



HAL
open science

An M-Estimator for Robust Centroid Estimation on the Manifold of Covariance Matrices

Ioana Ilea, Lionel Bombrun, Romulus Terebes, Monica Borda, Christian Germain

► **To cite this version:**

Ioana Ilea, Lionel Bombrun, Romulus Terebes, Monica Borda, Christian Germain. An M-Estimator for Robust Centroid Estimation on the Manifold of Covariance Matrices. IEEE Signal Processing Letters, 2016, IEEE Signal Processing Letters, 23 (9), pp.1255 - 1259. 10.1109/LSP.2016.2594149 . hal-01385302

HAL Id: hal-01385302

<https://hal.science/hal-01385302>

Submitted on 21 Oct 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.

An M-estimator for Robust Centroid Estimation on the Manifold of Covariance Matrices

Ioana Ilea, Lionel Bombrun, Romulus Terebes, Monica Borda, and Christian Germain

Abstract—This paper introduces a new robust estimation method for the central value of a set of N covariance matrices. This estimator, called the Huber’s centroid, is described starting from the expression of two well-known methods, that are the center of mass and the median. In addition, a computation algorithm based on the gradient descent is proposed. Moreover, the Huber’s centroid performances are analyzed on simulated data, to identify the impact of outliers on the estimation process. In the end, the algorithm is applied to brain decoding, based on magnetoencephalography (MEG) data. For both simulated and real data, the covariance matrices are considered as realizations of Riemannian Gaussian distributions and the results are compared to those given by the center of mass and the median.

Index Terms—centroid, classification, center of mass, median, Huber’s centroid.

I. INTRODUCTION

COVARIANCE matrices are used in a wide variety of applications in signal and image processing, including array processing [1], radar detection [2], [3], medical image segmentation [4], face detection [5], vehicle detection [6], etc. Another research direction concerns the signal and image classification, where covariance matrices can be used to model different kind of dependence, like spatial, temporal, spectral, polarimetric dependence, etc [7]–[10].

Recently, covariance matrices have been modeled as realizations of Riemannian Gaussian distributions (RGDs) and used in classification algorithms such as k-means or Expectation-Maximization (EM) [9]. This kind of classification procedures are based on the partition of the dataset in subsets, or clusters, characterized by their central values, also called centroids. The dataset’s partition is accomplished by assigning each observation to the closest cluster in terms of a predefined distance [11]. This is a recursive procedure and for each iteration, the centroid’s value is recomputed and the assignment step is repeated. Usually, the cluster’s centroid is the center of mass, computed by using the squared Euclidean distance. Despite its popularity, this method is not appropriate for covariance matrices having a Riemannian geometry. To solve this problem, the Euclidean distance can be replaced by an intrinsic metric such as the Riemannian distance. The main disadvantage of the center of mass is its non-robust behavior to outliers that can exist in the dataset [11]–[13]. A robust alternative for the centroid’s computation is the median, which has been also generalized for Riemannian manifolds [3], [14], [15]. This estimator is computed by using a gradient

descent algorithm. Nonetheless, in this algorithm, a division by the distance between each observed covariance matrix in the dataset and the median is needed. If those two points are too close, this distance tends toward zero and may lead to numerical instability. In such case, Yang propose to exclude those points, at each iteration of the algorithm [14]. Another possibility for determining robust centroids in the space of covariance matrices is the use of the trimming methods [16]. These algorithms imply the elimination of a fixed percentage of outliers, according to their distance with respect to the dataset’s mean or median, and the computation of the mean or the median on the remaining data. Nevertheless, the main difficulty of the trimmed estimators relies on the way to tune the percentage of discarded data.

The main contribution of the paper is to propose a novel centroid estimator, based on the theory of M-estimators. By considering the so-called Huber’s function [17], [18], we introduce the definition of this estimator and present an algorithm to estimate it from a sample of N covariance matrices. The proposed estimator is a trade-off between the center of mass and the median, where the former is efficient, while the latter is robust to outliers. Moreover, based on the median absolute deviation (MAD) concept, this paper presents a way to automatically determine the Huber’s threshold.

