

Trabalho final

Revisão Comentada de Artigo
Parameters estimate of Riemannian Gaussian
distribution in the manifold of covariance matrices

Guilherme Inácio Gonçalves



Instituto Politécnico – IPRJ
Universidade do Estado do Rio de Janeiro



Programa de Pós Graduação em Modelagem Computacional
Disciplina de Variedades Diferenciáveis
prof. Ricardo Fabbri

Nova Friburgo, 8 de Julho de 2021

Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices

Paolo Zanini and Marco Congedo and Christian Jutten
 GIPSA-Lab, CNRS, University of Grenoble Alpes,
 Grenoble Institute of Technology
 Grenoble, France
 Email: paolo.zanini@gipsa-lab.grenoble-inp.fr

Salem Said and Yannick Berthoumieu
 IMS, CNRS, University of Bordeaux
 Bordeaux, France

Abstract—The study of \mathcal{P}_m , the manifold of $m \times m$ symmetric positive definite matrices, has recently become widely popular in many engineering applications, like radar signal processing, mechanics, computer vision, image processing, and medical imaging. A large body of literature is devoted to the barycentre of a set of points in \mathcal{P}_m and the concept of barycentre has become essential to many applications and procedures, for instance classification of SPD matrices. However this concept is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, to represent the statistical variability of data in \mathcal{P}_m . Here we consider Gaussian distributions and mixtures of Gaussian distributions in \mathcal{P}_m . In particular we discuss parameter estimation of such distributions. This problem, while it is simple in the manifold \mathcal{P}_2 , becomes harder for higher dimensions, since there are some quantities involved whose analytic expression is difficult to derive. In this paper we introduce a smooth estimate of these quantities using convex cubic splines, and we show that in this case the parameters estimate is coherent with theoretical results. We also present some simulations and a real EEG data analysis.

I. INTRODUCTION

The study of Symmetric Positive Definite (SPD) matrices has been recently become widely popular in many engineering applications, like radar signal processing [3], mechanics [16], computer vision [23][24], image processing [9], and medical imaging (especially diffusion tensor imaging [17] and brain-computer interface [4]). In particular, great improvements come from information geometry, since it has been known for some time that the space \mathcal{P}_m of $m \times m$ SPD matrices can be equipped with a Riemannian metric. This metric, usually called Rao-Fisher or affine-invariant metric, gives it the structure of a Riemannian manifold (specifically a homogeneous space of non-positive curvature). Given two elements $C_1, C_2 \in \mathcal{P}_m$, the Riemannian distance $d_R : \mathcal{P}_m \times \mathcal{P}_m \rightarrow \mathbb{R}_+$ induced by the Rao-fisher metric is defined as

$$d_R^2(C_1, C_2) = \text{tr}(\log^2(C_1^{-1/2} C_2 C_1^{-1/2})) = \sum_{i=1}^m \log^2(\lambda_i), \quad (1)$$

where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of the matrix $C_1^{-1/2} C_2 C_1^{-1/2}$ or of a similar matrix $C_1^{-1} C_2$. An exhaustive description of the theoretic aspects of Rao-Fisher metric and Riemannian distance can be found in [6][22]. Here we remark only two useful properties. Affine-invariance: for any invertible

matrix B

$$d_R(BC_1B^T, BC_2B^T) = d_R(C_1, C_2), \quad (2)$$

and self-duality

$$d_R(C_1^{-1}, C_2^{-1}) = d_R(C_1, C_2). \quad (3)$$

Distance (1) is used to define the concept of Riemannian center of mass [1][15] (or barycentre, geometric mean, Karcher mean, etc.). Given a set $\{C_1, \dots, C_N\}$ of matrices in \mathcal{P}_m , the Riemannian center of mass \bar{C}_N is the unique global minimizer of:

$$\mathcal{E}_N(C) = \sum_{n=1}^N d_R^2(C_n, C) \quad (4)$$

and it is viewed as a representative of the points C_1, \dots, C_N in the manifold. A large literature is devoted to the problem of defining algorithms to find the minimizer of (4) in an efficient way (see [8] and [11] and references therein) or to make the concept of barycentre essential to many applications and procedures, like, for instance, classification of SPD matrices [4][13]. However this concept (or some robust alternatives, as the Riemannian median [2][3]) is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, able to represent the statistical variability of data in \mathcal{P}_m . The more common and natural statistical distribution is the Gaussian distribution. It can also be defined on the manifold \mathcal{P}_m and we call it Riemannian Gaussian distribution. It depends on two parameter $\bar{C} \in \mathcal{P}_m$ and $\sigma > 0$ and its probability density function is defined as

$$p(C|\bar{C}, \sigma) = \frac{1}{\zeta(\sigma)} \exp \left\{ -\frac{d_R^2(C, \bar{C})}{2\sigma^2} \right\}, \quad (5)$$

where $\zeta(\sigma)$ is a normalization factor depending only on σ . A Riemannian Gaussian distribution can be obviously used to define mixtures of Riemannian Gaussian distributions, in order to describe a wide variety of probability distributions. First attempts to describe (5) can be found in [18], even if the treatment there remains incomplete and is based on asymptotic results. Reference [20] provides a complete description of (5). Gaussian distribution and mixtures of Gaussian distributions in the Riemannian framework have been used to improve the performances of SPD matrices classification in the context of texture image [20][21]. Besides the good results obtained in

Figura 1: Primeira página do artigo sendo revisado. Acessível pela URL <https://hal.inria.fr/hal-01325055>.

1 Preliminares

Este trabalho consiste em expandir o artigo de Paolo Zaini *et al.* intitulado “*Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices*“ [1], cuja primeira página está reproduzida na Figura 1. O presente texto é uma versão comentada do artigo, expandindo o máximo possível os conceitos ligados a Variedades Diferenciáveis. Desta forma, o presente texto é um superconjunto do referido artigo. Ele contém todo o artigo, possivelmente em forma de recortes, com expansões e comentários, e conexões com outros conceitos.

Será utilizada uma mistura de línguas nesta revisão, sendo o inglês preferido sempre que possível. Sendo assim, não teremos o trabalho de traduzir do inglês algumas construções básicas do LATEX como *Theorem* ou *Definition*.

