Classification approach based on the product of riemannian manifolds from Gaussian parametrization space

Yannick Berthoumieu, Lionel Bombrun, Christian Germain, Salem Said

HAL Id: hal-01716919
https://hal.science/hal-01716919v1

Submitted on 25 Feb 2018

CLASSIFICATION APPROACH BASED ON THE PRODUCT OF RIEMANNIAN MANIFOLDS FROM GAUSSIAN PARAMETRIZATION SPACE

Yannick Berthoumieu[1,2], Lionel Bombrun[2], Christian Germain[2] and Salem Said[2],

Laboratoire IMS, Signal and Image Processing Group,
[1] Institut Polytechnique de Bordeaux, [2] Université de Bordeaux, France.

ABSTRACT

This paper presents a novel framework for visual content classification using jointly local mean vectors and covariance matrices of pixel level input features. We consider local mean and covariance as realizations of a bivariate Riemannian Gaussian density lying on a product of submanifolds. We first introduce the generalized Mahalanobis distance and then we propose a formal definition of our product-spaces Gaussian distribution on $\mathbb{R}^m \times \mathbb{SPD}(m)$. This definition enables us to provide a mixture model from a mixture of a finite number of Riemannian Gaussian distributions to obtain a tractable descriptor. Mixture parameters are estimated from training data by exploiting an iterative Expectation-Maximization (EM) algorithm. Experiments in a texture classification task are conducted to evaluate this extended modeling on several color texture databases, namely popular Vistex, 167-Vistex and CUReT. These experiments show that our new mixture model competes with state-of-the-art on the experimented datasets.

Index Terms— Classification, image local descriptors, generalized Mahalanobis distance, Product-spaces Riemannian Gaussian Mixture density.

1. INTRODUCTION

To solve various problems in image processing, multivariate Gaussian measures, or Gaussian laws, are often assumed to provide efficient solutions. Moreover, mixtures of Gaussian laws are considered as universal approximators of densities as long as there are enough Gaussian components. In the context of image or video indexing scheme, multivariate Gaussian probability distributions are widely used. Parametrized by mean vectors and covariance matrices, mixtures from random feature vectors associated to Fisher kernels have demonstrated significantly better performance in comparison to the Bag-of-Words approach [1]. Recent works have proposed to mix efficiently Fisher Vectors and Convolutional Neural Networks [2].

Beyond direct Gaussian measures, a trend in image processing over the years has emerged. It consists in modeling the image content with localized structured descriptors in the form of region covariances, i.e. symmetric positive definite (SPD) matrices or local covariance matrices (LCM). Various authors have shown the interest of the second-order statistical structure *i.e* covariance matrices. This approach has been proven to provide powerful representations for various image processing task including object or texture recognition [3]-[4], face recognition [5]-[6], human detection and tracking [7], visual surveillance [8]-[9]. In all these works, the main issue is how to take into account in intrinsic geometry of LCM space, i.e. the Riemannian manifold of SPD, to provide geometric means, k-means clustering or dictionary learning to classify the visual information. In [10], authors propose the intrinsic definition of the Riemannian Gaussian distributions for SPD matrices. The well-founded definition of the Riemannian probability density function enables the derivation and implementation of an expectation-maximization (EM) algorithm for the parameter estimation of mixture of Riemannian Gaussian distributions. The availability of the Riemannian Gaussian mixture model has led to an extension of Fisher vectors to the Riemannian case for texture and image classifications or indexing based on LCM descriptors [11].

The above works are mainly focused on the statistical analysis of LCM set for modeling the second order moment variability from local centered descriptors of the visual content. The considered descriptors are usually defined from detail wavelet coefficients or spatial derivative components for which, by construction, the multivariate mean is zero. In the context of full local Gaussian descriptors, some works consider jointly the LCM and local mean vector (LMV) descriptors to increase the capability of the image representation within the classification task [12]-[13]. For instance, LMV descriptors allow us to take into account the color mean to discriminate the visual content. In referenced works, some authors have proposed to transform local Gaussian model to an augmented SPD matrix based on the formal work described in [14]-[15]. This structured compact form, storing both the LMV and LCM descriptors in a unique SPD matrix, gives the

opportunity to exploit directly the manifold geometry of SPD matrices to perform the classification task. Considering the fusion approach using the approach proposed in [14]–[15], the main limitation of this compact representation consists in the ability to describe the variability of the visual content to a unique space, i.e. the SPD Riemannian manifold. In fact, if we consider the LMV and LCM descriptors as a pair of two random variables, the classification task can be viewed as an inference issue within a product of manifolds. Each manifold is geometrically specified by the underlying random variable, i.e. Euclidean and SPD matrix manifolds in the case of the localized structured descriptors based on the mean vector and covariance matrix.

