

EMBEDDED GEOMETRY OF THE SET OF SYMMETRIC POSITIVE SEMIDEFINITE MATRICES OF FIXED RANK

Bart Vandereycken¹, P.-A. Absil² and Stefan Vandewalle¹

¹ Department of Computer Science, Katholieke Universiteit Leuven, Belgium

² Department of Mathematical Engineering, Université catholique de Louvain, Belgium

ABSTRACT

This paper deals with the Riemannian geometry of the set of symmetric positive semidefinite matrices of fixed rank. This set is studied as an embedded submanifold of the real matrices equipped with the usual Euclidean metric. With this structure, we derive expressions of the tangent space and geodesics of the manifold, suitable for efficient numerical computations.

Index Terms— Symmetric positive semidefinite matrices, Riemannian geometry, geodesics, low rank

1. THE SET OF FIXED-RANK SYMMETRIC POSITIVE SEMIDEFINITE MATRICES

The set of symmetric positive definite matrices has a rich geometry which has been studied extensively in the literature. Computations with positive definite matrices consists of approximation, interpolation, estimation . . . They all depend on the chosen metric, where the Euclidean metric is perhaps the most intuitive thanks to its familiarity with $\mathbb{R}^{n \times n}$.

There is a recent interest for studying the set of symmetric positive-definite matrices of fixed rank; see, e.g., [1] and references therein. The rationale behind fixed rank matrices is that matrix algorithms are applied to problems of ever-increasing size. A matrix of fixed rank p then gives the possibility to lower the numerical burden considerably, typically from $O(n^3)$ to $O(np^c)$, with c small.

Let $p \leq n$ be positive integers. The focus of this paper is the embedded geometry of the set $S_+(p, n) \subset \mathbb{R}^{n \times n}$ of all real $n \times n$ symmetric positive semidefinite matrices of rank p . Let GL_n denote the general linear group, i.e., the set of all non-singular real $n \times n$ matrices. By $\mathbb{R}_*^{n \times p}$ we mean the set of all full-rank real $n \times p$ matrices. We have the following characterization of $S_+(p, n)$.

Proposition 1.1

$$S_+(p, n) = \{YY^T : Y \in \mathbb{R}_*^{n \times p}\}$$

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. B.V. is a Research Assistant of the Research Foundation – Flanders (FWO)

Proof. The inclusion $\{YY^T : Y \in \mathbb{R}_*^{n \times p}\} \subseteq S_+(p, n)$ is obvious. The inclusion $S_+(p, n) \subseteq \{YY^T : Y \in \mathbb{R}_*^{n \times p}\}$ can be shown using an eigenvalue decomposition. \square

2. EMBEDDED GEOMETRY

We will study the set $S_+(p, n)$ as a submanifold of $\mathbb{R}^{n \times n}$. Recall that an embedded (or regular) submanifold is a submanifold for which the inclusion map is a topological embedding, see e.g., [2].

Proposition 2.1 ([3, Ch. 5][4, Prop. 2.1]) *The set $S_+(p, n)$ is a smooth embedded submanifold of $\mathbb{R}^{n \times n}$ of dimension $pn - p(p - 1)/2$.*

Proof. Let us denote the $n \times n$ matrix

$$E = \begin{bmatrix} I_{p \times p} & 0_{p \times (n-p)} \\ 0_{(n-p) \times p} & 0_{(n-p) \times (n-p)} \end{bmatrix}.$$

Consider the transitive Lie group action

$$\Psi : GL_n \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n} : (M, N) \mapsto MNM^T, \quad (1)$$

then $S_+(p, n)$ is the orbit of E through Ψ . Since Ψ is a semi-algebraic action, all its orbits are smooth embedded submanifolds of $\mathbb{R}^{n^2} \simeq \mathbb{R}^{n \times n}$, see [5, App. B]. \square

2.1. Tangent space

Define the following surjective map

$$\psi : GL_n \rightarrow S_+(p, n), M \mapsto MEM^T := \Psi(M, E),$$

then the image of ψ is the orbit of E through Ψ of (1). The differential of ψ at M is given by

$$d\psi_M : T_M GL_n \rightarrow T_S S_+(p, n), \Delta \mapsto \Delta EM^T + ME \Delta^T.$$

