Matrix Extension for Pathological Radar Clutter Machine Learning

Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot

Matrix Extension for Pathological Radar Clutter Machine Learning

Yann Cabanes [1,2]
yann.cabanes@gmail.com

Frédéric Barbaresco [1]
frederic.barbaresco@thalesgroup.com

Marc Arnaudon [2]
marc.arnaudon@math.u-bordeaux.fr

Jérémie Bigot [2]
jeremie.bigot@math.u-bordeaux.fr

[1] Thales LAS, Advanced Radar Concepts, Limours, France
[2] Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, Talence, France

***Abstract*—This paper deals with radar clutter statistical learning based on spatial Doppler fluctuation. In articles [1]–[4], data is clustered cell by cell. In this article, we generalize the previous model to extract information not only from each cell independently, but also from the cells spatial correlation. We first introduce the radar data, then the model and efficient tools to estimate the model parameters. The model parameters will be shown to be Hermitian Positive Definite Block-Toeplitz matrices. Next we endow the manifold of Hermitian Positive Definite Block-Toeplitz matrices with a Riemannian metric coming from information geometry. Finally, we adapt a supervised classification algorithm (the k-Nearest Neighbors) and an unsupervised classification algorithm (the Agglomerative Hierarchical Clustering) to this Riemannian manifold.**

***Index Terms*—Radar clutter, multidimensional signals, spatio-temporal correlation, machine learning, Information geometry, Riemannian manifold, Block-Toeplitz matrices, Siegel disk.**

I. Introduction

Radar installation on a new geographical site is long and costly. We would like to shorten the time of deployment by automatically recognizing pathological clutters with past known diagnosed cases. This requirement will become more and more important as new generation radar sensors will be mobile and should work on the move and self-adapt to the environment. The objective is therefore to develop machine learning algorithms to recognize specific clutter characteristics from their Doppler spectrum statistical fluctuations.

To recognize pathological radar environments using a pulse-Doppler radar, we need to extract very precise Doppler information from a very small series of pulses (around 10). In this context, the classical FFT (Fast Fourier Transform) is not satisfactory due to its low resolution output for such small time series. To overcome this difficulty, we propose classification methods based on the signals spatio-temporal covariance matrices.

To begin with, we briefly introduce the radar data which we intend to analyze. For simplicity, we first consider one fixed direction in which a radar sends radio waves and we subdivide this direction into cells. The radar sends a burst of radio waves in a direction and then receives the echoes. For each echo we measure its amplitude r and phase ϕ, so it can be represented by a complex number $u = re^{i\phi}$. As a result, the original radar observation value of each cell is a complex vector $\mathbf{u} = [u(0), u(1), ..., u(n_{pulses} - 1)]^T$, where n_{pulses} is the number of radio waves emitted during the burst and $\cdot^T$ denotes the matrix transposition. We now group together spatially close cells, say $\mathbf{u}_0, ..., \mathbf{u}_{n_{cells}-1}$, and try to extract sharp Doppler information from each cell as well as spatial correlations.

Instead of using directly the original observation values recorded in an area, we implicitly vectorize the data registered in this zone: $\mathbf{u}_{zone} = \left[\mathbf{u}_0^T, ..., \mathbf{u}_{n_{cells}-1}^T\right]^T$ and estimate its covariance matrix. Under our model assumptions, this covariance matrix will be shown to be a Hermitian Positive Definite Block-Toeplitz matrix. In other words, the new observation value for each zone is a covariance matrix estimation, which is a Hermitian Positive Definite Block-Toeplitz matrix due to our model assumptions. Afterwards we define a metric on the manifold of Hermitian Positive Definite Block-Toeplitz matrices to consider it as a Riemannian manifold. The metric used here comes from the information geometry metric on the set of centered multidimensional complex Gaussian process with null relation matrix (such process can be uniquely determined by its covariance matrix). Then our classification problem can be summarized as follows: regroup in a same class the zones having close Hermitian Positive Definite Block-Toeplitz covariances matrices with respect to the previously mentioned metric.

We introduce the radar data in section II. We first present the one-dimensional temporal model of articles [1]–[4] in section III. Then we generalize this model to multidimensional signals in section IV. In section V, we detail a particular case of the multidimensional model presented in section IV based on separate estimations of the spatial and temporal correlation matrices. To illustrate our classification method, we use this last model in section VI to simulate radar data in order to perform the k-Nearest Neighbors classifier and the Agglomerative Hierarchical Clustering.