In the present paper, we propose to exploit the joint probabilistic modeling of the LMV and LCM descriptors for supervised classification task of image databases. The main contribution here deals with the extension of the Riemannian Gaussian distribution considering the random pair defined jointly by the mean vector and a covariance matrix. The random pair relies on the concept of product of spaces which correspond respectively to the Euclidean one for the mean vectors and to the space of SPD matrices for the covariance matrices. Moreover, based on this new joint Gaussian modeling for the augmented Riemannian space, we derive mixture models in order to develop dictionary learning method for classification task.

The paper is structured as follows. Section 2 gives a brief introduction on the concept of metric in the case of product of submanifolds. We introduce the generalized Mahalanobis distance leading to the definition of our product-spaces Gaussian distribution. Section 3 provides the definition of the corresponding mixture models. An expectation-maximization (EM) algorithm is presented to estimate the distribution parameters. Section 4 presents an application to image classification on various databases. We provide some comparisons with previous models in order 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.

2. PRODUCT OF SUBMANIFOLDS AND GAUSSIAN MODEL

In this section, we state the definition of the Riemannian Gaussian density for a set of several random variables. We consider the case of independent variables for which each variable belongs to a specific manifold.

2.1. Random Variables on the Riemannian Manifold $\mathcal{H}_m$

Let $\mathcal{H}_m$ be a complete and simply connected m-dimensional Riemannian manifold. Let $Y \in \mathcal{H}_m$ be a random variable

with probability density function $p(y)$. We have

$$\int_{\mathcal{H}_m} p(y)\, dv(y) = 1 \qquad (1)$$

where $v(y)$ is the volume measure induced by the Riemannian metric of $\mathcal{H}_m$. For some smooth functions $\mathcal{F}(Y)$ and if the density function exists, the expected value is defined as

$$\mathcal{E}[F(Y)] = \int_{\mathcal{H}_m} \mathcal{F}(y)\, p(y)\, dv(y). \qquad (2)$$

Numerous works have considered the question of moment estimation for different metrics such as Fisher-Rao or the Log-Euclidean one for the space of the SPD matrices $\mathbb{SPD}(m)$ [16]-[17]-[18]. More specifically, in [19]-[10], the Riemannian Gaussian distribution based on the Fisher-Rao metric is defined by

$$p_\Sigma(\Sigma|\,\bar{\Sigma}, \sigma) = \frac{1}{Z(\sigma)} \exp\left[-\frac{d_{FR}^2(\Sigma, \bar{\Sigma})}{2\sigma^2}\right] \qquad (3)$$

where $\bar{\Sigma} \in \mathbb{SPD}(m)$ and $\sigma > 0$ define respectively the mean and the dispersion. In this definition, $Z(\sigma)$ is the normalizing factor given by

$$Z(\sigma) = \int_{\mathbb{SPD}(m)} \exp\left[-\frac{d_{FR}^2(\Sigma, \bar{\Sigma})}{2\sigma^2}\right] dv(\Sigma). \qquad (4)$$

The Riemannian distance associated to the Fisher-Rao metric is the following

$$d_{FR}(\Sigma, \bar{\Sigma}) = \sqrt{\operatorname{tr}\left[\log\left(\Sigma^{-1/2}\bar{\Sigma}\Sigma^{-1/2}\right)\right]^2} \qquad (5)$$

