Lie Groups Statistics and Machine Learning for Military Sensors Based on Symplectic Structures of Information Geometry

Frédéric Barbaresco
THALES Land & Air Systems, 19/21, avenue Morane Saulnier, 78140 Vélizy-Villacoublay
FRANCE

frederic.barbaresco@thalesgroup.com

ABSTRACT

In a first part, we will present pioneering THALES Sensors/Radars algorithms on Geometric Matrix CFAR based on Jean-Louis Koszul's Information Geometry and its extension for STAP, and Complex-Valued Convolutional Neural Networks and Covariance-Matrix-Valued HPDNet for Micro-Doppler ATDR. In a second part, we will present Avant-Garde tools using statistics on Lie Groups for Radar Micro-Doppler and Kinematics Machine Learning. From French Jean-Marie Souriau's Symplectic Model of Statistical Physics and Russian Kirillov's Representation Theory of Lie Groups, we will introduce Gaussian statistical density for Lie Groups defined as Maximum Entropy Gibbs density on coadjoint orbits though moment map. This Symplectic foliation model of Information gives new geometric foundation of Radar Signal Processing.

1.0 STATISTICS OF RADAR ELECTROMAGNETIC WAVE FLUCTUATIONS

The main idea is to "code" the stationary state of a digital measurement of the electromagnetic wave (Direction, Doppler and Polarimetry) picked up by a Radar antenna by a point in a metric space. As it is a statistical measure, we have established that the most "natural" metric is given in Information Geometry by the Fisher metric of the stationary process.

The extension of statistics into metric spaces was introduced by Maurice Fréchet through the notion of "distanced space". The approach by metric spaces makes it possible to **define the average or the median of elements in a distanced space by the Fréchet geodesic barycenter calculated by the flow of gradient introduced in the Seventies by Hermann Karcher**. To define these metric spaces for an electromagnetic wave, we must first introduce a distance between numerical measurements of the states of the wave. The Fisher metric of the "Information Geometry" provides a natural distance, having the qualities of desired invariances (invariance by reparameterization, invariance linked to the symmetries of the electromagnetic signal). The Fisher metric can be generalized in the spaces describing the states of the electromagnetic wave from the geometric structures introduced by Jean-Louis Koszul (Koszul-Vinberg characteristic function, Koszul 2-form). The next step consists in generalizing **the notion of Gaussian density for the states of the electromagnetic wave, by generalizing the notion of maximum entropy densities (Gibbs density)**, which obliges us to generalize the definition of entropy. Starting from the François Massieu's characteristic function (introduced in thermodynamics by François Massieu), established by Jean-Marie Souriau in Statistical Mechanics within the framework of Geometric Mechanics, with a model which has been called "Lie groups Thermodynamics". **We deduced a density (of Gibbs) of probabilities and a metric of Fisher-Souriau to describe the average state (Fréchet barycenter) of the spatio-Doppler and polarimetric fluctuations of the digital measurements of the electromagnetic wave.**

In the case of a stationary signal, using the Toeplitz structures (respectively Blocs-Toeplitz) of the covariance matrices of the time [Doppler] or spatial [direction] (respectively time-space) signal, we demonstrate that **the natural metric is Kähler in the Poincaré polydisk productspace (coding of Doppler information) and the Siegel polydisk (coding of space-time information)**.

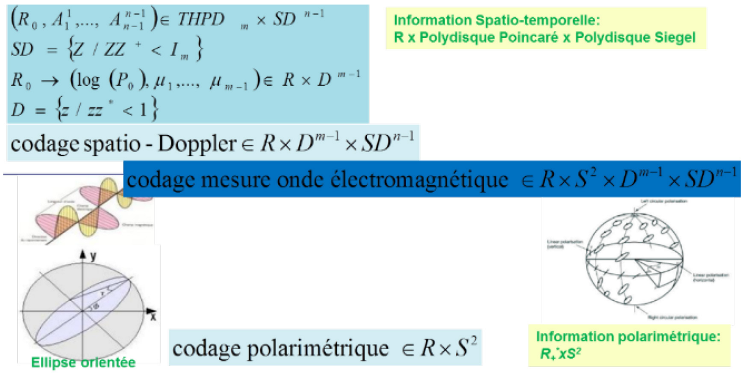


Figure 1: Coding of spatio -Doppler and polarimetric measurement of Radar EM wave in metric space.

In the case of a locally stationary signal, the electromagnetic spatio-temporal signal is characterized by a matrix autoregressive model (matrix extension of the Trench-Verblunsky theorem), based on the MOPUC theory (Matrix Orthogonal Polynomial on the Unit Circle). From the spatio-temporal numerical measurement of the electromagnetic wave, a series of N THDP covariance matrices (Toeplitz Hermitian Definite Positive) $\{R_0,...R_{N-1}\}$ is estimated. This time series is then parameterized by a matrix autoregressive model $\left(R_0, \{A_k^k\}_{k=1}^{N-1}\right) \in THPD \times SD^{N-1}$ with $R_0 \in THPD_n$ a THDP matrix, and $A_k^k \in SD$, $k = 1,..., N-1$ the matrix reflection/ Verblunsky coefficients in the unit Siegel disk SD, such that $A_k^k\left(A_k^k\right)^+ < I$ (where + stands for trans-conjugate and I the identity matrix). The matrix R_0 can itself be described by a complex classical autoregressive model, of parameters $\left(P_0, \{\mu_k\}_{k=1}^{m-1}\right) \in \Re^{+*} \times D^{m-1}$ with P_0 the power term, of which we classically take the logarithm $\log\left(P_0\right) \in \Re$ and $\mu_k \in D, k = 1,..., m-1$ the reflection/ Verblunsky coefficients in the Poincaré unit disk D, such as $\mu_k \mu_k^* = \left|\mu_k\right|^2 < 1$. We thus have for a locally stationary electromagnetic wave, a coding of the information in a metric product space $\left(\log(P_0), \{\mu_k\}_{k=1}^{m-1}, \{A_k^k\}_{k=1}^{N-1}\right) \in \Re \times D^{m-1} \times SD^{n-1}$, to which we can add $\mathbb{S}^2$: the Poincaré sphere for the polarimetric information. The signal being finally encoded on the product space: $\mathbb{R} \times \mathbb{S}^2 \times \boldsymbol{D}^{m-1} \times \boldsymbol{SD}^{n-1}$.