The paper is structured as follows. Section II recalls the definition of the centroid from a sample of N observations. A brief overview of the center of mass and the median are given. Next, we introduce the proposed Huber’s centroid estimator and present a gradient descent algorithm to estimate it. The performance of these estimators is then evaluated on simulated data. Section III introduces an application to brain decoding, based on MEG data. Finally, Section IV reports some conclusions and perspectives of this work.

II. THE HUBER’S ESTIMATOR FOR CLUSTER CENTROIDS

A. Centroids and estimation methods

Many signal and image processing applications including classification [9], segmentation [19], or filtering [3] require the computation of the central value of a covariance matrix dataset, which represents the subject of this section. Let $\{\mathbf{M}_1, \dots, \mathbf{M}_N\}$ be a random sample of N covariance matrices. The centroid estimator of this set, denoted $\widehat{\mathbf{M}}$, is defined as being the minimizer of the following cost function $f(\mathbf{M})$:

$$\widehat{\mathbf{M}} = \underset{\mathbf{M}}{\operatorname{argmin}} f(\mathbf{M}). \quad (1)$$

Depending on the choice of $f(\mathbf{M})$, different estimators of the centroids have been introduced in the literature. In the

I. Ilea, L. Bombrun and C. Germain are with Laboratoire IMS, Universit de Bordeaux, (e-mail: firstname.lastname@u-bordeaux.fr)

I. Ilea, R. Terebes and M. Borda are with Technical University of Cluj-Napoca, (e-mail: firstname.lastname@com.utcluj.ro)

following, we briefly recall the definition of the center of mass (CM) [20]–[22] and the median (Med) [2], [15] and next we introduce the proposed M-estimator.

a) *The center of mass* is one of the most popular estimators, for which the cost function is:

$$f_{CM}(\mathbf{M}) = \frac{1}{N} \sum_{i=1}^N d^2(\mathbf{M}, \mathbf{M}_i), \quad (2)$$

where $d(\cdot)$ represents the Rao's Riemannian distance between two covariance matrices defined as [23]:

$$d(\mathbf{M}_1, \mathbf{M}_2) = \left[\frac{1}{2} \sum_{i=1}^m (\ln \lambda_i)^2 \right]^{\frac{1}{2}}, \quad (3)$$

where $\lambda_i, i = 1 \dots m$ are the eigenvalues of $\mathbf{M}_2^{-1} \mathbf{M}_1$.

Even though this method is largely used, it has a major drawback: it is easily influenced by the outliers present in the dataset [14], [15]. In order to reduce the impact of aberrant data on the estimated centroid's value, several possibilities are available. Some authors have proposed in [15], [16] the use of some trimming based methods to remove the outliers before the computation of (2). By deleting the elements that differ from the rest of the dataset, some new ones will become outliers. If the removal procedure is repeated, the dataset may become too small for further reliable analysis. Therefore, a more appropriate solution is the use of robust methods for computing the centroid, like the median [15].

b) *The median* is defined by using the distance function:

$$f_{Med}(\mathbf{M}) = \frac{1}{N} \sum_{i=1}^N d(\mathbf{M}, \mathbf{M}_i). \quad (4)$$

It has to be mentioned that the estimation of the center of mass and the median from a set of covariance matrices have been recently studied in [3], [14], [15].

The center of mass and the median are two extreme solutions: the first one is efficient for datasets with no outliers, while the second one is robust to the presence of aberrant observations. In the following, we propose a trade-off between these two methods by introducing a Huber-like estimator.

