

Classification in the Siegel Space for Vectorial Autoregressive Data

Yann Cabanes^{1,2} and Frank Nielsen^{3,4*}

¹ Thales LAS Limours, France

² Institute of Mathematics of Bordeaux, France yann.cabanes@gmail.com

³ Sony Computer Science Laboratories Inc, Japan

⁴ École Polytechnique, LIX, France

Frank.Nielsen@acm.org

Abstract. We introduce new geometrical tools to cluster data in the Siegel space. We give the expression of the Riemannian logarithm and exponential maps in the Siegel disk. These new tools help us to perform classification algorithms in the Siegel disk. We also give the expression of the sectional curvature in the Siegel disk. The sectional curvatures are negative or equal to zero, and therefore the curvature of the Siegel disk is non-positive. This result proves the convergence of the gradient descent performed when computing the mean of a set of matrix points in the Siegel disk.

Keywords: Siegel space · Riemannian manifold · Riemannian exponential map · Riemannian logarithm map · sectional curvature · machine learning · Information geometry · VAR model.

1 Complex vectorial autoregressive Gaussian models

We present here the multidimensional linear autoregressive model which generalizes the one-dimensional model presented in [3].

1.1 The multidimensional linear autoregressive model

We assume that the multidimensional signal can be modeled as a centered stationary autoregressive multidimensional Gaussian process of order $p - 1$:

$$U(k) + \sum_{j=1}^{p-1} A_j^{p-1} U(k-j) = W(k) \quad (1)$$

where W is the prediction error vector which we assume to be a standard Gaussian random vector and the prediction coefficients A_j^{p-1} are square matrices.

* This work was initiated while the author was at École Polytechnique, France. This work was further developed during a thesis with Thales LAS France and the Institute of Mathematics of Bordeaux. We thank the French MoD, DGA / AID for funding (convention CIFRE AID N°2017.0008 & ANRT N°2017.60.0062)

1.2 Three equivalent representation spaces

There are at least three equivalent spaces to represent our model parameter, they are described in detail in [6]. The first one is the set of Hermitian Positive Definite Block-Toeplitz matrices corresponding to the covariance matrix of the large column vector U defined by:

$$U \stackrel{\text{def}}{=} [U(0)^T, \dots, U(p-1)^T]^T \quad (2)$$

The second one is a product space: a HPD matrix (which characterizes the average correlation matrix) and the coefficients $(A_i^i)_{i=1, \dots, p-1}$ (which characterize the multidimensional autoregressive model). The third representation space looks like the second one: the coefficients A_i^i are slightly modified to belong to the Siegel disk which metric has been studied in [6, 8].

1.3 A natural metric coming from information geometry

In [6], the three equivalent representation spaces presented in section 1.2 are endowed with a natural metric coming from information geometry. Indeed, the model assumptions done in section 1.1 are equivalent to the assumption that the large vector U described in equation (2) is the realisation of a Gaussian process with zero mean and an Hermitian Positive Definite Block-Toeplitz covariance matrix. We define a metric on this space as the restriction of the information geometry metric on the space of Gaussian processes with zero mean and an Hermitian Positive Definite covariance matrices. We finally transpose this metric to the equivalent representation space constituted of a HPD matrix and coefficients in the Siegel disks. Luckily, the metric in this space is a product metric. The metric on the HPD manifold is the information geometry metric [6]. The metric on the Siegel disk is described in detail in section 2.

2 The Siegel disk

In this section we present a Riemannian manifold named the Siegel disk and introduce the Riemannian logarithm and exponential maps. These tools will be very useful to classify data in the Siegel space, as shown in section 3. We also introduce the formula of the Siegel sectional curvature and prove it to be non-positive which proves the convergence of classification algorithms based on mean computations, such as the k-means algorithm. The Siegel disk generalizes the Poincaré disk described in [3], [4].

2.1 The space definition

Definition 1. *The Siegel disk is defined as the set of complex matrices M of shape $N \times N$ with singular values lower than one, which can also be written:*

$$SD_N = \{M \in \mathbb{C}^{N \times N}, I - MM^H > 0\} \quad (3)$$

or equally:

$$SD_N = \{M \in \mathbb{C}^{N \times N}, I - M^H M > 0\}. \quad (4)$$

We use the partial ordering of the set of complex matrices: we note $A > B$ when the difference $A - B$ is a positive definite matrix. As the Siegel disk is here defined as an open subset of the complex matrices $\mathbb{C}^{N \times N}$, its tangent space at each point can also be considered as $\mathbb{C}^{N \times N}$.

Note that another definition of the Siegel disk also exists in other papers [8], imposing an additional symmetry condition on the matrix M : $M = M^T$. We will not require the symmetry condition in our work.