The theory of "Information Geometry" introduced by Rao and Fréchet makes it possible to define a natural metric between probability densities in the space of parameters and which is invariant with respect to a change of parametrization. In particular, one can define the metric by the Hessian of the Entropy. Considering the stationary process describing the spatio-temporal state of the electromagnetic wave by the covariance matrix of the spatio-temporal vector and its associated Toeplitz-Blocks- Toeplitz covariance matrix: $R_{N,p} = \begin{bmatrix} R_0 & R_1 & \cdots & R_{N-1} \\ R_1^+ & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_{N-1}^+ & \cdots & R_1^+ & R_0 \end{bmatrix}$, the entropy is given by $S\left(R_{p,N}\right) = \tilde{\Phi}\left(R_{p,N}\right) = -\log\left(\det R_{p,N}\right) + cste$, which with the autoregressive matrix parameters: $S(\overline{R}_{p,N}) = -\sum_{k=1}^{N-1}(N-k).\log\det\left[1 - \overline{A}_k^k \overline{A}_k^{k+}\right] - N.\log\left[\pi.e.\det \overline{R}_0\right]$ where $\overline{X}$ represents the estimate of X. By considering, the parametrization $\theta^{(N-1)} = \left[\overline{R}_0 \quad \overline{A}_1^1 \quad \cdots \quad \overline{A}_{N-1}^{N-1}\right]^T$, the metric is then given by the Hessian of the Entropy which gives:

$$ds^2 = d\theta^{(N-1)+}\left[\frac{\partial^2 S}{\partial\theta_i^{(N-1)}\partial\theta_j^{(N-1)}}\right]_{i,j} d\theta^{(N-1)} = N.Tr\left[\left(\overline{R}_0^{-1} d\overline{R}_0\right)^2\right] + \sum_{k=1}^{N-1}(N-k)Tr\left[\left(I_n - \overline{A}_k^k \overline{A}_k^{k+}\right)^{-1} d\overline{A}_k^k \left(I_n - \overline{A}_k^{k+} \overline{A}_k^k\right)^{-1} d\overline{A}_k^{k+}\right] \tag{1}$$

We can integrate this metric to define a distance in parametrization $\left\{\overline{R}_0, \overline{A}_1^1, ..., \overline{A}_{N-1}^{N-1}\right\}$ allowing to determine a distance between 2 spatio-temporal covariance matrices of the electromagnetic wave $R_{(1)}$ et $R_{(2)}$:

$$\begin{cases} d^2\left(R_{(1)}, R_{(2)}\right) = N.\left\|\log\left(\overline{R}_{(1),0}^{-1/2}\overline{R}_{(2),0}\overline{R}_{(1),0}^{-1/2}\right)\right\|_F^2 + \sum_{k=1}^{N-1}(N-k)\log^2\left(\frac{1+\left\|\Phi_{A_{(1),k}^k}\left(\overline{A}_{(2),k}^k\right)\right\|}{1-\left\|\Phi_{A_{(1),k}^k}\left(\overline{A}_{(2),k}^k\right)\right\|}\right) \\ \Phi_Z(W) = \left(I - ZZ^+\right)^{-1/2}(W-Z)\left(I - Z^+W\right)^{-1}\left(I - Z^+Z\right)^{1/2} \end{cases} \quad (2)$$

In the following, we will describe a model to define a probability density of Maximum Entropy for measures states of space-time electromagnetic wave, which is covariant under the action of the Lie group which acts in the wave representation space (groups acting in the Poincaré and Siegel disks).

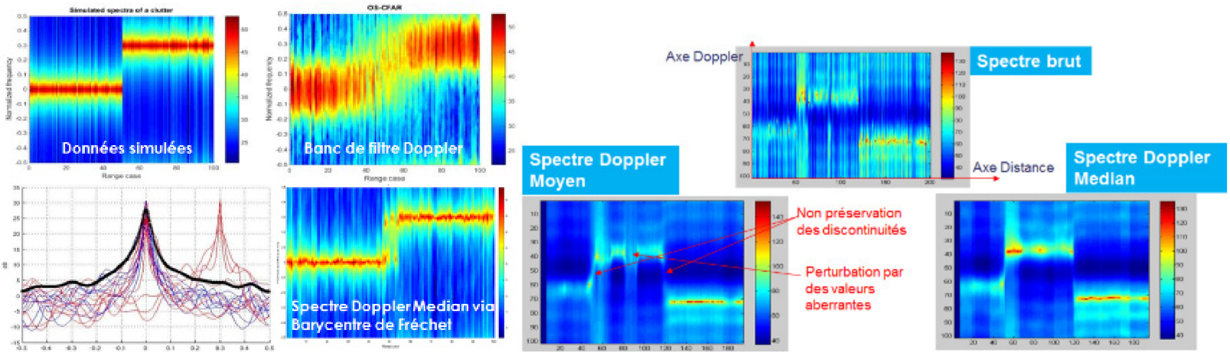


Figure 2: Example of calculation of the median Doppler spectrum (defined as Fréchet's geodesic barycenter via the Fisher metric) for the temporal covariance matrices of the Doppler signal of the electromagnetic wave (we note the robustness property of the "median" at atypical values).

2.0 MICRO-DOPPLER DEEP LEARNING WITH LEARNING OF DOPPLER FILTER BANKS AND SPDNET TYPE NETWORKS

2.1 Fully-Convolutional Neural Network and Fully-Temporal Covariance Network

THALES has developed in collaboration with the MLIA laboratory of Sorbonne University LIP6 department, new generations of neural networks to be able to classify in a robust way the micro-Doppler radar signatures of these objects to recognize their type and to differentiate them from other clutter sources. A first technique used by THALES was to develop fully convolutional networks FCN that preserve the time axis and provide a classification for each windowing along this time axis. These convolutional networks require the numerous parameters learning, the number of which depends on the depth of the layers of the network and the length of the convolutions. Under this constraint, THALES explored with LIP6 new neural network architectures more suited to Doppler radar signals. First approach consists in replacing the the Doppler spectrum computation by learning a complex filter bank by considering these filters as complex convolutions whose weight factors are learned. The second approach consists in not considering the raw signal, but only the covariance matrix of this signal which is assumed to be Hermitian Positive Definite, in case of a locally stationary signal. For this goal, THALES has designed a SPDNet (Symmetric Positive Definite Network) type network similar to an MLP (multi-layer perceptron) network but taking as input covariance matrices. This new type of network respects the Hermitian and positive definite structures, and is defined in 3 layers: a "BiMap" operator (bilinear) equivalent to a dense layer, a rectification operator "ReEig" acting on the eigenvalues (equivalent to a "ReLU" operation in a conventional network), and finally a logarithmic operator "LogEig" to move in SPD matrices tangent space tangent (locally Euclidean around the identity matrix). Other Improvements have been studied by considering the geodesic barycenter computation of SPD matrices (mean matrix in the geometric sense as defined by Maurice Fréchet for statistics in metric spaces) to align the tangent space used for the last operation, instead of the identity matrix.

