcal{M}' \to \tilde{\mathcal{M}}'$ is also a Riemannian submersion.

Now, let $\tilde{\gamma} : [a,b] \to \tilde{\mathcal{U}} \subseteq \tilde{\mathcal{M}}'$ be a geodesic, and let $\gamma \subseteq \mathcal{M}'$ be its horizontal lift by π. Then, γ is a geodesic in $\mathcal{U} \subseteq \mathcal{M}$ (see [3, Cor 2.109]) and hence $\log(g(\gamma(t)))$ is a convex function of t. Now,

$$\log(\tilde{g}(\tilde{\gamma}(t))) = \log(\tilde{g} \circ \pi(\gamma(t))) = \log(g(\gamma(t))),$$

so that $\tilde{g}$ is log-convex along $\tilde{\gamma}$, as wanted. $\square$

Corollary 6. *The function* $\tilde{g} : \mathbb{P}(\mathbb{GL}^{>}_{n,m}) \to \mathbb{R}$, $\tilde{g}(A) = \|A\|_F^2/\sigma_n^2(A)$ *is self-convex in* $\mathbb{P}(\mathbb{GL}^{>}_{n,m})$.

Proof. Note that $p : S_{\mathbb{GL}^{>}_{n,m}} \to \mathbb{P}(\mathbb{GL}^{>}_{n,m})$ is a Riemannian submersion and $\tilde{g} = g \circ p$ where g is as in Corollary 5. The corollary follows from Proposition 8. $\square$

5.2 The incidence variety.

Let us denote by p_1 and p_2 the canonical maps

$$S_1 \xrightarrow{p_1} \mathbb{P}\left(\mathbb{K}^{n\times(n+1)}\right) \quad \text{and} \quad S_2 \xrightarrow{p_2} \mathbb{P}\left(\mathbb{K}^{n+1}\right) = \mathbb{P}_n(\mathbb{K}),$$

where S_1 is the unit sphere in $\mathbb{K}^{n\times(n+1)}$ and S_2 is the unit sphere in $\mathbb{K}^{n+1}$. Consider the affine incidence variety,

$$\hat{\mathcal{W}}^{>} = \{(M,\zeta) \in S_1 \times S_2 \ : \ M \in \mathbb{GL}_{n,n+1} \text{ and } M\zeta = 0\}.$$

It is a Riemannian manifold equipped with the metric induced by the product metric on $\mathbb{K}^{n\times(n+1)} \times \mathbb{K}^{n+1}$. The tangent space to $\hat{\mathcal{W}}^{>}$ is given by

$$T_{(M,\zeta)}\hat{\mathcal{W}}^{>} = \left\{(\dot{M},\dot{\zeta}) \in T_M S_1 \times T_\zeta S_2 \ : \ \dot{M}\zeta + M\dot{\zeta} = 0\right\}.$$

The projective incidence variety considered here is

$$\mathcal{W}^{>} = \left\{(p_1(M), p_2(\zeta)) \in \mathbb{P}\left(\mathbb{K}^{n\times(n+1)}\right) \times \mathbb{P}_n(\mathbb{K}) \ : \ M \in \mathbb{GL}_{n,n+1} \text{ and } M\zeta = 0\right\},$$

that is also a Riemannian manifold equipped with the metric induced by the product metric on $\mathbb{P}\left(\mathbb{K}^{n\times(n+1)}\right) \times \mathbb{P}_n(\mathbb{K})$.

5.3 Self-convexity.

Let us denote by π_1 the restriction to $\hat{\mathcal{W}}^>$ of the first projection $S_1 \times S_2 \to S_1$, and by $R : \hat{\mathcal{W}}^> \to \mathbb{R}$, $R = \sigma_n \circ \pi_1$. We have