Property 1. The Siegel disk can also be defined as the set of complex matrices M with a linear operator norm lower than one: $SD_N = \{M \in \mathbb{C}^{N \times N}, \|M\| < 1\}$, where $\|M\| = \sup_{X \in \mathbb{C}^{N \times N}, \|X\|=1} (\|MX\|)$.

2.2 The metric

Square of the line element ds

$$ds^2 = \text{trace} \left((I - \Omega \Omega^H)^{-1} d\Omega (I - \Omega^H \Omega)^{-1} d\Omega^H \right) \quad (5)$$

The scalar product $\forall \Omega \in SD_N, \forall v, w \in \mathbb{C}^{N \times N}$:

$$\langle v, w \rangle_\Omega = \frac{1}{2} \text{trace} \left((I - \Omega \Omega^H)^{-1} v (I - \Omega^H \Omega)^{-1} w^H \right) \quad (6)$$

$$+ \frac{1}{2} \text{trace} \left((I - \Omega \Omega^H)^{-1} w (I - \Omega^H \Omega)^{-1} v^H \right) \quad (7)$$

The norm of a vector belonging to the tangent space is therefore:

$$\|v\|_\Omega^2 = \text{trace} \left((I - \Omega \Omega^H)^{-1} v (I - \Omega^H \Omega)^{-1} v^H \right) \quad (8)$$

The distance

$$d_{SD_N}^2(\Omega, \Psi) = \frac{1}{4} \text{trace} \left(\log^2 \left(\frac{I + C^{1/2}}{I - C^{1/2}} \right) \right) \quad (9)$$

$$= \text{trace} \left(\text{arctanh}^2 \left(C^{1/2} \right) \right) \quad (10)$$

with $C = (\Psi - \Omega) (I - \Omega^H \Psi)^{-1} (\Psi^H - \Omega^H) (I - \Omega \Psi^H)^{-1}$.

2.3 The isometry

In [6], the following function is said to be an isometry for the Siegel distance described in equation (9).

$$\Phi_{\Omega}(\Psi) = (I - \Omega\Omega^H)^{-1/2} (\Psi - \Omega) (I - \Omega^H\Psi)^{-1} (I - \Omega^H\Omega)^{1/2} \quad (11)$$

Property 2. The differential of the isometry Φ has the following expression:

$$D\Phi_{\Omega}(\Psi)[h] = (I - \Omega\Omega^H)^{1/2} (I - \Psi\Omega^H)^{-1} h (I - \Omega^H\Psi)^{-1} (I - \Omega^H\Omega)^{1/2} \quad (12)$$

Property 3. The inverse of the function Φ described in equation (11) is:

$$\Phi_{\Omega}^{-1}(\Psi) = \Phi_{\Omega}(-\Psi) \quad (13)$$

2.4 The Riemannian logarithm map

Riemannian logarithm map at 0 In [6], the logarithm map at 0 is given by the formula :

$$\log_0(\Omega) = \mathcal{V} \Omega \quad (14)$$

with:

$$\mathcal{V} = \mathcal{L} \left((\Omega\Omega^H)^{1/2}, \log \left(\frac{I + (\Omega\Omega^H)^{1/2}}{I - (\Omega\Omega^H)^{1/2}} \right) \right) \quad (15)$$

where $\mathcal{L}(A, Q)$ is defined as the solution of:

$$AZ + ZA^H = Q \quad (16)$$

However, the expression of the logarithm map at zero given in [6] can be greatly simplified.

Property 4. The Riemannian logarithm map of the Siegel disk at zero has the following expression:

$$\log_0(\Omega) = \operatorname{arctanh}(X) X^{-1} \Omega \quad \text{where } X = (\Omega\Omega^H)^{1/2} \quad (17)$$

Note that for $\|X\| < 1$, we can develop the function $\operatorname{arctanh}$ in whole series:

$$\operatorname{arctanh}(X) = \sum_{n=0}^{+\infty} \frac{X^{2n+1}}{2n+1} \quad (18)$$

Hence for all X in the Siegel disk, we can write the product $\operatorname{arctanh}(X) X^{-1}$ the following way:

$$\operatorname{arctanh}(X) X^{-1} = \sum_{n=0}^{+\infty} \frac{X^{2n}}{2n+1} \quad (19)$$

This new expression is also valid when the matrix X is not invertible.