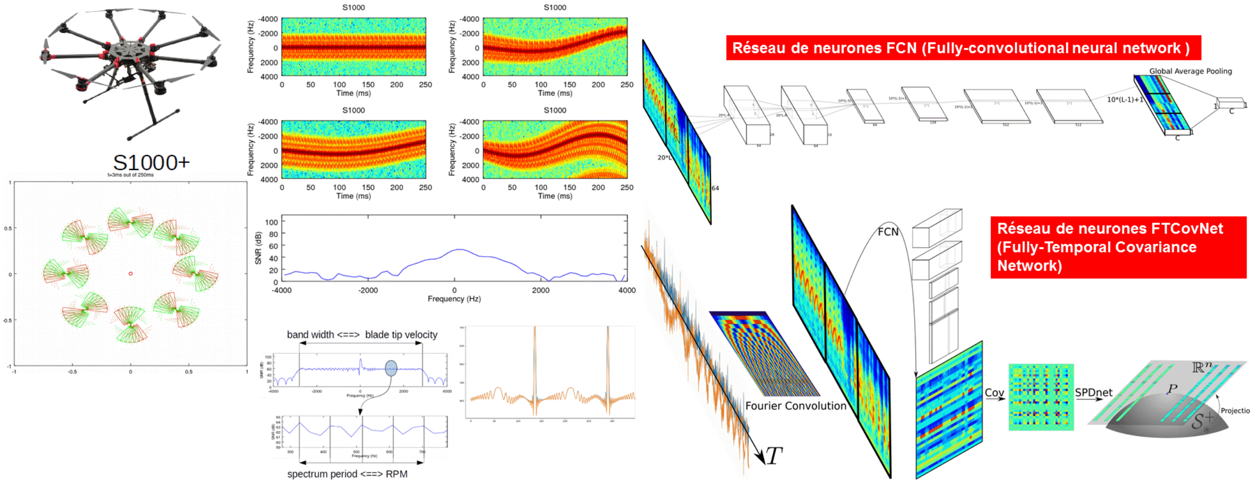


Figure 3: (Left) Radar Doppler signature simulation of a S1000+ drone from their blades mechanical and electromagnetic model developed in the HYBRID DIGITAL RADAR TWIN. (Right) Neural Networks FCN (Fully-convolutional neural network) and Neural Network FTCovNet (Fully - Temporal Covariance Network).

The method of coupling a learning complex filter bank followed by an FCN (Fully Convolutional Network) on the time-frequency image, followed by a SPDNet output FCN, with classification on the tangent plane at the geodesic barycenter geodesy is called the FTCovNet (Fully- Temporal Covariance Network) method. We have shown that in case of a low number of training data, the FTCovNet method retains good classification performance, greater than 90%, while the performance of the classical FCN method is more strongly degraded.

2.2 Machine Learning on Kinematics Criteria Based on Boosting Techniques and Ordered Statistics

To improve the drones classification performance when the Doppler signature of the blades is more difficult to characterize (fairing blades, carbon blades,…), we consider, in addition to Doppler signatures, kinematics features. Decision tree boosting methods (XGBOOST algorithm) are used by extracting statistical parameters over the time series from the following kinematic variables: speed / acceleration / jerk on the 3 axes, horizontal speed module, the 3D speed module, the 2D horizontal curvature, 3D curvature, the logarithm of the 3D trajectory torsion. The classification is done on ordered statistics parameters (median, quantile, L-moments ...) estimated on the time series of these parameters.

As in the Doppler case, there are few records for learning. The approach is to consider hybrid data generated by drones auto-pilot drones, as PAPARAZZI UAV autopilot (https://wiki.paparazziuav.org/wiki/Main_Page) developed by ENAC, to generate thousands realistic drone trajectories by varying the mass, the payload and the anemometric conditions. Learning takes place using simulated data. The list of drones used in the test is divided in 3 categories (later used for classification): fixedwing (EasyStar_ETS, Microjet_LisaM), rotorcraft (ardrone2, Bumblebee_Quad, LadyLisa, Quad_LisaMX, Quad_NavGo, bebop, bebop2), versatile (bixler) and others (birds). For performances assessment, we constitute two categories of datasets, and 3 radar probability of detection 0.90, 0.75 and 0.60. For the evaluation of the performances of the models, the confusion matrix will be used as a metric, rather than using Precision, Recall and F-measure. We wish indeed particularly know what are the risks that a drone was detected as a bird.

