

A Special Form of SPD Covariance Matrix for Interpretation and Visualization of Data Manipulated with Riemannian Geometry

Marco Congedo, Alexandre Barachant

▶ **To cite this version:**

Marco Congedo, Alexandre Barachant. A Special Form of SPD Covariance Matrix for Interpretation and Visualization of Data Manipulated with Riemannian Geometry. MaxEnt - 34th International Workshop on Bayesian Inference and Maximun Entropy Methods in Science and Engineering (MaxEnt'14), Sep 2014, Amboise, France. pp.495, 10.1063/1.4906015 . hal-01103344

HAL Id: hal-01103344
https://hal.science/hal-01103344

Submitted on 14 Jan 2015

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 Special Form of SPD Covariance Matrix for Interpretation and Visualization of Data Manipulated with Riemannian Geometry

Marco Congedo and Alexandre Barachant

GIPSA-lab, CNRS & Grenoble University, Grenoble, FRANCE

Abstract. Currently the Riemannian geometry of symmetric positive definite (SPD) matrices is gaining momentum as a powerful tool in a wide range of engineering applications such as image, radar and biomedical data signal processing. If the data is not natively represented in the form of SPD matrices, typically we may summarize them in such form by estimating covariance matrices of the data. However once we manipulate such covariance matrices on the Riemannian manifold we lose the representation in the original data space. For instance, we can evaluate the geometric mean of a set of covariance matrices, but not the geometric mean of the data generating the covariance matrices, the space of interest in which the geometric mean can be interpreted. As a consequence, Riemannian information geometry is often perceived by non-experts as a "black-box" tool and this perception prevents a wider adoption in the scientific community. Hereby we show that we can overcome this limitation by constructing a special form of SPD matrix embedding both the covariance structure of the data and the data itself. Incidentally, whenever the original data can be represented in the form of a generic data matrix (not even square), this special SPD matrix enables an exhaustive and unique description of the data up to second-order statistics. This is achieved embedding the covariance structure of both the rows and columns of the data matrix, allowing naturally a wide range of possible applications and bringing us over and above just an interpretability issue. We demonstrate the method by manipulating satellite images (pansharpening) and event-related potentials (ERPs) of an electroencephalography brain-computer interface (BCI) study. The first example illustrates the effect of moving along geodesics in the original data space and the second provides a novel estimation of ERP average (geometric mean), showing that, in contrast to the usual arithmetic mean, this estimation is robust to outliers. In conclusion, we are able to show that the Riemannian concepts of distance, geometric mean, moving along a geodesic, etc. can be readily transposed into a generic data space, whatever this data space represents.

Keywords: Covariance Matrix, Riemannian Geometry

DEFINITION OF DATA MATRICES

In many engineering applications we observe a number of realizations in the form of generic data matrices

$$\mathbf{X}_i \in \Re^{R\times C}, \tag{1}$$

where i is the index of the realization, R is the number of rows, indexed by $r \in \{1,\dots R\}$ and C is the number of columns, indexed by $c \in \{1,\dots C\}$. For example, $\mathbf{X}_i$ may be a black and white image of $R \times C$ pixels. Besides images, in this work we employ the example of data describing the realizations of R random variables of which C samples

are recorded, yielding again an $R \times C$ data matrix

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{1i} \\ \vdots \\ \mathbf{x}_{Ri} \end{bmatrix} \in \Re^{R \times C} \tag{2}$$

where $\mathbf{x}_{ri} \in \Re^C$ is the r^{th} row of the data matrix, that is, the i^{th} realization of the r^{th} random variable. We treat throughout the paper the real case, however the case of complex data is analogous. In general with real-life data we have $R \neq C$. In our first example (images) both the row and columns refer to a spatial dimension. Our second example concerns brain electrical event-related potentials. In this case we adopt the convention that the rows of $\mathbf{X}_i$ hold the potential recorded at electrodes placed on the scalp (spatial dimension) and the columns hold temporal samples (temporal dimension). As we will see, the form of covariance matrix we propose embeds both spatial covariance structures in the first example and both the spatial and temporal covariance structure in the second example. More in general, for any data matrix $\mathbf{X}_i$ the special form of covariance matrix we propose embeds both the row and column covariance structure, whatever they represent.

