Using the Bhattacharyya Mean for the Filtering and Clustering of Positive-Definite Matrices

Malek Charfi[1], Zeineb Chebbi[2], Maher Moakher[2], and Baba C. Vemuri[3]

[1] Tunisia Polytechnic School, Univ. Carthage, B.P. 743, 2078 La Marsa, Tunisia
[2] LAMSIN, ENIT, Univ. Tunis El Manar, B.P. 37, 1002 Tunis-Belvédère, Tunisia
[3] Dept. of CISE, Univ. of Florida, Gainesville, Florida 32611

Abstract. This work deals with Bhattacharyya mean, Bhattacharyya and Riemannian medians on the space of symmetric positive-definite matrices. A comparison between these averaging methods is given in two different areas which are mean (median) filtering to denoise a set of High Angular Resolution Diffusion Images (HARDI) and clustering data. For the second application, we will compare the efficiency of the Wishart classifier algorithm using the aforementioned averaging methods and the Bhattacharyya classifier algorithm.

1 Introduction

In recent years, need for the filtering and clustering positive-definite matrix data sets has increased considerably in various applications such as elasticity [13], radar signal processing [1,11], medical imaging [7,3,6,15] and image processing [12]. In these data processing tasks the concept of the average (mean or the median) of a set of positive-definite matrices plays a central role. The mean of a set of symmetric positive-definite (SPD) matrices is the minimizer of the sum of the squared distances between the mean and the members of the set while the median is the minimizer of the sum of the distances between the median and the elements of the set. Depending on the definition of the distance, one gets different kinds of means and medians. In Fletcher et al. [8], the authors defined the Riemannian median as the minimizer of the sum of the geodesic distances from the unknown median to every member of the set whose median is being sought. They proved existence and uniqueness of the Riemannian median and they conjecture that their proposed gradient descent algorithm is convergent for symmetric positive-definite matrices. They presented applications to show the robustness of the Riemannian median filtering compared to the Fréchet mean filtering. In [4], the authors defined the Bhattacharyya median and compared it to the Riemannian median in order to denoise a set of Diffusion Weighted Images (DWI).

In the present paper, we prove the existence and uniqueness of the Bhattacharyya median recently used in [4]. Then, we compare it with the Riemannian median and Bhattacharyya mean using the concept of median and mean filtering of a noisy synthetic HARDI data set. We also compare the use of these different averaging methods for clustering synthetic polarimetric SAR data.

F. Nielsen and F. Barbaresco (Eds.): GSI 2013, LNCS 8085, pp. 551–558, 2013.

2 Preliminaries

In this section, we review some background material and introduce the notations that will be used throughout this paper. Let $\mathcal{M}(n,\mathbb{R})$ be the set of $n \times n$ real matrices and $\mathcal{S}(n)$ its subspace of symmetric matrices, i.e., $\mathcal{S}(n) = \{\boldsymbol{A} \in \mathcal{M}(n,\mathbb{R}), \boldsymbol{A}^T = \boldsymbol{A}\}$, where the superscript ^{T} is the transpose operator. Let $\mathcal{P}(n) = \{\boldsymbol{A} \in \mathcal{S}(n), \boldsymbol{A} > 0\}$ be the set of all $n \times n$ SPD matrices where $\boldsymbol{A} > 0$ is equivalent to $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for any $\boldsymbol{x} \neq 0$. The Frobenius norm of a matrix $\boldsymbol{X} \in \mathcal{M}(n,\mathbb{R})$ is defined by $\|\boldsymbol{X}\|_F = \sqrt{\mathrm{trace}(\boldsymbol{X}\boldsymbol{X}^T)}$.