Lemma 1. *Let $w \in \hat{\mathcal{W}}^>$ and let γ be a geodesic in $\hat{\mathcal{W}}^>$, $\gamma(0) = w$. Then,*

$$D\sigma_n(\pi_1(w))(\pi_1 \circ \gamma)''(0) < 0.$$

Proof. Using unitary invariance we can take $M = (\Sigma, 0) \in \mathbb{GL}_{n,n+1}$, where $\Sigma = \text{diag}\ (\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n > 0) \in \mathbb{K}^{n\times n}$ and $\zeta = e_{n+1} = (0, \ldots, 0, 1)^T \in S_2$. As $\gamma = (M(t), \zeta(t))$ is a geodesic of $\hat{\mathcal{W}}^> \subseteq \mathbb{K}^{n\times(n+1)} \times \mathbb{K}^n$, $\gamma''(0)$ is orthogonal to $T_w\hat{\mathcal{W}}$, which contains all the pairs of the form $((A, 0), 0)$ where A is a $n \times n$ matrix, $\text{Re}\langle M, A\rangle = 0$. Hence, $M''(0)$ has the form

$$M''(0) = (a\Sigma, *),$$

for some real number $a \in \mathbb{R}$. Finally, $M(t)$ is contained in the sphere so

$$0 = (||M(t)||^2)''(0) = 2||M'(0)||^2 + 2\text{Re}\langle M(0), M''(0)\rangle = 2||M'(0)||^2 + 2a,$$

so that $a = -\|M'(0)\|^2$ and $(M''(0))_{nn} = -\|M'(0)\|^2\sigma_n$. From Proposition 5,

$$D\sigma_n(\pi_1(w))(\pi_1 \circ \gamma)''(0) = \text{Re}((\pi_1 \circ \gamma)''(0)_{nn}) = \text{Re}(M''(0))_{nn} < 0.$$

□

Theorem 3. *The map $g : \hat{\mathcal{W}}^> \to \mathbb{R}$ given by $g(M, \zeta) = 1/\sigma_n(M)^2$ is self-convex.*

Proof. Using unitary invariance we can take $M = (\Sigma, 0) \in \mathbb{GL}^>_{n,m}$, where $\Sigma = \text{diag}\ (\sigma_1 \geq \cdots \geq \sigma_{n-1} > \sigma_n > 0) \in \mathbb{K}^{n\times n}$ and $\zeta = e_{n+1} = (0, \ldots, 0, 1)^T \in S_2$. According to proposition 3 we have to prove that

$$2\, \|\dot{w}\|_w^2\, \|DR(w)\|^2 \geq D^2R^2(w)(\dot{w}, \dot{w})$$

for every $w \in \hat{\mathcal{W}}^>$ and $\dot{w} \in T_w\hat{\mathcal{W}}^>$. From Proposition 5 we have

$$DR(w)\dot{w} = D\sigma_n(\pi_1(w))(D\pi_1(w)\dot{w}) = \text{Re}(D\pi_1(w)\dot{w})_{nn},$$

so that $\|DR(w)\| = 1$. On the other hand, assume that $\dot{w} \neq 0$ and let γ be a geodesic in $\hat{\mathcal{W}}^{>}$, $\gamma(0) = w, \dot{\gamma}(0) = \dot{w}$. From Lemma 1,

$$D^2R^2(w)(\dot{w},\dot{w}) = (\sigma_n^2 \circ \pi_1 \circ \gamma)''(0) =$$
$$D^2\sigma_n^2(\pi_1(w))(D\pi_1(w)\dot{w}, D\pi_1(w)\dot{w}) + 2\sigma_n D\sigma_n(\pi_1(w))(\pi_1 \circ \gamma)''(0) <$$
$$D^2\sigma_n^2(\pi_1(w))(D\pi_1(w)(\dot{w}), D\pi_1(w)(\dot{w})).$$

Thus, we have to prove that for $\dot{y} \in \mathbb{K}^{n\times(n+1)}$,

$$2\left\|\dot{y}\right\|^2 \geq D^2\sigma_n^2(\pi_1(w))(\dot{y},\dot{y}).$$

which is a consequence of our Proposition 5. $\square$

Corollary 7. *The map $\tilde{g} : \mathcal{W}^{>} \to \mathbb{R}$ given by $\tilde{g}(M,\zeta) = \|M\|_F^2/\sigma_n^2(M)$ is self-convex.*

