A Majorization-Minimization Algorithm for the Karcher Mean of Positive Definite Matrices

Teng Zhang

February 3, 2014

Abstract

A majorization-minimization (MM) algorithm for the Karcher mean of \( n \times p \times p \) positive definite matrices is proposed and it is guaranteed to converge linearly. Simulations show that the MM algorithm performs faster than other current algorithms for the Karcher mean of positive definite matrices, including steepest descent, conjugate gradient descent and trust region methods.

1 Introduction

It is well-known that the geometric mean for a set of positive real numbers \((a_1, a_2, \cdots, a_n)\) is defined by \((a_1 a_2 \cdots a_n)^{\frac{1}{n}}\). However, this definition cannot be naturally generalized to the set of positive definite matrices \(A_1, A_2, \cdots, A_n \in \mathbb{R}^{p \times p}\), since \((A_1 A_2 \cdots A_n)^{\frac{1}{n}}\) is usually not symmetric. As for the geometric mean of positive definite matrices, the Karcher mean \([11, (6.24)]\) \([24, \text{Section 4}]\) defined as follows is commonly used:

\[
\hat{X} = \arg \min_{X} \text{dist}(X, A_i)^2 = \arg \min_{X} \sum_{i=1}^{n} \| \log(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) \|_{\text{F}}^2. \tag{1}
\]

Since \(\| \log(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) \|_{\text{F}}\) is the intrinsic distance between \(A_i\) and \(X\) on the manifold of positive definite matrices \([26, \text{pg} 326]\), the Karcher mean is the generalization of the Euclidean mean to Riemannian space.

As a manifold optimization problem, the algorithm for (1) has been investigated in many works. In particular, \([17, 30]\) apply gradient descent algorithms on general manifolds and (1) is treated as a special example. Since choosing step sizes in gradient descent methods by line search is expensive and heuristic methods of choosing step sizes leads to slow convergence, \([13]\) considers a linearization of gradient descent in the spirit of the Richardson iteration, and gives a criterion for choosing step sizes such that the algorithm is guaranteed to converge linearly to local minimum. \([21, 18]\) apply Newton’s algorithm to optimization on general manifolds, and the Karcher mean on manifolds is considered...
as a special example discussed under the name of “centroid computation”. [4] proposes gradient descent algorithm and stochastic algorithm for Riemannian $p$-means, and when $p = 2$ Riemannian $p$-means is equivalent to Karcher mean.

The most comprehensive work might be due to Jeuris et al. [21], which present a survey of various algorithms for Karcher mean of positive definite matrices, including first-order methods such as steepest descent method, conjugate descent method, and second-order methods such as trust region method and BFGS method.

A common issue of all current algorithms is the choice of step size in the update formula. As observed in [13], the line search strategy, which is commonly used, is computationally expensive. Another commonly used strategy of using constant step size lacks a theoretical guarantee on the convergence to the solution, unless the initialization is already close to the solution, see [2, Theorem 2.10] for gradient descent and [21, Theorem 5.2] for the case of Newton’s method. Besides, [13] observed that gradient descent with constant step size has slow empirical convergence in a large number of cases. There has been few attempts to improve the choice of step size. The only work in this direction is [13], which proposes a criterion of choosing step size in a modified gradient descent method. However this criterion only has the guarantee of convergence to local minimum.

The main contribution of this paper is to present and analyze a novel majorization-minimization (MM) method for the Karcher mean of positive definite matrices [1]. Compared to the previous methods based on manifold optimization, this algorithm fully explores the property of the positive definite matrices, and therefore it has many advantages. First, the implementation is simple and there is no need to choose step size. Second, the algorithm has theoretical guarantee on convergence for arbitrary initialization (unlike previous algorithms which require line search or good initialization). At last, without the computational burden of line search, the algorithm converges faster than current algorithms in simulation.