STANDARD COVARIANCE MATRICES

Let $\mathbf{1}_{(N)}$ the N-dimensional vector filled with one everywhere and $\mathbf{I}_{(N)}$ the $N \times N$ identity matrix. Given realization $\mathbf{X}_i$ there exist two possible sample covariance matrices, namely, the "row-covariance"

$$\mathbf{C}_{(R)} = \frac{1}{C}\left(\mathbf{X}_i \mathbf{H}_{(C)} \mathbf{H}_{(C)} \mathbf{X}_i^T\right) \in \Re^{R \times R}, \tag{3}$$

and the "column-covariance"

$$\mathbf{C}_{(C)} = \frac{1}{R}\left(\mathbf{X}_i^T \mathbf{H}_{(R)} \mathbf{H}_{(R)} \mathbf{X}_i\right) \in \Re^{C \times C}, \tag{4}$$

where

$$\mathbf{H}_{(N)} = \mathbf{I}_{(N)} - \left(\frac{1}{N}\mathbf{1}_{(N)}\mathbf{1}_{(N)}^T\right) \tag{5}$$

is named the centering matrix ([6], p. 66-68 and p. 349-352). The centering matrix $\mathbf{H}_{(N)}$ has $N-1$ equal eigenvalues and one null eigenvalue corresponding to eigenvector $\mathbf{1}_{(N)}$. It has also the following properties:

$$\mathbf{H} = \mathbf{H}^T = \mathbf{H}^2 \; ; \; \mathbf{H1} = \mathbf{H11}^T = \mathbf{11}^T\mathbf{H} = \mathbf{0} \tag{6}$$

As the name suggests, the centering matrix centers the signal around zero, that is, the centered signal has zero mean, i.e., it holds

$$\frac{1}{C}\left(\mathbf{H}_{(R)} \mathbf{X}_i \mathbf{1}_{(C)}\right) = \mathbf{0} \text{ and } \frac{1}{R}\left(\mathbf{H}_{(C)} \mathbf{X}_i \mathbf{1}_{(R)}\right) = \mathbf{0} \tag{7}$$

One may also double-center the data [5] such as

$$\mathbf{X}_i \leftarrow \mathbf{H}_{(R)}\mathbf{X}_i\mathbf{H}_{(C)}, \tag{8}$$

after which the two conditions $\mathbf{X}_i\mathbf{1}_{(C)} = 0$ and $\mathbf{X}_i^T\mathbf{1}_{(R)} = 0$ hold simultaneously. Notice that for the sequel data centering is not a necessary step and we may process as well the raw $\mathbf{X}_i$ as they are recorded. The important point here is that *if we shuffle the columns of* $\mathbf{X}_i$ *the row-covariance matrix (3) does not change. In the same way, if we shuffle the rows of* $\mathbf{X}_i$ *the column-covariance matrix (4) does not change; thus, neither covariance matrix taken alone may define exhaustively and uniquely the data matrix* $\mathbf{X}_i$. However, if we consider them simultaneously we can do so up to second order statistics. For instance, the definition is exhaustive if the data follows a multivariate Gaussian distribution.

