



HAL
open science

A GEOMETRIC LEARNING APPROACH ON THE SPACE OF COMPLEX COVARIANCE MATRICES

Hatem Hajri, Salem Said, Lionel Bombrun, Yannick Berthoumieu

► **To cite this version:**

Hatem Hajri, Salem Said, Lionel Bombrun, Yannick Berthoumieu. A GEOMETRIC LEARNING APPROACH ON THE SPACE OF COMPLEX COVARIANCE MATRICES. 2017. hal-01483449

HAL Id: hal-01483449

<https://hal.science/hal-01483449>

Preprint submitted on 5 Mar 2017

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 GEOMETRIC LEARNING APPROACH ON THE SPACE OF COMPLEX COVARIANCE MATRICES

Hatem Hajri, Salem Said, Lionel Bombrun and Yannick Berthoumieu

Laboratoire IMS (CNRS-UMR 5218), Université de Bordeaux, France

ABSTRACT

Many signal and image processing applications, including SAR polarimetry and texture analysis, require the classification of complex covariance matrices. The present paper introduces a geometric learning approach on the space of complex covariance matrices based on a new distribution called Riemannian Gaussian distribution. The proposed distribution has two parameters, the centre of mass \bar{Y} and the dispersion parameter σ . After having derived its maximum likelihood estimator and its extension to mixture models, we propose an application to texture recognition on the VisTex database.

Index Terms— Classification, Hermitian matrices, Information geometry, Riemannian centre of mass, EM algorithm.

1. INTRODUCTION

Covariance matrices are used as features for many signal and image processing applications, including biomedical image segmentation [1], radar detection [2, 3], texture analysis and scene recognition [4–6], etc. These covariance matrices are used to model different kinds of dependence such as the spatial, spectral and temporal ones to cite a few of them. An interesting space of covariance matrices is the manifold \mathcal{H}_m of $m \times m$ Hermitian positive definite matrices. This latter is involved for example in radar detection, polarimetric SAR image classification [7], texture recognition [8], etc.

Many works have been dedicated in the literature for the statistical modelling of covariance matrices. Due to its mathematical tractability, the Wishart distribution is certainly the most largely used model in the literature [9, 10]. Nevertheless, this model assumes Gaussian statistics for the observations which may not be realistic in practice. More advanced models have hence been proposed based on the so-called scalar product model. These compound models include the \mathcal{K} [11], \mathcal{G}^0 [12] and KummerU [13] distributions. They have shown promising results notably for the classification of high resolution polarimetric SAR images. Inspired from clustering approaches on Riemannian manifolds [14, 15], there is another

way to model covariance matrices. By considering Rao's distance on the manifold of covariance matrices, there is a canonical way to define the mean or barycentre of several covariance matrices in this manifold. Based on this concept, the Riemannian Gaussian distribution (RGD) has been introduced to model the statistical variability of real covariance matrices [16]. This distribution is represented by two parameters: its unique mode \bar{Y} (the unique Riemannian centre of mass) and σ its dispersion away from \bar{Y} . An important property of the RGD is that the maximum likelihood estimates of its parameters admit simple expressions and can be computed numerically using different algorithms. However, this RGD has only been proposed on the space \mathcal{P}_m of $m \times m$ real symmetric and strictly positive definite matrices. The main contribution of the paper is to extend this RGD to the space of Hermitian positive definite matrices. Moreover, based on mixture models of RGDs, a new geometric learning approach on \mathcal{H}_m is proposed. This approach is highlighted by an application on the classification of texture images.

The paper is structured as follows. Section 2 recalls some elements on the Riemannian geometry of \mathcal{H}_m . Section 3 introduces the proposed Riemannian Gaussian distribution on \mathcal{H}_m and its extension to mixture models. An expectation-maximization (EM) algorithm is presented to estimate the distribution parameters. Section 4 presents an application to image classification on the VisTex database. Some comparisons with the Wishart model are given to evaluate the potential of the proposed model in the context of texture image classification. Conclusions and future works are finally discussed in Section 5. Note that due to the restriction length, most of the mathematical proofs can not be detailed here and will be given in a forthcoming journal paper.