Observe that the differential at arbitrary M is related to the differential at I by a full-rank linear transformation:

$$\begin{aligned} d\psi_M(\Delta M) &= \Delta MEM^T + MEM^T \Delta^T \\ &= M(M^{-1} \Delta ME + EM^T \Delta^T M^{-T})M^T \\ &= Md\psi_I(M^{-1} \Delta M)M^T. \end{aligned}$$

So the rank of ψ is constant, which makes ψ a submersion [6, Th. 7.15]. As a consequence, the range of $d\psi_M$ is the whole tangent space of $S_+(p, n)$ at $S := \psi(M)$. This gives

$$T_S S_+(p, n) = \{\Delta S + S \Delta^T : \Delta \in \mathbb{R}^{n \times n}\}. \quad (2)$$

Since the dimension of the tangent space is $pn - p(p-1)/2$, this is an over-parametrization. A minimal parametrization is given by the following proposition.

Proposition 2.2 *The tangent space at $S = YY^T$ is given by*

$$\begin{aligned} T_{YY^T} S_+(p, n) &= \{(YH + Y_\perp K)Y^T + Y(YH + Y_\perp K)^T\} \\ &= \left\{ \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} 2H & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} Y^T \\ Y_\perp^T \end{bmatrix} \right\} \end{aligned} \quad (3)$$

with $H = H^T \in \mathbb{R}^{p \times p}$, $K \in \mathbb{R}^{(n-p) \times p}$ and $Y_\perp \in \mathbb{R}^{n \times (n-p)}$ the orthogonal complement of Y in GL_n , i.e., $Y_\perp^T Y = 0$.

Proof. The right-hand side of (3) has the correct number of degrees of freedom, it is a linear space and it is included in $T_S S_+(p, n)$ (take $\Delta = (YH + Y_\perp K)(Y^T Y)^{-1} Y^T$). \square

2.2. Riemannian metric

Endow $\mathbb{R}^{n \times n}$ with its canonical metric, given by

$$\langle Z_1, Z_2 \rangle_S = \text{tr}(Z_1^T Z_2), \quad Z_1, Z_2 \in T_S \mathbb{R}^{n \times n} \simeq \mathbb{R}^{n \times n}. \quad (4)$$

This structure turns $S_+(p, n)$ into a Riemannian submanifold of $\mathbb{R}^{n \times n}$. The Riemannian metric g on $S_+(p, n)$ is given by

$$g_S(Z_1, Z_2) = \text{tr}(Z_1^T Z_2), \quad Z_1, Z_2 \in T_S S_+(p, n).$$

With the parametrization of $T_S S_+(p, n)$ given in Prop. 2.2, we have

$$\text{tr}(Z_1^T Z_2) = 2\text{tr}(Y^T Y (H_1 Y^T Y H_2 + K_1^T Y_\perp^T Y_\perp K_2)).$$

The factor $Y_\perp^T Y_\perp$ disappears when Y_\perp is chosen orthonormal.

2.3. Normal space and projections

The normal space at S consists of all matrices perpendicular (w.r.t. g_S) to the tangent space at S , i.e.,

$$T_S^\perp S_+(p, n) = \{Z \in \mathbb{R}^{n \times n} : \text{tr}(Z^T T) = 0, \forall T \in T_S S_+(p, n)\}. \quad (5)$$

Using the form (2) for the tangent vectors T , we can write the orthogonality constraint as

$$\text{tr}(Z^T T) = \text{tr}(Z^T \Delta S + Z^T S \Delta^T) = \text{tr}(\Delta S (Z^T + Z)).$$

This expression has to vanish for all $\Delta \in \mathbb{R}^{n \times n}$, so we see that the normal space must have the following form

$$T_S^\perp S_+(p, n) = \{Z \in \mathbb{R}^{n \times n} : S(Z^T + Z) = 0\}. \quad (6)$$

Again, we can simplify this for a factorized matrix.