It should be noted that the differentiable manifold $\mathcal{P}(n)$ can be given a metric structure using several distances. In the sequel, we will use the following two distances on $\mathcal{P}(n)$:

1. The Riemannian distance:
$$d_R(\boldsymbol{X},\boldsymbol{Y}) := \|\operatorname{Log}_{\boldsymbol{X}}(\boldsymbol{Y})\|_F, \tag{1}$$
where $\operatorname{Log}_{\boldsymbol{X}}(\boldsymbol{Y}) = \boldsymbol{X}^{\frac{1}{2}}\operatorname{Log}(\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{Y}\boldsymbol{X}^{-\frac{1}{2}})\boldsymbol{X}^{\frac{1}{2}}$ is the log map on the manifold $\mathcal{P}(n)$.
2. The Bhattacharyya distance (also called log-determinant 0-distance) recently defined in [5]:
$$d_B(\boldsymbol{X},\boldsymbol{Y}) := 2\sqrt{\log\frac{\det\frac{1}{2}(\boldsymbol{X}+\boldsymbol{Y})}{\sqrt{\det(\boldsymbol{X})\det(\boldsymbol{Y})}}}. \tag{2}$$

We recall that the weighted median is defined as:
$$\mathrm{med}(\boldsymbol{P}_1,\ldots,\boldsymbol{P}_N) = \mathrm{argmin}_{\boldsymbol{X}\in\mathcal{P}(n)}\sum_{k=1}^{N}\omega_k d(\boldsymbol{P}_k,\boldsymbol{X}),$$
where $d(\cdot,\cdot)$ is the chosen distance, $\boldsymbol{P}_1,\ldots,\boldsymbol{P}_N$, are the given SPD matrices and $\omega_1,\ldots,\omega_N$ with $\sum_{i=1}^N \omega_i = 1$ are the corresponding positive real weights. Similarly, the weighted mean is defined as:
$$\mathrm{mean}(\boldsymbol{P}_1,\ldots,\boldsymbol{P}_N) = \mathrm{argmin}_{\boldsymbol{X}\in\mathcal{P}(n)}\sum_{k=1}^{N}\omega_k d(\boldsymbol{P}_k,\boldsymbol{X})^2.$$

3 Bhattacharyya Mean and Median

In [5], the authors introduced the concept of Bhattacharyya mean (BhMean) of a set of SPD matrices. That is, if $\boldsymbol{P}_1,\ldots,\boldsymbol{P}_N$ is a set of symmetric positive-definite matrices and ω_i, $i = 1,\ldots,N$ a set of real weights that sum to 1, the Bhattacharyya mean of $\boldsymbol{P}_1,\ldots,\boldsymbol{P}_N$ is the unique SPD matrix $\boldsymbol{P}$ satisfying:
$$\sum_{i=1}^{N}\omega_i\left(\frac{1}{2}\boldsymbol{P}_i + \frac{1}{2}\boldsymbol{P}\right)^{-1} = \boldsymbol{P}^{-1}. \tag{3}$$

The existence and uniqueness of equation (3) were discussed in [5] and its numerical solution can be computed using the following fixed-point algorithm:

Alg. 1: Fixed-point algorithm	
Start with an initial guess	$\boldsymbol{X}_0$
Repeat for $p = 0, 1, \ldots$	
Fixed-point iterations	$\boldsymbol{X}_{p+1} = g(\boldsymbol{X}_p)$
Until stopping criterion is satisfied	$\\|\boldsymbol{X}_{p+1} - \boldsymbol{X}_p\\| < \epsilon$

where $g(\boldsymbol{X}) := \left(\sum_{i=1}^{N} \omega_i \left(\frac{\boldsymbol{P}_i+\boldsymbol{X}}{2}\right)^{-1}\right)^{-1}$. Similarly, we can get the following result for the Bhattacharyya median (BhMedian):

Proposition 3.1. *The weighted Bhattacharyya median of N symmetric positive-definite matrices $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N$ with weights $\omega_1, \ldots, \omega_N$ is the unique symmetric positive-definite matrix $\boldsymbol{X}$ solution of the following equation:*

$$\sum_{i=1}^{N} \frac{\omega_i}{\mathrm{d}_B(\boldsymbol{P}_i, \boldsymbol{X})} \left(\frac{\boldsymbol{P}_i + \boldsymbol{X}}{2}\right)^{-1} = \sum_{i=1}^{N} \frac{\omega_i}{\mathrm{d}_B(\boldsymbol{P}_i, \boldsymbol{X})} \boldsymbol{X}^{-1}.$$