2. RIEMANNIAN GEOMETRY OF \mathcal{H}_m

This section recalls basic facts on the Riemannian geometry of \mathcal{H}_m , the space of Hermitian positive definite matrices of size $m \times m$. In many applications, \mathcal{H}_m is considered as a Riemannian manifold, equipped with the so-called affine-invariant Riemannian metric. For $Y \in \mathcal{H}_m$, this metric is given by the scalar product

$$g_Y(A, B) = \text{tr}(Y^{-1}AY^{-1}B) \quad (1)$$

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02).

where A, B are two Hermitian matrices of size $m \times m$. The Riemannian distance associated to g is the following

$$d(Y, Z) = \sqrt{\text{tr} [\log (Y^{-1/2} Z Y^{-1/2})]^2} \quad (2)$$

For more details, the interested reader is referred to [17–19].

The reason for using the Riemannian metric (1) with the resulting Riemannian distance (2) is that these provide particularly suitable mathematical properties [17, 18, 20] which are

- (i) invariance by all affine transformations.
- (ii) existence and uniqueness of Riemannian barycentres.

First, let us consider the property of invariance by affine transformations [17, 18]. An affine transformation of \mathcal{H}_m is a mapping $Y \mapsto G \cdot Y$, where G is an invertible complex matrix of size $m \times m$. This is given by

$$G \cdot Y = G Y G^* \quad (3)$$

where $*$ denotes the conjugate transpose. With regard to the Riemannian distance (2), invariance by affine transformations means that for all $Y, Z \in \mathcal{H}_m$

$$d(Y, Z) = d(G \cdot Y, G \cdot Z) \quad (4)$$

where $G \cdot Y$ and $G \cdot Z$ are defined by (3).

Affine transformations are of special importance in the study of the space \mathcal{H}_m since for any $Y \in \mathcal{H}_m$ there exists G such that $Y = G \cdot I_m$ where $I_m \in \mathcal{H}_m$ is the identity matrix. Replacing this transformation in (4) gives the following identity

$$d(Y, Z) = d(I_m, G^{-1} \cdot Z) \quad (5)$$

This will be used in the next section.

Consider the second property of existence and uniqueness of Riemannian barycentres [20]. The barycentre of N points $Y_1, \dots, Y_N \in \mathcal{H}_m$ is defined to be

$$\hat{Y}_N = \arg \min_Y \sum_{n=1}^N d^2(Y, Y_n) \quad (6)$$

where $d(Y, Y_n)$ is the Riemannian distance (2).

As a Riemannian manifold with non-positive sectional curvature [21], \mathcal{H}_m enjoys the important property of existence and uniqueness of barycentres as defined by (6) [20].

Before going on, let us consider the concept of invariant integration on \mathcal{H}_m . The invariant integral of a function $f : \mathcal{H}_m \rightarrow \mathbb{R}$ is given by

$$\int_{\mathcal{H}_m} f(Y) dv(Y) = \int \dots \int f(Y) \det(Y)^{-m} \prod_{i \leq j} d\text{Re}Y_{ij} \prod_{i < j} d\text{Im}Y_{ij} \quad (7)$$

where the subscripts i and j denote matrix elements, while Re and Im denote respectively the real and imaginary parts. In (7), $dv(Y)$ is the Riemannian volume on \mathcal{H}_m defined by

$$dv(Y) = \det(Y)^{-m} \prod_{i \leq j} d\text{Re}Y_{ij} \prod_{i < j} d\text{Im}Y_{ij} \quad (8)$$

The integral (7) is called an invariant integral because of the following property

$$\int_{\mathcal{H}_m} f(Y) dv(Y) = \int_{\mathcal{H}_m} f(G \cdot Y) dv(Y) \quad (9)$$

where $G \cdot Y$ is the affine transformation given by (3) and G is any invertible complex matrix.

The invariant integral (7) takes on a simplified form when the function f is a central function. This means that $f(Y)$ only depends on the eigenvalues of Y . Precisely, let the spectral decomposition of Y be given by $Y = U \text{diag}(e^{r_1}, \dots, e^{r_m}) U^*$, where U is a unitary matrix and e^{r_1}, \dots, e^{r_m} are the eigenvalues of Y . The function f is a central function if $f(Y) = f(r_1, \dots, r_m)$.

In this case, the invariant integral (7) reduces to

$$C \times \int_{\mathbb{R}^m} f(r_1, \dots, r_m) \prod_{i < j} \sinh^2 \left(\frac{|r_i - r_j|}{2} \right) dr \quad (10)$$

where $dr = dr_1 \dots dr_m$ and C is a constant [17].

3. RIEMANNIAN GAUSSIAN DISTRIBUTIONS ON \mathcal{H}_m

This section contains the definition and the main statistical properties of the Riemannian Gaussian distribution (RGD) on \mathcal{H}_m [17]. This distribution is introduced to give a statistical interpretation of the Riemannian barycentre on \mathcal{H}_m , which is used in many signal and image processing applications [18, 19, 22].

