

# A Fixed-Point Algorithm for Estimating Power Means of Positive Definite Matrices

Marco Congedo, Ronald Phlypo, Alexandre Barachant

► **To cite this version:**

Marco Congedo, Ronald Phlypo, Alexandre Barachant. A Fixed-Point Algorithm for Estimating Power Means of Positive Definite Matrices. 24th European Signal Processing Conference (EUSIPCO 2016), Aug 2016, Budapest, Hungary. pp.2016-2010, 24th European Signal Processing Conference (EUSIPCO). <hal-01363817>

**HAL Id: hal-01363817**

**<https://hal.archives-ouvertes.fr/hal-01363817>**

Submitted on 12 Sep 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Fixed-Point Algorithm for Estimating Power Means of Positive Definite Matrices

Marco Congedo, Ronald Phlypo

GIPSA-lab, CNRS, University of Grenoble-Alpes,  
Grenoble Institute of Technology  
Grenoble, France  
marco.congedo@gmail.com

Alexandre Barachant

Burke Medical Research Institute,  
White Plains, NY, USA

**Abstract** — The estimation of means of data points lying on the Riemannian manifold of symmetric positive-definite (SPD) matrices is of great utility in classification problems and is currently heavily studied. The power means of SPD matrices with exponent  $p$  in the interval  $[-1, 1]$  interpolate in between the Harmonic ( $p = -1$ ) and the Arithmetic mean ( $p = 1$ ), while the Geometric (Karcher) mean corresponds to their limit evaluated at  $0$ . In this article we present a simple fixed point algorithm for estimating means along this whole continuum. The convergence rate of the proposed algorithm for  $p = \pm 0.5$  deteriorates very little with the number and dimension of points given as input. Along the whole continuum it is also robust with respect to the dispersion of the points on the manifold. Thus, the proposed algorithm allows the efficient estimation of the whole family of power means, including the geometric mean<sup>1</sup>.

**Keywords:** Power Mean; Geometric Mean; High Dimension; Riemannian Manifold; Symmetric Positive-Definite Matrix.

## I. INTRODUCTION

The study of means (centers of mass) for a set of symmetric positive definite (SPD) matrices has recently attracted much attention, driven by practical problems in radar data processing, image and speech processing, computer vision, shape analysis, medical imaging (especially Diffusion Magnetic Resonance Imaging and Brain-Computer Interface), sensor networks, elasticity, mechanics, quantum entanglement, numerical analysis and machine learning (e.g., [1-6]). In many applications the observed data can be conveniently summarized by SPD matrices, for example, some form of their covariance matrix. In others, SPD matrices arise naturally as kernels, density matrices, elements of a search space, etc. Regardless the nature of the data generating them, SPD matrices may be treated as points on a smooth Riemannian manifold in which natural geometrical notions such as *distance* along the geodesic between two points and the *center of mass* (mean) between two or more points are naturally defined [2]. In turn, these notions allow useful operations such as interpolation, smoothing, filtering, approximation, averaging, signal detection, and classification. In the latter application, a simple minimum

distance to mean (MDM) classifier has been shown to perform as state-of-the-art classifiers in the challenging field of brain-computer interfaces [3]. Thence, given a cloud of data points, their mean on the manifold plays an important role as the natural point where a tangent space to manipulate them should be constructed, allowing the application of sophisticated non-Riemannian classification algorithms within the Riemannian framework. For these reasons, among many others, the numerical estimation of means on Riemannian manifold is currently a subject of intense investigation.

### A. The Manifold of Symmetric Positive-Definite matrices

In differential geometry, a smooth *Riemannian manifold*  $\mathcal{M}$  is a topological space equipped with an inner product on the tangent space defined at each point and varying smoothly from point to point. The tangent space  $T_G\mathcal{M}$  at point  $G$  is the Euclidean vector space containing the tangent vectors to all curves on  $\mathcal{M}$  passing through  $G$  (Fig. 1).