B. The Huber's estimator

1) Definition of the Huber's centroid

In this section, we introduce a novel centroid estimator on the manifold of covariance matrices, based on the theory of M-estimators [17], [18], [24]. In this case, the cost function in (1), denoted $f_u(\mathbf{M})$ for the M-estimator, can be expressed by means of a scalar weight function $u(\cdot)$, as follows:

$$f_u(\mathbf{M}) = \frac{1}{N} \sum_{i=1}^N u(d(\mathbf{M}, \mathbf{M}_i)) d^2(\mathbf{M}, \mathbf{M}_i), \quad (5)$$

where $u(\cdot)$ is a positive-valued function which gives a weight to each observation \mathbf{M}_i in the computation of the centroid. Obviously, the weight function $u(\cdot)$ should decrease to zero to ensure that the outliers have a smaller contribution to the centroid's estimate than the other observations. Note that even if the center of mass (2) and the median (4) have

expressions similar to (5) for respectively $u(d(\mathbf{M}, \mathbf{M}_i)) = 1$ and $u(d(\mathbf{M}, \mathbf{M}_i)) = \frac{1}{d(\mathbf{M}, \mathbf{M}_i)}$, they do not belong to the family of M-estimators since the regularity conditions of their corresponding weight function $u(\cdot)$ defined in [24] are not satisfied.

In [17], Huber introduces the so-called Huber's function $u(\cdot)$ defined as:

$$u(d(\mathbf{M}, \mathbf{M}_i)) = \min \left(1, \frac{T}{d(\mathbf{M}, \mathbf{M}_i)} \right) \quad (6)$$

where T is a threshold value controlling the contribution of outliers in the estimation. By combining (5) and (6), the proposed Huber's centroid estimator is the covariance matrix \mathbf{M} , which minimizes the following cost function:

$$f_H(\mathbf{M}) = \frac{1}{N} \sum_{i=1}^N d^2(\mathbf{M}, \mathbf{M}_i) \mathbb{1}_{\{d(\mathbf{M}, \mathbf{M}_i) \leq T\}} + \frac{T}{N} \sum_{i=1}^N d(\mathbf{M}, \mathbf{M}_i) \mathbb{1}_{\{d(\mathbf{M}, \mathbf{M}_i) > T\}}, \quad (7)$$

where $\mathbb{1}_{\{a \leq b\}}$ is the indicator function, which equals 1 if $a \leq b$ and 0 otherwise. Threshold T represents a measure for discriminating between normal and aberrant data and therefore, it controls the estimator's behavior. In other words, for a large value of T , the Huber's estimator behaves as the center of mass, while for a small value it is equivalent to the median.

In this paper, we propose an algorithm to estimate the Huber's centroid by means of a gradient descent algorithm which minimizes the distance function given in (7). The gradient of $f_H(\mathbf{M})$ with respect to \mathbf{M} that is $\nabla(f_H(\mathbf{M}))$ can be written as:

$$\nabla(f_H(\mathbf{M})) = -\frac{2}{N} \sum_{i=1}^N \text{Log}_{\mathbf{M}}(\mathbf{M}_i) \mathbb{1}_{\{d(\mathbf{M}, \mathbf{M}_i) \leq T\}} - \frac{T}{N} \sum_{i=1}^N \frac{\text{Log}_{\mathbf{M}}(\mathbf{M}_i)}{d(\mathbf{M}, \mathbf{M}_i)} \mathbb{1}_{\{d(\mathbf{M}, \mathbf{M}_i) > T\}}, \quad (8)$$

where $\text{Log}_{\mathbf{M}}$ is the Riemannian logarithm mapping [25], [26]. Once that this value is obtained, the centroid can be updated as:

$$\mathbf{M}_{it+1} = \text{Exp}_{\mathbf{M}_{it}}(-s_{it} \nabla(f_H(\mathbf{M}_{it}))), \quad (9)$$

with s_{it} being the descent step and $\text{Exp}_{\mathbf{M}}$ the Riemannian exponential mapping [25], [26]. In practice, the Armijo's backtracking procedure [27] is used to fix s_{it} at each iteration of the algorithm.