3.1. Definition

The probability density function of the RGD on \mathcal{H}_m , with respect to the Riemannian volume element (8), is given by

$$p(Y | \bar{Y}, \sigma) = \frac{1}{\zeta(\sigma)} \exp \left[-\frac{d^2(Y, \bar{Y})}{2\sigma^2} \right] \quad (11)$$

where $\bar{Y} \in \mathcal{H}_m$ and $\sigma > 0$. In this definition, $\zeta(\sigma)$ is the normalizing constant given by

$$\zeta(\sigma) = \int_{\mathcal{H}_m} \exp \left[-\frac{d^2(Y, \bar{Y})}{2\sigma^2} \right] dv(Y) \quad (12)$$

The fact that it does not depend on \bar{Y} can be derived from (5), by taking $Z = \bar{Y}$, and by replacing it in (9). Moreover, replacing \bar{Y} with the identity matrix I_m in (12) and using (10), $\zeta(\sigma)$ can be expressed as

$$\zeta(\sigma) = C \times \int_{\mathbb{R}^m} e^{-|r|^2/2\sigma^2} \prod_{i < j} \sinh^2 (|r_i - r_j|/2) dr \quad (13)$$

where $|r|$ is the Euclidean norm of r and C is the constant appearing in (10). In practice, the function $\zeta(\sigma)$ can be computed by Monte Carlo integration. Interestingly, for $m = 2$, $\zeta(\sigma)$ admits a closed-form expression which is proportional to $\sigma^2(e^{\sigma^2} - 1)$. In the following, the RGD (11) will be denoted by $\mathcal{G}(\bar{Y}, \sigma)$.

Algorithm 1 below describes a simple way to sample from $\mathcal{G}(\bar{Y}, \sigma)$. Note that step 2 can be achieved by a Metropolis-Hastings algorithm.

Algorithm 1 Sampling from $\mathcal{G}(\bar{Y}, \sigma)$

- 1: Generate U from a uniform distribution on the group of unitary matrices of size $m \times m$
 - 2: Generate $(r_1, \dots, r_m) \in \mathbb{R}^m$ from the distribution $\propto e^{-|r|^2/2\sigma^2} \prod_{i < j} \sinh^2(|r_i - r_j|/2)$
 - 3: $X \leftarrow U \cdot \text{diag}(e^{r_1}, \dots, e^{r_m})$
 - 4: $Y \leftarrow \bar{Y}^{1/2} \cdot X$
-

3.2. Maximum likelihood estimates of \bar{Y} and σ

Let Y_1, \dots, Y_N be N i.i.d samples from $\mathcal{G}(\bar{Y}, \sigma)$ and let \hat{Y}_N be the barycentre of Y_1, \dots, Y_N . Based on these observations, the maximum likelihood estimate (MLE) of \bar{Y} is the Riemannian barycentre \hat{Y}_N . Moreover, the MLE of σ is the solution $\hat{\sigma}_N$ of

$$\sigma^3 \times \frac{d}{d\sigma} \log \zeta(\sigma) = \frac{1}{N} \sum_{n=1}^N d^2(\hat{Y}_N, Y_n)$$

It can be proved that both \hat{Y}_N and $\hat{\sigma}_N$ exist and are unique [17].

For the main application of this paper, numerical computations of \hat{Y}_N and $\hat{\sigma}_N$ are needed. Computation of \hat{Y}_N consists in finding the Riemannian barycentre of Y_1, \dots, Y_N , defined by (6). This can be done using the Riemannian gradient descent algorithm, which may be found in [23]. The k^{th} iteration of this algorithm produces an approximation \hat{Y}_N^k of \hat{Y}_N in the following way. For $k = 1, 2, \dots$, let Δ_k be the Hermitian matrix $\Delta_k = \sum_{n=1}^N \text{Log}_{\hat{Y}_N^{k-1}}(Y_n)$ where Log denotes the Riemannian logarithm mapping, inverse to the exponential mapping, given by

$$\text{Exp}_Y(\Delta) = Y^{1/2} \exp\left(Y^{-1/2} \Delta Y^{-1/2}\right) Y^{1/2}$$

Then, \hat{Y}_N^k is defined to be $\hat{Y}_N^k = \text{Exp}_{\hat{Y}_N^{k-1}}(\tau_k \Delta_k)$ where $\tau_k > 0$ is a step size, which can be determined using a backtracking procedure.