The rest of the paper is organized as follows. Section 2 includes backgrounds on the properties of Karcher mean and the general framework of maximization-minimization (MM) algorithm. Section 3 presents the algorithm and analyzes its convergence and convergence rate. Section 4 compares the proposed MM algorithm with other algorithms in simulations, which shows the advantage of MM algorithm in terms of speed and accuracy.

2 Background

2.1 Matrix Geometric Mean

In this section, we explain the reason that Karcher mean is commonly used as the geometric mean for positive definite matrices: it is computationally inexpensive and also satisfies the desired properties. We also present the existence and the uniqueness of Karcher’s mean, based on the geodesic convexity of the objective
function in \(1\).

In certain physical applications one need to represent a set of positive definite matrix \(A_1, A_2, \cdots, A_n\) by a single average matrix \(X\), and the arithmetic mean \(\frac{1}{n} \sum_{i=1}^{n} A_i\) is not suitable since one of the required property is that, \(X^{-1}\) should also be the mean of the set \(A_1^{-1}, A_2^{-1}, \cdots, A_n^{-1}\) (see example of elasticity calculations of structures in \[27\]).

The matrix geometric mean has also been used in radar signal processing \[8, 5, 9\], medical imaging \[10, 19, 28\], brain computer interface \[7, 6\] and image processing \[29\].

A list of “good properties” for matrix geometric mean has been proposed in \[3, Section 1\]. While Karcher mean satisfies all the properties \[24, Section 4\], there are other ways to define a mean such as the conditions are satisfied, such as ALM mean \[3\] and NBMP mean \[14\]; however they are more computationally expensive \[24, Section 3.5\]. A computationally inexpensive mean, CHEAP \[12\] mean, does not satisfy the desired “good properties” and there is even no guarantee on the convergence of the algorithm. In summary, among all candidates that satisfy “good properties”, Karcher mean is the least computationally expensive one.

One fundamental property of Karcher mean is its existence and uniqueness, and here we prove the uniqueness of the Karcher mean by the geodesic convexity of the objective function. That is, if we write the objective function in \(1\) as

\[
F(X) = \sum_{i=1}^{n} \| \log(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) \|_F^2,
\]

and use \(X_1 \# X_2\) to denote the geometric mean between \(X_1\) and \(X_2\):

\[
X_1 \# X_2 = X_1^{\frac{1}{2}} (X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}})^{\frac{1}{2}} X_1^{\frac{1}{2}},
\]

then applying \[11, (6.16)\], we have

\[
F(X_1) + F(X_2) > 2F(X_1 \# X_2). \tag{2}
\]

It follows that the solution to \(1\) is unique. Since \(F(X)\) goes to infinity as the eigenvalues of \(X\) goes to 0 or infinity, the existence of the solution to \(1\) is proved.

2.2 Background on majorization-minimization

Majorization-minimization (MM) is a principle of designing algorithm. While the name “MM” and the analysis is due to recent works by Hunter and Lange \[22, 23\], the idea seems to have a longer history and MM principle has been used independently and implicitly in other works. For example, the MM principle has been used in the analysis of Weiszfeld’s algorithm \[31\] for finding the geometric median \[25, Section 3.1\], and in the analysis of iterative reweighted least square (IRLS) algorithms for sparse recovery and matrix completion \[16, 20\]. Indeed, when the objective function \(f(x)\) is convex and has a bounded second derivative
τ, MM principle can also be used to prove the convergence gradient descent algorithm with constant step size \(1/τ\).

Now we present the MM principle in detail. Given a function \(f(x) : \mathcal{A} \rightarrow \mathbb{R}\), the MM principle find its minimizer through a surrogate function \(g(x, x') : \mathcal{A}^2 \rightarrow \mathbb{R}\) that satisfies the following properties:

\[
g(x', x') = f(x') \quad \text{and} \quad g(x, x') \geq f(x).
\]

Then we obtain a sequence \(\{x_k\}_{k \geq 1}\) by

\[
x_{k+1} = \arg \min_{x \in \mathcal{A}} g(x, x_k)
\]

and output \(\lim_{k \to \infty} x_k\) as the possible minimizer of \(f(x)\).