This recursive process is repeated as long as the norm of $\nabla(f_H(\mathbf{M}_{it}))$, denoted D_{it} , is greater than a precision parameter ϵ , or until a maximum number of iterations N_{\max} is reached. Practically, D_{it} is given as:

$$D_{it} = \|\nabla(f_H(\mathbf{M}_{it}))\| = \text{tr}((\mathbf{M}_{it}^{-1} \nabla(f_H(\mathbf{M}_{it})))^2). \quad (10)$$

In the end, the Huber's centroid $\widehat{\mathbf{M}}_H$ estimator is obtained. A pseudo-code description of the Huber's centroid estimation is given in Algorithm 1.

As observed in (8), the first and the second terms correspond respectively to the gradient of the cost function for the center

Algorithm 1 Huber's centroid estimator

```

1: Input:  $\mathbf{M}_1, \dots, \mathbf{M}_N, T, \epsilon, N_{\max}$ 
2: Initialize  $\mathbf{M}$  using the sample mean
3:  $it = 1$ 
4: while ( $D_{it} > \epsilon$ ) and ( $it \leq N_{\max}$ ) do
5:   Estimate  $\mathbf{M}$  using one iteration of (9).
6:   Compute the gradient norm,  $D_{it}$ , according to (10).
7:    $it = it + 1$ 
8: end while
9: Output:  $\mathbf{M}$ 

```

of mass and median centroids. For the second term, it can be seen that the division by distance $d(\mathbf{M}_{it}, \mathbf{M}_i)$ is needed. In some cases, that is when an observation \mathbf{M}_i is close to the current centroid's estimate \mathbf{M}_{it} , their distance is close to 0 yielding to potential numerical unsuitability. To avoid this, in [14] the author proposes to exclude, at each iteration it , the observations \mathbf{M}_i that are too close from \mathbf{M}_{it} . By using the proposed Huber's centroid, this problem is solved automatically by considering the threshold T . In conclusion, by choosing an appropriate value for T , the division by zero in the gradient (8) will be avoided, which represents an important advantage of the proposed method.

2) Determination of an automatic Huber's threshold

As explained before, the performance of the Huber's centroid estimator depends greatly on the threshold T that discriminates between aberrant and normal data. There is hence a need to fix it automatically or at least to give an idea of the order of magnitude of T . In practice, T is application dependent and is related to the intrinsic variability of the observed data. By considering first and second order statistics, the Riemannian Gaussian distribution (RGD) has been introduced in [26]. This distribution is characterized by two parameters: the central value $\bar{\mathbf{M}}$ and the dispersion σ . Its probability density function of the RGD is given by

$$p(\mathbf{M}|\bar{\mathbf{M}}, \sigma) = \frac{1}{Z(\sigma)} \exp\left\{-\frac{d^2(\mathbf{M}, \bar{\mathbf{M}})}{2\sigma^2}\right\}, \quad (11)$$

where $Z(\sigma)$ is a normalization factor independent of the centroid $\bar{\mathbf{M}}$, and $d(\mathbf{M}, \bar{\mathbf{M}})$ is the Riemannian distance defined in (3).

In order to estimate the threshold's value, a robust estimator of the dispersion parameter σ is required. Inspired by previous works on robust statistics [28], we propose to extend the concept of median absolute deviation (MAD) to the case of covariance matrices which live in a Riemannian space. The MAD of the set $\mathbf{M}_1, \dots, \mathbf{M}_N$ is defined as the median of the Riemannian distances d computed between each sample \mathbf{M}_i and the Riemannian median (denoted $RMed(\mathbf{M})$):

$$MAD = \text{median}(d(\mathbf{M}_i, RMed(\mathbf{M}))). \quad (12)$$

Once the MAD is computed, the robust estimate $\hat{\sigma}$ of the RGD's dispersion can be obtained as:

$$\hat{\sigma} = \frac{K}{m} \times MAD, \quad (13)$$

where m is the size of covariance matrices and K is a constant depending on the distribution of $d(\mathbf{M}_i, RMed(\mathbf{M}))/\sigma$.