The Riemannian gradient descent algorithm is repeated as long as $\|\Delta_k\| > \epsilon$, where $\|\Delta_k\| = \sqrt{g_{\hat{Y}_N^{k-1}}(\Delta_k, \Delta_k)}$ and ϵ is a precision parameter. In practice ϵ is set to the machine precision. This algorithm is guaranteed to converge, when

a suitable backtracking procedure is used, regardless of the initialization \hat{Y}_N^0 .

On another hand, computation of $\hat{\sigma}_N$ requires solving a non-linear equation with one variable. This is readily done using Newton's method.

3.3. EM algorithm for mixture estimation

Mixture distributions are effective tools in statistical learning problems as they provide a good approximation to most distributions. The present paper applies this approach to learn distributions on \mathcal{H}_m . A mixture of RGD on \mathcal{H}_m is any probability distribution with density, with respect to the volume element (8), of the form

$$p(Y | (\varpi_\mu, \bar{Y}_\mu, \sigma_\mu)_{1 \leq \mu \leq M}) = \sum_{\mu=1}^M \varpi_\mu \times p(Y | \bar{Y}_\mu, \sigma_\mu) \quad (14)$$

where $\varpi_\mu \in (0, 1)$ are weights with sum one and where each $p(Y | \bar{Y}_\mu, \sigma_\mu)$ is a RGD on \mathcal{H}_m given by (11).

Let Y_1, \dots, Y_N be independent samples from (14). Based on these samples, the MLE of $\vartheta = \{(\varpi_\mu, \bar{Y}_\mu, \sigma_\mu)\}$ can be computed using an EM algorithm similar to [16, 17]. For this, define for all $\vartheta = \{(\varpi_\mu, \bar{Y}_\mu, \sigma_\mu)\}$,

$$\omega_\mu(Y_n, \vartheta) = \frac{\varpi_\mu \times p(Y_n | \bar{Y}_\mu, \sigma_\mu)}{\sum_{s=1}^M \varpi_s \times p(Y_n | \bar{Y}_s, \sigma_s)}, \quad N_\mu(\vartheta) = \sum_{n=1}^N \omega_\mu(Y_n)$$

The EM algorithm iteratively updates $\hat{\vartheta} = \{(\hat{\varpi}_\mu, \hat{Y}_\mu, \hat{\sigma}_\mu)\}$, which is an approximation of the MLE of $\vartheta = (\varpi_\mu, \bar{Y}_\mu, \sigma_\mu)$ as follows. Based on the current value of $\hat{\vartheta}$, it updates $\hat{\varpi}_\mu, \hat{Y}_\mu$ and $\hat{\sigma}_\mu$ as follows:

- Assign to $\hat{\varpi}_\mu$ the value $\hat{\varpi}_\mu = N_\mu(\hat{\vartheta})/N$
- Assign to \hat{Y}_μ the value

$$\hat{Y}_\mu = \arg \min_Y \sum_{n=1}^N \omega_\mu(Y_n, \hat{\vartheta}) d^2(Y, Y_n)$$

- Assign to $\hat{\sigma}_\mu$ the value

$$\hat{\sigma}_\mu = \Phi \left(N_\mu^{-1}(\hat{\vartheta}) \times \sum_{n=1}^N \omega_\mu(Y_n, \hat{\vartheta}) d^2(\hat{Y}_\mu, Y_n) \right)$$

where the function Φ is the inverse of $\sigma \mapsto \sigma^3 \times \frac{d}{d\sigma} \log \zeta(\sigma)$.

Consider now the problem of classifying new observations modeled as a realization of (14). Following [16, 17], a new observation Y_t is associated to the class C_{μ^*} of the closest cluster, i.e.

$$\mu^* = \underset{\mu}{\text{argmin}} \left\{ -\log \hat{\varpi}_\mu + \log \zeta(\hat{\sigma}_\mu) + \frac{d^2(Y_t, \hat{Y}_\mu)}{2\hat{\sigma}_\mu^2} \right\}$$

This classification rule is optimal in the sense of a Bayesian risk criterion given in [24].

4. APPLICATION TO TEXTURE IMAGE CLASSIFICATION

This section presents an application of the RGD on \mathcal{H}_m for the classification of texture images. The performance of this distribution is first compared with the RGD on \mathcal{P}_m . Second, it is compared with Wishart distribution. Third, the influence of the wavelet decomposition on the classification performance is evaluated. For this purpose, the following numerical experiment on the VisTex database [25] is carried out.