THE SPECIAL FORM OF COVARIANCE MATRIX

Given realization $\mathbf{X}_i \in \Re^{R\times C}$ consider first the following data expansion:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{X}_i & \alpha\mathbf{I}_{(R)} \\ \alpha\mathbf{I}_{(C)} & \mathbf{X}_i^T \end{bmatrix} \in \Re^{T\times T}, \tag{9}$$

where $T = R+C$. Now, consider its associated Gram matrix (covariance structure) scaled by $1/2\alpha$ such as

$$\mathbf{C}_i = \frac{1}{2\alpha}\mathbf{Y}_i\mathbf{Y}_i^T = \begin{bmatrix} \frac{1}{2\alpha}\left(\mathbf{X}_i\mathbf{X}_i^T + \alpha^2\mathbf{I}_{(R)}\right) & \mathbf{X}_i \\ \mathbf{X}_i^T & \frac{1}{2\alpha}\left(\mathbf{X}_i^T\mathbf{X}_i + \alpha^2\mathbf{I}_{(C)}\right) \end{bmatrix} \in \Re^{T\times T},\ \alpha > 0 \tag{10}$$

This matrix holds on the diagonal blocks a regularized and differently scaled version of the covariance structures (3) and (4) and on the off-diagonal block the original data matrix itself and its transpose (in the original scaling thanks to the division of the whole matrix by 2α). Since it holds both covariance structures, this matrix defines exhaustively $\mathbf{X}_i$ up to second-order statistics; now any shuffling of columns or rows would result in a different $\mathbf{X}_i$, therefore we have created a bijection between $\mathbf{C}_i$ and $\mathbf{X}_i$ such that to any $\mathbf{X}_i$ it corresponds one and only one $\mathbf{C}_i$. Matrix $\mathbf{C}_i$ is symmetric by construction, but not necessarily positive definite. However, there exists a real $\kappa > 0$ such that $\mathbf{C}_i$ is positive definite for any $\alpha \geq \kappa$. The larger is α, the stronger the regularization enforced on the covariance structures, thus in general we may want to choose α as small as possible provided that the resulting matrices $\mathbf{C}_i$ are positive definite. The selected value of α depends on the dynamic range of the original data. Once we have obtained symmetric positive-definite (SDP) matrices $\mathbf{C}_i$ we can manipulate them according to our purpose in the Riemannian manifold of SPD matrices. Now we are able to appreciate the effect of these manipulations also in the original space of the data. That is to say, *any manipulation of* $\mathbf{C}_i$ *on the Riemannian manifold will enforce a unique corresponding manipulation on its block* $\mathbf{X}_i$*, allowing the interpretation of our manipulations in the original space of the data matrix* $\mathbf{X}_i$.

ILLUSTRATIONS

We illustrate the effect of working with the special form of SPD matrix (10) with two examples.

Example 1

The first example illustrates the manipulation of color and black and white (B&W) images. A B&W image can be represented by a matrix $\mathbf{X}_i \in \Re^{R\times C}$ of $R \times C$ pixels that can take integer value 0 (black) to ρ (white), where ρ depends on the resolution of the image (255 in our example). A color image can be represented by a triplet of images of the same kind holding each the value of red (R), green (G) and blue (B), which combination defines the color image. In order not to distort the images we do not center them and we use (10) with $\alpha = \rho$, which ensures positive-definitiveness of matrices(10). As an example of manipulation in the Riemannian manifold of SPD matrices we show how images move along geodesics in this manifold. Given SPD matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ in the Riemannian manifold of SPD matrices we move on the geodesic connecting them by [2].

$$\mathbf{C}_{(\mathbf{C}_1 \to \mathbf{C}_2, \beta)} = \mathbf{C}_1^{1/2} \left(\mathbf{C}_1^{-1/2} \mathbf{C}_2 \mathbf{C}_1^{-1/2} \right)^{\beta} \mathbf{C}_1^{1/2}, \beta \in [0:1] \tag{11}$$