Proof. For all $\boldsymbol{Y}$ in $\mathcal{S}(n)$ we have [5]:

$$\frac{d}{dt} d_B(\boldsymbol{P}_i, \boldsymbol{X} + t\boldsymbol{Y})|_{t=0} = \frac{\mathrm{trace}\left(\left(\frac{\boldsymbol{P}_i+\boldsymbol{X}}{2}\right)^{-1} \boldsymbol{Y} - \boldsymbol{X}^{-1}\boldsymbol{Y}\right)}{d_B(\boldsymbol{P}_i, \boldsymbol{X})}.$$

Thus, the gradient of $f(\boldsymbol{X}) = \sum_{i=1}^{N} \omega_i \mathrm{d}_B(\boldsymbol{P}_i, \boldsymbol{X})$ at $\boldsymbol{X}$ is given by

$$\nabla f(X) = \sum_{i=1}^{N} \frac{\omega_i}{\mathrm{d}_B(\boldsymbol{P}_i, \boldsymbol{X})} \left(\left(\frac{\boldsymbol{P}_i + \boldsymbol{X}}{2}\right)^{-1} - \boldsymbol{X}^{-1}\right). \tag{4}$$

The equation $\nabla f(\boldsymbol{X}) = \boldsymbol{0}$ has a unique solution $\boldsymbol{X}$ on the space of symmetric positive-definite matrices. In fact, this equation is equivalent to

$$\boldsymbol{X} = F(\boldsymbol{X}) := \sum_{i=1}^{N} \frac{\omega_i}{d_B(\boldsymbol{P}_i, \boldsymbol{X})} \left(\sum_{i=1}^{N} \frac{\omega_i}{d_B(\boldsymbol{P}_i, \boldsymbol{X})} \left(\frac{P_i + X}{2}\right)^{-1}\right)^{-1}. \tag{5}$$

Analogously to the Bhattacharyya mean [5], one can prove that in $(\mathcal{P}(n), d_T)$, F defines a contraction mapping, where d_T is the Thompson's metric defined for two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{P}(n)$ by [14]

$$d_T(\boldsymbol{A}, \boldsymbol{B}) = \max\{\log M(\boldsymbol{A}, \boldsymbol{B}), \log M(\boldsymbol{B}, \boldsymbol{A})\},$$

where $M(\boldsymbol{A}, \boldsymbol{B}) = \inf\{\lambda > 0, \boldsymbol{A} \leq \lambda \boldsymbol{B}\} = \lambda_{\max}(\boldsymbol{B}^{-1}\boldsymbol{A})$.

Numerical solution can be found using a fixed-point algorithm like the Bhattacharyya mean (Alg. 1).

4 Riemannian Median

The weighted median of a set of SPD matrices, $\boldsymbol{P}_1, \dots, \boldsymbol{P}_N$, with respect to the Riemannian distance (1) (RiMedian), is defined as the unique SPD matrix $\boldsymbol{X}$, solution of the following equation [8]:

$$\sum_{i=1}^{N} \frac{\omega_i}{d_R(\boldsymbol{P}_i, \boldsymbol{X})} \operatorname{Log}_{\boldsymbol{X}} \boldsymbol{P}_i = \boldsymbol{0}. \tag{6}$$

In [8], the authors conjectured that the following gradient descent algorithm is convergent for $\alpha = 1$ in the case of symmetric positive-definite matrices:

Alg. 2: Gradient-descent algorithm	
Start with an initial guess	$\boldsymbol{X}_0$
Repeat for $p = 0, 1, \dots$	
Determine a descent direction	$\boldsymbol{D}_p := -\nabla f(\boldsymbol{X}_p)$
Choose a step	$\alpha_p > 0$
Update	$\boldsymbol{X}_{p+1} := \boldsymbol{X}_p + \alpha_p \boldsymbol{D}_p$
Until stopping criterion is satisfied	$\|\boldsymbol{D}_p\| < \epsilon$

where $f(\boldsymbol{X}) = \sum_{i=1}^{N} \omega_i \mathrm{d}_R(\boldsymbol{P}_i, \boldsymbol{X})$. Extensive numerical experiments show that for $0 < \alpha \leq 1$, the above algorithm generally converges. However, to ensure convergence, one should decrease α as the matrix dimension increases. In the table below, we compare the number of iterations and the CPU time needed by each of the following methods: Bhattacharyya mean / median and Riemannian median (for α equals 1 and 0.5) using different matrix dimensions.