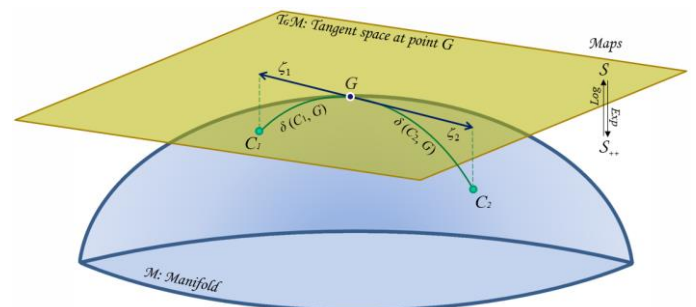


Figure 1: Schematic representation of the SPD manifold, the geometric mean  $G$  of two points and the tangent space at  $G$ . Consider two points (e.g., two covariance matrices)  $C_1$  and  $C_2$  on  $\mathcal{M}$ . The geometric mean of these points is the midpoint on the geodesic connecting  $C_1$  and  $C_2$ , i.e., it minimizes the sum of the two squared distances  $\delta^2(C_1, G) + \delta^2(C_2, G)$ . Now construct the tangent space  $T_G\mathcal{M}$  at  $G$ . There exists one and only one tangent vector  $\zeta_1$  (respectively  $\zeta_2$ ) departing from  $G$  and arriving at the projection of  $C_1$  (respectively  $C_2$ ) from the manifold onto the tangent space; we see that the geodesics on  $\mathcal{M}$  through  $G$  are transformed into straight lines in the tangent vector and that distances are mapped logarithmically; the map from the manifold (symmetric positive definite matrices  $S_{++}$ ) to the tangent space (symmetric matrices  $S$ ) is of logarithmic nature. The inverse map from the tangent space to the manifold is of exponential nature [2].

<sup>1</sup> This research has been partially supported by European project ERC-2012-AdG-320684-CHESS

### B. The Geodesic

The SPD manifold has nonpositive curvature and is complete; for any two points  $C_1$  and  $C_2$  on  $\mathcal{M}$ , a unique path on  $\mathcal{M}$  of minimal length (at constant velocity) connecting the two points always exists. The path is named the *geodesic* and the points along it satisfy

$$C_1 \#_t C_2 = C_1^{1/2} \left( C_1^{-1/2} C_2 C_1^{-1/2} \right)^t C_1^{1/2}, \quad t \in [0,1], \quad (1)$$

with  $t$  the arc-length parameter [2] (Fig. 1). With  $t=0$  we are at  $C_1$ , with  $t=1$  we are at  $C_2$  and with  $t=1/2$  we are at the *geometric mean* of the two points (denoted  $G$  in Fig. 1). As a special case we note  $I \# C = C^t$ . Geodesic equation (1) verifies  $C_1 \#_t C_2 = C_2 \#_{1-t} C_1$  and  $(C_1 \#_t C_2)^{-1} = C_1^{-1} \#_t C_2^{-1}$ . The points along the geodesic can be simply understood as the means of  $C_1$  and  $C_2$  weighted by  $t$  according to the Riemannian metric, in analogy with the weighted mean according to the Euclidean metric given by  $(1-t)C_1 + tC_2$ .

### C. The Distance

Given two matrices (points)  $C_1$  and  $C_2$  of dimension  $N \times N$  on  $\mathcal{M}$ , their *Riemannian distance* is the length of the geodesic (1) connecting them. It is given by [2] (Fig. 1)

$$\delta(C_1, C_2) = \left\| \text{Ln} \left( C_1^{-1/2} C_2 C_1^{-1/2} \right) \right\|_F = \sqrt{\text{tr} \text{Ln}^2(A)} = \sqrt{\sum_n \ln^2 \lambda_n}, \quad (2)$$

where  $A$  is the diagonal matrix holding the  $N$  eigenvalues  $\lambda_1, \dots, \lambda_N$  of matrix  $C_1^{-1/2} C_2 C_1^{-1/2}$  or of similar matrix  $C_1^{-1} C_2$ . This distance has a remarkable number of properties, some of which are listed in [7]. Besides the obvious *symmetry* and *positivity*, particularly useful in signal processing are the following invariance properties (for any invertible  $B$ ):

$$\text{Congruence} \quad \delta(BC_1 B^T, BC_2 B^T) = \delta(C_1, C_2), \quad (3)$$

$$\text{Self-Duality} \quad \delta(C_1^{-1}, C_2^{-1}) = \delta(C_1, C_2). \quad (4)$$