Proposition 2.3 *The normal space at $S = YY^T$ is given by*

$$T_{YY^T}^\perp S_+(p, n) = \left\{ \begin{bmatrix} Y & Y_\perp \end{bmatrix} \begin{bmatrix} \Omega & -L^T \\ L & M \end{bmatrix} \begin{bmatrix} Y^T \\ Y_\perp^T \end{bmatrix} \right\} \quad (7)$$

with $\Omega = -\Omega^T \in \mathbb{R}^{p \times p}$, $M = \mathbb{R}^{(n-p) \times (n-p)}$ and $L \in \mathbb{R}^{(n-p) \times p}$. The dimension of the normal space is $n^2 - pn + p(p-1)/2$.

Proof. The right-hand side of (7) has the correct number of degrees of freedom $n^2 - \dim T_{YY^T} S_+(p, n)$, it is a linear space and it consists of matrices that are perpendicular to the tangent space at YY^T (easily seen by working out the factorized expressions). \square

Proposition 2.4 *The orthogonal projections on $T_S S_+(p, n)$ and $T_S^\perp S_+(p, n)$ are given by*

$$\begin{aligned} P_{YY^T}^t(Z) &= \frac{1}{2}(P_Y(Z + Z^T)P_Y + P_Y^\perp(Z + Z^T)P_Y \\ &\quad + P_Y(Z + Z^T)P_Y^\perp), \\ P_{YY^T}^n(Z) &= Z - P_{YY^T}^t(Z) \\ &= \frac{1}{2}(P_Y^\perp(Z + Z^T)P_Y^\perp + Z - Z^T), \end{aligned}$$

with $P_Y Z = Y(Y^T Y)^{-1} Y^T Z$ and $P_Y^\perp Z = Z - P_Y Z$.

Proof. Express that $Z - (YH + Y_\perp K)Y^T - Y(YH + Y_\perp K)^T$, with H and K constrained as in Prop. 2.2, belongs to the normal space (6). \square

2.4. Riemannian connection

The canonical choice for the connection is the Riemannian connection. Since $S_+(p, n)$ is embedded in $\mathbb{R}^{n \times n}$, this connection equals the classical directional derivative followed by an orthogonal projection onto the tangent space.

Proposition 2.5 [7, Prop. 5.3.2] *Let ξ be a vector field on $S_+(p, n)$, then the Riemannian connection ∇ on $S_+(p, n)$ is given by*

$$\nabla_{Z_S} \xi = P^t_S(D\xi(S)[Z_S])$$

for all $Z_S \in T_S S_+(p, n)$.

3. GEODESICS

The (maximal) geodesic from S_0 along \dot{S}_0 is the curve $t \mapsto S(t)$ on $S_+(p, n)$ that satisfies the ODE

$$\dot{S}(t) \in T_S S_+(p, n) \quad (8)$$

$$\ddot{S}(t) \in T_S^\perp S_+(p, n) \quad (9)$$

for t in the (maximal) interval (T_-, T_+) , with the initial conditions $S(0) = S_0, \dot{S}(0) = \dot{S}_0$.

3.1. Derivation of the ODE

To derive this ODE for $S(t)$ such that (8)–(9) are satisfied, we make the substitution $S(t) = Y(t)Y(t)^T$ with $Y(t) \in \mathbb{R}_*^{n \times p}$ and derive an ODE directly for $Y(t)$. Now $S(t)$ belongs to $S_+(p, n)$ by construction, hence $\dot{S} = \dot{Y}Y^T + Y\dot{Y}^T$ lies in $T_{Y Y^T} S_+(p, n)$ by construction. From this we can observe that it is sufficient to consider

$$\dot{Y}(t) = Y(t)H(t) + P_{Y(t)}^\perp Z(t),$$

with $H = H^T \in \mathbb{R}^{p \times p}$ and $Z \in \mathbb{R}^{n \times p}$, because the skew-symmetric part of H would cancel in the expression of \dot{S} .