An important property of MM algorithm is that, the objective function \(f(x_k)\) does not increase. By applying (3) and (4),

\[
f(x_k) = g(x_k, x_k) \geq g(x_{k+1}, x_k) \geq f(x_{k+1}).
\]

Combining (5) with the properties of the function \(f(x)\) and \(g(x, x')\), one can usually prove stronger results on convergence of \(\{f(x_k)\}_{k \geq 1}\) or \(\{x_k\}_{k \geq 1}\). For example, the convergence of the objective function \(\{f(x_k)\}_{k \geq 1}\) is guaranteed if \(f(x)\) is bounded from below. In many cases (for example, in our case) one can show that \(x_k\) converges to the minimizer of \(f(x)\).

To design an MM algorithm, the important part is to find an appropriate surrogate function \(g(x, x')\). In particular, usually \(g(x, x')\) is chosen that \(\arg \min \ g(x, x')\) is computationally simple. A common choice of \(g(x, x')\) is the square function [25, 16, 20, Section 3.1] and therefore the solution to \(\arg \min \ g(x, x')\) is explicit. However, in our case we will use a different surrogate function, which allows us to use the properties of the manifold of positive definite matrices.

3 Our algorithm

We first present our majorization-minimization (MM) algorithm:

\[
X_{k+1} = f_2(X_k) - f_1(X_k) = f_1(X_k)^{\frac{1}{2}} (f_1(X_k)^{\frac{1}{2}} f_2(X_k) f_1(X_k)^{\frac{1}{2}})^{-\frac{1}{2}} f_1(X_k)^{\frac{1}{2}}.
\]

where

\[
f_1(X) = \sum_{i=1}^{n} A_i^{\frac{1}{2}} \left( \sqrt{\log^2(A_i^{-\frac{1}{2}} A_i^{-\frac{1}{2}}) + 1} - \log(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) \right) A_i^{-\frac{1}{2}} X,
\]

\[
f_2(X) = \sum_{i=1}^{n} A_i^{\frac{1}{2}} \left( \sqrt{\log^2(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) + 1} + \log(A_i^{-\frac{1}{2}} X A_i^{-\frac{1}{2}}) \right) A_i^{\frac{1}{2}} X^{-1}.
\]

For this algorithm we have the convergence guarantee as follows:
Theorem 3.1. The sequence \( \{X_k\}_{k \geq 1} \) generated by (6) converges to the solution of (1), and the objective function \( \{F(X_k)\}_{k \geq 1} \) converges linearly.

The main idea in the proof is based on the interpretation of (6) as a MM algorithm. That is, we need to show that there exists a surrogate function \( G(X, X') \) such that (3) is satisfied by \( G(X, X') \) and \( F(X) \), and \( X_{k+1} = \arg\min_X G(X, X_k) \). We will prove these two conditions in Lemmas 3.2 and 3.3 for \( G(X, X') \) defined by

\[
G(X, X') = \langle f_2(X'), X \rangle + \langle f_1(X'), X^{-1} \rangle + c_0(X'),
\]

where \( c_0(X') \) is a number depending on \( X' \) and will be specified later in (14).

The proofs of Lemmas are presented in Sections 3.1 and 3.2.

Lemma 3.2. \( G(X, X') \) and \( F(X) \) satisfy the conditions specified in (3):

\[
G(X', X') = F(X') \quad \text{and} \quad G(X, X') \geq F(X).
\]

Lemma 3.3. The minimizer of \( \langle C_1, X \rangle + \langle C_2, X^{-1} \rangle \) is explicitly given by

\[
C_1^{-1} \# C_2 = C_2^{-\frac{1}{2}} (C_2^{-\frac{1}{2}} C_1 C_2^{-\frac{1}{2}})^{-\frac{1}{2}} C_2^{-\frac{1}{2}}.
\]

Given the interpretation of (6) as a MM algorithm, the outline of the proof of Theorem 3.1 is as follows. Following the property (5) of MM algorithm, \( F(X_k) \) is a nonincreasing sequence. Since \( F(X) \) is bounded below by 0, \( F(X_k) \) converges. Then Theorem 3.1 can be proved by combining the convergence of \( F(X_k) \) and the properties of \( G(X, X') \) and \( F(X) \), such as continuity and differentiability. The full proof will be given in Section 3.3.