For each texture image of size 512×512 pixels, 196 subimages (128×128) are extracted with a 32 pixel overlap. For each subimage, 6 subbands were computed using the complex Gabor filter. Then, two spatial neighborhoods (i.e. horizontal dH and vertical dV) of one pixel are considered. Each subband s of subimage n is composed by a set of two bivariate populations $\Pi_{s,n,dH}$ and $\Pi_{s,n,dV}$, represented respectively by their Hermitian covariance matrices $Y_{s,n,dH}$ and $Y_{s,n,dV} \in \mathcal{H}_2$. These covariance matrices are features which describe the spatial dependence of subband coefficients. The size of the feature space is hence $F = 12$ (2 spatial neighborhood and 6 subbands). To summarize, each subimage n is represented by a set of F covariance matrices denoted $Y_{f,n}$. Then, the database is equally divided for the classification, where half is used for training and the other half for testing. For each training class, a mixture model is learned by considering the EM algorithm described in Section 3.3. In this experiment, the number of clusters per class (denoted K) is set to 1, 3 or automatically fixed by optimizing the Bayesian information criterion (BIC) [26]. It results that each texture is represented by a set of K clusters containing the maximum likelihood estimates $(\hat{\omega}_a, \hat{Y}_{f,a}, \hat{\sigma}_{f,a})$, for $a = 1, \dots, K$ and $f = 1, \dots, F$. Then, as the subband are assumed independent, a test subimage t represented by a set of F covariance matrices $Y_{f,t}$ in \mathcal{H}_2 is associated to the class of the closest cluster C_* , i.e. the one realizing the minimum over c of

$$-\log \hat{\omega}_c + \sum_{f=1}^F \log Z(\hat{\sigma}_{f,c}) + \sum_{f=1}^F \frac{d^2(Y_{f,t}, \hat{Y}_{f,c})}{2\hat{\sigma}_{f,c}^2}$$

Table 1 displays the classification performance in terms of kappa accuracy (mean \pm standard deviation) on the VisTex database. The first line gives the results for the RGD with respectively $K = 1$, $K = 3$ and K fixed by the BIC criterion. The second line shows the corresponding results for the Wishart model. Moreover, the second column in Table 1 shows the results with the Hermitian covariance matrices in \mathcal{H}_m , while the first column exhibits the results obtained for real covariance matrices (obtained by taking the absolute value of the Gabor subbands) in \mathcal{P}_m .

As shown in Table 1, for each model, a significant gain of more than 2% is observed by working on \mathcal{H}_m rather than \mathcal{P}_m . Note also, that the use of the BIC criterion improved the classification performance compared to a fixed number of clusters

model	number of modes (K)	on \mathcal{P}_m	on \mathcal{H}_m
RGD	K=1	78.24 \pm 0.36	80.62 \pm 0.46
	K=3	90.24 \pm 0.40	92.24 \pm 0.32
	BIC	91.91 \pm 0.45	94.29 \pm 0.30
Wishart	K=1	81.93 \pm 0.51	82.15 \pm 0.50
	K=3	86.07 \pm 0.61	87.27 \pm 0.71
	BIC	86.14 \pm 0.44	87.64 \pm 1.06

Table 1: Classification performance (κ accuracy) on the VisTex database.

Wavelet decomposition	RGD	Wishart
Gabor	94.29 \pm 0.30	87.64 \pm 1.06
CWT	98.03 \pm 0.25	97.61 \pm 0.26
DT-CWT	98.79 \pm 0.17	98.32 \pm 0.30

Table 2: Influence of the wavelet decomposition on the classification performance (κ accuracy) on the VisTex database.

($K = 1$ or 3). Finally, while having the same complexity in terms of number of parameters, the mixture models for the proposed RGD have a significant gain of 4% to 7% compared to the mixture of Wishart model.

A second experiment is carried out on the same database to evaluate the influence of the wavelet decomposition on the classification performance. For that, three decompositions have been considered, which are the Gabor filters, the complex wavelet transform (CWT) and the dual-tree complex wavelet transform (DT-CWT). Here, the number of modes is fixed by optimizing the BIC criterion. As observed in Table 2, the best results are obtained with the DT-CWT. Note also that for each wavelet decomposition, the proposed RGD model leads to better performances compared to the Wishart model. This illustrates the potential of the proposed geometric learning approach on \mathcal{H}_m .