2 Tabela de Notação

m	dimensão da variedade
\mathcal{P}_m	variedade das SPD de dimensão m
\mathcal{E}_N	erro médio quadrático
d_R	distância Riemanniana
Γ_m	distribuição Gamma multivariada
$p(C \bar{C}, \sigma)$	distribuição Gaussiana Riemanniana
tr	traço

3 Commented and Expanded Abstract

A figura a seguir mostra o *Abstract* do paper. Os autores abordam o estudo da variedade \mathcal{P}_m das matrizes simétricas positivas-definidas de dimensão $m \times m$. Este campo de pesquisa ultimamente tem se tornado amplamente empregado em várias áreas da engenharia, como processamento de sinais, visão computacional, processamento de imagens, valendo destacar a área de imagiologia médica. Esta última é uma especialidade médica que se ocupa do uso das tecnologias de imagem para realização de diagnósticos. As análises realizadas neste trabalho utilizam dados simulados e dados reais de Eletroencefalografia ou EEG, que é um método de monitoramento eletrofisiológico utilizado para registrar a atividade elétrica do cérebro.

Abstract—The study of \mathcal{P}_m , the manifold of $m \times m$ symmetric positive definite matrices, has recently become widely popular in many engineering applications, like radar signal processing, mechanics, computer vision, image processing, and medical imaging. A large body of literature is devoted to the barycentre of a set of points in \mathcal{P}_m and the concept of barycentre has become essential to many applications and procedures, for instance classification of SPD matrices. However this concept is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, to represent the statistical variability of data in \mathcal{P}_m . Here we consider Gaussian distributions and mixtures of Gaussian distributions on \mathcal{P}_m . In particular we deal with parameter estimation of such distributions. This problem, while it is simple in the manifold \mathcal{P}_2 , becomes harder for higher dimensions, since there are some quantities involved whose analytic expression is difficult to derive. In this paper we introduce a smooth estimate of these quantities using convex cubic splines, and we show that in this case the parameters estimate is coherent with theoretical results. We also present some simulations and a real EEG data analysis.

No contexto de álgebra linear, uma matriz positiva-definida (SPD, da sigla em inglês *Symmetric Positive Definite*) é análoga a um número real positivo

em muitos aspectos. Uma matriz simétrica M com entradas reais é positiva-definida se o número real $z^T M z$ é positivo para cada vetor coluna real não-nulo z , onde z^T é o seu transposto [2]. Esta definição ainda pode ser estendida para o domínio dos números complexos, no caso de matrizes Hermitianas.

Os autores citam que grande parte da literatura está focada no estudo do baricentro, ou centro de massa riemanniano, de um conjunto de pontos na variedade de matrizes simétricas positivas-definidas \mathcal{P}_m . O conceito de baricentro é de notável importância para muitas aplicações, como a classificação de SPD's. O problema do centro de massa riemanniano também é conhecido na literatura como *média riemanniana de tensores* [3].

Entretanto, os autores indagam que o conceito de baricentro tipicamente tem sido utilizado apenas para a caracterização do conjunto de pontos. Não obstante, pouca atenção tem sido empregada na caracterização da forma das amostras na própria variedade ou ainda na definição do modelo probabilístico que representa e governa a variabilidade estatística dos dados em \mathcal{P}_m .

Neste trabalho, portanto, o problema de estimação de parâmetros de distribuições Gaussianas em \mathcal{P}_m será explorado para altas dimensões, onde a tratabilidade do problema é mais difícil comparada ao caso bidimensional.

4 Commented and Expanded Section 1: Introduction

The study of Symmetric Positive Definite (SPD) matrices has been recently become widely popular in many engineering applications, like radar signal processing [3], mechanics [16], computer vision [23][24], image processing [9], and medical imaging (especially diffusion tensor imaging [17] and brain-computer interface [4]). In particular, great improvements come from information geometry, since it has been known for some time that the space \mathcal{P}_m of $m \times m$ SPD matrices can be equipped with a Riemannian metric. This metric, usually called Rao-Fisher or affine-invariant metric, gives it the structure of a Riemannian manifold (specifically a homogeneous space of non-positive curvature). Given two elements $C_1, C_2 \in \mathcal{P}_m$, the Riemannian distance $d_R : \mathcal{P}_m \times \mathcal{P}_m \rightarrow \mathbb{R}_+$ induced by the Rao-fisher metric is defined as

$$d_R^2(C_1, C_2) = \text{tr}(\log^2(C_1^{-1/2} C_2 C_1^{-1/2})) = \sum_{i=1}^m \log^2(\lambda_i), \quad (1)$$

where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of the matrix $C_1^{-1/2} C_2 C_1^{-1/2}$ or of a similar matrix $C_1^{-1} C_2$. An exhaustive description of the theoretic aspects of Rao-Fisher metric and Riemannian distance can be found in [6][22]. Here we remark only two useful properties. Affine-invariance: for any invertible matrix B

$$d_R(BC_1B^T, BC_2B^T) = d_R(C_1, C_2), \quad (2)$$

and self-duality

$$d_R(C_1^{-1}, C_2^{-1}) = d_R(C_1, C_2). \quad (3)$$

A modelagem matemática do problema abordada pelos autores deriva os conceitos da Geometria da Informação, área da matemática que utiliza ferramentas geométricas no estudo de modelos estatísticos. Parte-se da variedade de matrizes simétricas positivas-definidas \mathcal{P}_m de dimensão $m \times m$, sobre a qual define-se uma métrica Riemanniana. Introduzida em 1945 por Rao, utilizando os fundamentos desenvolvidos por Ronald Fisher em 1921, esta métrica tipicamente chamada de Rao-Fisher, define uma distância entre duas distribuições de probabilidade, geodésicas, curvaturas e outras propriedades do espaço [4]. Com esta métrica, definida pela Equação (1), a variedade adquire uma estrutura Riemanniana de um espaço homogêneo de curvatura não-positiva.

Os autores destacam duas importantes propriedades da métrica de Rao-Fisher $d_R^2 : \mathcal{P}_m \times \mathcal{P}_m \rightarrow \mathbb{R}_+$. A primeira delas é a *affine-invariance*, ou seja, a distância entre dois elementos $C_1, C_2 \in \mathcal{P}_m$ não se altera o aplicar um conjunto de transformações afins. A segunda propriedade é a *self-duality*, onde a distância entre C_1^{-1} e C_2^{-1} é a mesma que entre C_1 e C_2 .

