Classification of multivariate signals using Riemannian geometry and the Hilbert transform

Martin Gemborn Nilsson[1] and Bo Bernhardsson[1]

[1]Department of Automatic Control, Lund University

April 23, 2025

1 Introduction

We use the Hilbert transform to augment multivariate signals and compute the corresponding complex-valued covariance matrices. This augmented covariance matrix representation is then combined with the geometrically intuitive minimum distance to mean classifier, working directly on a Riemannian manifold of Hermitian positive definite matrices. We see that this combination, without the need of any additional hyper-parameters, can lead to improved classification of multivariate signals, compared to using standard covariance matrices.

2 Methods

For a n-variate wide sense stationary (WSS) stochastic process $\mathbf{x}(t)$ we denote the covariance as $X := C_{\mathbf{xx}} := \mathrm{Cov}(\mathbf{x}(t), \mathbf{x}(t)) \in R^{n\times n}$. We also introduce the auto-covariance function $r_{\mathbf{xx}}(\tau) = \mathrm{Cov}(\mathbf{x}(t+\tau), \mathbf{x}(t))$ and notice that $r_{\mathbf{xx}}(0) = C_{\mathbf{xx}}$. Covariance matrices, or estimated sample covariance matrices thereof, can be identified with the symmetric positive definite (SPD) matrix of size n, here denoted $\mathcal{S}^n_{++}$. We use $\mathcal{S}^n$ to denote the set of symmetric matrices of size n.

2.1 Classification of SPD matrices using Riemannian Geometry

The set $\mathcal{S}^n_{++}$ constitutes a differentiable manifold and the set of tangent vectors to $\mathcal{S}^n_{++}$ at a point $P \in \mathcal{S}^n_{++}$ can be identified with $\mathcal{S}^n$. When equipped with a Riemannian metric, g, the pair $(\mathcal{S}^n_{++}, g)$ constitutes a Riemannian manifold. A commonly used metric is the so called affine invariant Riemannian (AIR) metric

$$\begin{aligned}\langle V, W\rangle_X &= \langle X^{-\frac{1}{2}} V X^{-\frac{1}{2}}, X^{-\frac{1}{2}} W X^{-\frac{1}{2}}\rangle \\ &= \mathrm{tr}(X^{-1} V X^{-1} W),\end{aligned} \tag{1}$$

where $X \in \mathcal{S}^n_{++}$ and $V, W \in \mathcal{S}^n$ are vectors in the tangent-space at X. The induced geodesic distance between two points $X, Y \in \mathcal{S}^n_{++}$ using the AIR-metric is [1]

$$d_{AIR}(X, Y) = ||X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}||_F = \left[\sum_i^n \ln^2(\lambda_i)\right]^{\frac{1}{2}}, \tag{2}$$

where λ_i are the eigenvalues of $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$.

The minimum distance to mean (MDM) classifier is a simple and geometrically intuitive classifier that can be used directly on Riemannian manifolds $(\mathcal{M}, g_R)$. The MDM classifier is trained by finding a mean for data from each class-condition. Often the sample Fréchet mean is used, defined as

$$\bar{X} = \underset{X}{\mathrm{argmin}} \sum_{i=1}^N d_R^2(X_i, X), \tag{3}$$

where $d_R(\cdot,\cdot)$ is the geodesic distance induced from the metric. Unseen data is assigned to the class for which it has the shortest distance to the class mean.

2.2 Hilbert transform

The Hilbert transform for functions on R is defined as

$$\hat{f}(t) = \frac{1}{\pi}\,\mathrm{p.v.}\int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau} d\tau, \tag{4}$$

where p.v. denotes an extension of the regular integral definition called the Cauchy principal value [2].

For a component-wise WSS stochastic process, $\mathbf{x}(t)$, the Hilbert transform is defined as the component-wise output of the linear filter with frequency function $g(\omega) = -i\,\mathrm{sgn}(\omega)$ applied to $\mathbf{x}(t)$ [3].

For a signal $\mathbf{x}(t)$ we denote its component-wise Hilbert transform as $\hat{\mathbf{x}}(t)$. The corresponding analytic signal of $\mathbf{x}(t)$ is denoted as $\mathbf{x}_a(t) = \mathbf{x}(t) + i\hat{\mathbf{x}}(t)$. No matter if we consider the signal to be a WSS stochastic process, a (truncated) realization thereof, or two functions of certain function classes, one can show that

$$r_{\mathbf{x}\hat{\mathbf{y}}}(\tau) = -r_{\hat{\mathbf{x}}\mathbf{y}}(\tau) = -r^T_{\mathbf{y}\hat{\mathbf{x}}}(-\tau). \tag{5}$$

The complex valued covariance matrix corresponding to $\mathbf{x}_a(t)$ is $X_a := \mathrm{Cov}(\mathbf{x}_a, \mathbf{x}_a)$. Expanding X_a using Equation 5, we get the following real and imaginary components

$$X_a = C_{\mathbf{xx}} + iC_{\hat{\mathbf{x}}\mathbf{x}} - iC_{\mathbf{x}\hat{\mathbf{x}}} + C_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = 2C_{\mathbf{xx}} + 2iC_{\hat{\mathbf{x}}\mathbf{x}}. \tag{6}$$

Such matrices X_a are Hermitian positive definite (HPD) matrices, here denoted as $\mathcal{H}^n_{++}$. At a point

$X \in \mathcal{H}_{++}^n$, the corresponding tangent space is the set of Hermitian matrices, here denoted $\mathcal{H}^n$. Pairing $\mathcal{H}_{++}^n$ and the AIR-metric from Equation 1 yields another Riemannian manifold. The geodesic distance between two such elements is again given by Equation 2.