More precisely, K is obtained by studying the statistics of $z = \frac{d(\mathbf{M}, \bar{\mathbf{M}})}{m\sigma}$ since by definition of the MAD, we have:

$$\frac{1}{2} = p(d(\mathbf{M}, \bar{\mathbf{M}}) \leq MAD) = p\left(\frac{d(\mathbf{M}, \bar{\mathbf{M}})}{m\sigma} \leq \frac{MAD}{m\sigma}\right). \quad (14)$$

In practice, it has been observed on simulated data that the distribution of z is independent of $\bar{\mathbf{M}}$ and σ ¹. By combining (13) and (14), the constant $K = 1/(\phi^{-1}(0.5))$, knowing that ϕ^{-1} is the inverse of the cumulative distribution function of z . Experiments have shown that $K \approx 1.312$. Finally, the Huber's threshold is obtained by multiplying the estimated standard deviation $\hat{\sigma}$ by a constant c , which will give $T = c \times \hat{\sigma}$. A common value for c is 1.5 as recommended in [28].

C. Performance Analysis

In the following, the influence of outliers on the proposed Huber's centroid estimator is studied. The obtained results are presented in this section and they are compared to those given by the center of mass and the median.

For this experiment, covariance matrices are generated as realizations of RGDs. For more information concerning the generation of samples from an RGD, the interested reader is referred to section III-A of [26]. In our case, the simulated covariance matrix datasets are obtained for centroids $\bar{\mathbf{M}}$ of size $m \times m$ having the form $\bar{\mathbf{M}}(i, j) = \rho^{|i-j|}$ for $i, j \in \llbracket 1, m \rrbracket$.

Since the centroid is a covariance matrix, the manifold of the space of covariance matrices should be taken into account for the estimators' performance evaluation. In the literature, many authors have proposed to define the concept of intrinsic analysis for statistical estimation [29]–[31]. To this aim, the notions of intrinsic root-mean square error (RMSE) and intrinsic bias vector field have been introduced. We briefly recall here their definitions.

Let $\hat{\mathbf{M}}$ be the estimated centroid of the dataset, that is the estimate of the centroid $\bar{\mathbf{M}}$. The intrinsic RMSE is given by [29]–[31]:

$$RMSE = \sqrt{E[d^2(\hat{\mathbf{M}}, \bar{\mathbf{M}})]}, \quad (15)$$

where $d(\cdot)$ is the Riemannian distance defined in (3). In addition, the bias vector field $\mathbf{b}(\hat{\mathbf{M}})$ of $\hat{\mathbf{M}}$ is given by [29]–[31]:

$$\mathbf{b}(\hat{\mathbf{M}}) = \text{Log}_{\hat{\mathbf{M}}} E_{\hat{\mathbf{M}}}[\hat{\mathbf{M}}] = E[\text{Log}_{\hat{\mathbf{M}}}\hat{\mathbf{M}}], \quad (16)$$

knowing that $E_{\hat{\mathbf{M}}}[\hat{\mathbf{M}}] = \text{Exp}_{\hat{\mathbf{M}}} E[\text{Log}_{\hat{\mathbf{M}}}\hat{\mathbf{M}}]$. Since the bias vector field $\mathbf{b}(\hat{\mathbf{M}})$ in (16) is a covariance matrix, we compute its norm according to (10) to plot it in the following figures.