3.1 Proof of Lemma 3.2

We start the proof by the following lemma:

Lemma 3.4. \( X' \) is the unique minimizer for

\[
g_{X'}(X) = \left( \sqrt{\log^2 X' + 1 + \log X'} \right) X'^{-1} + \left( \sqrt{\log^2 X' + 1 - \log X'} \right) X^{-1} - \| \log X \|_F^2.
\]

Proof. For the simplicity of the proof we write \( Y = \left( \sqrt{\log^2 X' + 1 + \log X'} \right) X'^{-1} \), and notice that \( \left( \sqrt{\log^2 X' + 1 - \log X'} \right) X' = Y^{-1} \).

Any minimizer of \( g_{X'}(X) \) must satisfies that \( \frac{d}{dX} g_{X'}(X) = 0 \), which is

\[
Y - X^{-1} Y^{-1} X^{-1} - 2 X^{-1} \log X = 0.
\]

When \( X \) is fixed, then

\[
Y - X^{-1} Y^{-1} X^{-1} - 2 X^{-1} \log X = \frac{d}{dY^{-1}} h(Y^{-1}),
\]

5
where \( h(Y) = \log \det(Y) - \frac{1}{2}\|YX^{-1}\|_F^2 - \text{tr}(2X^{-1}\log XY) \) is a concave function with respect to \( Y \). Therefore there is at most one \( Y \) that satisfies (11).

By direct calculation, it is easy to verify that the unique solution \( Y = g_0(X) \), where
\[
g_0(x) = \frac{\log x + \sqrt{\log x^2 + 1}}{x}. \tag{12}
\]

Since \((1-y)\sqrt{y^2 + 1} + (y-y^2)(y^2 - 1) = -y^2 < 0 \) (with equality holds only when \( y = 0 \)) and \( y - y^2 - 1 < 0 \),
\[
(1-y)\sqrt{y^2 + 1} + y - y^2 - 1 \leq 0, \text{ and equality holds only when } y = 0.
\]

Therefore, \( g_0(x)' = 1 - \log x + \frac{\log x}{\sqrt{\log x^2 + 1}} - \sqrt{\log x^2 + 1} \leq 0 \), \( g_0(x)' = 0 \) only when \( x = 1 \), and we proved that \( g_0(x) \) is monotonically decreasing and \( g_0^{-1}(x) \) exists.

Now we assume that \( Y \) is fixed, and applying \( Y = g_0(X') \) and (12), then the solution to (11) is \( X = g_0^{-1}(Y) = X' \).

Since the minimizer of \( g_X(X) \) exists \( g_X(X) \) tends to infinity when eigenvalues of \( X \) tends to zero of infinity) and any minimizer of \( g_X(X) \) satisfies (11), \( X' \) is the unique minimizer of \( g_X'(X) \).

Now we are ready to prove Lemma 3.2. Applying Lemma 3.4, we have
\[
g_X(X) - g_X(X') \geq 0, \quad g_X(X) - g_X(X') = 0. \tag{13}
\]

Replace \( X, X' \) in (13) by \( A_i^{-\frac{1}{2}}X_i A_i^{-\frac{1}{2}}, A_i^{-\frac{1}{2}}X_i' A_i^{-\frac{1}{2}} \), and summing it over \( 1 \leq i \leq n \), we obtain (8) with \( c_0(X') \) in (7) defined by
\[
c_0(X') = -\sum_{i=1}^{n} g_{A_i^{-\frac{1}{2}}X_i' A_i^{-\frac{1}{2}}} (A_i^{-\frac{1}{2}}X_i' A_i^{-\frac{1}{2}}). \tag{14}
\]