For instance, with $\beta = 0$ we are at point $\mathbf{C}_1$. With $\beta = 1$ we are at $\mathbf{C}_2$. With $\beta = 0.5$ we reach the geometric mean of $\mathbf{C}_1$ and $\mathbf{C}_2$. Suppose we have a color image at low (spatial) resolution and the same image in black and white at high resolution. Suppose our goal is to generate a color image with high resolution (HR) starting only from the color image at low resolution (LR) and the B&W image at HR. This problem is known in the literature as *pansharpening* and arises in satellite imaging. Satellites have several cameras, of which one camera integrates over a wide light wavelength spectrum, achieving HR but no color information, and several cameras integrate over a narrow-band light wavelength spectrum, returning color information at LR. We use a very simple idea: we decompose the LR color image in its red (R), green (G) and blue (B) component and compute the special form of covariance (10) for the three LR R,G,B images as well as for the B&W HR image. Then we move along the geodesic the three R,G,B matrices all toward the direction of the B&W matrix, as per (11) with a step size close enough to 1 (0.95 in this example). Finally, we assign to each pixel the saturation of the LR color image. Figure 1 shows the result of these manipulations. Figure 2 shows the special form of the covariance matrix (10) for the HR B&W image (top-left), the red component of the LR image (top-right), the green component of the LR image (bottom-left) and the blue component of the LR image. In these images the diagonal blocks are the row and column covariance structures, while the off-diagonal blocks are the image itself and its transpose, as per construction of (10). Note that since the original picture is mainly red, the red LR special form of covariance matrix has more detailed covariance structures as compared to the green and blue components, but still much less detailed covariance structure as compared to the HR B&W image. We see that moving the special

form of covariance matrix of the R,G,B components along the geodesic toward the HR B&W image amounts to progressively matching the covariance structures of the three R,G,B components with the covariance structures of the B&W image, i.e., in this case, it amounts to increasing the spatial resolution loosing progressively color information; by construction of (10) such matching is reflected in the image itself (top-right off-diagonal block).



FIGURE 1. Simulation of satellite images of a car parking. From left to right and top to bottom: a) a color low-resolution image, b) its B&W high resolution version, c) its color high-resolution version (here used as a benchmark, but it cannot be generated by the satellite) and d) the color image we have reconstructed using a) and b) only.

Example 2

The second example estimates for the first time the "geometric mean" of event-related potentials (ERP). ERP data were gathered using the Brain Invaders brain-computer interface [4]. $\mathbf{X}_i$ in this case holds the brain electrical signal recorded for a short time period (one second in this example) starting at time "zero" corresponding to an external stimulation sent to the brain (a visual stimulation in this case) and it is named a "trial". Each row $\mathbf{x}_{ri} \in \Re^C$ of $\mathbf{X}_i$ as per (2) is the electric potential difference recorded at the r^{th} electrode with respect to a cephalic reference. These potential changes are stereotypical in a given experimental condition and are time- and phase-locked to the stimulus, i.e., they have similar form and amplitude in different trials. However, an inter-trial variability is present. Moreover, potential changes are very small in amplitude (a few μV) as compared to the spontaneous electroencephalography in the background (tens of μV), hence the signal-to-noise variance ratio (SNR) for a single trial is -20dB or smaller,