To study the influence of outliers on the centroid's estimation, a dataset containing 1000 matrices of size 2×2 is created. These matrices have an RGD distribution of dispersion $\sigma = 0.1$ and centroid $\bar{\mathbf{M}}$ obtained for $\rho = 0.7$. To this original data set, some outliers are added. They are i.i.d. covariance matrices samples issued from an RGD of centroid $10 \times M_o$, with M_o obtained for $\rho_o = 0.1$. Here, the dispersion for the outlier samples σ_o is set to 0.1.

¹The use of z is equivalent to the standardization step $z = \frac{x-\mu}{\sigma}$ for a univariate normal distribution.

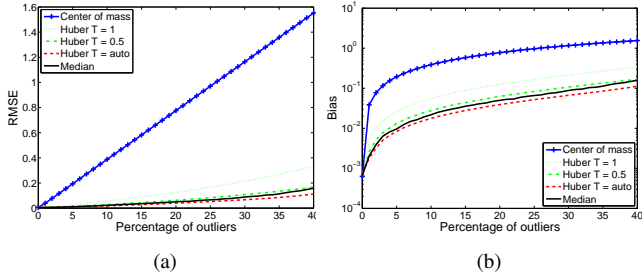


Fig. 1. The RMSE (a) and the bias vector field (b) as functions of the outlier percentage.

Figs. 1 draws the results obtained for the intrinsic RMSE (a) and for the intrinsic bias vector field (b) as functions of the percentage of outliers, knowing that 5000 Monte Carlo runs have been used to evaluate the estimators' performance. The behavior of the center of mass (in blue), the median (in black) and the Huber's centroid with fixed threshold $T = 1$ and $T = 0.5$ (in green) and automatically computed value for T (in red) are analyzed, when the percentage of aberrant data varies from 0 to 40%. As observed, the center of mass is clearly influenced by the presence of outliers while for robust estimators, like the median or the Huber's centroid, this influence is less important.

III. APPLICATION TO MEG BASED BRAIN DECODING

In this section, we apply the proposed centroid estimator to brain decoding, based on MEG data. The database used for the Biomag 2014 Decoding Challenge: Brain Decoding Across Subjects (DecMeg2014) [32] has been considered. The idea of brain decoding consists in predicting the stimulus presented to the subject from the concurrent brain activity [33]. For this experiment, two categories of visual stimulus have been considered: face and scrambled face. Therefore, the problem to solve can be viewed as a two-class classification task. A detailed description of the neuroscientific experiment implemented to collect the data can be found in [34].

The database contains 16 training and 7 testing subjects. For each training subject, approximately 580 trials have been considered, giving a training set of 9414 trials. Next, for each trial, covariance matrices of size 16×16 have been extracted, as described in [35]. Further on, a modified version of the unsupervised classification method presented in [35] has been implemented. First, a regularized logistic regression model has been trained to obtain the initial labels for the unsupervised classification algorithm (k-means). Second, the centroids of each class (face or scrambled face) are computed. For this step, several estimators have been considered: the center of mass, the median, the Huber's estimator with both fixed ($T = 0.2$ and $T = 0.5$) and automatically computed thresholds and also the trimmed based methods [16], when $d = 5\%$ of discarded extreme data. For this latter, only the best result has been retained, that is the mean-based trimmed median. Next, for each testing subject, covariance matrices have been computed and the classification has been performed by two approaches. First, the winner method of the DecMeg2014 competition has been implemented, for which the test trials have been assigned to the closest class, by using the minimum distance to mean

TABLE I
CLASSIFICATION RESULTS FOR MEG BASED BRAIN DECODING.

Estimator	MDM	MGD
CM	74.106	73.845
Med	73.627	74.150
Huber $T = 0.2$	74.847	75.109
Huber $T = 0.5$	74.063	73.976
Huber $T = auto$	74.455	74.106
Trimming ($d = 5\%$) [16]	74.412	74.542

(MDM) Riemannian classifier [36]. Second, the covariance matrices have been modeled as RGDs and each trial has been assigned to the centroid maximizing the log-likelihood criterion derived from (11).