2.2. Product of spaces and Generalized Mahalanobis Distance

Let $Y = (Y_1, Y_2) \in \mathcal{H}_m$ be a variable composed of two independent terms Y_1 and Y_2. In the Riemannian geometry, some well-known results concern the product of spaces and the associated metric. In [20], the twisted product manifolds is defined as follows: let $(\mathcal{H}_n^1, g_1)$ and $(\mathcal{H}_p^2, g_2)$ be two irreducible simply connected manifolds specified respectively by two Riemannian metrics with $m = n + p$. Let $\pi_1 : (\mathcal{H}_n^1 \times \mathcal{H}_p^2) \to \mathcal{H}_n^1$ and $\pi_2 : (\mathcal{H}_n^1 \times \mathcal{H}_p^2) \to \mathcal{H}_p^2$ be the canonical projections and let $\rho_1 : (\mathcal{H}_n^1 \times \mathcal{H}_p^2) \to [0, +\infty)$ and $\rho_2 : (\mathcal{H}_n^1 \times \mathcal{H}_p^2) \to [0, +\infty)$ be positive differentiable functions. The product manifold $\mathcal{H}_m = \mathcal{H}_n^1 \times \mathcal{H}_p^2$ is equipped with the metric tensor g given by

$$g = \rho_1^2 \pi_1^* g_1 + \rho_2^2 \pi_2^* g_2. \qquad (6)$$

The definition (6) underlines a direct analogy to the generic Mahalanobis distance. The quantities ρ_1 and ρ_2 can be seen as functions of the dispersion parameters of the random variables Y_1 and Y_2 away from their mode $\bar{y}_1$ and $\bar{y}_2$ respectively

within their submanifold. Due to their independence, the Riemannian generic Mahalanobis distance is expressed as

$$d^2(y, \bar{y}) = \rho_1^2(\sigma_1) d^2_{\mathcal{H}_n^1}(y_1, \bar{y}_1) + \rho_2^2(\sigma_2) d^2_{\mathcal{H}_p^2}(y_2, \bar{y}_2) \quad (7)$$

where the quantity $d^2_{\mathcal{H}_i}(.,.)$ depends directly on the metric considered on $\mathcal{H}_i$.

Remark. In the deterministic case, it is noticeable that if we consider the pair $(Y_1, Y_2) = (\mu, \Sigma)$, we can observe that the Wasserstein distance between two multivariate Gaussian laws, $\mathcal{N}_1(\mu_1, \Sigma_1)$ and $\mathcal{N}_2(\mu_2, \Sigma_2)$ is clearly a Riemannian distance within the general class of distance associated to the equation (7). In [21] from the optimal transport theory, the Wasserstein distance is given by the following formula

$$W^2(\mathcal{N}_1, \mathcal{N}_2)) = \| \mu_1 - \mu_2 \|_2^2 + \| \Sigma_1^{1/2} - \Sigma_2^{1/2} \|_F^2 \quad (8)$$

where $\| A \|_F^2 = trace(AA^T)$ is the Frobenius norm.

2.3. Product-spaces Gaussian Distribution from Gaussian Measures

Let us now pay attention to the random pair $Y = (\mu, \Sigma)$ which corresponds to $\mathbb{R}^m \times \mathbb{SPD}(m)$ of dimension $m + m(m+1)/2$. This section provides definition of the product-spaces Riemannian Gaussian distributions. We choose the following form

$$(\mu, \Sigma) \sim p(\mu | \bar{\mu}, \Gamma) \times p(\Sigma | \bar{\Sigma}, \sigma) \quad (9)$$

- For the variable μ

The conventional Euclidean multivariate Gaussian law is selected
$$p(\mu | \bar{\mu}, \Gamma) = \frac{1}{(2\pi)^{m/2} |\Gamma|^{1/2}} \exp\left[-\frac{1}{2}(\mu - \bar{\mu})^T \Gamma^{-1}(\mu - \bar{\mu})\right]$$

- For the variable Σ

We propose to use the Riemannian Gaussian distribution based on the Fisher-Rao metric given by (3). As shown in [10], the term $Z(\sigma)$ can be expressed as follows

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

where $|r|$ is the Euclidean norm of r, C is a constant and $dr = dr_1 \cdots dr_m$. Let the spectral decomposition of Σ be given by $\Sigma = U \text{diag}(e^{r_1}, \cdots, e^{r_m}) U^*$, where U is an unitary matrix and $e^{r_1}, \cdots, e^{r_m}$ are the eigenvalues of Σ. In practice, the function $Z(\sigma)$ can be computed by Monte Carlo integration, as reported in [10].