Table 1. Comparison of the number of iterations (N_{iter}) and the CPU time (T in seconds) between different algorithms and different matrix dimensions ($n = 3, 4, 6$ and 10) for a set of 200 SPD matrices

	$n=3$		$n=4$		$n=6$		$n=10$	
	N_{iter}	T	N_{iter}	T	N_{iter}	T	N_{iter}	T
RiMedian $\alpha = 1$	12	9.1886	41	14.7577	83	65.9260	-	diverges
RiMedian $\alpha = .5$	10	22.3705	13	32.1986	19	50.2635	24	100.6673
BhMedian	37	13.9465	42	14.8201	47	17.1913	60	194065
BhMean	63	0.3120	73	0.4056	86	0.5772	107	0.8736

5 Application of Smoothing Magnetic Resonance Imaging Data

We apply the different algorithms described above to denoise a set of Magnetic Resonance Imaging (MRI) data. Experiments were performed using synthetic DW-MRI dataset. We use an image region of size 32×32 [2] of fourth-order

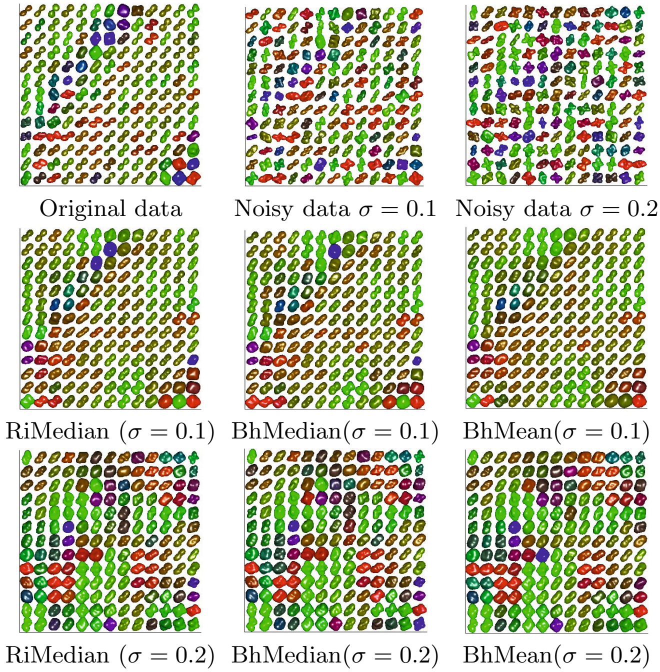

Fig. 1. Comparison of the robustness of the different median (mean) filtering methods

diffusion tensors as presented in Fig. 1. Then, we add to it various levels of Rician noise [2] of standard deviation $\sigma = 0.1$, 0.2 and 0.5.

To denoise this data, we assign to each voxel (i, j) the median (or the mean) of a window W centered at the voxel (i, j) of square shape whose sides are an odd number of pixels, e.g., (3×3). Fig. 1 shows the results of denoising our data (for Rician noise of $\sigma = 0.1$, 0.2) using the different methods presented above. Each tensor is colored according to the orientation of maximum value. The orientation components X, Y and Z are assigned to the color components R, G and B, respectively. In Table 2 we give the error over the whole image. The error is calculated as the sum of the differences between the norm of the initial data and the norm of the denoised data.

Table 2. Error values of the different filtering algorithms

Error	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.5$
Original data	0	0	0
Noisy data	0.7091	1.0898	1.2782
Riemannian median	0.5474	0.8780	1.0403
Bhattacharyya median	0.5477	0.8849	1.0582
Bhattacharyya mean	0.5498	0.8627	1.0350

We remark that the three filtering algorithms are close in terms of robustness but taking into account the CPU time, the Bhattacharyya mean algorithm is preferred.