Proof. Consider the Riemannian submersion

$$p_1 \times p_2 : S_1 \times S_2 \longrightarrow \mathbb{P}\left(\mathbb{K}^{n\times(n+1)}\right) \times \mathbb{P}_n\left(\mathbb{K}\right),\ p_1 \times p_2(M,\zeta) = (p_1(M), p_2(\zeta)).$$

Note that $T_{(M,\zeta)}\hat{\mathcal{W}}^{>}$ contains the kernel of the derivative $d_{(M,\zeta)}(p_1 \times p_2)$. Thus, the restriction $p_1 \times p_2 : \hat{\mathcal{W}}^{>} \to \mathcal{W}^{>}$, is also a Riemannian submersion. The corollary follows combining Proposition 8 and Theorem 3. $\square$

6 Self-convexity of the distance from a submanifold

Let $\mathcal{N}$ be a complete C^k submanifold without boundary $\mathcal{N} \subset \mathbb{R}^j$, $k \geq 2$. Let us denote by

$$\rho(x) = d(x,\mathcal{N}) = \min_{y\in\mathcal{N}} \|x - y\|$$

the distance from $\mathcal{N}$ to $x \in \mathbb{R}^j$ (here $d(x,y) = \|x - y\|$ denotes the Euclidean distance). Let $\mathcal{U}$ be the largest open set in $\mathbb{R}^j$ such that, for any $x \in \mathcal{U}$, there is a unique closest point from $\mathcal{N}$ to x. This point is denoted by $K(x)$ so that we have a map defined by

$$K : \mathcal{U} \to \mathcal{N},\ \rho(x) = d(x, K(x)).$$

Classical properties of ρ and K are given in the following (see also Foote [2], Li and Nirenberg [4]).

Proposition 9. *1. ρ is $1-$Lipschitz on $\mathbb{R}^j$,*

2. *K is continuous on $\mathcal{U}$,*

3. *For any $x \in \mathcal{U}$, $x - K(x)$ is a vector normal to $\mathcal{N}$ at $K(x)$ i.e. $x - K(x) \in \left(T_{K(x)}\mathcal{N}\right)^{\perp}$,*

4. *K is C^{k-1} on $\mathcal{U}$,*

5. *ρ^2 is C^k on $\mathcal{U}$, $D\rho^2(x)\dot{x} = 2\langle x - K(x), \dot{x}\rangle$ and $D\rho^2(x)(\dot{x}, \dot{x}) = 2\|\dot{x}\|^2 - 2\langle DK(x)\dot{x}, \dot{x}\rangle$*

6. *ρ is C^k on $\mathcal{U} \setminus \mathcal{N}$,*

7. *$\langle DK(x)\dot{x}, \dot{x}\rangle \geq 0$ for every $x \in \mathcal{U}$ and $\dot{x} \in \mathbb{R}^j$.*

Proof. 1. For any x and y one has $\rho(x) = d(x, K(x)) \leq d(x, K(y)) \leq d(x, y) + d(y, K(y)) = d(x, y) + \rho(y)$. Since x and y play a symmetric role we get $|\rho(x) - \rho(y)| \leq d(x, y)$.

2. For any sequence $x_k \to x$ in $\mathcal{U}$ we have $d(K(x_k), x) \leq d(K(x_k), x_k) + d(x_k, x) = d(x_k, \mathcal{N}) + d(x_k, x) \leq d(x, \mathcal{N}) + 2d(x, x_k)$ so that the sequence $K(x_k)$ is bounded. Let $y \in \mathcal{N}$ be a limit point of $(K(x_k))$. From the last inequality we get $d(y, x) \leq d(K(x), x)$ so that $y = K(x)$. Thus $K(x_k)$ converges to $K(x)$.