Distance (1) is used to define the concept of Riemannian center of mass [1][15] (or barycentre, geometric mean, Karcher mean, etc.). Given a set $\{C_1, \dots, C_N\}$ of matrices in \mathcal{P}_m , the Riemannian center of mass \hat{C}_N is the unique global minimizer of:

$$\mathcal{E}_N(C) = \sum_{n=1}^N d_R^2(C_n, C) \quad (4)$$

and it is viewed as a representative of the points C_1, \dots, C_N in the manifold. A large literature is devoted to the problem of defining algorithms to find the minimizer of (4) in an efficient way (see [8] and [11] and references therein) or to make the concept of barycentre essential to many applications and procedures, like, for instance, classification of SPD matrices [4][13]. However this concept (or some robust alternatives, as the Riemannian median [2][3]) is often used alone in order to define and characterize a set of points. Less attention is paid to the characterization of the shape of samples in the manifold, or to the definition of a probabilistic model, able to represent the statistical variability of data in \mathcal{P}_m . The more common and natural statistical distribution is the Gaussian distribution. It can also be defined on the manifold \mathcal{P}_m and we call it Riemannian Gaussian distribution. It depends on two parameters $\bar{C} \in \mathcal{P}_m$ and $\sigma > 0$ and its probability density function is defined as

$$p(C|\bar{C}, \sigma) = \frac{1}{\zeta(\sigma)} \exp \left\{ -\frac{d_R^2(C, \bar{C})}{2\sigma^2} \right\}, \quad (5)$$

where $\zeta(\sigma)$ is a normalization factor depending only on σ .

A distância d_R^2 expressa pela Equação (1), portanto, define o conceito de centro de massa Riemanniano (ou baricentro, centro geométrico), o qual é utilizado para a classificação de matizes SPD, dentre outras aplicações. Este centro de massa, denotado por \hat{C}_N , é único e é dado pelo mínimo global da Equação (4). Várias técnicas tem sido empregadas para se encontrar este minimizador. Como métodos de primeira ordem, pode-se citar o *Steepest Descent Method* e o *Conjugate Gradient Method*. Também destacam-se métodos de segunda ordem baseados na região de confiança (*Trust region method*) com suas variantes *Exact Hessian*, *Hessian by decomposition*, *Hessian by approximation* e ainda o *Broyden–Fletcher–Goldfarb–Shanno* (BFGS).

Entretanto, o conceito de centro de massa Riemannian tem sido amplamente utilizado para definir e caracterizar um conjunto de pontos. Pouca atenção tem-se dado para a caracterização da forma (ou *shape*) das amostras na própria variedade Riemanniana, o que possibilitaria caracterizar o modelo probabilístico que governa os dados em \mathcal{P}_m .

Neste contexto de estatística, o modelo de distribuição de probabilidades mais comum e naturalmente utilizado nas mais diversas áreas da ciência é o modelo Gaussiano. Este modelo é dotado de particular notoriedade em virtude do *Teorema Central do Limite*, o qual estabelece que quando variáveis aleatórias

independentes são somadas, a distribuição de probabilidades resultante tende a uma distribuição normal, ou Gaussiana [5]. Esta distribuição é parametrizada por apenas duas quantidades, a média e a variância da variável aleatória.

Definido este modelo de distribuição de probabilidades na variedade \mathcal{P}_m , tem-se então uma distribuição Gaussiana Riemanniana dada pela Equação (5), a qual depende da média $\tilde{C} \in \mathcal{P}_m$ e da variância $\sigma > 0$.

A Riemannian Gaussian distribution can be obviously used to define mixtures of Riemannian Gaussian distributions, in order to describe a wide variety of probability distributions. First attempts to describe (5) can be found in [18], even if the treatment there remains incomplete and is based on asymptotic results. Reference [20] provides a complete description of (5). Gaussian distribution and mixtures of Gaussian distributions in the Riemannian framework have been used to improve the performances of SPD matrices classification in the context of texture image [20][21]. Besides the good results obtained in [20], this analysis is limited to the manifold \mathcal{P}_2 . However in real applications, like the analysis of Electroencephalography (EEG) covariance matrices, we deal with matrices of high dimension. Increasing the dimension m of the manifold, the estimation of the parameters becomes problematic. Our goal in this paper is to overcome this drawback. In Section II we describe more in details the Riemannian Gaussian distribution and the possible problems which can arise in parameter estimation. Then, in Section III, we present a technique to overcome this drawback, through a smooth approximation of the normalization factor $\zeta(\sigma)$. Finally, in Section IV we show some results through numerical examples and a real application with EEG data.

O modelo de distribuição Gaussiana Riemanniana pode estendido para combinações de distribuições Gaussianas Riemannianas, útil para descrever uma ampla variedade de modelos probabilísticos. Estas construções tem sido utilizadas para melhorar a performance de classificação de matrizes SPD no contexto de imagens texturizadas, porém apesar dos bons resultados apresentados na literatura, as análises até aqui tem se limitado em variedades de apenas duas dimensões (\mathcal{P}_2). Contudo, para muitos outros problemas práticos, como a análise de matrizes de covariância em Eletroencefalografia, lida-se com fenômenos de alta dimensão onde o problema de estimativa de parâmetros se torna mais desafiador. Este trabalho se insere neste contexto, apresentando técnicas para superar estas limitações.

5 Commented and Expanded Section 2: Riemannian Gaussian Distribution

Nesta seção os conceitos de Distribuição Gaussiana Riemanniana serão derivados bem como os possíveis problemas que podem surgir na estimativa de parâmetros.

A first important issue for a complete description of (5) is the explicit definition of the normalization factor $\zeta(\sigma)$. In [20] it is shown that an analytic expression can be derived only for $m = 2$, while, for $m > 2$, $\zeta(\sigma)$ is defined through the formula

$$\zeta(\sigma) = k_m \int_{\mathbb{R}^m} e^{-\frac{r_1^2 + \dots + r_m^2}{2\sigma^2}} \prod_{i < j} \sinh\left(\frac{|r_i - r_j|}{2}\right) \prod_{i=1}^m dr_i, \quad (6)$$

where e^{r_1}, \dots, e^{r_m} are the eigenvalues of $C \in \mathcal{P}_m$, and

$$k_m = \frac{\pi^{\frac{m^2}{2}} 8^{\frac{m(m-1)}{4}}}{m! \Gamma_m(m/2)} \quad (7)$$

with Γ_m the multivariate Gamma function. Even if formula (6) seems quite complicated, it can be evaluated using Monte Carlo integration, allowing to build tables of $\zeta(\sigma)$ as a function of σ , for different values of m . The evaluation of $\zeta(\sigma)$ is necessary for parameters estimation of distribution (5), specifically for the estimate of σ . In [20] it is shown that \widehat{C}_N , the Maximum Likelihood Estimate (MLE) of \overline{C} , corresponds to the well known Riemannian center of mass of C_1, \dots, C_N , while $\widehat{\sigma}_N$, the MLE of σ , is the solution of equation

$$\sigma^3 \frac{d}{d\sigma} \log \zeta(\sigma) = \mathcal{E}_N(\widehat{C}_N). \quad (8)$$

To solve equation (8) the tabulated values of $\zeta(\sigma)$ are considered and $\widehat{\sigma}_N$ is taken as that σ which better solve (8). However, Monte Carlo integration provides unstable estimates of $\zeta(\sigma)$ (and, hence, of $\log \zeta(\sigma)$) as the dimension m and σ increase. This reflects in an unexpected behavior of $\widehat{\sigma}_N$, shown in the top panels of Figure 1. Indeed, $\widehat{\sigma}_N$ does not result to be continuous with respect to the mean square error \mathcal{E}_N , and this problem seems to be more prominent as m and σ increase.