6 Application to Data Classification

Another area of application of symmetric positive-definite matrix averaging is the clustering polarimetric SAR data [9]. For polarimetric non Gaussian model [16], the data in each pixel is a compound Gaussian vector defined as the product of a positive scalar random variable τ (generally chosen as Gamma distributed) and complex Gaussian vector $\boldsymbol{x}$. Then, the polarimetric data in each pixel follows a $\mathcal{K}$ distribution with parameters τ and $\boldsymbol{C}$, where $\boldsymbol{C}$ is the covariance matrix of $\boldsymbol{x}$. Gini et al. [10], derived the Maximum Likelihood estimate of the covariance matrix, which enables the parametrization of the polarimetric data in each pixel as a Hermitian positive definite matrix.

In the literature, the well known classification model for polarimetric SAR data is the Wishart classifier which is described by:

Alg. 3: The Wishart classifier algorithm
1. Start with an initial classification of the image.
2. Compute the class centers H_i as the mean (or median) of the class elements.
3. Reassign the pixels to the corresponding class that minimizes the Wishart distance measure defined by $\mathrm{d}_W(\boldsymbol{C}, \boldsymbol{H_i}) = \log\det(\boldsymbol{H_i}) + \mathrm{trace}({\boldsymbol{H_i}}^{-1}C)$.
4. Repeat steps 2-3 until a stopping criterion is met.

To compare the efficiency of the three averaging methods, we constructed a simulated image (Fig. 2 (a)) as in [9]. The image is divided into four equal quadrants $A_1, \ldots, A_4$. Each $A_i, i = 1, \ldots, 4$ is also divided in four small quadrants $A_{ij}, j = 1, \ldots, 4$. Polarimetric information in each pixel of the region A_{ij} is randomly chosen following the $\mathcal{K}$ distribution [9].

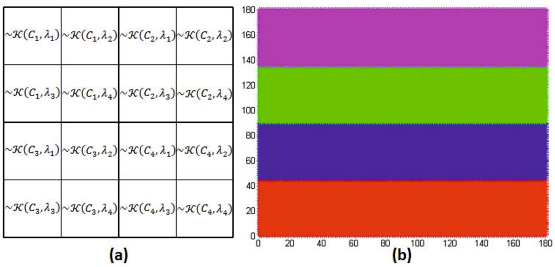

Fig. 2. In (a) we present the shape of a constructed image as in [9], while in (b) we show the initialization clusters for both the Wishart and Bhattacharyya classifier algorithms

After only three iterations of the Wishart classifier algorithm we obtain the results shown in Fig. 3. One can easily note the robustness of the classification based on the Bhattacharyya mean when compared with the classification based on the two other averaging methods.

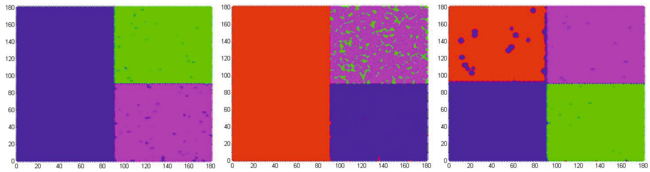

Fig. 3. Results of Wishart classifier algorithm for Riemannian median, Bhattacharyya median and Bhattacharyya mean, respectively

Now, we replace the Wishart distance by the Bhattacharyya distance in the Wishart classifier algorithm. This new algorithm will be called Bhattacharyya classifier algorithm. We apply this new classifier algorithm to the same data used before. After three iterations, we obtain the results shown in Fig. 4.