Remark. In this paper, we select the Fisher-Rao metric. However, any metric (or distance) can be used provided that the corresponding normalizing constant does not depend on $\bar{\Sigma}$. This property is guaranteed if the selected distance is invariant by affine transformation [10].

3. MIXTURES OF PRODUCT-SPACES GAUSSIAN DISTRIBUTIONS

In the context of visual content classification or indexation, various works have focused on dictionary learning in the descriptor space which can be Euclidean or a more constrained Riemannian manifolds [2, 3, 4, 11, 12, 13]. For this, mixture models have shown nice performance as tool for statistical learning.

3.1. Definition

From the definition of our product-spaces Gaussian distributions versus $\mathbb{R}^m \times \mathbb{SPD}(m)$, the mixture is defined as:

$$p(\mu, \Sigma | (\varpi_k, \bar{\mu}_k, \Gamma_k, \bar{\Sigma}_k, \sigma_k)_{1 \le k \le K}) = \sum_{k=1}^K \varpi_k \times p(\mu | \bar{\mu}_k, \Gamma_k) \times p(\Sigma | \bar{\Sigma}_k, \sigma_k) \quad (10)$$

where $\varpi_\mu \in (0, 1)$ are the weights with sum one and where each Gaussian law is given by (9).

3.2. EM algorithm for mixture parameter estimation

Let $(\mu_1, \Sigma_1), \ldots, (\mu_N, \Sigma_N)$ be independent samples from (9). Based on these samples, the maximum likelihood estimate (MLE) of $\vartheta = (\vartheta_{1 \le k \le K}) = (\{\varpi_k, \bar{\mu}_k, \Gamma_k, \bar{\Sigma}_k, \sigma_k\}_{1 \le k \le K})$ can be computed using an EM algorithm. For this, let us consider for ϑ,
$$\omega_{n,k}(\vartheta) = \frac{\varpi_k \times p(\mu_n, \Sigma_n | \vartheta_k)}{\sum_{s=1}^K \varpi_s \times p(\mu_n, \Sigma_n | \vartheta_s)} \quad N_k(\vartheta) = \sum_{n=1}^N \omega_{n,k}(\vartheta)$$

The EM algorithm iteratively updates $\hat{\vartheta} = (\hat{\vartheta}_{1 \le k \le K})$, which is an approximation of the MLE of ϑ as follows. Based on the current value of $\hat{\vartheta}$, it updates $\hat{\vartheta}_k^{new}$ as follows:

▶ Assign to $\hat{\varpi}_k$ the value $\hat{\varpi}_k^{new} = N_k(\hat{\vartheta})/N$

▶ Assign to $\hat{\bar{\mu}}_k$ the value $\hat{\bar{\mu}}_k^{new} = 1/N_k(\hat{\vartheta}) \sum_{n=1}^N \omega_{n,k}(\hat{\vartheta})\mu_n$

▶ Assign to $\hat{\Gamma}_k$ the value

$$\hat{\Gamma}_k^{new} = 1/N_k(\hat{\vartheta}) \sum_{n=1}^N \omega_{n,k}(\hat{\vartheta})(\mu_n - \hat{\bar{\mu}}_k^{new})(\mu_n - \hat{\bar{\mu}}_k^{new})^t$$

▶ Assign to $\bar{\Sigma}_k$ the value as defined in [10]

$$\bar{\Sigma}_k^{new} = \arg\min_{\Sigma} \left(1/N_k(\hat{\vartheta}) \sum_{n=1}^N \omega_{n,k}(\hat{\vartheta}) d^2_{FR}(\Sigma_n, \Sigma)\right)$$

▶ Assign to $\hat{\sigma}_k$ the value as defined in [10]

$$\hat{\sigma}_k^{new} = \Phi\left(N_k^{-1}(\hat{\vartheta}) \times \sum_{n=1}^N \omega_{n,k}(\hat{\vartheta}) d^2_{FR}(\bar{\Sigma}_k^{new}, \Sigma_n)\right)$$

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

4. APPLICATION TO TEXTURE IMAGE CLASSIFICATION

In this section, the product-spaces Riemannian Gaussian model we have proposed is exercised for the classification of texture images. For this experiment, three databases are considered, namely the VisTex, the VisTex complete and the CUReT database composed respectively of 40, 167 and 62 classes.