We will impose condition (9) on $\ddot{S} = \ddot{Y}Y^T + Y\ddot{Y}^T + 2\dot{Y}\dot{Y}^T$ and see what this means for $Y(t)$. First working out

$$\begin{aligned} \frac{d}{dt}(P_{Y^\perp}^\perp Z) &= \frac{d}{dt}(Z - Y(Y^T Y)^{-1}Y^T Z) \\ &= P_{Y^\perp}^\perp \dot{Z} - \dot{Y}(Y^T Y)^{-1}Y^T Z - Y(Y^T Y)^{-1}\dot{Y}^T Z \\ &\quad + Y(Y^T Y)^{-1}[\dot{Y}^T Y + Y^T \dot{Y}](Y^T Y)^{-1}Y^T Z \\ &= P_{Y^\perp}^\perp \dot{Z} - P_{Y^\perp}^\perp Z(Y^T Y)^{-1}Y^T \dot{Z} - Y(Y^T Y)^{-1}Z^T P_{Y^\perp}^\perp \dot{Z}, \end{aligned}$$

we get

$$\begin{aligned} \ddot{Y} &= \dot{Y}H + Y\dot{H} + \frac{d}{dt}(P_{Y^\perp}^\perp Z) \\ &= YH^2 + P_{Y^\perp}^\perp ZH + Y\dot{H} + P_{Y^\perp}^\perp \dot{Z} \\ &\quad - P_{Y^\perp}^\perp Z(Y^T Y)^{-1}Y^T \dot{Z} - Y(Y^T Y)^{-1}Z^T P_{Y^\perp}^\perp \dot{Z}. \end{aligned}$$

Noting that \ddot{S} is symmetric by construction and projecting \ddot{S} onto the four subspaces that make up the tangent and normal spaces, we get

$$\ddot{S} = YX_{YY}Y^T + P_{Y^\perp}^\perp X_{\perp Y}Y^T + YX_{Y\perp}P_{Y^\perp}^\perp + P_{Y^\perp}^\perp X_{\perp\perp}P_{Y^\perp}^\perp$$

with

$$\begin{aligned} X_{YY} &= 4H^2 + 2\dot{H} \\ &\quad - (Y^T Y)^{-1}Z^T P_{Y^\perp}^\perp \dot{Z} - Z^T P_{Y^\perp}^\perp \dot{Z}(Y^T Y)^{-1} \\ X_{\perp Y} &= 3ZH + \dot{Z} - Z(Y^T Y)^{-1}Y^T \dot{Z} \\ X_{Y\perp} &= X_{\perp Y}^T \\ X_{\perp\perp} &= 2ZZ^T. \end{aligned}$$

In order that the acceleration \ddot{S} would belong to the normal space, it is sufficient that both $X_{YY} = 0$ and $X_{\perp Y} = 0$. Finally we arrive at the ODE for a geodesic $S = YY^T$.

Proposition 3.1 *The geodesic $S(t) = Y(t)Y(t)^T$ with foot $S(0) = Y_0 Y_0^T$ and direction $\dot{S}(0) = 2Y_0 H_0 Y_0^T + Z_0 Y_0^T + Y_0 Z_0^T$ satisfies the ODE*

$$\begin{aligned} \dot{Y} &= YH + P_{Y^\perp}^\perp Z, \\ \dot{H} &= -2H^2 + \frac{1}{2}(Y^T Y)^{-1}Z^T P_{Y^\perp}^\perp \dot{Z} + \frac{1}{2}Z^T P_{Y^\perp}^\perp \dot{Z}(Y^T Y)^{-1}, \\ \dot{Z} &= -3ZH + Z(Y^T Y)^{-1}Y^T \dot{Z}, \end{aligned}$$

with initial conditions $Y(0) = Y_0$, $H(0) = H_0$ and $Z(0) = \tilde{Z}_0$ where $P_{Y_0}^\perp \tilde{Z}_0 = Z_0$.