### D. Means of a Matrix Set: Variational Definition

Let  $\mathbf{C} = \{C_1, \dots, C_K\}$  be a set of  $K$  SPD matrices and its associated  $K$  positive weights  $w = \{w_1, \dots, w_K\}$  with  $\sum_k w_k = 1$ . Typically, in signal processing the elements of  $\mathbf{C}$  are noisy data points (recordings, observations, etc.) or quantities derived thereof. Following Fréchet's variational approach, the center of mass  $G$  of set  $\mathbf{C}$  given a distance function  $d(\cdot, \cdot)$  is the point  $G$  minimizing the *variance (dispersion)* of points:  $\sum_k w_k d^2(G, C_k)$ . For instance, the  $w$ -weighted *Arithmetic* and *Harmonic Mean* are defined, respectively, as

$$G_A(w; \mathbf{C}) = \arg \min_G \sum_k w_k \|C_k - G\|_F^2 = \sum_k w_k C_k, \quad (5)$$

$$G_H(w; \mathbf{C}) = \arg \min_G \sum_k w_k \|C_k^{-1} - G^{-1}\|_F^2 = \left( \sum_k w_k C_k^{-1} \right)^{-1}. \quad (6)$$

### E. The Geometric Mean of a Matrix Set

In  $\mathcal{M}$  the  $w$ -weighted *Geometric Mean*  $G_G(w; \mathbf{C})$  is the point realizing the minimum of  $\sum_k w_k \delta^2(C_k, G_G)$ , where the Riemannian distance function  $\delta(\cdot, \cdot)$  acting on  $\mathcal{M}$  has been defined in (2). The geometric mean is the unique point on  $\mathcal{M}$  satisfying non-linear matrix equation [8, 9]

$$\sum_k w_k \text{Ln} \left( G_G^{-1/2} C_k G_G^{-1/2} \right) = 0. \quad (7)$$

In general, it has closed-form solution only for  $K=2$ , in which case it is indeed equal to  $C_1 \#_{1/2} C_2$  (indicated shortly as  $C_1 \# C_2$ ) - see (1) and Fig. 1 - furthermore, for  $K=2$  it is the unique solution to Riccati equation  $(C_1 \# C_2) C_2^{-1} (C_1 \# C_2) = C_1$  [2] and is also equal to  $B^{-1} D_1^{1/2} D_2^{1/2} B^{-T}$  for whatever joint diagonalizer  $B$  of  $C_1$  and  $C_2$ , i.e., for whatever matrix  $B$  satisfying  $BC_1 B^T = D_1$  and  $BC_2 B^T = D_2$ , with  $D_1, D_2$  invertible diagonal matrices [7]. The geometric mean enjoys all 10 properties postulated in the seminal work [10]. These properties can be found listed in [9,11]. We remind here the congruence invariance (8) and self-duality (9), inherited directly from the corresponding properties of its associated distance function (see (3) and (4)):

$$BG_G(w; C_1, \dots, C_K) B^T = G_G(w; BC_1 B^T, \dots, BC_K B^T), \quad (8)$$

$$G_G^{-1}(w; C_1^{-1}, \dots, C_K^{-1}) = G_G(w; C_1, \dots, C_K). \quad (9)$$

### F. Power Means

In the same spirit of section D. and E., given again  $\mathbf{C} = \{C_1, \dots, C_K\}$  and  $w = \{w_1, \dots, w_K\}$ , we can generalize to SPD matrices the power mean of real numbers (generalized means) in the continuum  $p \in [-1, 1]$  such as [9, 11-12],

$$\begin{cases} G_p(w; \mathbf{C}; p) = \sum_k w_k (G_p \#_p C_k), & p \in (0, 1] \\ G_p(w; \mathbf{C}; p) = G_p^{-1}(w; \mathbf{C}^{-1}; -p), & p \in [-1, 0) \\ G_p(w; \mathbf{C}; 0) = G_G(w; \mathbf{C}), & p = 0 \end{cases} \quad (10)$$