5. CONCLUSION

In this paper, the Riemannian Gaussian distribution (RGD) has been introduced on the space \mathcal{H}_m of Hermitian positive definite matrices. This distribution has a one to one similarity with the RGD on the space \mathcal{P}_m , of $m \times m$ positive definite matrices, previously introduced in [16]. The proposed RGD has two parameters: the unique Riemannian centre of mass \bar{Y} and the dispersion parameter σ . Once having presented the maximum likelihood estimators and the extension of the RGD to mixture models, an experiment on the VisTex database has been carried out to illustrate the potential of the proposed model for texture image classification. A significant gain has been observed compared to the state-of-the-art Wishart model.

Further works will focus on the generalization of the RGD model to structured covariance matrices (e.g. Toeplitz, block Toeplitz matrices).

6. REFERENCES

- [1] 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.
- [2] F. Gini and M. V. Greco, “Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter,” *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [3] Le Yang, M Arnaudon, and F Barbaresco, “Riemannian median, geometry of covariance matrices and radar target detection,” *European Radar Conference (EuRAD)*, pp. 415–418, 2010.
- [4] O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via classification on riemannian manifolds,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, Oct 2008.
- [5] K. Mader and G. Reese, “Using covariance matrices as feature descriptors for vehicle detection from a fixed camera,” *ArXiv e-prints*, Feb. 2012.
- [6] M. Faraki, M.T. 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.
- [7] J.S. Lee, M.R. Grunes, T.L. Ainsworth, L.J. Du, D.L. Schuler, and S.R. Cloude, “Unsupervised classification using polarimetric decomposition and the complex Wishart classifier,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2249–2258, 1999.
- [8] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.
- [9] J. Wishart, “The generalised product moment distribution in samples from a normal multivariate population,” *Biometrika*, vol. 20A, no. 1/2, pp. 32–52, 1928.
- [10] N. R. Goodman, “Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction),” *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 03 1963.
- [11] J.S. Lee, D.L. Schuler, R.H. Lang, and K.J. Ranson, “K-distribution for multi-look processed polarimetric SAR imagery,” in *Geoscience and Remote Sensing, IGARSS '94, Pasadena, California, United States*, 1994, pp. 2179–2181.
- [12] C.C. Freitas, A.C. Frery, and A.H. Correia, “The polarimetric G distribution for SAR data analysis,” *Environmetrics*, vol. 16, pp. 13–31, 2005.
- [13] L. Bombrun and J.-M. Beaulieu, “Fisher distribution for texture modeling of polarimetric SAR data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 512–516, July 2008.
- [14] 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, April 2012.
- [15] F. Nielsen, *Pattern learning and recognition on statistical manifolds: an information-geometric review*, pp. 1–25, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [16] S. Said, L. Bombrun, Y Berthoumieu, and J. H. Manton, “Riemannian Gaussian distributions on the space of symmetric positive definite matrices,” *available on arxiv via <http://arxiv.org/abs/1507.01760>*, 2015.
- [17] S. Said, H. Hajri, L Bombrun, and Baba C. Vemuri, “Gaussian distributions on Riemannian symmetric spaces: statistical learning with structured covariance matrices,” *available on arxiv via <https://arxiv.org/pdf/1607.06929.pdf>*, 2016.
- [18] M. Charon and F. Barbaresco, “Une nouvelle approche pour la détection de cibles dans les images radar basée sur des distances et moyennes dans des espaces et moyennes dans des espaces de matrices de covariance,” *Traitement du signal*, vol. 26, no. 4, pp. 269–278, 2009.
- [19] X. Pennec, P. Fillard, and N. Ayache, “A Riemannian framework for tensor computing,” *Int. J. Comput. Vision*, vol. 66, no. 1, pp. 41–66, 2006.
- [20] B. Afsari, “Riemannian L^p center of mass: existence, uniqueness and convexity,” *Proc. Amer. Math. Soc.*, vol. 139, no. 2, pp. 655–673, 2011.
- [21] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, American Mathematical Society, 2001.
- [22] Z. Chebbi and M. Moakher, “Means of Hermitian positive-definite matrices based on the log-determinant α -divergence function,” *Linear Algebra and its Applications*, vol. 436, no. 7, pp. 1872–1889, 2012.
- [23] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras, “Statistics on the manifold of multivariate Normal distributions,” *J. Math. Imaging Vis.*, vol. 25, no. 3, pp. 423–444, 2006.
- [24] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2009.
- [25] “VisTex: Vision texture database,” MIT Media Lab Vision and Modeling group, 1995.
- [26] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.