II. The radar data

We introduce the input radar data in the form of the two-dimensional complex matrix U:

$$U = \begin{bmatrix} u_{0,0} & u_{0,1} & \dots & u_{0,n_{pulses}-1} \\ u_{1,0} & u_{1,1} & \dots & u_{1,n_{pulses}-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n_{cells}-1,0} & u_{n_{cells}-1,1} & \dots & u_{n_{cells}-1,n_{pulses}-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n-1,0} & u_{n-1,1} & \dots & u_{n-1,n_{pulses}-1} \end{bmatrix} \tag{1}$$

Here the first coordinate corresponds to the spatial axis (index close to zero corresponds to cells close to the radar); the second coordinate represents the temporal axis (pulse index in the burst). In articles [1]–[4], the input data are the cells, i.e. the rows of the matrix U. The input data is now a group of spatially close cells: we gather several rows of the matrix U and try to extract spatial correlation information in addition to the Doppler information.

III. The one-dimensional temporal model

In this section, data is clustered cell by cell as in articles [1]–[4]. Instead of using directly the original observation vector $\mathbf{u}$ of each cell, we assume it to be a realization of a centered stationary complex Gaussian process and identify it with its covariance matrix $\mathbf{R} = \mathbb{E}[\mathbf{u}\,\mathbf{u}^H]$, where $\cdot^H$ denotes the complex matrix conjugate transpose. The new observation value for each cell is therefore a covariance matrix estimation, which is Toeplitz due to the assumption of stationarity of the process. We then develop efficient tools to estimate the Toeplitz Hermitian Positive Definite covariance matrix of the temporal signal $\mathbf{u}$ registered in each cell. Afterwards we endow the set of THPD matrices with a Riemannian metric coming from information geometry. Our classification problem can be summarized as follows: regroup in a same class the cells having close Toeplitz covariance matrices.

A. The model

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

$$u(k) + \sum_{i=1}^{n-1} a_i^{n-1}\, u(k-i) = w(k) \tag{2}$$

where a_i^{n-1} are the prediction coefficients and w is the prediction error.

B. Equivalent representations

According to [9], [10], based on Verblunsky-Trench algorithm, in case of locally stationary signal, the Levinson algorithm gives us the following bijection:

$$\begin{aligned} \mathcal{T}_n^+ &\to \mathbb{R}_+^* \times \mathbb{D}^{n-1} \\ R_n &\mapsto (p_0, \mu_1, ..., \mu_{n-1}) \end{aligned} \tag{3}$$

where $\mathcal{T}_n^+$ denotes the set of THPD matrices of size n; p_0 is the averaged quadratic power and $\mu_j = a_j^j$ $(1 \le j \le n-1)$ are the reflection coefficients, where a_i^j denotes the i^{th} coefficient of the Gaussian autoregressive model of order j. It is therefore equivalent to estimate the coefficients $(p_0, \mu_1, ..., \mu_{n-1})$ and R_n.

In practice, we use the regularized Burg algorithm [12] described in algorithm 1 to estimate the reflection coefficients.

Algorithm 1 The regularized Burg algorithm

Initialization:

$$f_{0,k} = b_{0,k} = u_k \quad k = 0, ..., n-1 \tag{4}$$

$$a_{0,k} = 1 \quad k = 0, ..., n-1 \tag{5}$$

$$p_0 = \frac{1}{n} \sum_{k=0}^{n-1} |u_k|^2 \tag{6}$$

for $i = 1, ..., M$: **do**

$$\mu_i = -\frac{\frac{2}{n-i} \sum_{k=i}^{n-1} f_{i-1,k} \bar{b}_{i-1,k-1} + 2 \sum_{k=1}^{i-1} \beta_{k,i} a_{k,i-1} a_{i-k,i-1}}{\frac{1}{n-i} \sum_{k=i}^{n-1} |f_{i-1,k}|^2 + |b_{i-1,k-1}|^2 + 2 \sum_{k=0}^{i-1} \beta_{k,i} |a_{k,i-1}|^2} \tag{7}$$

where:

$$\beta_{k,i} = \gamma (2\pi)^2 (k-i)^2 \tag{8}$$

$$\begin{cases} a_{k,i} &= a_{k,i-1} + \mu_i \bar{a}_{i-k,i-1} \quad k = 1, ..., i-1 \\ a_{i,i} &= \mu_i \end{cases} \tag{9}$$

$$\begin{cases} f_{i,k} &= f_{i-1,k} + \mu_i b_{i-1,k-1} \quad k = i, ..., n-1 \\ b_{i,k} &= b_{i-1,k-1} + \bar{\mu}_i f_{i-1,k} \quad k = i, ..., n-1 \end{cases} \tag{10}$$

end for

return $(p_0, \mu_1, ..., \mu_{n-1})$

C. The Kähler metric

A centered complex Gaussian distribution with null relation matrix is characterized by its covariance matrix, which belongs to the manifold P_n of HPD matrices of dimension n. Hence the Fisher metric naturally endows P_n with a Riemannian metric. The restriction of the Fisher metric on the submanifold $\mathcal{T}_n^+$ of THPD matrices allows us to consider $\mathcal{T}_n^+$ as a Riemannian manifold.

According to bijection (3), we can represent a THPD matrix T_i by the corresponding coefficients $(p_{0,i}, \mu_{1,i}, ..., \mu_{n-1,i})$. The following distance has been introduced by F. Barbaresco in [10] on the set $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ to make this bijection an isometry:

$$\begin{aligned} & d^2_{\mathcal{T}_n^+}(T_1, T_2) \\ =& d^2_{\mathcal{T}_n^+}\left((p_{0,1}, \mu_{1,1}, ..., \mu_{n-1,1}), (p_{0,2}, \mu_{1,2}, ..., \mu_{n-1,2})\right) \\ =& n \log^2\left(\frac{p_{0,2}}{p_{0,1}}\right) + \sum_{l=1}^{n-1} \frac{n-l}{4} \log^2\left(\frac{1+\left|\frac{\mu_{l,1}-\mu_{l,2}}{1-\mu_{l,1}\mu_{l,2}^*}\right|}{1-\left|\frac{\mu_{l,1}-\mu_{l,2}}{1-\mu_{l,1}\mu_{l,2}^*}\right|}\right) \end{aligned} \quad (11)$$

Note that the metric on the product space $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ is a product metric, which greatly simplifies the computations. The equations of the geodesics of the set $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$ endowed with the Kähler metric are described in [8].

The model presented in this section is based on Toeplitz covariance matrices of one-dimensional temporal signals. We now generalize this model to multidimensional signals.

IV. Generalization to multidimensional signals

We present here the multidimensional linear autoregressive model which generalize the one-dimensional model presented in section III.

A. The multidimensional linear autoregressive model

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

$$U(k) + \sum_{j=1}^{n} A_i^{n-1}\ U(k-j) = W(k) \quad (12)$$

where W is the prediction error vector and the prediction coefficients A_j^n are square matrices.

B. Three equivalent representation spaces

There is at least three equivalent spaces to represent our model parameter. The first one is the set of Hermitian Positive Definite Block-Toeplitz matrices corresponding to the covariance matrix of the vectorized data $U = \left[U(0)^T, \ldots, U(n-1)^T\right]^T$. The second one is a product space: a HPD matrix (which characterize the average temporal correlation matrix registered in a cell) and the coefficients $\left(A_i^i\right)_{i=1,\ldots,n-1}$ (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 [5], [14].

C. The Yule-Walker equations

The Yule-Walker equations link the first two representation spaces described previously. We define the autocorrelation coefficient R_t of a stationary signal $U(k)$ as $R_t \stackrel{\text{def}}{=} \mathbb{E}\left[U(k+t)U(k)^H\right]$. By applying $\mathbb{E}\left[(\cdot)U(k)^H\right]$ on each side of the equation (12) for different values of k, we obtain the Yule-Walker equations:

$$\begin{aligned} \widetilde{A}_n\ \widetilde{R}_n &= -\widetilde{V}_n \\ \widetilde{A}_n &= [A_1^n, \ldots, A_n^n] \\ \widetilde{V}_n &= [R_1, \ldots, R_n] \end{aligned} \quad (13)$$

$$\widetilde{R}_n = \begin{bmatrix} R_0 & R_1 & R_2 & \ldots & R_{n-1} \\ R_1^H & R_0 & R_1 & \ldots & R_{n-2} \\ R_2^H & R_1^H & R_0 & \ldots & R_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{n-1}^H & R_{n-2}^H & R_{n-3}^H & \ldots & R_0 \end{bmatrix} \quad (14)$$

Note that $\widetilde{R}_n$ is almost the covariance matrix of the vectorized data $U = \left[U(0)^T, \ldots, U(n-1)^T\right]^T$.