No contexto de estimação de parâmetros de variáveis aleatórias Gaussianas, busca-se estimar a média e a variância σ da distribuição, uma vez que este modelo pode ser totalmente caracterizado por estas duas quantidades [6]. Para este fim, diversos estimadores podem ser encontrados na literatura, tanto de natureza linear (*Best Linear Unbiased Estimator*, *Least Squares*, *Wiener Filter*), quanto não-linear (*General Minimum Variance Estimator*, *Bayesian Estimators*)[6, 7].

Entretanto, para problemas de alta dimensão, encontrar uma formulação fechada para a distribuição Gaussiana em função do parâmetro desconhecido σ torna-se um desafio. A busca por estimadores ótimos naturalmente é desenvolvida na direção de se encontrar um minimizador para uma função de custo que, tipicamente, depende do modelo da distribuição dos dados[6]. Esta busca é em muito facilitada quando a função de custo é mais matematicamente tratável. Contudo, dependendo da natureza do problema, nem sempre é possível obter uma formulação analítica simplificada como entrada para o desenvolvimento de estimadores, recorrendo-se à alternativas numéricas para contornar esta limitação.

Os autores evidenciam este problema na Equação (6), onde pode-se notar a

complexidade da formulação do fator de normalização $\zeta(\sigma)$. A técnica utilizada para avaliar $\zeta(\sigma)$ foi através da tabulação do valor da função para alguns valores de σ utilizando integrações de Monte Carlo. Este método de integração numérica utiliza números aleatórios para aproximar o valor da integral multivariada ([8]).

A avaliação do valor de $\zeta(\sigma)$ é de fundamental importância para estimar o parâmetro desconhecido σ (variância da distribuição). Um estimador tipicamente empregado para estimação de parâmetros de distribuições normais é o Estimador de Máxima Verossimilhança (MLE), o qual estima os valores dos diferentes parâmetros do modelo estatístico de maneira a maximizar a probabilidade dos dados observados, ou seja, busca-se parâmetros que maximizem a função de verossimilhança. Geralmente, a função de verossimilhança é dada pelo logaritmo da função dos parâmetros desconhecidos. A estimativa MLE da variância $\hat{\sigma}_N$ é dada pela solução da Equação (8). Note que esta equação depende do valor de $\zeta(\sigma)$, o qual foi tabelado pela integração de Monte Carlo. Entretanto, esta aproximação apresenta instabilidades que degradam a eficiência da estimativa $\hat{\sigma}_N$. O problema fica ainda pior quando se aumenta o valor da variância σ ou da dimensão da variedade m .

If we consider the parameter $\eta = -\frac{1}{\sigma^2}$, it can be shown that the function $\psi(\eta) = \log \zeta$ is a strictly convex function. Then, $\psi'(\eta)$ is an increasing function. Since problem (8) can be reformulated to find $\hat{\eta}_N$, the MLE of η , as the solution of

$$\psi'(\eta) = \frac{\mathcal{E}_N(\hat{C}_N)}{2}, \quad (9)$$

it is clear that $\hat{\eta}_N$, and then $\hat{\sigma}_N$, need to be a continuous function of $\mathcal{E}_N(\hat{C}_N)$. In the next Section we present the main contribution of this paper, i.e., exploiting the properties of $\psi(\eta)$ in order to avoid this continuity problem. Our goal is thus to provide a tractable instrument to describe and analyze Riemannian distributions even for relatively high dimension of the manifold \mathcal{P}_m , thus for a wide variety of real applications.

Aplicando-se uma transformação não-linear no parâmetro desconhecido σ , pode-se obter um função estritamente convexa sobre o parâmetro de forma com que sua derivada é uma função crescente (Equação (9)). Esta manipulação matemática apresenta algumas vantagens do ponto de vista de otimização que tornam o problema mais tratável. Em especial, esta reformulação irá prover um instrumental matemático que possibilitará a análise de distribuições Riemannianas em variedades de altas dimensões, com ampla aplicação em problemas reais.

6 Commented and Expanded Section 3: Parameters Estimate

Nesta seção é apresentada uma técnica de estimativa de parâmetros de distribuições Gaussianas multivariadas aplicadas em variedades Riemannianas.

The procedure we propose here is to consider the convexity property of $\psi(\eta)$. We evaluate $\log \zeta$ for different values of η using a classic Monte Carlo integration technique, recovering $\log \zeta$ as a sample average using, for each η , 10^5 observations, randomly generated according to the function in 6. Then, we use these points as starting estimates to fit a curve, constrained to be convex, which will represent $\psi(\eta)$. In order to build a convex approximation we consider splines functions.

Vale reforçar que o grande desafio de se estimar parâmetros de distribuições Gaussianas Riemannianas de alta dimensão é devido à complexidade do modelo analítico que descreve matematicamente esta distribuição. Tipicamente, quando a solução analítica é impraticável ou demasiadamente complexa, recorre-se à técnicas numéricas que possibilitam encontrar aproximações para a solução do problema em questão.

Para estimar a variância σ da distribuição Gaussian Riemanniana de alta dimensão, os autores propõem uma técnica numérica que proverá uma aproximação para esta quantidade. Parte-se inicialmente de uma transformação não-linear $\eta = \frac{-1}{\sigma^2}$ que possibilita a construção de uma função de verossimilhança $\psi(\eta) = \log \zeta(\eta)$ estritamente convexa, ou seja, $\psi''(\eta) > 0, \forall \eta$. Posteriormente, a função $\log \zeta(\eta)$ é avaliada numericamente para alguns valores de η , utilizando integrações de Monte Carlo. Em posse destas aproximações, a função $\phi(\eta)$ pode ser finalmente interpolada através de *splines cúbicos*.