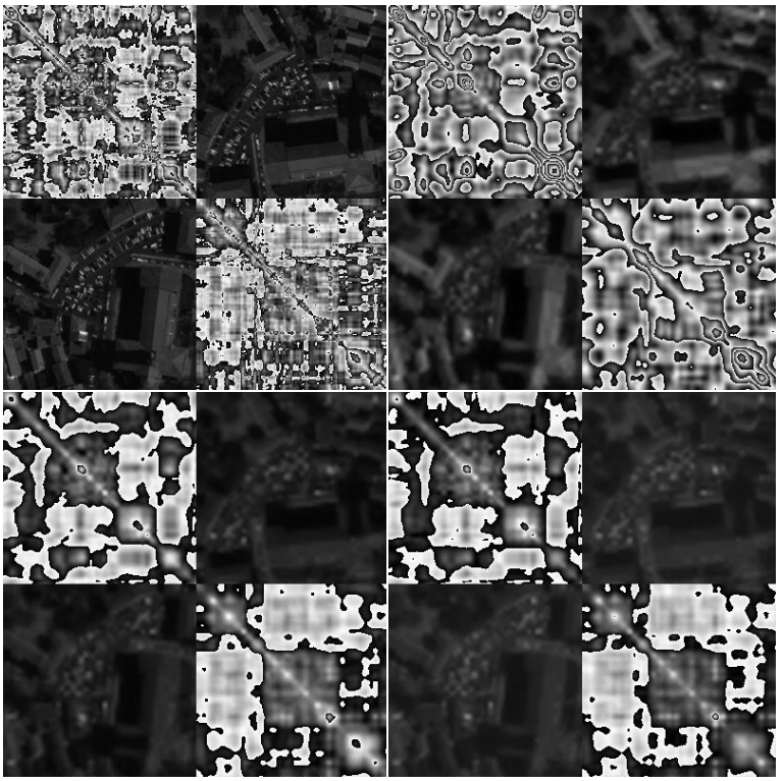


FIGURE 2. Color-coded (white=0, black=max) images of the special form of covariance matrix (10) computed from the high resolution B&W image (top-left), the red component of the low resolution color image (top-right), the green component of the low resolution color image (bottom-left) and the blue component of the low resolution color image (bottom-right).

that is to say, the ERP is invisible on raw data. To increase the SNR the ERPs are usually averaged (sample-by-sample arithmetic mean) in order to enhance the stereotypical response, while spontaneous activity and artifacts, being not time- and -phase locked to the stimulus and oscillating around zero, cancel out across trials in the summation. In this example the spatial structure of a trial depends on what areas of the brain contribute to the signal and on the electrode placement on the scalp, whereas the temporal structure depends on the timing of their activity. The data has been presented in [3], to which the reader is referred for details on data acquisition and experimental procedures. We present data obtained on one subject for which 58 ERPs were recorded. 16 electrodes ($R = 16$) were used. As only preprocessing, data was band-pass filtered in the frequency region 1-20Hz and down-sampled at 128 samples per second ($C = 128$). Each trial holds one second of data post-stimulus, therefore our data matrices $\mathbf{X}_i$ have dimension 16×128.

In order to compute the "geometric mean" of ERPs we use (10) with double-centering (8) as PSD characterization of the trials. Before double-centering the data and applying (10) (with $\alpha = 1$), in this example we reduce the dimension of the data by "reducing" each trial $\mathbf{X}_i$ such as

$$\mathbf{X}_i \leftarrow \mathbf{U}^T \mathbf{X}_i \mathbf{V}, \tag{12}$$

where $\mathbf{U} \in \Re^{R \times P}$ and $\mathbf{V} \in \Re^{C \times Q}$ are taken, respectively, as the first $P = Q = 8$ left

and right singular vectors of the arithmetic mean of target trials $\mathbf{X}_i$. That is, if $\mathbf{U}^T\mathbf{W}\mathbf{V}$ is the singular value decomposition of the average ERP, the columns of $\mathbf{U}$ and $\mathbf{V}$ span an orthogonal basis of the signal subspace for the spatial ($\mathbf{U}$) and temporal ($\mathbf{V}$) covariance of the average ERP. Once data dimension have been reduced by (12) matrices $\mathbf{C}_i$ (10) have dimension 16×16 only. As algorithm for computing the geometric mean of trials (10) we use the well-known gradient descent algorithm that can be found, for instance, in [3]. The geometric mean is known to be more robust to outliers as compared to the arithmetic mean. In order to verify this feature on ERPs, on one trial only (out of 58 available) we added -500 μV (an aberrant voltage value) to samples 5 to 15 of the first electrode. This simulates a strong square-wave transient artifact. See caption of Fig. 3 for the results. Note that the filtered arithmetic mean is very similar to the arithmetic mean obtained on the raw data. This merely states that the generation of the ERP involves no more than eight active dipoles (row, spatial components) acting on no more than eight temporal windows (column, temporal components). This is well-known in the ERP literature. Processing the data without the simulated outlier the arithmetic mean (AM) and the geometric mean (GM) are similar, but not identical; in both cases we can observe three major peaks of the global field power (see pointing arrows in Fig. 3): a positive deflection at 280 ms, a negative deflection at 450 ms and a positive deflection at 650 ms. For the GM a forth positive deflection at 580 ms is better visible. Also, the first two GFP peaks are sharper and better separated for the GM mean.