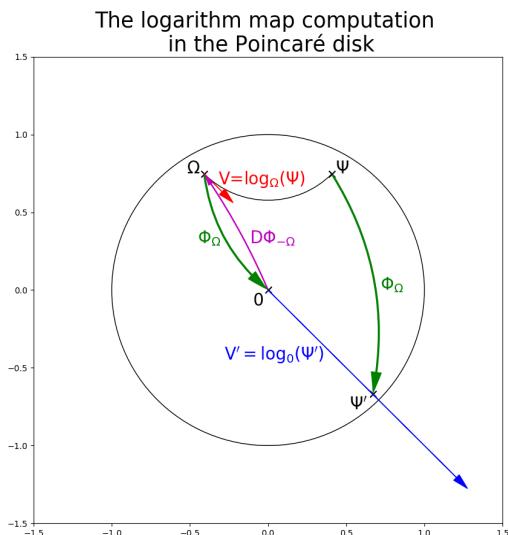


Fig. 1. The Poincaré disk logarithm map computation

2.5 The Riemannian logarithm map at any point

To compute the Riemannian logarithm map at a point Ω the key idea here is to transport the problem at zero, compute a certain logarithm at zero and transport the result back to Ω . If we want to compute the logarithm map: $\log_{\Omega}(\Psi)$, we first transport both Ω and Ψ using the isometry Φ_{Ω} given in equation (11). The point Ω is sent to zero, and we denote Ψ' the image of Ψ by Φ_{Ω} :

$$\Psi' \stackrel{\text{def}}{=} \Phi_{\Omega}(\Psi) = (I - \Omega\Omega^H)^{-1/2} (\Psi - \Omega) (I - \Omega^H\Psi)^{-1} (I - \Omega^H\Omega)^{1/2} \quad (20)$$

Then we compute the logarithm map at zero $\log_0(\Psi')$:

$$V' \stackrel{\text{def}}{=} \log_0(\Psi') = \operatorname{arctanh}(X) X^{-1} \Psi' \quad \text{where } X = (\Psi' \Psi'^H)^{1/2} \quad (21)$$

And finally, we transport back the logarithm to the point Ω using the differential of the isometry Φ given in equation (12):

$$V \stackrel{\text{def}}{=} \log_{\Omega}(\Psi) = D\Phi_{-\Omega}(0) [V'] = (I - \Omega\Omega^H)^{1/2} V' (I - \Omega^H\Omega)^{1/2} \quad (22)$$

2.6 The Riemannian exponential map

Riemannian exponential map at 0

Property 5. The Riemannian exponential map of the Siegel disk at zero has the following expression:

$$\exp_0(V) = \tanh(Y) Y^{-1} V \quad \text{where } Y = (VV^H)^{1/2} \quad (23)$$

Note that for $\|X\| < \frac{\pi}{2}$, we can develop the function \tanh in whole series:

$$\tanh(X) = \sum_{n=1}^{+\infty} \frac{2^{2n} (2^{2n} - 1)}{(2n)!} B_{2n} X^{2n-1} \quad (24)$$

where B_{2n} are the Bernoulli numbers.

Hence for all X in the Siegel disk, we can write the product $\tanh(X) X^{-1}$ the following way:

$$\tanh(X) X^{-1} = \sum_{n=1}^{+\infty} \frac{2^{2n} (2^{2n} - 1)}{(2n)!} B_{2n} X^{2n-2} \quad (25)$$

This new expression is also valid when the matrix X is not invertible.

2.7 The Riemannian exponential map at any point

To compute the Riemannian exponential map at a point Ω the key idea here is to transport the problem at zero (as for the logarithm), compute a certain exponential at zero and transport the result back to Ω . If we want to compute the exponential map: $\exp_\Omega(V)$, we first transport the vector V at zero using the differential of the isometry Φ given in equation (12):

$$V' \stackrel{\text{def}}{=} D\Phi_\Omega(\Omega)[V] \quad (26)$$

$$= (I - \Omega\Omega^H)^{1/2} (I - \Omega\Omega^H)^{-1} V (I - \Omega^H\Omega)^{-1} (I - \Omega^H\Omega)^{1/2} \quad (27)$$

$$= (I - \Omega\Omega^H)^{-1/2} V (I - \Omega^H\Omega)^{-1/2} \quad (28)$$

Then we compute the exponential map at zero $\exp_0(V')$:

$$\Psi' \stackrel{\text{def}}{=} \exp_0(V') = \tanh(Y) Y^{-1} V' \quad \text{where } Y = (V'V'^H)^{1/2} \quad (29)$$

And finally, we transport back the exponential to the point Ω using the isometry $\Phi_{-\Omega}$ which is the inverse of isometry Φ_Ω (see property 3) and transport the point 0 back to Ω and the point Ψ' back to $\exp_\Omega(V)$:

$$\Psi \stackrel{\text{def}}{=} \exp_\Omega(V) \quad (30)$$

$$= \Phi_{-\Omega}(\Psi') \quad (31)$$

$$= (I - \Omega\Omega^H)^{-1/2} (\Psi' + \Omega) (I + \Omega^H\Psi')^{-1} (I - \Omega^H\Omega)^{1/2} \quad (32)$$