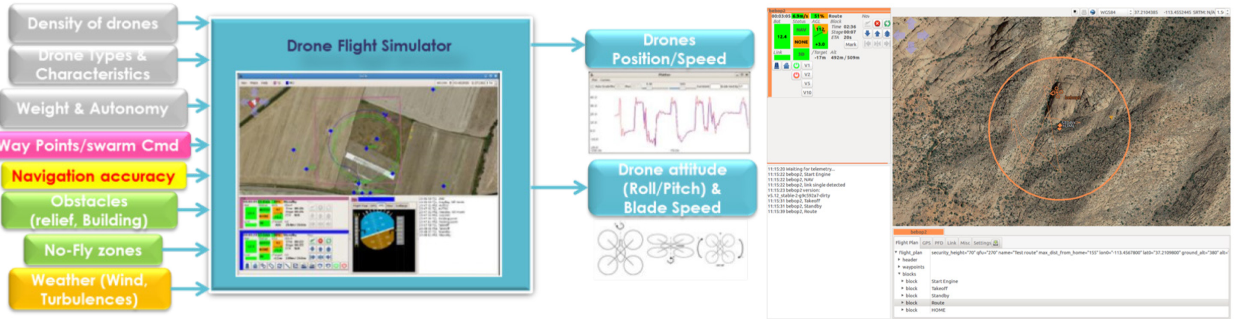
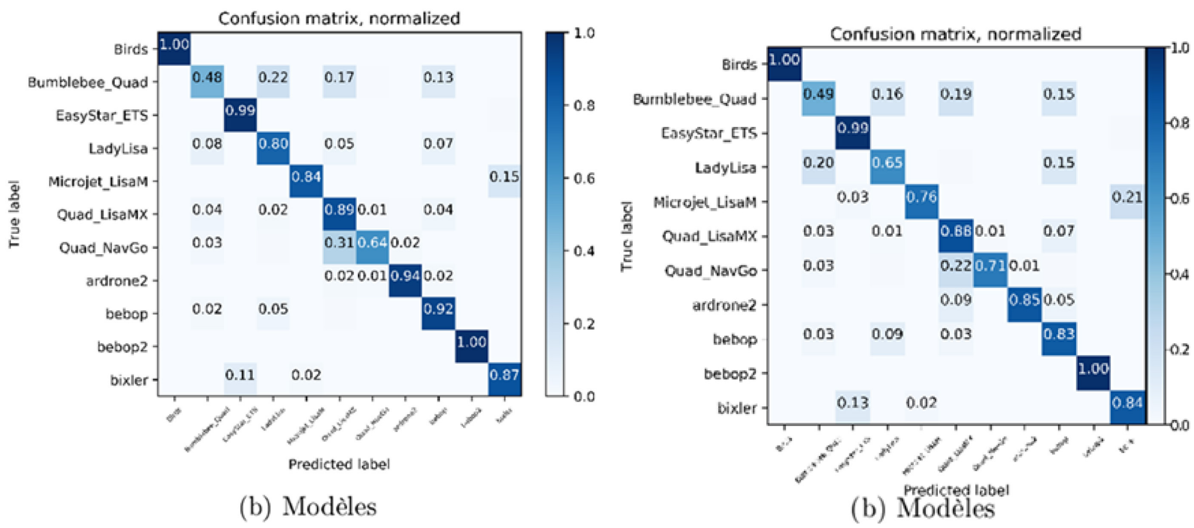


Figure 4: (Left) Kinematic trajectory simulator based on UAV PAPARAZZI autopilot. (Right) A rotorcraft trajectory generation under PAPARAZZI.

Table 1: Results obtained for P_{detect} = 0.9 (at left), P_{detect} = 0.6 (at right),



3.0 MICRO-DOPPLER AND KINEMATICS ML BASED ON LIE GROUPS

We will conclude by extending previous approaches to deal with coordinate-free method based on Lie Group Machine Learning. Drone recognition on Radar micro-Doppler data could be modelled by a problem of classification of dataset considered as elements of SU(1,1) Lie Group. Second example will consider flying drone recognition on their kinematics coded in SE(3) Lie Groups. We will use Souriau method to define covariant Gibbs density on Lie groups to be able to make statistics on Lie groups.

We will illustrate two problems of Machine Learning on Lie groups for Radar applications. Target recognition on Radar micro-Doppler data could be modeled by a problem of classification of dataset considered as elements of SU(1,1) Lie Group. Radar complex time series of micro-Doppler observation of data are classically processed on sliding time window to estimate their associated covariance matrices that are characterized by a Toeplitz Hermitian Positive-definitiveness structure. Using a well-known Verblunsky/Trench Theorem, we can parametrized all Toeplitz Hermitian Positive Definite Covariance matrices of stationary Radar Time series in a product space with a real positive axis (for signal power) and a Poincaré polydisk (for Doppler Spectrum shape). If we consider the Poincaré Unit Disk as an homogeneous space where SU(1,1) Lie Group acts transitively. Each data in Poincaré unit disk of this polydisk could be then coded by SU(1,1) matrix Lie group element. We have transform the problem into a statistical learning challenge processing data of SU(1,1) matrix Lie group. Another example consider flying object recognition on their kinematics coded in SE(3) Lie Groups. 3D trajectories could be coded by SE(3) Lie group time series provided through Invariant Extended Kalman Filter (IEKF) Radar Tracker, that locally estimate displacement of Frenet-Seret frame. Object kinematics will be then coded by time series of SE(3) matrix Lie Groups characterizing local rotation/translation of Frenet frame along the drone 3D trajectory. Statistics of this SE(3) Lie group elements with characterize flight mechanics of different kinds of object (birds, drones).

3.1 Matrix Lie Groups Structure for Doppler Radar Data

Lie Group structure appears naturally on Doppler data, if we consider time series of locally stationary signal and their associated covariance matrix. Covariance matrix is Toeplitz Hermitian Positive Definite. We can then use a Theorem due to Verblunsky and Trench, that this structure of covariance matrix could be coded in product space involving the Poincaré unit Polydisk:

$$\begin{aligned}\varphi : THDP(n) &\rightarrow R_+^* \times D^{n-1} \\ R_n &\mapsto \left(P_0, \mu_1, ..., \mu_{n-1}\right)\end{aligned} \tag{3}$$

where D is the Poincaré Unit Disk:

$$D = \left\{ z = x + iy \in C \,/\, \left|z\right| < 1 \right\} \tag{4}$$

The Poincaré unit disk is an homogeneous bounded domain where the Lie Group SU(1,1) act transitively. This Matrix Group is given by:

$$SU(1,1) = \left\{ \begin{bmatrix} a & b \\ b^* & a^* \end{bmatrix} / \left|a\right|^2 - \left|b\right|^2 = 1,\ a, b \in C \right\} \tag{5}$$

where SU(1,1) acts on the Poincaré Unit Disk by:

$$g \in SU(1,1) \Rightarrow g.z = \frac{az+b}{b^* z + a} \tag{6}$$

with Cartan Decomposition of SU(1,1):

$$\begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix} = \left|a\right| \begin{pmatrix} 1 & z \\ z^* & 1 \end{pmatrix} \begin{pmatrix} a/\left|a\right| & 0 \\ 0 & a^*/\left|a\right| \end{pmatrix}$$
$$\text{with } z = b\left(a^*\right)^{-1}, \left|a\right| = \left(1 - \left|z\right|^2\right)^{-1/2} \tag{7}$$

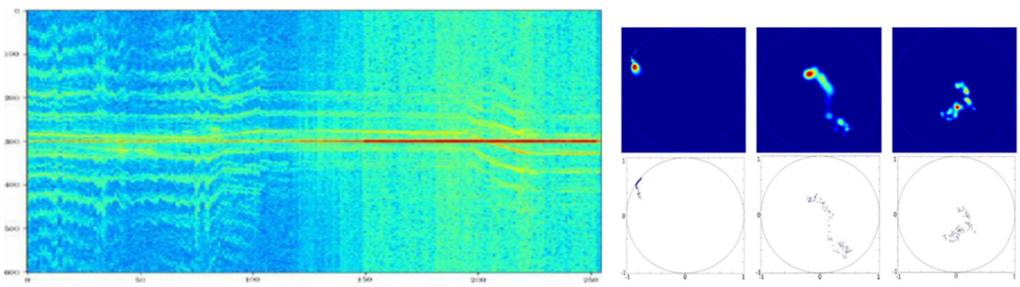


Figure 5: (at left) Time-Doppler Spectrum, (at right) Verblunsky coefficient density in the Poincaré Polydisk.