3.2. Analytical solution of a straight line

We did not find an analytical solution for this ODE for all initial conditions. However, in case $P_{Y(0)}^\perp Z(0) = 0$ the geodesic reduces to a straight line $S(t) = S_0 + t\dot{S}_0$ with $S_0 = Y_0 Y_0^T$ and $\dot{S}_0 = 2Y_0 H_0 Y_0^T$ for all t where $S(t)$ remains in $S_+(p, n)$. This can be seen in the equations. By assumption

$$S(t) = S_0 + t\dot{S}_0 = Y_0(I + 2tH_0)Y_0^T = Y(t)Y(t)^T,$$

and so $Y(t) := Y_0(I + 2tH_0)^{1/2}$ for all t where $I + 2tH_0 \succ 0$. Since H_0 commutes with all powers of $I + 2tH_0$, differentiating $Y(t)$ reduces to scalar differentiation. This gives,

$$\dot{Y}(t) = Y_0(I + 2tH_0)^{-1/2}H_0 = Y(t)(I + 2tH_0)^{-1}H_0$$

and so $H(t) := (I + 2tH_0)^{-1}H_0$. Finally,

$$\dot{H}(t) = -(I + 2tH_0)^{-2}(2H_0^2) = -2H^2(t)$$

and $Z(t) = 0$.

The geodesic $S(t)$ is only defined on the interval (T_-, T_+) where it remains of rank p . It is clear that this interval can be finite which shows that the geodesics are not complete. For the same reason, Euclidean geodesics in $S_+(n, n)$ are not complete, see [8].

3.3. Well-conditioned ODE

As we will see in section 3.4, the numerical integration of the ODE in Prop. 3.1 tends to fail due to blowup of Z . We will therefore modify the equation of motion for Z such that the resulting ODE is much better conditioned.

It is obvious from the ODE that only $P_{Y^\perp}^\perp Z$ has an influence on S and Y , so there is some freedom for Z as long as $P_{Y^\perp}^\perp Z$ remains the same. Indeed, replacing the equation of motion for Z by

$$\dot{Z} = -3ZH + Z(Y^T Y)^{-1}Y^T \dot{Z} + YM$$

with $M \in \mathbb{R}^{p \times p}$ arbitrary will still satisfy the conditions that both $P_{Y^\perp}^\perp X_{\perp Y}Y$ and X_{YY} are zero and this will give the same geodesic $S = YY^T$. We can use this to our benefit to keep the factor Z well behaved, that is orthogonal to Y .

Suppose Z is orthogonal to Y for all t where the geodesic exists, then $\langle Y(t), Z(t) \rangle = 0$. Differentiating gives

$$\langle \dot{Y}(t), Z(t) \rangle + \langle Y(t), \dot{Z}(t) \rangle = 0$$

Substituting \dot{Y} and the modified \dot{Z} , we get

$$\langle P_{Y(t)}^\perp Z(t), Z(t) \rangle + \langle Y(t), Y(t)M(t) \rangle = 0.$$

This reduces to

$$M(t) = -(Y(t)^T Y(t))^{-1}(Z(t)^T Z(t)).$$

Finally, we get the following well-conditioned ODE.

Proposition 3.2 *The geodesic $S(t) = Y(t)Y(t)^T$ with foot $S(0) = Y_0Y_0^T$ and direction $\dot{S}(0) = 2Y_0H_0Y_0^T + Z_0Y_0^T + Y_0Z_0^T$ satisfies the ODE*

$$\begin{aligned}\dot{Y} &= YH + Z, \\ \dot{H} &= -2H^2 + \frac{1}{2}(Y^TY)^{-1}(Z^TZ) + \frac{1}{2}(Z^TZ)(Y^TY)^{-1}, \\ \dot{Z} &= -3ZH - Y(Y^TY)^{-1}(Z^TZ),\end{aligned}$$

with initial conditions $Y(0) = Y_0$, $H(0) = H_0$ and $Z(0) = Z_0$ where $Z_0^TY_0 = 0$.

This ODE has the invariant that $Z(t)$ stays perpendicular to $Y(t)$ for all t where the geodesic exists.