D. The transformation

The following bijection is defined in [5]:

$$\begin{aligned} \mathcal{B}_{n,N}^+ &\to P_N \times \mathcal{SD}_N^{n-1} \\ \widetilde{R}_n &\mapsto (P_0, \Omega_1, ..., \Omega_{n-1}) \end{aligned} \quad (15)$$

where $\mathcal{B}_{n,N}^+$ denotes the set of Hermitian Positive Definite Block-Toeplitz matrices of size $n \times N$ (n blocks of size N), P_N is the set of HPD matrices and $\mathcal{SD}_N = \{\Omega \in \mathbb{C}^{N\times N} | I - \Omega\Omega^H > 0\}$ is called the Siegel disk of dimension N (warning: there also exists a more restrictive definition of the Siegel disk).

The bijection is given explicitly by $P_0 = R_0$ and for $l = 1, ..., n-1$ we compute:

$$\Omega_l := -L_{l-1}^{-1/2}(R_l - M_{l-1})K_{l-1}^{-1/2} \quad (16)$$

$$L_{l-1} = R_0 - [R_1, ..., R_{l-1}]\ \widetilde{R}_{l-1}^{-1}[R_1, ..., R_{l-1}]^H \quad (17)$$

$$K_{l-1} = R_0 - [R_{l-1}^H, ..., R_1^H]\ \widetilde{R}_{l-1}^{-1}[R_{l-1}^H, ..., R_1^H]^H \quad (18)$$

$$M_{l-1} = [R_1, ..., R_{l-1}]\ \widetilde{R}_{l-1}^{-1}[R_{l-1}^H, ..., R_1^H]^H \quad (19)$$

A recursive algorithm can be found in [5], [9] taking advantage of the Block-Toeplitz structure of matrix $\widetilde{R}_n$.

E. The model parameters estimation

We use algorithm 2 to estimate the multidimensional reflection coefficients of the linear autoregressive model performed on the sequence of vectors $U_0, \ldots, U_{n-1}$.

In our study, we choose to perform the linear autoregressive model along the spatial axis: the input sequence will therefore be the vectors $\mathbf{u}_0, ..., \mathbf{u}_{n_{cells}-1}$ from a single zone.

For any positive-definite matrix R, we denote $R^{\frac{1}{2}}$ the lower triangular matrix satisfying $R = R^{\frac{1}{2}}\left(R^{\frac{1}{2}}\right)^H$ (Cholesky decomposition of R). $R^{\frac{1}{2}}$ can be made unique by requiring the diagonal elements to be positive. We also denote $R^{-\frac{1}{2}} \stackrel{\text{def}}{=} \left(R^{\frac{1}{2}}\right)^{-1}$.

Algorithm 2 Estimate multidimensional reflection coefficients

Input A vector sequence $U_0, \ldots, U_{n-1}$.

Initialization:

$$\boldsymbol{F}_{0,k} = \boldsymbol{B}_{0,k} = U_k \quad k = 0, ..., n-1 \tag{20}$$

$$P_0 = \frac{1}{n} \sum_{k=0}^{n-1} U_k \; U_k^H \tag{21}$$

for $i = 1, ..., n-1$: **do**

$$R_{i-1}^F = \sum_{k=i}^{n-1} \boldsymbol{F}_{i-1,k} {\boldsymbol{F}_{i-1,k}}^H \tag{22}$$

$$R_{i-1}^B = \sum_{k=i}^{n-1} \boldsymbol{B}_{i-1,k-1} {\boldsymbol{B}_{i-1,k-1}}^H \tag{23}$$

$$R_{i-1}^{FB} = \sum_{k=i}^{n-1} \boldsymbol{F}_{i-1,k} {\boldsymbol{B}_{i-1,k-1}}^H \tag{24}$$