We can observe that $z=b\left(a^*\right)^{-1}$ could be considered as action of $g\in SU(1,1)$ on the centre on the unit disk $z=g.0=b\left(a^*\right)^{-1}$. The principal idea is that we can code any point $z=b\left(a^*\right)^{-1}$ in the unit disk by an element of the Lie Group SU(1,1). Main advantage is that the point position is no longer coded by coordinates but intrinsically by transformation from 0 to this point. Finally, a covariance matrix of a stationary signal could be coded by (n-1) Matrix SU(1,1) Lie Group elements:

$$\begin{aligned} &\text{THPD} \to R_+^* \times D^{n-1} \qquad \to R_+^* \times SU(1,1)^{n-1} \\ &R_n \mapsto \left(P_0,\mu_1,\ldots,\mu_{n-1}\right) \mapsto \left(P_0,\begin{bmatrix} a_1 & b_1 \\ b_1^* & a_1^* \end{bmatrix},\ldots,\begin{bmatrix} a_{n-1} & b_{n-1} \\ b_{n-1}^* & a_{n-1}^* \end{bmatrix}\right) \end{aligned} \tag{8}$$

3.2 Matrix Lie Groups Structure for Kinematics Data

When we consider a 3D trajectory of a mobile target, we can describe this curve by a time evolution of the local Frenet–Serret frame (local frame with tangent vector, normal vector and binormal vector). This frame evolution is described by the Frenet-Serrtet formula that gives the kinematic properties of the target moving along the continuous, differentiable curve in 3D Euclidean space $\mathbb{R}^3$. More specifically, the formulas describe the derivatives of the so-called tangent, normal, and binormal unit vectors.

$$\frac{d}{dt}\begin{pmatrix} \vec{t} \\ \vec{n} \\ \vec{b} \end{pmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \gamma \\ 0 & -\gamma & 0 \end{bmatrix}\begin{pmatrix} \vec{t} \\ \vec{n} \\ \vec{b} \end{pmatrix} \quad \text{with} \quad \begin{cases} \kappa : \text{curvature} \\ \gamma : \text{torsion} \end{cases} \tag{9}$$

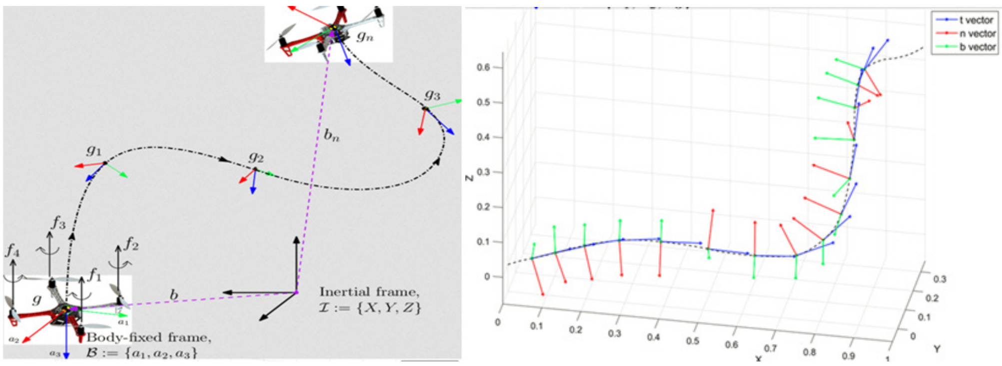


Figure 6: Inertial Frenet-Serret frame motion along a 3 D trajectory based on a local rotation and translation than will be coded in SE(3) Lie groups time series.

we will consider motions determined by exponentials of paths in the Lie algebra. Such a motion is determined by a unit speed space-curve $\tau(t)$. Now in a Frenet-Serret motion a point in the moving body moves along the curve and the coordinate frame in the moving body remains aligned with the tangent $\vec{t}$, normal $\vec{n}$, and binormal $\vec{b}$, of the curve. Using the 4-dimensional representation of the Lie Group SE(3), the motion can be specified as :

$$G(t)=\begin{pmatrix} R(t) & \tau(t) \\ 0 & 1 \end{pmatrix} \in SE(3) \tag{10}$$

where $\tau(t)$ is the curve and the rotation matrix $R(t)$ has the unit vectors $\vec{t}$, $\vec{n}$, and $\vec{b}$ as columns:

$$R(t) = \begin{pmatrix} \vec{t} & \vec{n} & \vec{b} \end{pmatrix} \in SO(3) \tag{11}$$

If we introduce the Darboux vector $\vec{\omega} = \gamma\vec{t} + \kappa\vec{b}$ that we can rewritte from Frenet-Serret Formulas :

$$\frac{d\vec{t}}{dt} = \vec{\omega} \times \vec{t} \quad , \quad \frac{d\vec{n}}{dt} = \vec{\omega} \times \vec{n} \quad , \quad \frac{d\vec{b}}{dt} = \vec{\omega} \times \vec{b} \tag{12}$$

Then, we can write with Ω is the 3×3 anti-symmetric matrix corresponding to $\vec{\omega}$:

$$\frac{dR}{dt} = \Omega R \tag{13}$$

We note that $\frac{d\tau(t)}{dt} = \vec{t}$ and $\frac{d\vec{\omega}}{dt} = \frac{d\gamma}{dt}\vec{t} + \frac{d\kappa}{dt}\vec{b}$. The instantaneous twist of the motion $G(t)$ is given by:

$$S_d = \frac{dG(t)}{dt} G^{-1}(t) = \begin{pmatrix} \Omega & \upsilon \\ 0 & 0 \end{pmatrix} \tag{14}$$