It is possible to compute H using a smaller ODE than the one from Prop. 3.2. Introducing $A(t) = Y(t)^TY(t)$ and $B(t) = Z(t)^TZ(t)$, we see that $H(t)$ satisfies the ODE

$$\begin{aligned}\dot{A} &= AH + HA, \\ \dot{B} &= -3(BH + HB), \\ \dot{H} &= -2H^2 + \frac{1}{2}A^{-1}B + \frac{1}{2}BA^{-1}.\end{aligned}$$

Once the matrices $A(t)$, $B(t)$ and $H(t)$ are obtained, the geodesic satisfies a linear homogeneous ODE

$$\begin{bmatrix} \dot{Y} & \dot{Z} \end{bmatrix} = \begin{bmatrix} Y & Z \end{bmatrix} \begin{bmatrix} H & -A^{-1}B \\ I_p & -3H \end{bmatrix}.$$

3.4. Numerical example

In Fig. 1 we see some properties of the geodesic for

$$Y(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H(0) = -1, \quad Z(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The ODE from Prop. 3.1 (indicated with circles) as well as the ODE from Prop. 3.2 (in full line) were integrated with `ode45` from MATLAB. The absolute and relative error tolerances were 10^{-10} and 10^{-7} respectively.

The three unique elements of geodesic $S(t)$ are depicted in Fig. 1(a). Similarly, the two elements of $Z(t)$ are shown in Fig. 1(b). We can see that, at around $t = 1$, the integration of the ODE from Prop. 3.1 breaks down due to blowup of Z . Although the obtained geodesics S are the same for both ODE's, only the one from Prop. 3.2 can be integrated for all times.

By construction, the matrix $Z(t)$ in Prop. 3.2 should be orthogonal to $Y(t)$ for all t . This is indeed the case, as can be seen in Fig. 1(c), since $\|Z^TY\|_F \simeq 10^{-10}$, the absolute error tolerance for integration.

Fig. 1(d) verifies that the velocity of the geodesic, $\|\dot{S}\|_F$, is indeed constant. In addition, we see that, in the beginning, the geodesic is curved due to a non-vanishing acceleration $\|\ddot{S}\|_F$. For $t \rightarrow \infty$, $\ddot{S} = 0$ and the geodesic becomes a straight line. This behavior can be seen also in Fig. 1(a).

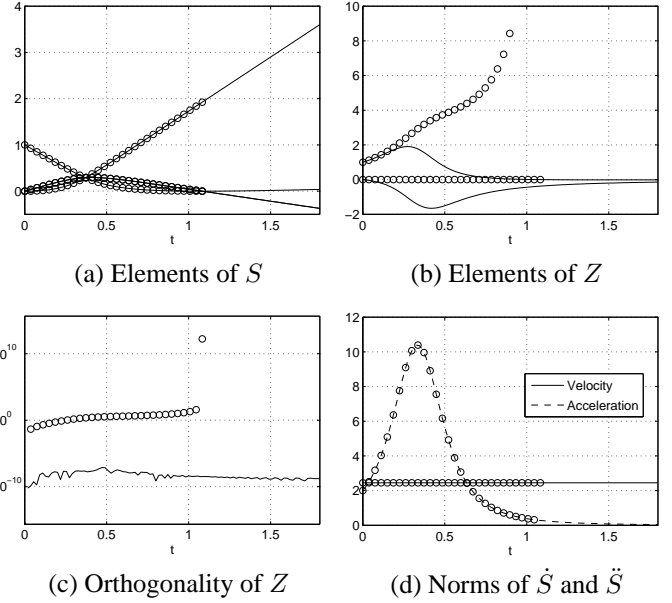


Fig. 1. The geodesic on $S_+(2, 1)$ from § 3.4.

4. REFERENCES

- [1] S. Bonnabel and R. Sepulchre, “Geometric distance and mean for positive semi-definite matrices of fixed rank,” Accepted for publication in *SIAM J. Matrix Anal. Appl.* <http://arxiv.org/abs/0807.4462>, 2008.
- [2] W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, second edition, 1986.
- [3] U. Helmke and J.B. Moore, *Optimization and Dynamical Systems*, Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994.
- [4] U. Helmke and M.A. Shayman, “Critical points of matrix least squares distance functions,” *Linear Algebra Appl.*, vol. 215, pp. 1–19, 1995.
- [5] C. G. Gibson, *Singular Points of Smooth Mappings*, Research Notes in Mathematics. Pitman, 1979.
- [6] John M. Lee, *Introduction to smooth manifolds*, vol. 218 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, 2003.
- [7] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, January 2008.
- [8] M. Moakher, “A differential geometric approach to the geometric mean of symmetric positive-definite matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.