$$M_i = -{R_{i-1}^F}^{-\frac{1}{2}} \; R_{i-1}^{FB} \; \left({R_{i-1}^B}^{-\frac{1}{2}}\right)^H \tag{25}$$

$$\left\{ \begin{array}{rcll} \boldsymbol{F}_{i,k} & = & \boldsymbol{F}_{i-1,k} + M_i \boldsymbol{B}_{i-1,k-1} & k = i, ..., n-1 \\ \boldsymbol{B}_{i,k} & = & \boldsymbol{B}_{i-1,k-1} + {M_i}^H \boldsymbol{F}_{i-1,k} & k = i, ..., n-1 \end{array} \right. \tag{26}$$

end for

return $(P_0, M_1, ..., M_{n-1})$

Theorem 1. *The matrices M_i of algorithm 2 satisfies, $\forall i \in [\![1, n-1]\!]$:*

$$I - M_i \; {M_i}^H \geq 0 \tag{27}$$

i.e. the matrix M_i has singular values of magnitude less or equal to one [6].

Proof. Let $i \in [\![1, n-1]\!]$, we first define the matrices E_f and E_b containing respectively the forward and backward prediction errors:

$$E_f \stackrel{\text{def}}{=} [F_{i-1,i}, \ldots, F_{i-1,n-1}] \tag{28}$$

$$E_b \stackrel{\text{def}}{=} [B_{i-1,i-1}, \ldots, B_{i-1,n-2}] \tag{29}$$

and

$$\widetilde{E}_f \stackrel{\text{def}}{=} \left(E_f \; E_f^H\right)^{-\frac{1}{2}} \; E_f, \quad \left(\widetilde{E}_f \widetilde{E}_f^H = I\right) \tag{30}$$

$$\widetilde{E}_b \stackrel{\text{def}}{=} \left(E_b \; E_b^H\right)^{-\frac{1}{2}} \; E_b, \quad \left(\widetilde{E}_b \widetilde{E}_b^H = I\right) \tag{31}$$

With these definitions, we can write:

$$M_i = \widetilde{E}_f \; \widetilde{E}_b^H \tag{32}$$

since

$$R_{i-1}^F = E_f E_f^H \quad R_{i-1}^B = E_b E_b^H \quad R_{i-1}^{FB} = E_f E_b^H. \tag{33}$$

We now consider:

$$0 \leq \left[\begin{array}{c} \widetilde{E}_f \\ \widetilde{E}_b \end{array} \right] \left[\widetilde{E}_f^H \quad \widetilde{E}_b^H \right] \tag{34}$$

$$= \left[\begin{array}{cc} I & M_i \\ M_i^H & I \end{array} \right] \tag{35}$$

$$= \left[\begin{array}{cc} I & M_i \\ 0 & I \end{array} \right] \left[\begin{array}{cc} I - M_i M_i^H & 0 \\ 0 & I \end{array} \right] \left[\begin{array}{cc} I & 0 \\ M_i^H & I \end{array} \right] \tag{36}$$

$$\iff I - M_i M_i^H \geq 0 \tag{37}$$

□

Note that if the number of linearly independent columns of the matrix $\left[\begin{array}{c} \widetilde{E}_f \\ \widetilde{E}_b \end{array} \right]$ is greater or equal to its number of rows, then we have:

$$I - M_i M_i^H > 0. \tag{38}$$

Under our model hypotheses, it occurs almost surely when: $i \leq n_{cells} - 2 * n_{pulses}$.

F. The metric

The metric of the space $P_N \times \mathcal{SD}_N^{n-1}$ is well described in [5], [9], we only provide an overview here.

We define the distance between two points of our representation space as follows:

$$d_{BT}^2 \left(\widetilde{T}_1, \widetilde{T}_2\right)$$

$$= d_{BT}^2 \left((P_{0,1}, \Omega_{1,1}, ..., \Omega_{n-1,1}), (P_{0,2}, \Omega_{1,2}, ..., \Omega_{n-1,2})\right) \tag{39}$$

$$= n \; \log \left(P_{0,1}^{-1/2} \; P_{0,2} \; P_{0,1}^{-1/2}\right)_F^2 + \tag{40}$$

$$\sum_{l=1}^{n-1} \frac{n-l}{4} \; \text{trace} \left(\log^2 \left(\frac{1 + C_l^{1/2}}{1 - C_l^{1/2}} \right) \right)$$

$$\left[\begin{array}{c} C_l = (\Omega_{l,2} - \Omega_{l,1}) \left(I - \Omega_{l,1}^H \; \Omega_{l,2}\right)^{-1} \\ \left(\Omega_{l,2}^H - \Omega_{l,1}^H\right) \left(I - \Omega_{l,1} \; \Omega_{l,2}^H\right)^{-1} \end{array} \right.$$

A gradient descent tool is described in [5], [9] to estimate the mean of a set of points on the Siegel disk.