This is the Lie algebra element corresponding to the tangent vector to the curve $G(t)$. It is well known that elements of the Lie algebra $se(3)$ can be described as lines with a pitch. The fixed axode of a motion $G(t) \in SE(3)$ is given by the axis of S_d as t varies. The instantaneous twist in the moving reference frame is given by $S_b = G^{-1}(t) S_d G(t)$, that is, by the adjoint action on the twist in the fixed frame. The instantaneous twist S_b can also be found from the relation:

$$S_b = G^{-1}\frac{dG}{dt} = \begin{pmatrix} R^T & -R^T\tau \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Omega R & \vec{t} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} R^T \Omega R & R\vec{t} \\ 0 & 0 \end{pmatrix} \tag{15}$$

We can observe that we could describe a 3D trajectory by a time series of SE(3) Lie group elements:

$$SE(3) = \left\{ \begin{bmatrix} R & \tau \\ 0 & 1 \end{bmatrix} / R \in SO(3), \tau \in R^3 \right\} \text{ with } SO(3) = \left\{ R / R^T R = RR^T = I, \det{}^2 R = 1 \right\} \tag{16}$$

Then, the trajectory will be given by the following time series:

$$\left\{ \begin{bmatrix} R_1 & \tau_1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} R_2 & \tau_2 \\ 0 & 1 \end{bmatrix}, \ldots, \begin{bmatrix} R_n & \tau_n \\ 0 & 1 \end{bmatrix} \right\} \in SE(3)^n \tag{17}$$

3.3 Statistics on Lie Groups Based on Souriau Gibbs Density

In this chapter, we will explain how to define statistics on Lie Groups, how to adapt Machine Learning for Lie Groups, and more especially how to define extension of Gauss density as Gibbs density in the Framework of Geometric Statistical Mechanics.

Classically, to optimize the parameter θ of a probabilistic model, based on a sequence of observations y_t, is an online gradient descent:

$$\theta_t \leftarrow \theta_{t-1} - \eta_t \frac{\partial l_t(y_t)^T}{\partial \theta} \tag{18}$$

with learning rate η_t, and the loss function $l_t = -\log p\left(y_t / \hat{y}_t\right)$. This simple gradient descent has a first drawback of using the same non-adaptive learning rate for all parameter components, and a second drawback of non invariance with respect to parameter re-encoding inducing different learning rates. Amari has introduced the natural gradient to preserve this invariance to be insensitive to the characteristic scale of each parameter direction. The gradient descent could be corrected by $I(\theta)^{-1}$ where I is the Fisher information matrix with respect to parameter θ, given by:

$$I\left(\theta\right) = \left[g_{ij}\right] \text{ with } g_{ij} = \left[-E_{y \approx p(y/\theta)}\left[\frac{\partial^2 \log p\left(y/\theta\right)}{\partial\theta_i \partial\theta_j}\right]\right]_{ij} \tag{19}$$

with natural gradient:

$$\theta_t \leftarrow \theta_{t-1} - \eta_t I(\theta)^{-1} \frac{\partial l_t\left(y_t\right)^T}{\partial\theta} \tag{20}$$

Amari has proved that the Riemannian metric in an exponential family is the Fisher information matrix:

$$g_{ij} = -\left[\frac{\partial^2 \Phi}{\partial\theta_i \partial\theta_j}\right]_{ij} \quad \text{with } \Phi(\theta) = -\log \int_R e^{-\langle\theta, y\rangle} dy \tag{21}$$

and the dual potential, the Shannon entropy, is given by the Legendre transform:

$$S(\eta) = \langle\theta, \eta\rangle - \Phi(\theta) \text{ with } \eta_i = \frac{\partial\Phi(\theta)}{\partial\theta_i} \text{ and } \theta_i = \frac{\partial S(\eta)}{\partial\eta_i} \tag{22}$$

We can observe that $\Phi(\theta) = -\log \int_R e^{-\langle\theta, y\rangle} dy = -\log \psi\left(\theta\right)$ is related to the classical cumulant generating function. These relations have been extended by Jean-Marie Souriau in geometric statistical mechanics, where he developed a “Lie groups thermodynamics” of dynamical systems where the (maximum entropy) Gibbs density is covariant with respect to the action of the Lie group. In the Souriau model, previous structures of information geometry are preserved:

$$\begin{gathered} I(\beta) = -\frac{\partial^2 \Phi}{\partial\beta^2} \text{ with } \Phi(\beta) = -\log \int_M e^{-\langle U(\xi), \beta\rangle} d\lambda_\omega \\ \text{and } U : M \rightarrow \mathfrak{g}^* \end{gathered} \tag{23}$$

We preserve the Legendre transform:

$$\begin{gathered} S(Q) = \langle Q, \beta\rangle - \Phi(\beta) \\ \text{with } Q = \frac{\partial\Phi(\beta)}{\partial\beta} \in \mathfrak{g}^* \text{ and } \beta = \frac{\partial S(Q)}{\partial Q} \in \mathfrak{g} \end{gathered} \tag{24}$$

Souriau Foundamental Theorem is that « Every symplectic manifold on which a Lie group acts transitively by a Hamiltonian action is a covering space of a coadjoint orbit ». We can observe that for Souriau model, Fisher metric is an extension of this 2-form in non-equivariant case:

$$g_\beta\left([\beta, Z_1], [\beta, Z_2]\right) = \tilde{\Theta}\left(Z_1, [\beta, Z_2]\right) + \left\langle Q, \left[Z_1, [\beta, Z_2]\right] \right\rangle \tag{25}$$

The Souriau additional term $\tilde{\Theta}(Z_1,[\beta,Z_2])$ is generated by non-equivariance through Symplectic cocycle. The tensor $\tilde{\Theta}$ used to define this extended Fisher metric is defined by the moment map $J(x)$, application from M (homogeneous symplectic manifold) to the dual space of the Lie algebra $\mathfrak{g}^*$, given by:

$$\tilde{\Theta}(X,Y)=J_{[X,Y]}-\{J_X,J_Y\}$$
$$\text{with } J(x):M\to\mathfrak{g}^* \text{ such that } J_X(x)=\langle J(x),X\rangle,\ X\in\mathfrak{g} \tag{26}$$

Souriau has then defined a Gibbs density that is covariant under the action of the group:

$$p_{Gibbs}(\xi)=e^{\Phi(\beta)-\langle U(\xi),\beta\rangle}=\frac{e^{-\langle U(\xi),\beta\rangle}}{\int_M e^{-\langle U(\xi),\beta\rangle}d\lambda_\omega}$$
$$\text{with } \Phi(\beta)=-\log\int_M e^{-\langle U(\xi),\beta\rangle}d\lambda_\omega \tag{27}$$

$$Q=\frac{\partial\Phi(\beta)}{\partial\beta}=\frac{\int_M U(\xi)e^{-\langle U(\xi),\beta\rangle}d\lambda_\omega}{\int_M e^{-\langle U(\xi),\beta\rangle}d\lambda_\omega}=\int_M U(\xi)p(\xi)d\lambda_\omega \tag{28}$$

We can express the Gibbs density with respect to Q by inverting the relation $Q=\frac{\partial\Phi(\beta)}{\partial\beta}=\Theta(\beta)$.