where, for any SPD matrices  $C_k$ ,  $G_p \#_p C_k$  with  $p \in (0, 1]$  is the mean of  $G_p$  and  $C_k$  weighted by  $p$  (1),  $\mathbf{C}^{-1} = \{C_1^{-1}, \dots, C_K^{-1}\}$  and  $G_G(w; \mathbf{C})$  is the geometric mean of section E.  $G_p(w; \mathbf{C}; p)$  is named the  $w$ -weighted *power mean of order  $p$*  [9, 11-12]. Following (10), the pair of power means obtained at symmetric values of  $p$  around zero are the *dual* of each other; for a negative value of  $p$  the mean is defined as the inverse of the mean for  $-p$  as applied on the inverted input matrices  $\mathbf{C}^{-1}$ . Power means interpolate continuously in between the harmonic mean ( $p = -1$ ), the geometric mean ( $p = 0$ , evaluated at the limit) and the arithmetic mean ( $p = 1$ ). Thereupon, the family of the power means encompasses and generalizes all Pythagorean means we have encountered so far. All of them enjoy the congruence invariance as the geometric mean (8), but their duality (second line of (10)) coincides with the self-duality property (9) only for  $p = 0$ . The numerous properties of the power means can be found in [12].

## II. ALGORITHMS FOR POWER MEANS

### A. Motivation

We sought a general algorithm for computing the  $w$ -weighted power mean of order  $p$ , with  $p \in (-1, 1) \setminus \{0\}$ . We are also interested in an effective algorithm for estimating the geometric mean ( $p=0$ ). The most popular algorithm for computing the geometric mean is its gradient descent flow with fixed step size, however its convergence rate deteriorates rapidly as the SNR decreases. The same is true for the method based on approximate joint diagonalization in [7]. Second order methods have complexity growing very fast with the size of the input matrices, thus they are little useful in practical applications [13]. The algorithm proposed in [14] has high complexity per iteration and slow linear convergence rate. For a review see both [13] and [7]. In contrast, the complexity per iteration of the algorithm we propose is low and its (linear) convergence rate is fast and almost *uniform* for  $p = \pm 1/2$ . Here *uniform convergence rate* means that the number of iterations does not increase with the size nor with the number of the input set matrices, a very useful property in practical applications. Importantly, the convergence rate is robust with respect to the dispersion of the points (noise) in the whole interval  $p \in (-1, 1) \setminus \{0\}$ , as we will see.

### B. A General Multiplicative Fixed-Point Algorithm

Hereafter it will be convenient to lighten notation; let us denote the weighted power mean of order  $p$  as  $P$ , which by (10) is equal to  $G_\varphi(w; \mathbf{C}; p)$  if  $p \in (0, 1]$  or to  $(G_\varphi(w; \mathbf{C}^{-1}; -p))^{-1}$  if  $p \in [-1, 0)$ . This way we will need to handle only one expression for whatever value of  $p \in (-1, 1) \setminus \{0\}$ , such as

$$P^* = G_\varphi(w; \mathbf{C}^*; |p|) \quad (11)$$

where  $* = \text{sgn}(p)$  is the *dual operator*. As initialization we take as customary the closed form solution of the mean in the case when all matrices in set  $\mathbf{C}$  all pair-wise commute. It has been proved in [12] (Property 1, p. 1502) that the mean is then

$$P_0 = \left( \sum_k w_k C_k^p \right)^{1/p}. \quad (12)$$

Let us now turn to the algorithm. We write (11) using (1) to obtain

$$P^* = P^{1/2} \left( \sum_k w_k \left( P^{-1/2} C_k^* P^{-1/2} \right)^{|p|} \right) P^{1/2}. \quad (13)$$

Equation (13) as a unique SPD solution [12]. Numerical experiments show that iterating this expression as it is (hereafter referred to as “naive fixed-point”) results in a rather slow convergence rate, maximal for  $\text{abs}(p)=1/2$ , but slower and slower as  $\text{abs}(p)$  get closer to 0 or to 1. In order to fasten convergence we design a multiplicative algorithm as it follows: post-multiplying both sides of (13) by  $P^{*/2}$  and taking the inverse at both sides we obtain