The obtained results are shown in Table I and several remarks can be made. By analyzing the above table, it can be seen that the use of Huber's estimator may increase the classification performance. The obtained values are comparable or higher to those given by the other robust estimators, but without their disadvantages: division by zero for the median, or choice of the percentage d of discarded observation for the trimmed estimators. Interestingly, note that the estimated Huber's threshold T is recomputed at each k-means iteration. And in this experiment, it varies between 0.38 and 0.46 across the test subjects and the classes. Moreover, the proposed estimated value of T by the MAD gives an order of magnitude of the threshold we may consider in the Huber estimation algorithm. This value can be readjusted to improve the classification performance as observed in Table I.

IV. CONCLUSION

In this article, a new method called the Huber's centroid, for the estimation of the central value of a covariance matrix dataset has been introduced. This estimator is a trade-off between the center of mass and the median. The definition of the Huber's centroid and its computational algorithm have been detailed. In addition, an algorithm for choosing the appropriate threshold value for the Huber's estimator has been developed. Further on, the Huber's centroid, has been applied to the case of covariance matrices representing realizations of Riemannian Gaussian distributions. The robustness to outlier values has been studied on simulated data, but also in the context of brain decoding, that is a two-class classification experiment. The results have been compared to those given by two well-known estimators that are the center of mass and the median but also to those given by trimmed based methods.

Further works will include the statistical modeling of $z = d(\mathbf{M}_i, \bar{\mathbf{M}})/m\sigma$ to derive the analytical expression of K . In addition, the proposed centroid will be used to build the codebook for patch-based image classification algorithms.

ACKNOWLEDGMENT

This study has been carried out in the frame of the Investments for the future Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02) of the French National Research Agency (ANR). It has also been supported by the French Foreign Affairs and International Development Ministry and by the Executive Agency for Higher Education, Research, Development and Innovation Funding Romania, under the projects 32619VL and PNII Capacitati 779/27.06.2014.