### 3.2 Proof of Lemma 3.3
Since \( X^{-1} \) is operator convex [15, Theorem 2.6], i.e.,
\[
(X + Y)^{-1} + (X - Y)^{-1} - 2X^{-1}
\]
is positive definite, \( \langle C_2, X^{-1} \rangle \) is convex:
\[
\langle C_2, (X + Y)^{-1} \rangle + \langle C_2, (X - Y)^{-1} \rangle = \langle C_2, (X + Y)^{-1} + (X - Y)^{-1} \rangle \geq 2 \langle C_2, X^{-1} \rangle. \tag{15}
\]

Therefore \( \langle C_1, X \rangle + \langle C_2, X^{-1} \rangle \) is also convex and the unique minimizer is root of its derivative:
\[
C_1 - X^{-1}C_2X^{-1} = 0.
\]

Lemma 3.3 is then proved by verifying that \( X = C_2^{\frac{1}{2}}(C_2^{\frac{1}{2}}C_1C_2^{\frac{1}{2}})^{-\frac{1}{2}}C_2^{\frac{1}{2}} \) satisfies the equation above.
3.3 Proof of Theorem 3.1

Proof. For simplicity we denote the algorithm by $X_{k+1} = T(X_k)$, where $T : \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ is the operation defined by (6).

Since $\{X_k\}_{k \geq 1}$ are bounded in the bounded set $\{X : F(X) \leq F(X_1)\}$, any subsequence of $\{X_k\}_{k \geq 1}$ contains a subsubsequence of $\{X_k\}_{k \geq 1}$ that converging to some positive definite matrix $\tilde{X}$. Then we have

$$F(T(\tilde{X})) = F(\tilde{X}).$$

If $\tilde{X} \neq \arg \min_{X} f(X)$, then $F'(\tilde{X}) \neq 0$ (by geodesic convexity, any stationary point is the global minimizer). Due to (4) and the continuity of $F$ and $G$, $\left. \frac{d}{dt} G(X, \tilde{X}) \right|_{t=\hat{t}} = F'(\tilde{X}) \neq 0$. Since the second derivative of $G(X, X')$ is well defined, $G(X - \eta F'(\tilde{X}), \tilde{X}) < G(X, \tilde{X})$ for some small $\eta$, and therefore $F(T(\tilde{X})) = \min_{X} G(X, \tilde{X}) \leq G(\tilde{X} - \eta F'(\tilde{X}), \tilde{X}) < G(\tilde{X}, \tilde{X}) = F(\tilde{X})$ is a contradiction to (16).

Since any subsequence of $\{X_k\}_{k \geq 1}$ contains a subsubsequence of $\{X_k\}_{k \geq 1}$ that converging to $\tilde{X}$, the minimizer of $\{1\}, \{X_k\}_{k \geq 1}$ converges to $\tilde{X}$.

Now we will prove the linear convergence of $\{F(X_k)\}_{k \geq 1}$. To prove it, assume that the second derivative of $G(X, X')$ for $X, X'$ in a local neighborhood of $\tilde{X}$, $B(\tilde{X}, \varepsilon)$, is bounded above by $b$, and this neighborhood contain the sequence $\{X_k\}_{k \geq K}$. Then for $k > K$, in each minimization step $F(X_k)$ is reduced by $G(X_k, X_k) - \arg \min_{X} G(X, X_k)$ which is at least

$$\left\| G'(X, X_k) \right\|_{X=X_k}^2 / 2b = \| F'(X_k) \|^2 / 2b.$$

That is,

$$F(X_{k+1}) \leq F(X_k) - \| F'(X_k) \|^2 / 2b.$$  \hspace{1cm} (17)

Since in the small neighborhood $B(\tilde{X}, \varepsilon)$, $F(X)$ is twice-differentiable and continuous, $F''(X)$ can be bounded above by a positive definite matrix $c_1 I$. Therefore

$$\| F'(X_k) \| < \sqrt{\frac{F(X_k) - F(\tilde{X})}{c_1}},$$

and therefore the combination of (17) and (18) shows that $\{F(X_k)\}_{k \geq 1}$ converge linearly. \hfill \Box