V. The spatio-temporal model

A. The model

The spatio-temporal model presented here is a particular case of the multidimensional model presented in section IV. We assume the matrix signal Z, without noise from the radar measurement, has the following spatio-temporal correlation:

$$Z = \underbrace{T^{1/2} R_s^{1/2} N R_t^{T\,1/2}}_{\substack{\text{information coming from} \\ \text{the environment}}} \tag{41}$$

where R_s and R_t are the spatial and temporal autocorrelation matrices. We assume them to be THPD covariance matrices,

i.e. the signal is assumed to be stationary spatially and temporally. This model is more restrictive than the model presented in section IV: the spatio-temporal autocorrelation matrix $\widetilde{R}_{st}$ of the vectorized data will be shown to be a tensor product between the spatial correlation matrix R_s and the transpose of the temporal correlation matrix R_t. An almost similar model is described in [7].

B. The simulation model

The simulation model is the following:

$$Z = \underbrace{T^{1/2} R_s^{1/2} N {R_t^T}^{1/2}}_{\substack{\text{information coming from} \\ \text{the environment}}} + \underbrace{B_{radar}}_{\substack{\text{thermal noise coming} \\ \text{from the radar itself}}} \tag{42}$$

with:

- Z: input radar data of size (n_{cells}, n_{pulses}).
- T: clutter texture; it is a diagonal matrix of size (n_{cells}, n_{cells}). Its diagonal coefficients are independent positive real random variables. They are also independent from N and B_{radar}.
- R_s: spatial autocorrelation matrix of size (n_{cells}, n_{cells}). It is a THPD matrix since the signal is assumed to be stationary on the spatial axis. Its diagonal coefficients are equal to 1.
- R_t: temporal autocorrelation matrix of size (n_{pulses}, n_{pulses}). It is a THPD matrix since the signal is assumed to be stationary on the temporal axis. Its diagonal coefficients are equal to 1.
- N, B_{radar}: matrices of size (n_{cells}, n_{pulses}). They are filled with independent standard complex Gaussian random variables.

C. Equivalent simulation model

We now present an equivalent simulation model:

$$\widetilde{Z} = \underbrace{\widetilde{T}^{1/2} \widetilde{R}_{st}^{1/2} \widetilde{N}}_{\substack{\text{information coming from} \\ \text{the environment}}} + \underbrace{\widetilde{B}_{radar}}_{\substack{\text{thermal noise coming} \\ \text{from the radar itself}}} \tag{43}$$

with:

- $\widetilde{Z}$: input radar data of size $(n_{cells} * n_{pulses}, 1)$.
- $\widetilde{T}$: clutter texture; it is a diagonal matrix of size $(n_{cells} * n_{pulses}, n_{cells} * n_{pulses})$. Its diagonal coefficients are independent positive real random variables. They are also independent of $\widetilde{N}$ and $\widetilde{B}_{radar}$.
- $\widetilde{R}_{st}$: spatio-temporal autocorrelation matrix of size $(n_{cells} * n_{pulses}, n_{cells} * n_{pulses})$. Its diagonal coefficients are equal to 1.
- $\widetilde{N}$, $\widetilde{B}_{radar}$: matrices of size $(n_{cells} * n_{pulses}, 1)$. They are filled with independent standard complex Gaussian random variables.