3 Results

3.1 Classification example

Artificial data for two classes C_1, C_2 are generated as

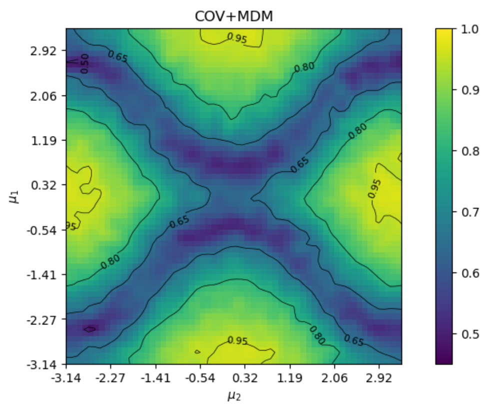
$$\mathbf{x}_{C_i}(t) = \begin{bmatrix} \cos(2\pi f t + \psi) \\ \cos(2\pi f t + \psi + \phi_i) \end{bmatrix} + \mathbf{e}(t), \tag{7}$$

where $\psi \sim \mathcal{U}(0, 2\pi)$, $\phi_i \sim \mathcal{N}(\mu_i, \sigma_i)$ (for $i = 1, 2$), $t \in [0, 1]$ is linearly spaced with $T = 1000$ samples, $f = 5$, and $\mathbf{e}(t)$ is white noise with covariance $\Sigma_{\mathbf{e}}$. The corresponding standard and analytic sample covariance matrix of such data are approximately

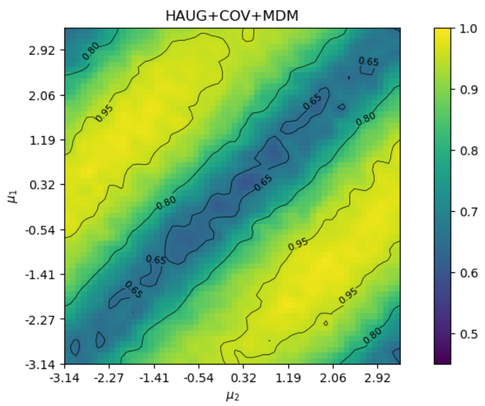
$$X(\phi_i) = \frac{1}{2}\begin{bmatrix} 1 & \cos(\phi_i) \\ \cos(\phi_i) & 1 \end{bmatrix} + \Sigma_e, \tag{8}$$

$$X_a(\phi_i) = \begin{bmatrix} 1 & e^{-j\phi_i} \\ e^{j\phi_i} & 1 \end{bmatrix} + 2\Sigma_e. \tag{9}$$

For datasets generated with different class-parameters ($\mu_1, \mu_2 \in [-\pi, \pi]$, $\sigma_1 = 0.45$, $\sigma_2 = 0.9$), MDM-classifiers using the two covariance matrix representation are evaluated. The corresponding accuracies are reported in Figure 1.



(a) Accuracies using standard covariance matrices.



(b) Accuracies using analytic covariance matrices.

Figure 1: Average classification accuracies for the MDM-classifiers on the artificial datasets.

3.2 Classification of EEG-data:

We also evaluate the two different covariance matrix representations for classification of EEG-data. The dataset includes data from 9 subjects running two sessions each. Each session consists of 288 trials, where the subject performs motor imagery (MI) during 4 seconds for one out of 4 different MI-conditions: *left hand*, *right hand*, *feet*, or *tongue*. The EEG-device used in the experiments have 22 channels and data was sampled 250 Hz. The data is band-pass filtered between 8-35 Hz which is common practice for MI decoding. The used performance metric is the average accuracy of a 5-fold cross-validation, evaluated for each experiment session individually. The results for each individual subjects are reported in Figure 2.

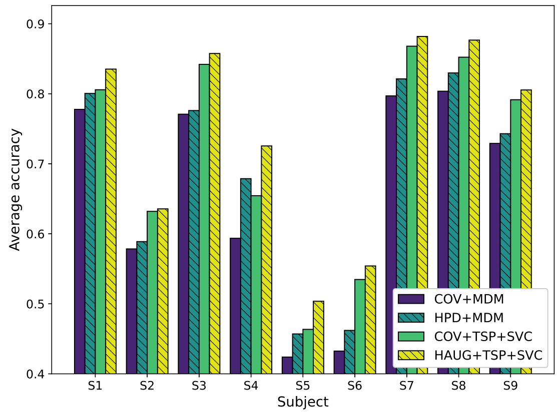


Figure 2: Average accuracy of each subject using the MDM-classifiers (as well as a tangent space based classifier) when using the analytic and standard covariance matrices respectively.

4 Conclusions

The Hilbert transform is used to construct analytic covariance matrices estimated from multivariate signals. Without introducing any new hyperparameters, we display how this way of augmenting covariance matrices can differentiate cases where the standard covariance matrix can not. We show how the analytic version of covariance matrices can improve the performance of a geometrically intuitive machine learning algorithm (MDM), working directly on the differentiable manifold of SPD and HPD matrices.

References

[1] Moakher M. A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. SIAM Journal on Matrix Analysis and Applications. 2005;26(3):735-47.

[2] King FW. Hilbert Transforms. Encyclopedia of Mathematics and its Applications. Cambridge University Press; 2009.

[3] Lindgren G. Stationary stochastic processes: Theory and applications. Texts in Statistical Science. Chapman and Hall; 2012.