4 Simulations

In the simulation we compare MM algorithm (6) with algorithms from [24, Section 4], where the paper surveyed steepest descent method (SD), conjugate gradient method (CG), Riemannian BFGS method (RBFGS), trust region method by exact Hessian (TR), trust region method with Hessian by decomposition (TR-DECOMP), trust region method with Hessian by approximation (TR-APPROX). By considering the Euclidean metric in general matrices $\mathbb{R}^{p \times p}$ or the metric of positive definite matrices, each algorithm comes
with two versions denoted by ‘SYMM’ or ‘SPD’. We use the code available at [http://people.cs.kuleuven.be/~raf.vandebril/]. We did not include Newton’s method since trust region method is understood as an enhancement of Newton’s method [1, page 136].

We compare these algorithms with for three cases: (a) 40 5 × 5 random positive definite matrices, (b) 40 7 × 7 random positive definite matrices, (c) 40 10 × 10 random positive definite matrices. We initialize all algorithms with the same randomly chosen matrix. As for the measurement of accuracy, we use

\[
\max \left( \left\| \frac{d}{dX} F(X) \right\|_F, \left\| \frac{d}{dX^{-1}} F(X) \right\|_F \right)
\]

Indeed, the measure \( \| \frac{d}{dX} F(X) \|_F \) used in [24, Figure 4.6(c)] is problematic: when the magnitude \( X \) is arbitrarily large, the gradient \( 2 \left\| \sum_{i=1}^{n} X^{-1} \log(\frac{X}{A_i^{-1}}) \right\|_F \) converges to 0, but obviously \( X \) is not even close to the Karcher mean.

The performance of these algorithms are recorded in Figures 1. The performance of MM algorithm corresponds to the estimation of convergence rate in Theorem 3.1. Besides, Figures 1 shows that MM algorithm performs better than other algorithms, and the advantage over other first order methods (steepest descent, conjugate gradient) is very significant. As for second-order methods including trust region methods and BFGS methods, the closest competitor is Riemannian BFGS-SPD method, where the quadratic convergence (which is the advantage of second-order methods) is observed in the cases of 5 × 5 and 7 × 7 matrices. However, its convergence is not as fast as MM in the first few iterations. Besides, as shown in the case of 10 × 10 matrices and also remarked in [24, Section 4.3.5], second order methods usually do not perform as well for larger matrices as for small matrices. In conclusion, MM algorithm wins since it is fast in each iteration and it does not require line search.

5 Conclusion

This paper has presented a novel algorithm for the Karcher mean of positive definite matrices, based on the majorization-minimization (MM) principle. The proposed MM algorithm is simple to implement and performs faster than other current algorithms, and has a linear convergence guarantee. The advantage of MM algorithm is that, it used the property of positive definite matrices while designing the surrogate function. By contrast, other current algorithms such as gradient descent or Newton’s method are only generalizations of their Euclidean counterparts. With the superior performance of MM algorithm for Karcher mean on the space of positive definite matrices, a natural next direction is to extend MM algorithms to other optimization problems on manifold.
Figure 1: The performance of algorithms for matrix geometric mean, where \( x \)-axis is the running time and \( y \)-axis is the accuracy.
References

[1] P. Absil, R. Mahony, and R. Sepulchre. Optimization Algorithms on Matrix Manifolds. Princeton University Press, 2009.

[2] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the riemannian center of mass. SIAM Journal on Control and Optimization, 51(3):2230–2260, 2013.

[3] T. Ando, C.-K. Li, and R. Mathias. Geometric means. Linear Algebra and its Applications, 385(0):305 – 334, 2004. Special Issue in honor of Peter Lancaster.

[4] M. Arnaudon, F. Barbaresco, and L. Yang. Medians and means in riemannian geometry: Existence, uniqueness and computation. In F. Nielsen and R. Bhatia, editors, Matrix Information Geometry, pages 169–197. Springer Berlin Heidelberg, 2013.