For the two Vistex databases, each image is first split into 169 patches of 128×128 pixels, with an overlap of 32 pixels. For the CUReT database, each class is composed by a set of 92 patches of 200×200 pixels.

For the classification algorithm, each patch is represented by a set of F mean vectors μ_f and covariance matrices Σ_f describing the color dependence in the wavelet domain. Each color channel is first filtered with the Daubechies db4 wavelet, with 2 scales and 3 orientations. The color information is then modeled for each wavelet subband by a multivariate Gaussian distribution. For each wavelet subband (detail and approximation), the sample mean vector and the sample covariance matrix are computed resulting in a feature space of $F = 13$ features.

In this first experiment, the considered databases are first split in order to obtain the training and the testing sets. In order to model the within-class diversity of textures, each class in the training set is considered as a realization of a mixture of product-spaces Gaussian distributions. But, since we estimate one sample mean vector and one sample covariance matrix per subband, the mixture model defined in (10) is generalized to :

$$p\left((\mu_f, \Sigma_f)_{1 \leq f \leq F} \mid (\varpi_k, \bar{\mu}_k, \Gamma_k, \bar{\Sigma}_k, \sigma_k)_{1 \leq k \leq K, 1 \leq f \leq F}\right) =$$

$$\sum_{k=1}^{K} \varpi_k \times \prod_{f=1}^{F} p(\mu_f \mid \bar{\mu}_{k,f}, \Gamma_{k,f}) \, p(\Sigma_f \mid \bar{\Sigma}_{k,f}, \sigma_{k,f})$$

$$(11)$$

Next, for each class, an EM algorithm is run to estimate the MLE of $\vartheta = (\varpi_k, \bar{\mu}_k, \Gamma_k, \bar{\Sigma}_k, \sigma_k)_{1 \leq k \leq K, 1 \leq f \leq F}$. In the end, each test patch, in the testing set, is assigned to the class of the closest cluster k, *i.e.* the one maximizing the conditional probability. In the following, classification performances are averaged on 10 Monte Carlo runs.

Table. 1 shows the average classification rate on the VisTex, VisTex complete and CUReT databases. Here half of the samples are used for learning while the other half is used for testing. Classification results are computed for $K = 1$ and $K = 3$ in (10). Column "Mu" (respectively "Sigma") corresponds to a model where only the mean μ (respectively on the covariance matrix "Sigma") is used for modeling the color wavelet coefficients. In column "Mu Sigma", the results with the proposed mixtures of product-spaces Gaussian distributions. Finally, the column "Sigma augmented" corresponds to a model used in [12]-[13] where the mean is embedded in a

Database	Number of modes	Mu	Sigma	Mu Sigma	Sigma augmented
VisTex	K=1	89.27	99.30	99.48	98.25
	K=3	88.22	**99.96**	**99.96**	99.68
VisTex complete	K=1	64.72	83.75	86.88	80.98
	K=3	67.20	92.20	**93.75**	90.52
CUReT	K=1	64.51	45.03	67.90	37.93
	K=3	66.15	68.90	**87.82**	57.56

Table 1. Classification results on the VisTex, VisTex complete and CUReT databases.
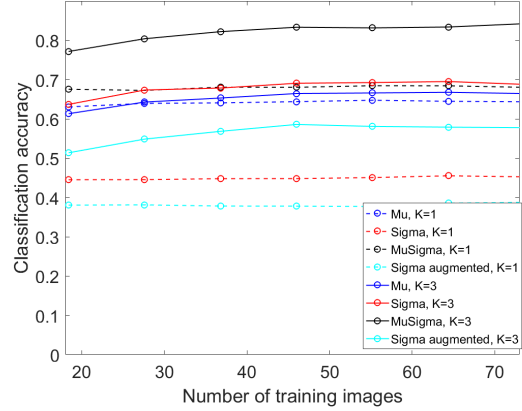
Fig. 1. Influence of the number of training samples on classification accuracy for the CUReT database.

larger covariance matrix of dimension $(m+1) \times (m+1)$ [14] :

$$\Sigma_{augmented} = |\Sigma|^{-\frac{1}{m+1}} \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix} \qquad (12)$$

As observed, the proposed product-spaces Gaussian model offers the best performance on these three databases.