CONCLUSIONS

We have presented a special form of covariance matrix that can be computed starting from any generic data matrix. Such special covariance matrix de facto allows the manipulation of generic data matrices using the tools of Riemannian geometry in the SPD manifold. Furthermore, this special form of covariance matrix is unique for each data matrix (for a given α), thus may be used, for instance, to classify directly any kind of data matrix using Riemannian methods as shown for example in [1] and in [3]. In summary, the presented special form of covariance matrix has two interesting properties:

1. it allows the visualization of manipulations on the Riemannian manifold of SPD matrices in the original space of the data that have generated the SDP matrices.
2. It allows an exhaustive and unique characterization of a data matrix by describing both the covariance structure of its rows and of its columns.

The only practical limitation of the method is that each data matrix $\mathbf{X}_i$ analyzed must be of the same size. Our first example (satellite images) has illustrated how any data matrix can be moved along geodesics toward any other data matrix. This may be very useful in some applications, however the method used in the example is not meant to represent a competitive option for pansharpening. In the second example we have estimated the "geometric mean" of 58 ERPs. This is the first time the concept of geometric mean on the SPD manifold is translated to ERP data. The example demonstrates that the geometric mean may be robust to outliers, thence such estimation deserves further investigation in the field of neuroscience. The choice of parameter α is not trivial. It should be high enough so that the $\mathbf{C}_i$ matrices are positive and well enough conditioned. On the other