3. This is the classical first order optimality condition in optimization.

4. This classical result may be derived from the inverse function theorem applied to the canonical map defined on the normal bundle to $\mathcal{N}$

$$\mathrm{can} : \mathrm{N}\mathcal{N} \to \mathbb{R}^j, \ \mathrm{can}(y, n) = y + n,$$

for every $y \in \mathcal{N}$ and $n \in N_y\mathcal{N} = (T_y\mathcal{N})^{\perp}$. The normal bundle is a C^{k-1} manifold, the canonical map is a C^{k-1} diffeomorphism when restricted to the set $\{(y, n) : y + tn \in \mathcal{U}, \ \forall\ 0 \leq t \leq 1\}$ and $K(x)$ is easily given from can^{-1}.

5. The derivative of ρ^2 is equal to $D\rho^2(x)\dot{x} = 2\langle x - K(x), \dot{x} - DK(x)\dot{x}\rangle = 2\langle x - K(x), \dot{x}\rangle$ because $DK(x)\dot{x} \in T_{K(x)}\mathcal{N}$ and $x - K(x) \in \left(T_{K(x)}\mathcal{N}\right)^{\perp}$. Thus $D\rho^2(x) = 2(x - K(x))$ is C^{k-1} on $\mathcal{U}$ so that ρ^2 is C^k. The formula for $D^2\rho^2$ follows.

6. Obvious.

7. Let $x(t)$ be a curve in $\mathcal{U}$ with $x(0) = x$. Let us denote $\frac{dx(t)}{dt} = \dot{x}(t)$, $\frac{d^2x(t)}{dt^2} = \ddot{x}(t)$, $y(t) = K(x(t))$, $\frac{dy(t)}{dt} = \dot{y}(t)$ and $\frac{d^2y(t)}{dt^2} = \ddot{y}(t)$. From the first order optimality condition we get

$$\langle x(t) - y(t), \dot{y}(t) \rangle = 0$$

whose derivative at $t = 0$ is

$$\langle \dot{x} - \dot{y}, \dot{y} \rangle + \langle x - y, \ddot{y} \rangle = 0.$$

Thus

$$\langle DK(x)\dot{x}, \dot{x} \rangle = \langle \dot{y}, \dot{x} \rangle = \langle \dot{y}, \dot{y} \rangle - \langle x - y, \ddot{y} \rangle .$$

This last quantity is equal to $\frac{1}{2}\frac{d^2}{dt^2} \|x - y(t)\|^2 \Big|_{t=0}$. It is nonnegative by the second order optimality condition.

□

Proof of Theorem 2 and Corollary 1 We are now able to prove our second main theorem. Let us denote $g(x) = 1/\rho(x)^2$. We shall prove that g is self-convex on $\mathcal{U}$. From proposition 3 it suffices to prove that, for every $\dot{x} \in \mathbb{R}^j$,

$$2\|\dot{x}\|^2\|D\rho(x)\|^2 \geq D^2(\rho^2)(x)(\dot{x}, \dot{x})$$

or, in other words, that

$$2\|\dot{x}\|^2 \geq 2\|\dot{x}\|^2 - 2\langle DK(x)\dot{x}, \dot{x} \rangle .$$

This is obvious from proposition 9-7.

Now we prove Corollary 1. Let S be the sphere of radius 1 in $\mathbb{R}^j$. As in the proof of Corollary 5, the mapping $1/\rho(x)^2$ is self-convex in the set $S \cap \mathcal{U}$. Now, apply Proposition 8 to the Riemannian submersion $p : S \to \mathbb{P}(\mathbb{R}^j)$ to conclude the corollary. □

References

[1] Beltrán C., and M. Shub, *Complexity of Bézout's Theorem VII: Distances Estimates in the Condition Metric.* Foundations of Computational Mathematics, to appear. DOI 10.1007/s10208-007-9018-5

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

[3] GALLOT S., D. HULIN AND J. LAFONTAINE, *Riemannian Geometry*, Springer, 2004.

[4] LI Y. AND L. NIRENBERG, *Regularity of the distance function to the boundary*, Rendiconti Accad. Naz. delle Sc. 123 (2005) pp 257-264.

[5] SHUB M., *Complexity of Bézout's Theorem VI: Geodesics in the Condition Metric.* Foundations of Computational Mathematics, to appear. DOI 10.1007/s10208-007-9017-6

[6] UDRISTE, C., *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer, 1994.