A second experiment is carried out to evaluate the influence of the number of training samples on classification accuracy for the CUReT database. As observed in Fig. 1, the best performance are retrieved for the proposed product-spaces Gaussian model. A significant gain of about 30% is observed compared to [14] who considers an augmented covariance matrix.

5. CONCLUSION

In this paper, we introduce the product-spaces Gaussian probability distribution: a novel parametric modeling of the random pair defined by a mean vector and a covariance matrix. We show that the mixture Riemaniann Gaussian probability distribution on $\mathbb{R}^m \times \mathbb{SPD}(m)$ is a serious competitor with the state of the art in terms on modeling image from local multiple input-features. The product-spaces Gaussian probability distribution tackles supervised learning for content textured image classification.

6. REFERENCES

[1] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *Proceedings of the 11th European Conference on Computer Vision: Part IV*, Berlin, Heidelberg, 2010, ECCV'10, pp. 143–156, Springer-Verlag.

[2] F. Perronnin and D. Larlus, "Fisher vectors meet neural networks: A hybrid classification architecture," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3743–3752.

[3] B. C. Lovell, S. Shirazi, C. Sanderson, and M. T. Harandi, "Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 00, no. undefined, pp. 2705–2712, 2011.

[4] M. Faraki, M. T. Harandi, and F. Porikli, "More about vlad: A leap from euclidean to riemannian manifolds," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4951–4960.

[5] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on riemannian manifolds with gaussian rbf kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2464–2477, Dec 2015.

[6] J. Zhang, L. Wang, L. Zhou, and W. Li, "Learning discriminative stein kernel for spd matrices and its applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 5, pp. 1020–1033, May 2016.

[7] 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, 2008.

[8] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 989 – 993, 2008.

[9] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, 2012, pp. 11–pages.

[10] S. Said, L. Bombrun, Y. Berthoumieu, and J. Manton, "Riemannian gaussian distributions on the space of symmetric positive definite matrices," *to appear in IEEE Transactions on Information Theory, https://arxiv.org/abs/1507.01760*, 2016.

[11] I. Ilea, L. Bombrun, C. Germain, R. Terebes, M. Borda, and Y. Berthoumieu, "Texture image classification with riemannian fisher vectors," in *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, 2016, pp. 3543–3547.

[12] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 2015, pp. 720–729.

[13] M. T. Harandi, M. Salzmann, and R. I. Hartley, "Dimensionality reduction on spd manifolds: The emergence of geometry-aware methods," *CoRR*, vol. abs/1605.06182, 2016.

[14] M. Lovric, M. Min-Oo, and E. A Ruh, "Multivariate normal distributions parametrized as a riemannian symmetric space," *Journal of Multivariate Analysis*, vol. 74, no. 1, pp. 36 – 48, 2000.

[15] R. Hosseini and S. Sra, "Matrix manifold optimization for gaussian mixtures," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, 2015, pp. 910–918.

[16] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imaging Vis.*, vol. 25, no. 1, pp. 127–154, July 2006.

[17] M. Moakher, "On the averaging of symmetric positive-definite tensors," *Journal of Elasticity*, vol. 82, no. 3, pp. 273–296, 2006.

[18] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 1, pp. 328–347, 2007.

[19] G. Cheng and B. C. Vemuri, "A novel dynamic system in the space of SPD matrices with applications to appearance tracking," *SIAM J. Imaging Sciences*, vol. 6, no. 1, pp. 592–615, 2013.

[20] B.-Y. Chen, "Warped products in real space forms," *Rocky Mountain J. Math.*, vol. 34, no. 2, pp. 551–563, 06, 2004.

[21] A. Takatsu, "Wasserstein geometry of gaussian measures," *Osaka J. Math.*, vol. 48, no. 4, pp. 1005–1026, 12, 2011.