Then $p_{Gibbs,Q}(\xi)=e^{\Phi(\beta)-\langle U(\xi),\Theta^{-1}(Q)\rangle}$ with $\beta=\Theta^{-1}(Q)$.

3.4 GIBBS Density for SU(1,1) Lie Group for Ml on Doppler Data

We will introduce Souriau moment map for SU(1,1)/K group that acts transitively on Poincaré Unit Disk, based on moment map. Considering the Lie group:

$$SU(1,1)=\left\{\begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix} / a,b\in C,\ |a|^2-|b|^2=1\right\} \tag{29}$$

and its Lie algebra given by elements:

$$su(1,1)=\left\{\begin{pmatrix} ir & \eta \\ \eta^* & -ir \end{pmatrix} / r\in R,\eta\in C\right\} \tag{30}$$

A basis for this Lie algebra $su(1,1)$ is $(u_1,u_2,u_3)\in\mathfrak{g}$ with

$$u_1=\frac{i}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ u_2=-\frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } u_3=\frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \text{ with } [u_1,u_3]=-u_2,[u_1,u_2]=u_3,[u_2,u_3]=-u_1 \tag{31}$$

The compact subgroup is generated by u_1, while u_2 and u_3 generate a hyperbolic subgroup. The dual space of the Lie algebra is given by:

$$su(1,1)^*=\left\{\begin{pmatrix} z & x+iy \\ -x+iy & -z \end{pmatrix} / x,y,z\in R\right\} \tag{32}$$

with the basis

$$\left(u_1^*,u_2^*,u_3^*\right)\in\mathfrak{g}^* \text{ with } u_1^*=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\ u_2^*=\begin{pmatrix}0 & i\\ i & 0\end{pmatrix} \text{ and } u_3^*=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \tag{33}$$

The associated moment map $J:D\to su^*(1,1)$ defined by $J(z).u_i=J_i(z,z^*)$, maps D into a coadjoint orbit in $su^*(1,1)$. Then, we can write the moment map as a matrix element of $su^*(1,1)$:

$$J(z)=J_1\left(z,z^*\right)u_1^*+J_2\left(z,z^*\right)u_2^*+J_3\left(z,z^*\right)u_3^*\Rightarrow J(z)=\rho\begin{pmatrix}\dfrac{1+|z|^2}{1-|z|^2} & -2\dfrac{z^*}{1-|z|^2}\\ 2\dfrac{z}{1-|z|^2} & -\dfrac{1+|z|^2}{1-|z|^2}\end{pmatrix}\in\mathfrak{g}^* \tag{34}$$

The moment map J is a diffeomorphism of D onto one sheet of the two-sheeted hyperboloid in $su^*(1,1)$, determined bythe following equation $J_1^2-J_2^2-J_3^2=\rho^2$, $J_1\geq\rho$ with $J_1u_1^*+J_2u_2^*+J_3u_3^*\in su^*(1,1)$.

We can the write the covariant Gibbs density in the unit disk given by moment map of the Lie group $SU(1,1)$ and geometric temperature in its Lie algebra $\beta\in\Lambda_\beta$:

$$p_{Gibbs}(z)=\frac{e^{-\langle J(z),\beta\rangle}}{\int_D e^{-\langle J(z),\beta\rangle}d\lambda(z)} \text{with } d\lambda(z)=2i\rho\frac{dz\wedge dz^*}{\left(1-|z|^2\right)^2} \tag{35}$$

$$p_{Gibbs}(z)=\frac{e^{-\left\langle\rho\left(2\Im bb^+-Tr\left(\Im bb^+\right)I\right),\beta\right\rangle}}{\int_D e^{-\langle J(z),\beta\rangle}d\lambda(z)}=\frac{e^{-\left\langle\rho\begin{pmatrix}\frac{1+|z|^2}{\left(1-|z|^2\right)} & \frac{-2z^*}{\left(1-|z|^2\right)}\\ \frac{2z}{\left(1-|z|^2\right)} & -\frac{1+|z|^2}{\left(1-|z|^2\right)}\end{pmatrix},\begin{pmatrix}ir & \eta\\ \eta^* & -ir\end{pmatrix}\right\rangle}}{\int_D e^{-\langle J(z),\beta\rangle}d\lambda(z)} \tag{36}$$

To write the Gibbs density with respect to its statistical moments, we have to express the density with respect to $Q=E\left[J(z)\right]$. Then, we have to invert the relation between Q and β, to replace this last variable $\beta=\begin{pmatrix}ir & \eta\\ \eta^* & -ir\end{pmatrix}\in\Lambda_\beta$ by $\beta=\Theta^{-1}(Q)\in\mathfrak{g}$ where $Q=\dfrac{\partial\Phi(\beta)}{\partial\beta}=\Theta(\beta)\in\mathfrak{g}^*$ with $\Phi(\beta)=-\log\int_D e^{-\langle J(z),\beta\rangle}d\lambda(z)$, deduce from Legendre tranform. The mean moment map is given by:

$$Q=E\left[J(z)\right]=E\left[\rho\begin{pmatrix}\dfrac{1+|w|^2}{\left(1-|w|^2\right)} & \dfrac{-2w^*}{\left(1-|w|^2\right)}\\ \dfrac{2w}{\left(1-|w|^2\right)} & -\dfrac{1+|w|^2}{\left(1-|w|^2\right)}\end{pmatrix}\right] \text{ where } w\in D \tag{37}$$

3.5 GIBBS Density for SE(2) Lie Group for ML on Kinematics Data

We will restrict the study to the 2D case with SE(2), but it could be extended to SE(3). We will consider Souriau model for $SE(2)$ Lie group with non-null cohomology and then with introduction of Souriau one-cocycle. We consider

$$SE(2)=SO(2)\times R^2:\ SE(2)=\left\{\begin{bmatrix} R_\varphi & \tau \\ 0 & 1\end{bmatrix} / R_\varphi \in SO(2), \tau \in R^2\right\} \tag{38}$$