A splines function $f(x)$ is a piecewise polynomial function defined on the interval $[a, b]$. It depends on the number of knots $a = x_0 < x_1 < \dots < x_N = b$. A splines function of degree K is composed by a polynomial of order $K - 1$ on each interval, and the connections at the knots are made such that the splines function is globally continuous up to the derivative of order $K - 2$. The most common splines function used in the application are that of order 4, named cubic splines. Cubic splines are very important in the approximation of a set of data points through a smooth function (where in general smooth means that the second derivative is quite regular). A strong result says that, given a sample of points $(x_i, y_i), \dots, (x_N, y_N)$ the function f that minimizes the quantity

$$\sum_{i=1}^N (f(x_i) - y_i)^2 + \lambda \int f''(x)^2 dx$$

is a cubic splines function [5].

Em matemática, os *Splines* são um tipo especial de função definida por dois ou mais pontos de controle, cujo grau é escolhido por projeto. Através desta técnica, é possível construir uma curva que passa por todos os pontos de controle, ou seja, uma curva que interpola os intervalos entre estes pontos [9]. Nas splines

cúbicas, determina-se um polinômio interpolador ou aproximador de terceiro grau com restrições que garantam uma curva suave, ou seja, onde a primeira e segunda derivadas são contínuas.

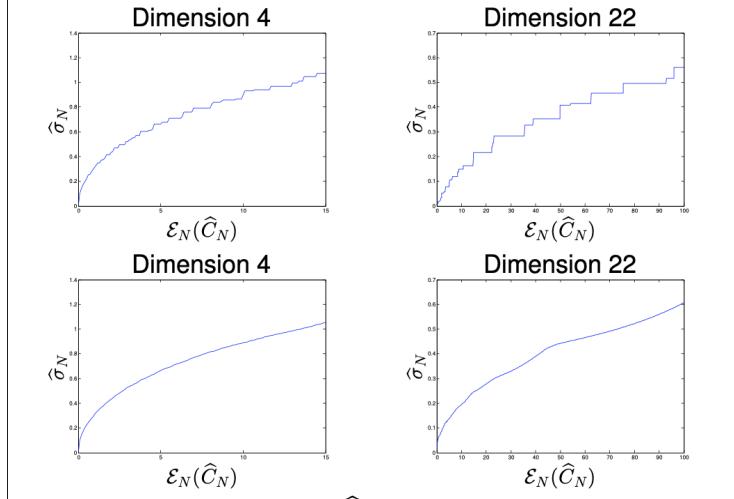


Fig. 1. $\hat{\sigma}_N$ as a function of $\mathcal{E}_N(\hat{C}_N)$ for dimension 4 (left panels) and 22 (right panels). On the top panels $\hat{\sigma}_N$ is found using the tabulated value of $\log \zeta$ provided by the Monte Carlo estimates, on the bottom panels the smooth splines approximation of $\log \zeta$ is exploited.

A figura acima mostra o valor estimado da variância da distribuição Gaussiana Riemanniana $\hat{\sigma}_n$ em função do erro médio quadrático (SMR) $\mathcal{E}_N(\hat{C}_N)$ para quatro condições diferentes. Os gráficos a esquerda referem-se à uma variedade \mathcal{P}_m de dimensão $m = 4$ e os gráficos a direita de dimensão $m = 22$. Os gráficos superiores utilizam os valores tabulados através do estimador de Monte Carlo. Os gráficos inferiores, os valores foram aproximados através dos splines cúbicos.

Para ambos os estimadores, pode-se notar diretamente que o erro médio quadrático cresce em função da variância. Entretanto, a função variância aproximada pelo estimador de Monte Carlo é descontínua dentro do intervalo de operação, apresentando variações abruptas ao longo da curva. Para alta dimensão, este problema é ainda mais evidente e intensificado. Em contra partida, a aproximação por splines cúbicos resulta em uma função aproximada mais suave em toda a faixa de operação, tanto em baixa quanto em alta dimensão. Este resultado se deve às restrições impostas na técnica dos splines cúbicos, que garantem a existência e a continuidade da primeira e segunda derivadas.

Cubic splines are not necessarily globally convex. Nevertheless it is possible to define, imposing some constraints, convex cubic splines [7][19]. To implement convex cubic splines we use the Shape Language Modeling (SLM) Matlab toolbox. In Figure 2 the splines approximation of $\psi(\eta)$ in the case of \mathcal{P}_{22} is depicted¹. The knots (green vertical line in the figures) used to build splines function have been chosen to be equally spaced in σ , hence they are not in η .

Por definição, os splines cúbicos não são globalmente convexos, ou seja, de forma geral, não existem restrições quanto ao módulo da segunda derivada em todo o domínio da interpolação. Portanto, pode-se construir os splines cúbicos através de convoluções de forma a preservar a convexidade global do polinômio approximador (*fit*) [10]. Os autores implementaram esta técnica de splines por convoluções com o auxílio da ferramenta SML do Matlab[11], a qual proporciona a construção de curvas a partir de pontos de forma mais simplificada.

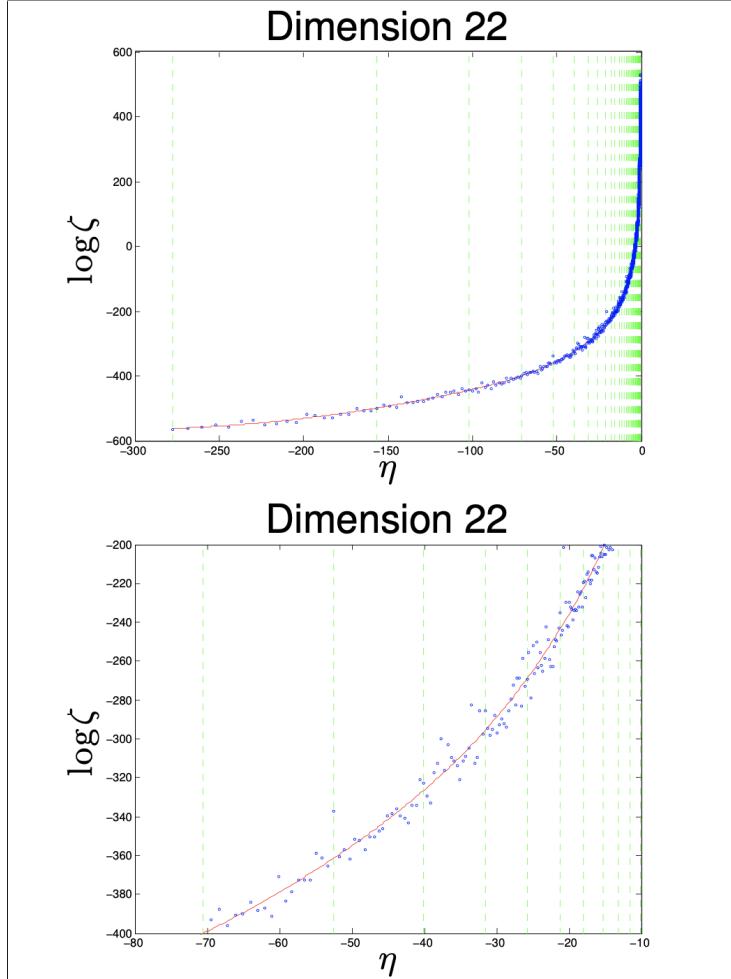


Fig. 2. Fitting of $\log \zeta$ as a function of η . Convex cubic splines (red line) fitted on the Monte Carlo estimates (blue points) for \mathcal{P}_{22} . Vertical green lines represent the knots used to build the splines function. Figure on the bottom is a zoom of the figure on the top for η varying between -70 and -15.