We have the following correspondence between the two models:

$$Z = \begin{bmatrix} \boxed{Z_0^T} \\ \boxed{Z_1^T} \\ \vdots \\ \boxed{Z_{n_{cells}-1}^T} \end{bmatrix} \quad \widetilde{Z} = \begin{bmatrix} \underline{Z_0} \\ \underline{Z_1} \\ \vdots \\ \underline{Z_{n_{cells}-1}} \end{bmatrix} \tag{44}$$

$$\widetilde{R}_{st} = R_s \otimes R_t^T \tag{45}$$

$$\widetilde{T} = T \otimes I_{n\_pulses} \tag{46}$$

The first model is used to simulate the data, the second model is used to represent the structure of the simulated data. The matrix $\widetilde{R}_{st}$ is a Block-Toeplitz matrix. If we assume the signal to be stationary in space, then the matrix $\widetilde{R}_{st}$ is a Block-Toeplitz matrix with Toeplitz structured blocks.

D. The model parameters estimation

We estimate the spatial and temporal correlation matrices R_s and R_t performing respectively the regularized Burg algorithm (1) on the spatial and temporal axis of the observation matrix Z presented in equation (42).

We estimate the temporal correlation matrix R_t performing the regularized Burg algorithm on the columns matrix Z. In practice, we perform the regularized Burg algorithm on each column of the matrix Z to estimate the reflection coefficients of each cell. Then we compute the average reflection coefficients on the manifold $\mathbb{D}^{n-1}$ endowed with the Kähler metric described in section III-C. We can then estimate the temporal correlation matrix R_t using the inverse of bijection (3), i.e. the inverse of the Levinson algorithm described in [5].

We estimate the spatial correlation matrix using the same method, but performing the regularized Burg algorithm on the rows of the matrix Z instead of its columns.

We can now estimate the spatio-temporal correlation matrix using formula (45): it is the tensor product of the spatial correlation matrix and the transpose of the temporal correlation matrix.

Finally, we use bijection (15) to represent our data in $P_N \times \mathcal{SD}_N^{n-1}$. We prefer this last representation space because the associated metric is a product metric (see section IV-F). When performing classification algorithms, for computational reasons it is preferable to work in $P_N \times \mathcal{SD}_N^{n-1}$ which is a product of several low dimensional spaces endowed with a product metric rather than in the single higher dimensional space as $\mathcal{B}_{n,N}^+$.

VI. Classification

The difficulty here is to adapt classification algorithms to the Riemannian manifold described in section IV. We present a supervised classification algorithm: the k-Nearest

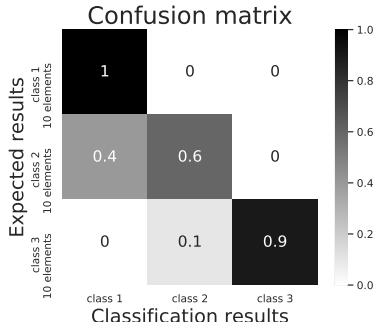

Fig. 1. Confusion matrix for the KNN classifier

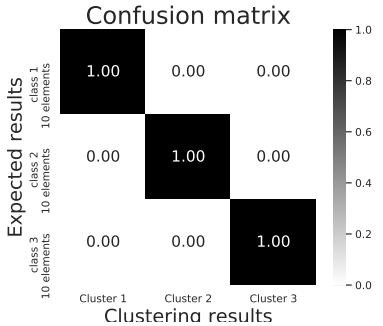

Fig. 2. Confusion matrix for the AHC clustering

Neighbors and an unsupervised classification algorithm: the Agglomerative Hierarchical Clustering. They both only use a distance as a geometrical tool to perform the classification. For more details on these algorithms, refer to the Python public package for machine learning on Riemannian manifolds Geomstats [15].

For these classification experiments, we simulated a training and a target dataset using the spatio-temporal model described in equation (42). The training and target datasets, simulated with the same model parameters, use a single temporal correlation matrix R_t (in order to show the interest of our spatio-temporal model compared with the one-dimensional temporal model) and 3 different spatial correlation matrices R_s: the identity matrix $R_1 = I$ and the geometrical correlation matrices $R_2 = \left(\left(\frac{1}{3}\right)^{|i-j|}\right)_{i,j}$ and $R_3 = \left(\left(\frac{2}{3}\right)^{|i-j|}\right)_{i,j}$. The AHC, as unsupervised, is performed directly on the target dataset.