The Lie algebra $se(2)$ of $SE(2)$ has underlying vector space R^3 and Lie bracket:

$$(\xi,u)\in se(2)=R\times R^2, \begin{bmatrix} -\xi\Im & u \\ 0 & 0\end{bmatrix}\in se(2) \text{ with } \Im=\begin{bmatrix} 0 & 1 \\ -1 & 0\end{bmatrix} \tag{39}$$

Coadjoint action of $SE(2)$ is given by:

$$Ad^*_{(R_\varphi,\tau)}(m,\rho)=\left(m+\Im R_\varphi \rho.\tau, R_\varphi \rho\right) \tag{40}$$

The moment map $J:R^2\to se^*(2)$ of $SE(2)$ is defined by: $J_{(\xi,u)}(x)=J(x).(\xi,u)$ with the right action of $SE(2)$ on R^2:

$$J_{(\xi,u)}(x)=-2\left(\frac{1}{2}\xi\|x\|^2+\Im u.x\right)=-2\left(\frac{1}{2}\|x\|^2,-\Im x\right).(\xi,u)=J(x).(\xi,u)\Rightarrow J(x)=-2\left(\frac{1}{2}\|x\|^2,-\Im x\right)\ ,\ x\in R^2 \tag{41}$$

With the expression of moment map, we can compute Souriau covariant Gibbs density of Maximum Entropy. Considering the symplectic form $\omega(\zeta,\upsilon)=\zeta.\Im\upsilon$ with $\Im=\begin{bmatrix} 0 & 1 \\ -1 & 0\end{bmatrix}$ on R^2, we have seen that the action of SE(2) is symplectic and admits the momentum map, $J(x)=-\left(\frac{1}{2}\|x\|^2,-\Im x\right)\ ,\ x\in R^2$.

Souriau Gibbs density is defined for generalized temperature $\beta\in\Omega=\left\{(b,\mathrm{B})\in se(2)/b<0,\mathrm{B}\in R^2\right\}$ and given by:

$$p_{Gibbs}(x)=\frac{e^{-\langle J(x),\beta\rangle}}{\int_{R^2} e^{-\langle J(x),\beta\rangle}d\lambda(x)}=\frac{e^{\frac{1}{2}b\|x\|^2-\mathrm{B}.\Im x}}{\int_{R^2} e^{\frac{1}{2}b\|x\|^2-\mathrm{B}.\Im x}d\lambda(x)} \tag{42}$$

The Massieu Potential could be computed:

$$\Phi(\beta)=\log\int_{R^2} e^{\frac{1}{2}b\|x\|^2-\mathrm{B}.\Im x}d\lambda(x)=\log\left(-\frac{2\pi}{b}e^{-\frac{1}{2b}\|B\|^2}\right) \tag{43}$$

By derivation of Massieu potential, we can deduce expression of Heat:

$$Q\in\Omega^*=\left\{(m,M)\in se^*(2)/m+\frac{\|M\|^2}{2}<0\right\};Q=\frac{\partial\Phi(\beta)}{\partial\beta}=\left(\frac{1}{b}-\frac{\|\mathrm{B}\|^2}{2b^2},\frac{1}{b}\mathrm{B}\right)=\Theta(\beta) \tag{44}$$

We can the inverse this relation to express generalized temperature with respect to the heat:

$$\beta=\Theta^{-1}(Q)=\left(\left(m+\frac{1}{2}\|M\|^2\right)^{-1},\left(m+\frac{1}{2}\|M\|^2\right)^{-1}M\right) \tag{45}$$

We can the express the Gibbs density with respect to the Heat Q which is the mean of moment map:

$$p_{Gibbs}(x)=\frac{e^{\frac{\frac{1}{2}\|x\|^2-M.\Im x}{\left(m+\frac{1}{2}\|M\|^2\right)}}}{\Gamma}\ \text{ with } \Gamma=\int_{R^2} e^{\frac{\frac{1}{2}\|x\|^2-M.\Im x}{\left(m+\frac{1}{2}\|M\|^2\right)}}d\lambda(x)\ \text{ with } (m,M)=E(J(x))=\left[-E\left(\|x\|^2\right),2\Im E(x)\right] \tag{45}$$

So we can rewrite the Gibbs density:

$$p_{Gibbs}(x)=\frac{e^{\frac{\frac{1}{2}\|x\|^2+2E(x).Ix}{\left(-E\left(\|x\|^2\right)+2\|E(x)\|^2\right)}}}{\int_{R^2} e^{\frac{\frac{1}{2}\|x\|^2+2E(x).Ix}{\left(-E\left(\|x\|^2\right)+2\|E(x)\|^2\right)}}d\lambda(x)} \tag{46}$$

4.0 REFERENCES

[1] F. 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 (IJETSP), March 2013.

[2] Y. Cabanes, F. Barbaresco, M. Arnaudon, and J. Bigot. Non-Supervised High Resolution Doppler Machine Learning for Pathological Radar Clutter. IEEE, RADAR 2019, Toulon, France, Sept. 2019.

[3] F. Barbaresco, Coding & Statistical Characterization of Radar Signal Fluctuation for Lie Group Machine Learning, 2019 International Radar Conference (RADAR), 2019.

[4] D. Brooks, O. Schwander, F. Barbaresco, J.-Y. Schneider and M. Cord, Deep Learning and Information Geometry for Drone Micro-Doppler Radar Classification, IEEE RadarConf20, 2020.

[5] P.Y. Lagrave, Y. Cabanes, F. Barbaresco, SU(1,1) Equivariant Neural Networks and Application to Robust Toeplitz Hermitian Positive Definite Matrix Classification. LNCS vol 12829. Springer, 2021.

[6] Barbaresco F., Radar Micro-Doppler Signal Encoding in Siegel Unit Poly-Disk for Machine Learning in Fisher Metric Space, International Radar Symposium, Bonn, June 2018.

[7] Barbaresco F., Symplectic theory of heat and information geometry, chapter 4, Handbook of Statistics, Volume 46, Pages 107-143, Elsevier, 2022, https://doi.org/10.1016/bs.host.2022.02.003 ; https://www.sciencedirect.com/science/article/abs/pii/S0169716122000062

[8] Barbaresco F., Symplectic Foliation Structures of Non-Equilibrium Thermodynamics as Dissipation Model: Application to Metriplectic Non-Linear Lindblad Quantum Master Equation, submitted to MDPI special Issue "Geometric Structure of Thermodynamics: Theory and Applications", 2022.