A figura acima mostra em azul o conjunto de pontos da função $\psi(\eta) = \log \zeta(\eta)$ aproximados pelo método de Monte Carlo. A função que aproxima a distribuição dos pontos (função de *fitting*) está representada em vermelho. Vale notar que, para este caso, a dimensão da variedade P_m é $m = 22$. Esta função de fitting foi construída através dos splines cúbicos a partir de nós de controle igualmente espaçados (linhas verticais em verde), utilizando a restrição de convexidade global citada anteriormente. Nota-se visualmente que a função de fitting consegue modelar bem conjunto de dados, porém os autores não apresentam nenhuma medida de concordância ou alguma análise de erro de aproximação. Este tipo de análise qualitativa é fundamental para se verificar o quanto bem um modelo representa os dados estudados.

Now we can use this approximation of $\log \zeta$ in order to obtain the $\hat{\sigma}_N$. As can be appreciated in the bottom panels of Figure 1, $\hat{\sigma}_N$ is now a continuous function of $\mathcal{E}_N(\hat{C}_N)$, as expected. We now can use Riemannian distributions to analyze several statistical problems, and we might consider any dimension of \mathcal{P}_m . In the next Section we focus on a numerical example and a real EEG data analysis regarding the classification problem of covariance matrices.

Relembrando, o objetivo é estimar a variância $\hat{\sigma}_N$ da distribuição Gaussiana Riemannian $\zeta(\sigma)$ (Equação (6)) através da maximização da função de verossimilhança $\psi(\eta)$ (Equação (9)), a qual depende do conhecimento de $\log \zeta(\eta)$, cuja solução analítica é impraticável. Para contornar o problema, a função $\log \zeta(\eta)$ foi aproximada para alguns valores de η através do método de Monte Carlo e, em posse destes pontos aproximados, uma função de fitting contínua foi construída pelo métodos dos splines cúbicos globalmente convexos. Desta forma, tem-se todos os elementos para estimar a variância $\hat{\sigma}_N$.

7 Commented and Expanded Section 4: Results

Nesta seção os resultados obtidos nas análises de sinais com EEG utilizando técnicas baseadas em Variedades Riemannianas são apresentados.

Riemannian Gaussian distributions can be used, for instance, for the classification of data lying in \mathcal{P}_m . Classification in \mathcal{P}_m is a challenging problem in various applications based on signal and image analysis ranging from medical applications [4][14], computer vision [23][10] to remote sensing [9][12]. In this paper, we are interested in a motor imagery experiment for which a spatial mapping of the brain is analyzed from recorded EEG signals, with electrodes placed along the scalp [4]. The work proposed in [4] has shown that the structure of the covariance matrix between electrodes is an excellent feature to recognize mental actions. Due to the fact that, for all these previous works, a probabilistic model is not available, authors proposed a classification procedure based on Minimum Distance to Mean (MDM) classifier. From [4], MDM is defined as follows. Given K groups and a training phase where the barycentres $\hat{C}(k)$ of the groups ($k = 1, \dots, K$) are estimated, a new observation C_i is assigned to the k group according to the classification rule

$$\hat{k} = \arg \min_{k \in \{1, \dots, K\}} \{d_R(C_i, \hat{C}(k))\}. \quad (10)$$

Como ambiente experimental, os autores utilizam sinais de Eletroencefalografia (EEG) capturados através de eletrodos localizados ao longo do couro

cabeludo de um indivíduo. Através destes sinais, busca-se analisar o mapeamento espacial através da matriz de covariância entre os eletrodos, a qual pode ser utilizada para caracterizar padrões mentais. Portanto, trata-se de um problema de classificação onde o modelo de distribuição Gaussiana Riemanniana pode ser útil.

A geometria Riemanniana é um ramo da geometria diferencial que estuda variedades suaves, espaços curvos com geometrias peculiares. Nestes espaços as noções de ângulos, caminho mais curto entre dois pontos, distâncias, centro de massa de vários pontos, etc., permitem estudar propriedades analíticas de operadores matemáticos de uma perspectiva geométrica, tornando-os acessíveis à intuição [12, 13].

No contexto de Eletroencefalografia multivariada, a variedade de matrizes simétricas positivas-definidas \mathcal{P}_m tem se mostrado muito útil, uma vez que dados de EEG em janelas de tempo finitas podem ser efetivamente mapeados como pontos nesta variedade através da estimação de parâmetros de sua matriz de covariância. Esta abordagem levou à introdução de classificadores com características notáveis em comparação com o estado da arte [13].

Devido à complexidade de se determinar um modelo probabilístico que governa o processo estocástico associado aos sinais de EEG, tipicamente adota-se um procedimento de classificação baseada na Mínima Distância para a Média (da sigla em inglês, MDM). Em particular, este classificado, embora não seja o mais preciso entre os classificadores Riemannianos, destaca-se por sua simplicidade e eficiência computacional. É um classificador totalmente determinístico e livre de parâmetros, portanto, nenhum parâmetro precisa ser ajustado por validação cruzada ou outros métodos que possam comprometer sua generalização [13].

O MDM é definido pela Equação (10), onde dados os grupos K e uma fase de treinamento onde os baricentros $\hat{C}(k)$ dos grupos ($k = 1, \dots, K$) são estimados, associa-se uma nova observação C_i ao grupo \hat{K} de acordo com a regra de classificação baseada na mínima distância d_R .

In order to allow more general distributions, in [20] the authors, starting from Gaussian distribution (5), define mixtures of Gaussian distributions in a straightforward way such as

$$p(C) = \sum_{j=1}^M \omega_j p(C|\bar{C}_j, \sigma_j), \quad (12)$$

where ω_j are positive weight adding up to 1. To estimate the parameters of mixtures model, in [20] an extension of EM algorithm for data lying in \mathcal{P}_m is proposed. Here this algorithm is integrated with the splines approximation of the different normalization constants. Using this class of distributions, the Bayesian classification rule is updated accordingly.