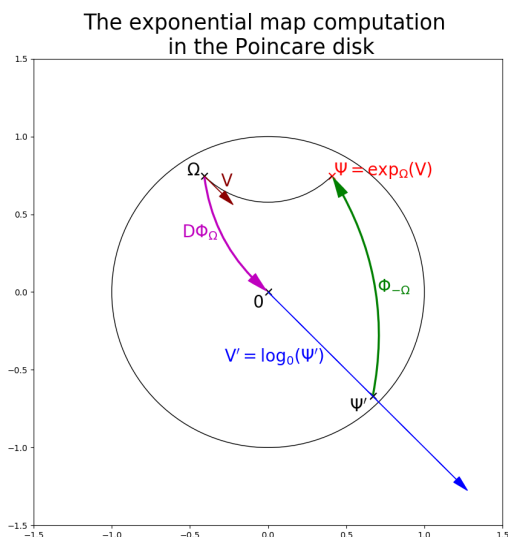


Fig. 2. The Poincaré disk exponential map computation

2.8 The geodesics

The expression the geodesics can be obtained using the exponential map: the geodesic starting from Ω with velocity V is given by the application:

$$\zeta(t) : t \mapsto \exp_{\Omega}(tV) \quad (33)$$

2.9 Sectional curvature of the Siegel space

We first focus on the sectional curvature at 0. We can then obtain the sectional curvature at any point using the isometry Φ defined in equation (11).

Let σ be a section defined by the two first vectors of an orthogonal basis (E_1, \dots, E_n) of the tangent space of the Siegel disk at the point $\Omega = 0$.

Theorem 1. *The sectional curvature at zero of the plan σ defined by E_1 and E_2 has the following expression:*

$$K(\sigma) = -\frac{1}{2} \left(\|E_1 E_2^H - E_2 E_1^H\|^2 + \|E_1^H E_2 - E_2^H E_1\|^2 \right) \quad (34)$$

As a consequence, we have:

$$-4 \leq K(\sigma) \leq 0 \quad \forall \sigma \quad (35)$$

The Siegel disk is therefore a Hadamard manifold. The bounds of the sectional curvature provides a proof of the convergence of certain algorithms calculating the Riemannian p-mean [1] or the circumcenter [2] of a set of points on a manifold.

2.10 The symmetric Siegel disk

We defined the Siegel disk in definition 1 as the set of complex matrices with singular values lower than one: $SD_N = \{M \in \mathbb{C}^{N \times N}, I - MM^H > 0\}$. We recall that another definition of the Siegel disk also exists in other papers [8], imposing an additional symmetry condition on the matrix M : $M = M^T$. Note that the symmetric Siegel disk is a totally flat submanifold of the Siegel disk. Hence the formula of the logarithm map 2.5, the exponential map 2.7 and the sectional curvature 2.9 computed in previous sections are still meaningful when working in the submanifold of symmetric matrices.

3 Application to stationary signals classification

An effective algorithm to estimate the model parameters from the raw data is described in [7], [5]. The classification model presented in this article has been applied to radar clutter classification in [4], [3] in the special case of one-dimensional complex signals and in [5] in the case of simulated multidimensional radar signals. This model will be applied in future work to stereo audio signals classification in the case of two-dimensional real signals.

References

1. Arnaudon, M., Barbaresco, F., Yang, L.: Riemannian Medians and Means With Applications to Radar Signal Processing. IEEE (August 2013)
2. Arnaudon, M., Nielsen, F.: On Approximating the Riemannian 1-Center. arXiv:1101.4718v3 <https://arxiv.org/pdf/1101.4718.pdf>
3. Cabanes, Y., Barbaresco, F., Arnaudon, M., Bigot, J.: Toeplitz Hermitian Positive Definite Matrix Machine Learning based on Fisher Metric. IEEE, GSI 2019, Toulouse, France (August 2019)
4. Cabanes, Y., Barbaresco, F., Arnaudon, M., Bigot, J.: Unsupervised Machine Learning for Pathological Radar Clutter Clustering: the P-Mean-Shift Algorithm. IEEE, C&ESAR 2019, Rennes, France (November 2019)
5. Cabanes, Y., Barbaresco, F., Arnaudon, M., Bigot, J.: Matrix Extension for Pathological Radar Clutter Machine Learning. hal-02875440 (2020)
6. Jeuris, B., Vandrebil, R.: The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks (2016)
7. Martin Morf, Augusto Vieira, D.T.L.L., Kailath, T.: Recursive Multichannel Maximum Entropy Spectral Estimation. IEEE Transactions on Geoscience Electronics, vol. GE-16, NO. 2 (April 1978)
8. Nielsen, F.: The Siegel-Klein Disk: Hilbert Geometry of the Siegel Disk Domain. Entropy **22**(9), 1019 (2020)