REFERENCES

- [1] E. Ollila and V. Koivunen, "Robust antenna array processing using M-estimators of pseudo-covariance," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, vol. 3, Sept 2003, pp. 2659–2663.
- [2] L. Yang, M. Arnaudon, and F. Barbaresco, "Riemannian median, geometry of covariance matrices and radar target detection," *European Radar Conference (EuRAD)*, pp. 415–418, 2010.
- [3] F. Barbaresco, M. Arnaudon, and L. Yang, "Riemannian medians and means with applications to radar signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 4, pp. 595–604, 2013.
- [4] R. de Luis-García, C.-F. Westin, and C. Alberola-López, "Gaussian mixtures on tensor fields for segmentation: Applications to medical imaging," *Computerized Medical Imaging and Graphics*, vol. 35, no. 1, pp. 16–30, 01 2011.
- [5] J. Robinson, "Covariance matrix estimation for appearance-based face image processing," *Proceedings of the British Machine Vision Conference 2005*, pp. 389–398, 2005. [Online]. Available: <http://www.intuac.com/userport/john/pubs/covestbmvc.pdf>
- [6] K. Mader and G. Reese, "Using covariance matrices as feature descriptors for vehicle detection from a fixed camera," *ArXiv e-prints*, Feb. 2012.
- [7] P. Formont, F. Pascal, G. Vasile, J. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric SAR images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 567–576, June 2011.
- [8] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *NeuroComputing*, vol. 112, pp. 172–178, 2013.
- [9] S. Said, L. Bombrun, and Y. Berthoumieu, "Texture classification using Rao's distance on the space of covariance matrices," in *Geometric Science of Information (GSI)*, 2015.
- [10] M. Faraki, M. Harandi, and F. Porikli, "More about VLAD: A leap from Euclidean to Riemannian manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015*, June 2015, pp. 4951–4960.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Springer, Oct. 2007.
- [12] B. Afsari, "Riemannian lp center of mass: existence, uniqueness and convexity," *Proceedings of the American Mathematical Society*, vol. 139, no. 2, pp. 655–673, 2011.
- [13] P. Formont, J.-P. Ovarlez, and F. Pascal, "On the use of matrix information geometry for polarimetric SAR image classification," in *Matrix Information Geometry*, F. Nielsen and R. Bhatia, Eds. Springer Berlin Heidelberg, 2013, pp. 257–276.
- [14] L. Yang, "Riemannian median and its estimation," *LMS Journal of Computation and Mathematics*, vol. 13, pp. 461–479, 2010.
- [15] P. T. Fletcher, S. Venkatasubramanian, and S. C. Joshi, "Robust statistics on Riemannian manifolds via the geometric median," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA, 2008*. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587747>
- [16] T. Uehara, T. Tanaka, and S. Fiori, *Advances in Cognitive Neurodynamics (V): Proceedings of the Fifth International Conference on Cognitive Neurodynamics - 2015*. Singapore: Springer Singapore, 2016, ch. Robust averaging of covariance matrices by Riemannian geometry for motor-imagery brain-computer interfacing, pp. 347–353.
- [17] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [18] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 03 1987. [Online]. Available: <http://dx.doi.org/10.1214/aos/1176350263>
- [19] X. Gu, J. Deng, and M. Purvis, "Improving superpixel-based image segmentation by incorporating color covariance matrix manifolds," in *IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 4403–4406.
- [20] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977. [Online]. Available: <http://dx.doi.org/10.1002/cpa.3160300502>
- [21] F. Nielsen and R. Bhatia, *Matrix Information Geometry*. Springer Berlin Heidelberg, 2012.
- [22] S. Fiori, "Learning the Fréchet mean over the manifold of symmetric positive-definite matrices," *Cognitive Computation*, vol. 1, no. 4, pp. 279–291, 2009.
- [23] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, 2012.
- [24] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, Jan. 1976.
- [25] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006. [Online]. Available: <http://dx.doi.org/10.1007/s11263-005-3222-z>
- [26] S. Said, L. Bombrun, and Y. Berthoumieu, "Riemannian Gaussian distributions on the space of symmetric positive definite matrices," available on arxiv via <http://arxiv.org/abs/1507.01760>, 2015.
- [27] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966. [Online]. Available: <http://projecteuclid.org/euclid.pjm/1102995080>
- [28] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. John Wiley & Sons, Inc., 2009.
- [29] J. M. Oller and J. M. Corcuera, "Intrinsic analysis of statistical estimation," *The Annals of Statistics*, vol. 23, no. 5, pp. 1562–1581, 1995. [Online]. Available: <http://dx.doi.org/10.2307/2242534>
- [30] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, May 2005.
- [31] G. Garcia and J. M. Oller, "What does intrinsic mean in statistical estimation?" *Statistics and Operations Research Transactions*, vol. 30, no. 2, pp. 125–170, 2006.
- [32] "DecMeg2014 - Decoding the Human Brain," Biomag 2014 Decoding Challenge: Brain Decoding Across Subjects. Available: <https://www.kaggle.com/c/decoding-the-human-brain>.
- [33] E. Olivetti, S. M. Kia, and P. Avesani, "MEG decoding across subjects," in *International Workshop on Pattern Recognition in Neuroimaging, PRNI, 2014, Tübingen, Germany, June 4-6, 2014*, 2014, pp. 1–4.
- [34] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston, "A parametric empirical bayesian framework for the EEG/MEG inverse problem: generative models for multisubject and multimodal integration," *Frontiers in Human Neuroscience*, vol. 5, no. 76, 2011.
- [35] A. Barachant, "MEG decoding using Riemannian geometry and unsupervised classification," *Technical Report*. [Online]. Available: <http://alexandre.barachant.org/wp-content/uploads/2014/08/documentation.pdf>
- [36] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by riemannian geometry," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.