Com o objetivo de possibilitar modelos de distribuição de probabilidade mais gerais, os autores utilizam uma combinação linear de várias distribuições Gaussianas (*Gaussian Mixtures (GM)*), definida pela Equação (12). Uma extensão

do algoritmo *Expectation-Maximisation* (EM) é utilizada para calcular as estimativas de máxima verossimilhança dos parâmetros da distribuição Gaussiana mista. Este algoritmo utiliza uma regra de classificação ótima de Bayes, usando probabilidades de classificação a posteriori para melhorar a eficiência de estimação.

We consider here a simulated example where two classes (i.e., $K = 2$) are generated according to the Gaussian law (5) in \mathcal{P}_{22} . To generate random samples from (5) we follow the procedure described in [20]. The two groups are characterized by different parameters σ , specifically $\sigma(1) = 0.15$ and $\sigma(2) = 0.17$. It is worth to point out that for Riemannian Gaussian distribution σ is a dispersion parameter, and these two values allow here a different dispersion between the two groups. $\bar{C}(1)$ corresponds to the identity matrix and then we consider four different frameworks where we gradually increase $d_R(\bar{C}(2), \bar{C}(1))$. We generate 60 SPD matrices for each sample and then we repeat the same experiment five times for each framework. In Table I we display the mean classification accuracy (expressed in percent) over the five runs for two classifiers, the MDM and the Bayes classifier using simple Gaussian distribution. In the following we consider also Gaussian Mixtures (GM) with M modes. The case of simple Gaussian can be seen as a Gaussian mixture with $M = 1$. Classification accuracy is evaluated through a 30-fold cross-validation procedure. This means that we randomly divide the dataset in 30 parts. Then we repeat the classification procedure 30 times using each part once as test set, and the others as training set.

Looking at the results in Table I we can observe that, when the two barycentres are very close, the different variability of the classes is the only way to classify correctly. As the distance increases, the performances of the two classifiers become closer, until the limit situation of framework 4, where the barycentres are very far and the variability is no longer essential to obtain good classification performances.

Os autores utilizaram dois grupos ($K = 2$), os quais são caracterizados por valores de σ distintos, sendo $\sigma(1) = 0.15$ e $\sigma(2) = 0.17$. De forma análoga à distribuições Gaussianas em geral, o parâmetro σ refere-se a dispersão, ou variância, dos dados na variedade Riemanniana. $\bar{C}(1)$ corresponde à matriz identidade e quatro diferentes cenários serão considerados, incrementando gradualmente a distância $d_R(\bar{C}(2), \bar{C}(1))$. Os autores geram 60 matrizes SPD para cada amostra e repetem o mesmo experimento por cinco vezes para cada cenário.

$d_R(\bar{C}(2), \bar{C}(1))$	MDM	GM ($M = 1$)
0.26	59.5	75.2
0.36	67.5	75.3
0.6	84.5	95.5
0.91	97.2	98.2

TABLE I. SIMULATED EXAMPLE - CLASSIFICATION ACCURACY IN
30-FOLD CROSS-VALIDATION VARYING THE DISTANCE OF THE
BARYCENTRES OF THE TWO CLASSES.

A tabela acima mostra a acurácia de classificação média em porcentagem das cinco rodadas de cada um dos quatro cenários. São apresentados os resultados de dois classificadores, o MDM e o classificador de Bayes utilizando uma distribuição Gaussiana simples, ou seja, GM com $M = 1$. Os resultados foram obtidos através de um processo de validação cruzada, subdividindo o conjunto de dados em 30 partes e repetindo o processo de classificação 30 vezes usando cada parte como conjunto de teste e as demais como conjunto de treinamento. Pode-se notar que o classificador de Bayes apresenta melhor eficiência em termos de acurácia média para toda a faixa de distância avaliada, em especial para os valores de baixa amplitude. Também vale notar que quando os dois baricentros estão muito perto, ou seja, quando o valor de d_R é pequeno, as variantes de cada grupo são a única forma de classificar corretamente. Conforme a distância d_R aumenta, o desempenho dos dois classificadores se aproxima, até a situação limite do cenário 4, onde os baricentros estão muito distantes e a variabilidade não é mais essencial para obter bons desempenhos de classificação.

To conclude, we analyze a real EEG motor imagery problem. The analyzed dataset is the one from BCI competition analyzed in [4]. It contains EEG data from nine subjects who perform 4 kinds of motor imagery (right hand, left hand, foot, and tongue imagined movements). Hence in this experiment we have $K = 4$. A total of 576 trials per subject are available, each trial corresponding to a movement (balanced experiment, i.e., 144 trials per class). EEG signals are recorded using 22 electrodes, hence covariance matrices here belong to \mathcal{P}_{22} . In Table II the classification accuracies obtained through a 30-fold cross-validation procedure are displayed. In this case we considered also Bayesian classification using Gaussian mixtures, varying the number of modes from 1 to 4.

Agora serão analisados dados reais de EEG de nove indivíduos que executam quatro tipos de imagens motoras (movimentos imaginados da mão direita, mão esquerda, pé e língua). Desta forma, neste experimento há quatro grupos de classificação, ou seja, $K = 4$. O conjunto de dados contém 144 amostras de cada classe para cada um dos indivíduos. Estes sinais de EEG são capturados através de 22 eletrodos, e portanto, as matrizes de covariância pertencem à variedade \mathcal{P}_{22} .

Subject	MDM	$M = 1$	$M = 2$	$M = 3$	$M = 4$	M best
1	79.17	78.99	74.48	76.74	77.26	78.99
2	47.74	48.96	50.69	50.17	45.49	50.69
3	76.91	76.91	78.30	78.30	80.38	80.38
4	59.72	59.90	59.90	57.29	55.03	59.90
5	40.62	40.62	39.41	38.37	38.54	40.62
6	46.88	46.35	38.89	37.15	42.01	46.35
7	75.52	75.00	79.51	79.51	77.08	79.51
8	77.26	77.78	80.38	83.33	84.20	84.20
9	77.08	77.26	78.99	80.90	80.90	80.90
Tot	64.54	64.64	64.51	64.64	64.54	66.84

TABLE II. REAL EEG MOTOR IMAGERY PROBLEM - CLASSIFICATION ACCURACY (IN PERCENT) BY 30-FOLD CROSS-VALIDATION. THE BEST ACCURACY FOR EACH SUBJECT IS PRINTED IN BOLD

A tabela acima mostra a acurácia obtida através de um processo de validação cruzada com 30 subdivisões, considerando o classificador MDM e o classificador de Bayes utilizando uma distribuição Gaussiana Mista (GM) com M variando de 1 a 4.