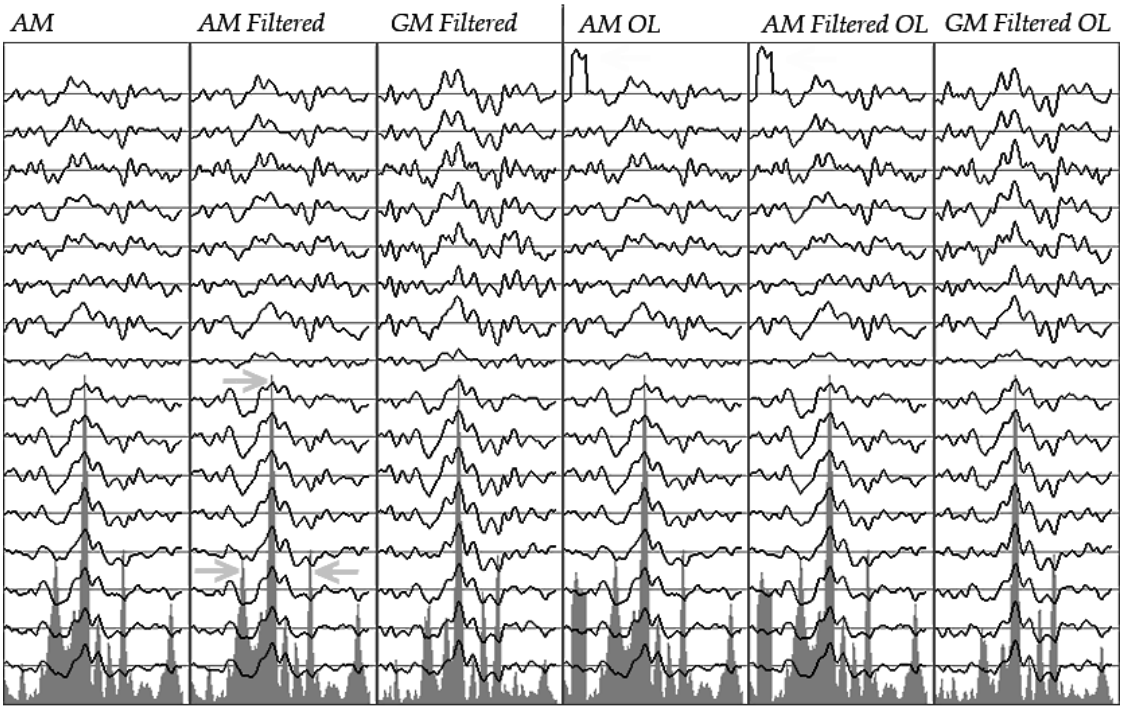


FIGURE 3. 1-sec average ERP estimations for one subject recorded at 16 electrodes. For consistency with neurological convention, a positive voltage deflection is plotted downward. The voltage scale of each plot is arbitrary but consistent within each plot. The shaded grey area is the global field power (GFP), that is, the total energy (sum of the squared potential across electrodes) at each sample. The GFP is scaled arbitrarily in each plot. The first three plots are, from left to right, the arithmetic mean of the raw ERPs $\mathbf{X}_i$, the arithmetic mean of the ERPs after filtering of the trials as per (12) (given by $\mathbf{X}_i \leftarrow \mathbf{U}^T\mathbf{X}_i\mathbf{V}$) and the geometric mean of the trials obtained as the top-right off-diagonal block of the geometric mean of trials represented by (10), ri-expanded in the original dimension by $\mathbf{X}_i \leftarrow \mathbf{U}\mathbf{X}_i\mathbf{V}^T$. The next three plots are the same three mean estimations obtained with the outlier, which is well visible at the beginning of the trace of the first electrode only in the arithmetic means.

hand, increasing α amounts to move the points $\mathbf{C}_i$ on the manifold all toward the point $\alpha\mathbf{I}$. Such region of the SPD manifold (the hyper plane of SPD diagonal matrices) is the least curved therein - in fact, this hyper plane is log-Euclidean, see [3] - . Thus, increasing α will not only progressively reduce the radius of the ball containing the points $\mathbf{C}_i$, but also as α increases we expect to observe less and less difference in between the geometric and arithmetic mean of the data matrix blocks $\mathbf{X}_i$ embedded in these points. This suggests that several "mean ERP" may be estimated varying α within a suitable range leaving to the investigator the choice of the most suitable mean. Further research will elucidate the implications on the choice of α we have touched on.

ACKNOWLEDGMENTS

This work has been partly supported by the European Project ERC-2012-AdG-320684-CHESS

REFERENCES

1. A. Barachant, S. Bonnet, M. Congedo and C. Jutten, *Multi-Class Brain Computer Interface Classification by Riemannian Geometry*, IEEE Transactions on Biomedical Engineering 59(4), 2012, pp. 920-928.
2. R. Bhatia and J. Holbrook J, *Riemannian geometry and matrix geometric mean*, Linear Algebra and its applications, 413, 2006, pp. 594-618.
3. M. Congedo, *Riemann Geometry: an Universal BCI Classification Framework (Ch. VIII)*, In: EEG Source Analysis (Congedo M), HDR presented at doctoral School EDISCE, Grenoble University, 2013, pp. 194-222.
4. M. Congedo, M. Goyat, N. Tarrin , L. Varnet , B. Rivet, G. Ionescu, et al. *"Brain Invaders": a prototype of an open-source P300-based video game working with the OpenViBE platform*. Proc of the 5th Int BCI Conference, Graz, Austria, 2011, pp. 280-283.
5. R.D. Pascual-Marqui , D. Lehmann, K. Kochi, T. Kinoshita, N. Yamada : *A measure of association between vectors based on "similarity covariance"*, 2013-01-21, arXiv:1301.4291.
6. S.S. Searle, *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York, 1982.