$$P^{-1/2} = H^{-1} P^{-1/2}, \quad (14)$$

where

$$H = \sum_k w_k \left( P^{-1/2} C_k^* P^{-1/2} \right)^{|p|}. \quad (15)$$

From (13) we see that upon convergence  $H=I$ .  $H$  here plays the role of the origin in the SPD manifold  $\mathcal{M}$  for data linearly transformed by  $P^{-*/2}$ . In particular, the identity matrix  $I$  is the point of symmetry in  $\mathcal{M}$  corresponding to 0 in the Euclidean space; as  $P^{-1/2}$  is a whitening matrix for the arithmetic mean ( $p=1$ ), so  $P^{-*/2}$  is a whitening matrix for the whole family of power means along continuum  $p \in [-1, 1]$ . We wish to proceed by multiplicative updates according to (14). Rather than to  $P^*$  itself, we thus seek an algorithm converging to  $P^{*/2}$ , which is its inverse square root for  $* = 1$ , i.e., when  $p \in (0, 1]$  and its square root for  $* = -1$ , i.e., when  $p \in [-1, 0)$ . The numerical stability of fixed-point iterates (14) is ensured by the fact that  $H$  converges toward  $I$ . Moreover, using our update rule any update matrix with form  $H^{-\varphi}$  in (14) is equivalent to  $H^{-1}$  upon convergence. We have observed that by using a fixed exponent  $\varphi$  from the first iteration we can reach convergence rate faster than the one achieved by the naive fixed point in the whole interval  $p \in (-1, 1) \setminus \{0\}$ . For doing so, we have to take  $\varphi$  inversely proportional to  $p$  as per

$$\varphi = \frac{1}{2} \varepsilon^{-1} / |p|, \quad \varepsilon \in [1, 2], \quad (16)$$

where  $\varepsilon$  is a constant eccentricity parameter for hyperbola (16) delimiting an acceptable interval for  $\varphi$  as a function of  $|p|$ . At the lower limit  $\varepsilon = 1$  we have a unit hyperbola  $\varphi = 1/2 / \text{abs}(p)$ . By increasing  $\varepsilon$  toward 2 we obtain faster convergence, up to a certain value, which depends upon the signal to noise ratio. In this study we take  $\varepsilon$  as 4/3 and we keep it fixed in all simulations, obtaining quasi-uniform convergence rate for  $p = \pm 1/2$  in a wide range of situations, as we will show next by means of simulations.

#### Algorithm MPM (Multiplicative Power Means)

**INPUT:**  $p \in (-1, 1) \setminus \{0\}$ ,  $K$  positive weights  $w = \{w_1, \dots, w_K\}$  such that  $\sum_k w_k = 1$  and  $K$  SPD matrices  $\mathbf{C}^* = \{C_1^*, \dots, C_K^*\}$ , with  $* = \text{sgn}(p)$ .

**OUTPUT:**  $P$ , the  $w$ -weighted Power Mean of order  $p$ .

#### BEGIN

Initialize  $X$  as the principal square root inverse of (12) if  $p \in (0, 1]$  or as its principal square root if  $p \in [-1, 0)$ .

Set  $\zeta$  equal to a small floating precision number (e.g.,  $10^{-10}$ )

Set  $\varphi = 0.375 / \text{abs}(p)$

#### REPEAT

$$H \leftarrow \sum_k \left[ w_k \left( X C_k^* X^T \right)^{|p|} \right]$$

$$X \leftarrow H^{-\varphi} X$$

UNTIL  $\frac{1}{\sqrt{N}} \|H - I\|_F < \zeta$

$$\text{RETURN } P = Y^T Y, \quad Y = \begin{cases} X^{-T} & \text{if } p \in (0, 1] \\ X & \text{if } p \in [-1, 0) \end{cases}$$

#### END.

### III. SIMULATION STUDIES

In many engineering applications, the matrix condition number of the SPD matrices summarizing the data (observations, recordings,...) tends to be positively correlated with the number of sensors. Also, the dispersion in the manifold of the matrices is proportional to the noise level. The following generative model for input data matrices  $\{C_1, \dots, C_K\}$  of size  $N \times N$  is able to reproduce these properties:

$$C_k = UD_kU^T + \nu(V_kE_kV_k^T) + \alpha I, \quad (17)$$

where

- the **signal part** is given by  $UD_kU^T$ , where  $U$  is a matrix with elements drawn at random at each simulation from a uniform distribution in  $[-1,1]$  and then normalized so as to have columns with unit norm and  $D_k$  are  $K$  diagonal matrices with diagonal elements  $d_{k,n}$  randomly drawn at each simulation from a squared Gaussian distribution with expectation  $=1/2^n$ , where  $n \in \{1, \dots, N\}$  is the index of the  $N$  diagonal elements, thus forming elements of a well-known geometrical series absolutely converging to 1. The elements of the series represent the energy of  $N$  source processes, thus their sum is supposed finite.
- The **uncorrelated noise part** is given by  $\alpha I$ , where  $I$  is the identity matrix and  $\alpha$  here is taken as  $10^{-6}$ ;
- The **structured noise part** is given by  $V_kE_kV_k^T$ , where the  $V_k$  matrices are generated as  $U$  above, the  $E_k$  matrices are generated as  $D_k$  above and  $\nu$  is a constant controlling the SNR of the generated points through

$$SNR = \frac{E\left[\text{tr}\left(\sum_k UD_kU^T\right)\right]}{\nu E\left[\text{tr}\left(\sum_k (V_k E_k V_k^T + \alpha I)\right)\right]}, \quad (18)$$

where  $E$  denotes expectation. In the ensuing simulations we study relevant outcome parameters as a function of the SNR, which is inversely proportional to noise level as per (18), as well as a function of the size ( $N$ ) and number ( $K$ ) of input matrices. We compare the gradient descent algorithm for estimating the geometric mean (section II.A), the naive fixed point algorithm for power means given in [12] (section II.B) and the MPM algorithm here presented, the latter for several values of  $p$ . The gradient descent algorithm is known to converge only up to a certain value of dispersion, i.e., when the SNR is sufficiently high [15]. The stopping criterion is chosen identical for all algorithms: the relative error of matrix  $P$  with respect to a reference matrix  $P_{ref}$  is a dimensionless measure defined as:

$$\frac{\|P - P_{ref}\|_F^2}{\|P_{ref}\|_F^2}. \quad (19)$$

As stopping criterion, considering two successive iterates  $P(i-1)$  and  $P(i)$ , we use

$$\frac{1}{N} \left\| P_{(i)}^{-1} P_{(i-1)} - I \right\|_F^2, \quad (20)$$

whose magnitude does not depend on the size nor to the norm of the matrices. Figure 2 shows the typical convergence behavior for the geometric mean gradient descent algorithm, the naive algorithm with  $p=0.5$  and the MPM algorithm ( $p=0.5$  and  $p=0.001$ ), for  $N=20$ ,  $K=100$  and  $SNR=\{100, 10, 1, 0.1\}$ .

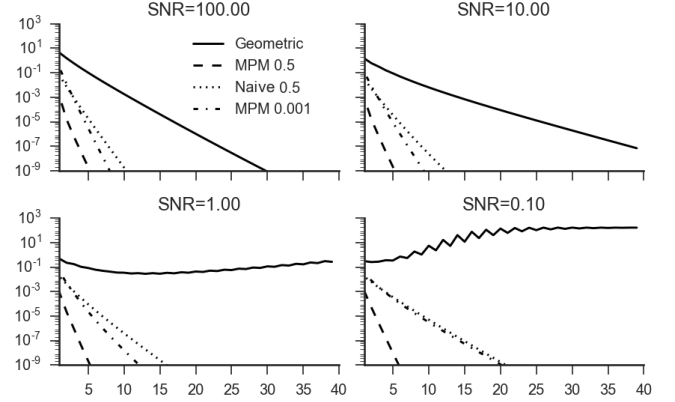


Figure 2: Typical convergence behavior (on abscissa the number of iterations and on the ordinate the convergence as defined in (20) expressed in dB) for the geometric mean gradient descent, naive fixed point with  $p=0.5$  and the MDM algorithm with  $p=\{0.5, 0.001\}$ , for  $N=20$  (dimension of input matrices),  $K=100$  (number of input matrices) and  $SNR=\{100, 10, 1, 0.1\}$  (18).