A. Supervised classification: the k-Nearest Neighbors

The k-Nearest Neighbors (KNN) classifier principle is quite simple: each point of the target dataset is labeled according to the k nearest points of the training set, i.e. a majority vote is performed on their labels. Figure 1 shows the result the KNN performed on the Riemannian manifold $P_N \times \mathcal{SD}_N^{n-1}$ on the simulated dataset described previously for $k = 5$.

B. Unsupervised classification: the Agglomerative Hierarchical Clustering

The Agglomerative Hierarchical Clustering (AHC) works as follows: at the initial state, each point corresponds to a cluster. Then, until the desired number of cluster is reached, merge the two closest clusters. Here we defined the distance between two clusters as the average distance between two points belonging to each of these clusters (other choices are possible). The clustering result of the AHC performed on our simulated dataset is shown on figure 2. To plot the confusion matrix of an unsupervised algorithm, we performed all possible permutations to find the best matching between the clustering result and the true labels.

ACKNOWLEDGMENT

We thank the French MoD, DGA / AID for funding (convention CIFRE AID $N°2017.0008$ & ANRT $N°2017.60.0062$).

REFERENCES

[1] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Unsupervised Machine Learning for Pathological Radar Clutter Clustering: the P-Mean-Shift Algorithm", IEEE, C&ESAR 2019, Rennes, France, 19-21 November 2019

[2] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Non-Supervised High Resolution Doppler Machine Learning for Pathological Radar Clutter", IEEE, RADAR 2019, Toulon, France, 23-27 September 2019

[3] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Toeplitz Hermitian Positive Definite Matrix Machine Learning based on Fisher Metric", IEEE, GSI 2019, Toulouse, France, 27-29 August 2019

[4] Yann Cabanes, Frédéric Barbaresco, Marc Arnaudon, Jérémie Bigot, "Non-supervised Machine Learning Algorithms for Radar Clutter High-Resolution Doppler Segmentation and Pathological Clutter Analysis", IEEE, Ulm, Germany, 26-28 June 2019

[5] Ben Jeuris and Raf Vandrebril, "The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks", 2016

[6] Martin Morf, Augusto Vieira, Daniel T. L. Lee and Thomas Kailath, "Recursive Multichannel Maximum Entropy Spectral Estimation", IEEE Transactions on Geoscience Electronics, vol. GE-16, NO. 2, April 1978

[7] Romain Couillet, Maria Sabrina Greco, Jean-Philippe Ovarlez and Frédéric Pascal, "RMT for Whitening Space Correlation and Applications to Radar Detection", IEEE, 2015

[8] Marc Arnaudon, Frédéric Barbaresco, Le Yang, "Riemannian Medians and Means With Applications to Radar Signal Processing", IEEE journal, August 2013.

[9] Frédéric Barbaresco, "Information Geometry Manifold of Toeplitz Hermitian Positive Definite Covariance Matrices: Mostow/Berger Fibration and Berezin Quantization of Cartan-Siegel Domains", International Journal of Emerging Trends in Signal Processing, Volume 1, pp. 111, March 2013

[10] Frédéric Barbaresco, "Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median", Springer, 2012, chapter 9, pp. 199255.

[11] Frédéric Barbaresco, "Robust statistical radar processing in Frchet metric space: OS-HDR-CFAR and OS-STAP processing in Siegel homogeneous bounded domains", in IRS 2011, International Radar Conference, Leipzig, 2011, pp. 639644

[12] Frédéric Barbaresco, "Super-resolution spectrum analysis regularization: Burg, Capon & AGO-antagonistic algorithms", 8^{th} European Signal Processing Conference (EUSIPCO 1996), Trieste, Italy, 1996

[13] Michel-Marie Deza and Elena Deza, "Encyclopedia of Distances", Springer, 2016

[14] Frank Nielsen, "Hilbert geometry of the Siegel disk: The Siegel-Klein disk model", arXiv, 2020

[15] Nina Miolane, Alice Le Brigant, Johan Mathe, Benjamin Hou, Nicolas Guigui, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, Hatem Hajri, Yann Cabanes, Thomas Gerald, Paul Chauchat, Christian Shewmake, Bernhard Kainz, Claire Donnat, Susan Holmes and Xavier Pennec, "Geomstats: A Python Package for Riemannian Geometry in Machine Learning", working paper or preprint, April, 2020, https://hal.inria.fr/hal-02536154/file/main.pdf, https://github.com/geomstats/geomstats