A first thing we can observe from Table II is that classification accuracies of the MDM and the simple Gaussian ($M = 1$) classifiers provide comparable results. This is due to the fact that in this specific example the four classes present a similar variability. In this case, as already mentioned, rules (10) and (11) coincide. Secondly, classification accuracy of the best Bayesian classifier is, for all subjects, at least as performant as the MDM. The best classifier has not always the same number of modes. This implies that, even if the experiment is common for every subject, the shapes of data clouds is highly variable between individuals. This aspect is not usually studied in the literature of Riemannian geometry applied to EEG data, and can be useful and interesting in this domain. The example described here is a first attempt of an analysis in this direction.

Os resultados apresentados na Tabela 2 são interessantes em dois aspectos. Primeiramente, os dados mostram que para todos os indivíduos, o classificador de Bayes se mostrou pelo menos tão performático quanto o MDM. Segundo, os resultados também mostram que o melhor classificador não está sempre associado ao mesmo valor de M . Isto significa que, mesmo que o procedimento experimental seja o mesmo para cada indivíduo, a variabilidade é alta, ou seja, a forma da nuvem de dados na variedade Riemanniana é bem distinta entre os indivíduos. Este aspecto ainda não foi amplamente estudado na literatura de Geometria Riemanniana no contexto de EEG e pode ser um campo promissor para novos estudos. Vale ressaltar que os resultados mostrados neste trabalho foram os primeiros nesta direção.

It is worth to point out a comment on computational costs of the illustrated procedures. First of all, Monte Carlo integration and splines approximation done to obtain a table of $\zeta(\sigma)$ are offline operations which have to be performed only once. Then, when a table for $\zeta(\sigma)$ is available, classification algorithms can be implemented without worrying about the computational time needed to build the table. Secondly, we can observe in Table III the computational time needed to perform the 30-fold cross-validation for subject 1 for the the classification algorithms analyzed. It is interesting to observe that the Bayes classifier with Gaussian distributions requires the same time as the reference Minimum Distance to Mean classifier. On the contrary, considering mixtures of Gaussian distributions makes the classification algorithm more time consuming.

Outro ponto que merece destaque é relacionado ao custo computacional dos procedimentos aqui apresentados. A análise do tempo de processamento depende de muitos fatores, como a especificação técnica do hardware utilizado para executar a rotina computacional, o sistema operacional que gerencia os recursos de máquina, a linguagem de programação, entre outros. Os autores não detalharam no paper qual é a característica do maquinário utilizado para obter os resultados apresentados. Porém, com o intuito de comparar o tempo de processamento dos classificadores executados no mesmo maquinário sob condições semelhantes, abordagem apresentada pelos autores pode oferecer uma boa medida.

Method	MDM	$M = 1$	$M = 2$	$M = 3$	$M = 4$
Time (s)	40.8	44	361.9	427	412.7

TABLE III. COMPUTATIONAL TIME NEEDED TO PERFORM 30-FOLD CROSS-VALIDATION FOR SUBJECT 1.

A tabela acima mostra o tempo de processamento em segundos de cada um dos classificadores avaliados através do processo de validação cruzada com 30 partes considerando apenas um único indivíduo. Vale notar que a integração de Monte Carlo e a aproximação por splines são feitas offline e, portanto, sem influência no tempo de processamento de classificação. Pode-se observar que para $M = 1$ (Gaussiana simples), os tempos de processamento dos classificadores MDM e Bayes são muito próximos. Contudo, incrementando o valor de M (Gaussiana Mista), o tempo de processamento também aumenta.

To conclude, in this paper we analyzed Gaussian distribution and mixtures of Gaussian distribution on the manifold \mathcal{P}_m of SPD matrices. In particular we introduced an approximation technique to derive the normalization constant of the distribution, based on convex cubic splines. This allows to solve a problem in the parameter estimation, making the estimate coherent with the theoretical results presented in [20]. In this way Riemannian distribution can be used for applications in any dimension m , for classification analysis, introducing Bayesian classifiers, or for the observation of the shape of data clouds in the Riemannian manifold.

Concluindo, neste trabalho técnicas de estimação de parâmetros baseadas em Geometria Riemanniana foram aplicadas para o problema de classificação no contexto de EEG. Em particular, foi introduzida uma abordagem de aproximação por splines cúbicos convexos que se mostrou coerente com resultados consolidados na literatura. Logo, distribuições Riemannianas podem ser empregadas em qualquer dimensão, para análise de classificação, introduzindo classificadores Bayesianos, ou observando a estrutura da nuvens de dados na variedade Riemanniana.

Referências

- [1] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthonnieu, “Parameters estimate of Riemannian Gaussian distribution in the manifold of covariance matrices,” in *SAM 2016 - 9th IEEE Sensor Array and Multi-channel Signal Processing Workshop*, (Rio de Janeiro, Brazil), July 2016.
- [2] *Appendix C: Positive Semidefinite and Positive Definite Matrices*, pp. 259–263. John Wiley & Sons, Ltd, 2007.
- [3] M. Moakher, “A differential geometric approach to the geometric mean of symmetric positive-definite matrices,” *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005.
- [4] J. P. S. Porto *et al.*, “Geometria da informação: métrica de fisher,” 2013.
- [5] H. Fischer, *A history of the central limit theorem: from classical to modern probability theory*. Springer Science & Business Media, 2010.
- [6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, Inc., 1993.
- [7] S. Haykin, *Adaptive Filter Theory*. Pearson, 2014.
- [8] R. E. Caflisch, “Monte carlo and quasi-monte carlo methods,” *Acta numerica*, vol. 7, pp. 1–49, 1998.
- [9] M. Ruggiero and V. da Rocha Lopes, *Cálculo numérico: aspectos teóricos e computacionais*. Pearson Makron Books, 1996.

- [10] D. Russell, “Computing convex spline approximations,” *International Journal of Information & Systems Sciences*, vol. 5, 01 2009.
- [11] J. D’Errico, “Slm-shape language modeling,” *SLM-Shape Language Modeling..* <http://www.mathworks.com/matlabcentral/fileexchange/24443-slm-shape-language-modeling>: Mathworks, 2009.
- [12] T. Levi-Civita, *Lezioni di calcolo differenziale assoluto*. Zanichelli, 1925.
- [13] M. Congedo, P. L. C. Rodrigues, and C. Jutten, “The riemannian minimum distance to means field classifier,” in *8th Graz Brain-Computer Interface Conference 2019*, 2019.