Figure 3 shows the main effects (bars) and their standard deviation (sd: lines) across 50 simulations of  $N=\{10, 25, 50\}$ ,  $K=\{10, 100, 500\}$  and  $SNR=\{100, 1, 0.01\}$  on the number of iterations. ‘‘Main effects’’ means that for each level of  $N$ ,  $K$  and  $SNR$  the average and sd are computed across all levels of the other two variables, as in a classical analysis of variance (ANOVA). We see that the number of iterations required by the MPM algorithm is comparable in all situations for  $p=0.5$  (quasi-uniform convergence behavior), whereas for the geometric mean gradient descent it grows rapidly with  $N$  and with noise, as it is well known [7]. Even with  $p=0.001$ , which is a very small value, the convergence rate of the MPM is superior to the convergence rate of the geometric mean gradient descent flow, which often do not converge at all (the maximum number of iterations allowed was fixed to 50 for all algorithms. See Fig. 2 for examples of divergence).

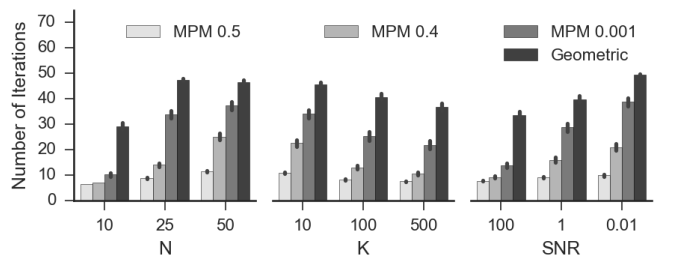


Figure 3: main effects average (50 simulations) and sd number of iterations for  $N=\{10, 25, 50\}$ ,  $K=\{10, 100, 500\}$  and  $SNR=\{100, 1, 0.01\}$  for the geometric mean gradient descent and the MPM algorithm with  $p=\{0.5, 0.4, 0.001\}$ .

The simulation in Fig. 4 (left) shows the relative error (19) of the power mean with respect to the geometric mean ( $p=0$ ) for several small values of  $p$ ,  $N=10$ ,  $K=100$  and  $\text{SNR}=\{100, 1, 0.01\}$ . We see that the approximation to the geometric mean in term of log-relative error is linear in  $\log-p$ , ceiling to  $10^{-5}$  with high noise. Figure 4 (right) shows the main effects (across 50 simulations) of  $N=\{10, 25, 50\}$ ,  $K=\{50, 100, 500\}$  and  $\text{SNR}=\{100, 10, 1\}$  on the error for  $p=0.001$ . We see that the error decreases with  $K$  and increases with  $N$  and with noise. Notice that an error much lower than this can be achieved by computing the power mean for two small symmetric values around zero and then computing their geometric mean by the closed-form solution for two matrices (see section E.); in practice a relative error  $< 10^{-3}$  is fully satisfactory, since the noise level in challenging data is much higher than that.

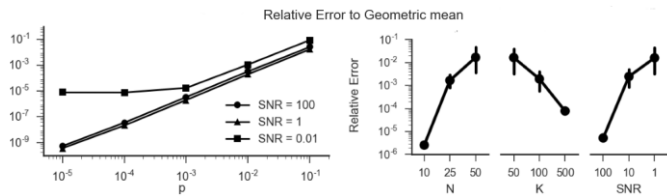


Figure 4: Relative Error to the geometric mean obtained with the MPM algorithm. Left:  $N=10$ ,  $K=100$ ,  $\text{SNR}=\{100, 1, 0.01\}$  and  $p=\{10^{-5}, \dots, 10^{-1}\}$ . Right: Main effects (bars) and sd (lines) obtained across 50 repetitions for  $p=0.001$ ,  $N=\{10, 25, 50\}$ ,  $K=\{10, 100, 500\}$  and  $\text{SNR}=\{100, 1, 0.01\}$ .

Figure 5 is a *TraDe* plot (log-trace vs. log-determinant) for a sampling of power means along continuum  $p \in [-1, 1]$ , showing the log-log linear relation of the trace and determinant of power means along the Pythagorean continuum.

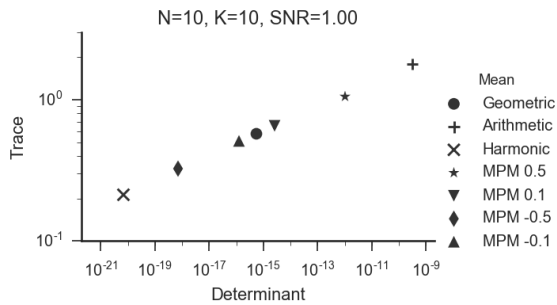


Figure 5: *TraDe* plot obtained with  $N=10$ ,  $K=10$  and  $\text{SNR}=1$ , for power means corresponding to  $p=1$ (Arithmetic), 0.5, 0.1, 0 (Geometric), -0.1, -0.5 and -1 (Harmonic). The relationship between the trace and the determinant of power means is log-log linear.