[5] M. Arnaudon, F. Barbaresco, and L. Yang. Riemannian medians and means with applications to radar signal processing. Selected Topics in Signal Processing, IEEE Journal of, 7(4):595–604, 2013.

[6] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Riemannian geometry applied to bci classification. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, editors, Latent Variable Analysis and Signal Separation, volume 6365 of Lecture Notes in Computer Science, pages 629–636. Springer Berlin Heidelberg, 2010.

[7] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. A Brain-Switch using Riemannian Geometry. In Proceedings of the 5th International BCI Conference 2011, pages 64–67, Graz, Austria, Sept. 2011.

[8] F. Barbaresco. Innovative tools for radar signal processing based on cartan’s geometry of spd matrices & information geometry. In Radar Conference, 2008. RADAR ’08. IEEE, pages 1–6, 2008.

[9] F. Barbaresco. Information geometry of covariance matrix: Cartan-siegel homogeneous bounded domains, mostow/berger fibration and fréchet median. In F. Nielsen and R. Bhatia, editors, Matrix Information Geometry, pages 199–255. Springer Berlin Heidelberg, 2013.

[10] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. Magnetic Resonance in Medicine, 53(1):221–225, 2005.

[11] R. Bhatia. Positive Definite Matrices. Princeton Series in Applied Mathematics. Princeton University Press, 2007.

[12] D. Bini and B. Iannazzo. A note on computing matrix geometric means. Advances in Computational Mathematics, 35(2-4):175–192, 2011.
[13] D. A. Bini and B. Iannazzo. Computing the karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4):1700–1710, 2013.

[14] D. A. Bini, B. Meini, and F. Poloni. An effective matrix geometric mean satisfying the Ando-Li-Mathias properties. *Mathematics of Computation*, 79:437–452, 2010.

[15] E. A. Carlen. Trace inequalities and quantum entropy: An introductory course. *Contemporary Mathematics*, 2010.

[16] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63:1–38, 2010.

[17] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso. Newton method for riemannian centroid computation in naturally reductive homogeneous spaces. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 3, pages III–III, 2006.

[18] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso. Newton algorithms for riemannian distance related problems on connected locally symmetric manifolds. *Selected Topics in Signal Processing, IEEE Journal of*, 7(4):634–645, 2013.

[19] P. T. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

[20] M. Fornasier, H. Rauhut, and R. Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. on Optimization*, 21(4):1614–1640, Dec. 2011.

[21] D. Groisser. Newton’s method, zeroes of vector fields, and the riemannian center of mass. *Advances in Applied Mathematics*, 33(1):95–135, 2004.

[22] D. R. Hunter and K. Lange. Quantile regression via an mm algorithm. *Journal of Computational and Graphical Statistics*, 9(1):60–77, 2000.

[23] D. R. Hunter and K. Lange. A tutorial on mm algorithms. *The American Statistician*, 58(1):pp. 30–37, 2004.

[24] B. Jeuris, R. Vandebril, and B. Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39:379–402, 2012.

[25] H. W. Kuhn. A note on Fermat’s problem. *Mathematical Programming*, 4:98–107, 1973. 10.1007/BF01584648.

[26] S. Lang. *Fundamentals of differential geometry*. Number v. 160 in Graduate texts in mathematics. Springer, 1999.
[27] M. Moakher. On the averaging of symmetric positive-definite tensors. *Journal of Elasticity*, 82(3):273–296, 2006.

[28] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006. 10.1007/s11263-005-3222-z.

[29] Y. Rathi, A. Tannenbaum, and O. Michailovich. Segmenting images on the tensor manifold. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.

[30] Q. Rentmeesters and P.-A. Absil. Algorithm comparison for karcher mean computation of rotation matrices and diffusion tensors. In *Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011)*, pages 2229–2233. EURASIP, 2011.

[31] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donne’s est minimum. *Tohoku Math. J.*, 43:35–386, 1937.