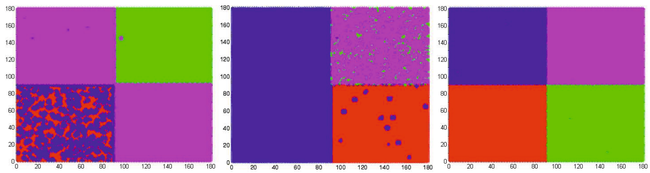

Fig. 4. Results of Bhattacharyya classifier algorithm for the Riemannian median, Bhattacharyya median and Bhattacharyya mean, respectively

We can easily notice that the classification results have been clearly improved by using the Bhattacharyya distance instead of the Wishart distance.

7 Conclusion

In this work, we have compared three averaging methods which are the Riemannian median, Bhattacharyya median and Bhattacharyya mean. The results showed that these methods when used for smoothing DW-MRI data are equivalent in terms of robustness but the difference appears in terms of computational time which leads us to prefer the Bhattacharyya mean algorithm. On the other hand, it is clear based on the clustering results that the Wishart classifier and its modified version are much more robust when using the Bhattacharyya mean. In forthcoming works, we will try to apply these results to real data. We believe that the Bhattacharyya classifier algorithm will give good results for real data.

Acknowledgments. This research was in part supported by the Tunisian Ministry of Higher Education and Scientific Research to Moakher and by the NIH grant RO1 NS066340 to Vemuri.

References

1. Barbaresco, F.: New foundation of Radar Doppler signal processing based on advanced differential geometry of symmetric spaces: Doppler matrix CFAR and Radar application. In: International Radar Conference (2009)
2. Barmpoutis, A., Vemuri, B.C.: A unified framework for estimating diffusion tensors of any order with symmetric positive definite constraints. In: IEEE International Symposium on Biomedical Imaging. From Nano to Macro, pp. 1385–1388 (2010)
3. Batchelor, P.G., Moakher, M., Atkinson, D., Calamante, F., Connelly, A.: A rigorous framework for diffusion tensor calculus. Magnetic Resonance in Medicine 53(1), 221–225 (2005)
4. Charfi, M., Chebbi, Z., Moakher, M., Vemuri, B.C.: Bhattacharyya median of symmetric positive-definite matrices and application to the denoising of diffusion-tensor fields. In: IEEE 10th International Symposium on Biomedical Imaging: From Nano to Macro, San Francisco, CA, USA, April 7-11, pp. 1215–1218 (2013)
5. Chebbi, Z., Moakher, M.: Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function. Linear Algebra and its Applications 436, 1872–1889 (2012)
6. Fillard, P., Pennec, X., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vision 66, 41–66 (2006)
7. Fletcher, P.T., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing 87, 250–262 (2007)
8. Fletcher, P.T., Venkatasubramanian, S., Joshi, J.: Robust statistics on Riemannian manifolds via the geometric median. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
9. Formont, P., Ovarlez, J.P., Pascal, F.: On the use of matrix information geometry for polarimetric SAR image classification. In: Nielsen, F., Bhatia, R. (eds.) Matrix Information Geometry, pp. 257–276. Springer, Heidelberg (2013)
10. Gini, F., Greco, M.V.: Covariance matrix estimation for CFAR detection in correlated heavy-tailed clutter. Signal Processing 82(12), 1847–1859 (2002)
11. Lapuyade-Lahorgue, J., Barbaresco, F.: Radar detection using Siegel distance between autoregressive processes, application to HF and X-band Radar. In: IEEE RADAR 2008, Rome (May 2008)
12. Michailovich, O., Rathi, Y., Tannenbaum, A.: Segmenting images on the tensor manifold. In: Computer Vision and Pattern Recognition, pp. 1–8 (2007)
13. Moakher, M.: On the averaging of symmetric positive-definite tensors. J. Elasticity 82(3), 273–296 (2006)
14. Thompson, A.C.: On certain contraction mappings in a partially ordered vector space. Proc. Amer. Math. Soc. 14, 438–443 (1963)
15. Wang, Z., Vemuri, B.C.: DTI segmentation using an information theoretic tensor dissimilarity measure. IEEE Trans. Med. Imag. 24 (2005)
16. Yao, K.: A representation theorem and its applications to spherically invariant random processes. IEEE Transactions on Information Theory 19(5), 600–608 (1973)