#### IV. CONCLUSIONS

Power means are generalized means interpolating continuously in the interval  $p \in [-1, 1]$ , with  $p=1$  yielding the arithmetic mean,  $\lim p \rightarrow 0$  yielding the geometric mean and  $p=-1$  yielding the harmonic mean. We have presented a new multiplicative algorithm for estimating power means in the

interval  $p \in [-1, 1] \setminus \{0\}$ . A numerical analysis shows that its convergence rate is very fast and quasi-uniform for values of  $p$  close to  $\frac{1}{2}$  and  $-\frac{1}{2}$ , while for values of  $p$  close to 0 it is still faster as compared to the gradient descent with fixed step-size used for estimating the geometric mean. Furthermore, it converges also in low SNR situations. We can therefore prefer the proposed algorithm also for approximating the geometric mean. In conjunction with expressions (5) and (6), with the MPM algorithm we can now estimate a number of means sampling along the continuum  $p \in [-1, 1]$ , as in Fig. 5. We name such a sampling a Pythagorean *mean field*. Applications of mean fields include the possibility to evaluate the most appropriate choice of mean depending on its use and on the data at hand. Mean fields also allow the extension of current Riemannian classifiers, such as in [1, 3-5], for instance combining current classifiers as applied to all the means of the field. The application of mean fields to real data will be the object of future investigations.

#### REFERENCES

- [1] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, "Multi-Class Brain Computer Interface Classification by Riemannian Geometry", IEEE Trans. Biomed. Eng., vol. 59(4), 2012, pp. 920-928.
- [2] R. Bhatia, Positive Definite Matrices. New Jersey: Princeton University Press, 2007.
- [3] M. Congedo, "Riemann Geometry: an Universal BCI Classification Framework. In: EEG Source Analysis, Ch. VIII, M. Congedo, HDR presented at doctoral School EDISCE, Grenoble University, 2013, pp. 194-222.
- [4] P. Fillard, V. Arsigny, N. Ayache, X. Pennec, "A Riemannian Framework for the Processing of Tensor-Valued Images". DSSCV, 2005, pp. 112-123.
- [5] Y. Li, K.M. Wong, "Riemannian Distances for Signal Classification by Power Spectral Density", IEEE J. Sel. Top. Signal Process., vol. 7(4), 2013, pp. 655-669.
- [6] M. Moakher, M. Zéraï, "The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data", J. Math. Imaging Vis., vol. 40, 2011, pp. 171-187.
- [7] M. Congedo, B. Afsari, A. Barachant, M. Moakher, "Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices", PLoS ONE, vol. 10(4), 2015, e0121423.
- [8] M. Moakher, "A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces". SIAM. J. Matrix Anal. Appl., vol. 26 (3), 2005, pp. 735-747.
- [9] J. Lawson, Y. Lim, "Weighted means and Karcher equations of positive operators", PNAS, vol. 110(39), 2013, pp 15626-32.
- [10] T. Ando, C.K Li, R. Mathias, "Geometric means", Linear Algebra Appl., vol. 385, 2004, pp. 305-334.
- [11] J. Lawson, Y. Lim, "Monotonic properties of the least squares mean", Math Ann, vol. 351(2), 2011, pp 267-279.
- [12] Y. Lim and M. Pálfi, "Matrix Power means and the Karcher mean", J. Funct. Anal., vol. 262, 2012, pp 1498-1514.
- [13] B. Jeuris, R. Vanderbril, B. Vandereycken, "A survey and comparison of contemporary algorithms for computing the matrix geometric mean". Elec. Trans. Numer. Anal., vol. 39, 2012, pp. 379-402.
- [14] Zhang T, "A Majorization-Minimization Algorithm for the Karcher Mean of Positive Definite Matrices", arXiv, 2014, 1312.4654.
- [15] Afsari B, Tron R, Vidal R, "On the convergence of gradient descent for finding the Riemannian center of mass", SIAM J. Control Optim., 51 (3), 2013, 2230-2260.