The Abdus Salam
International Centre
for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2470-2

Advanced School and Workshop on Matrix Geometries and Applications

1 - 12 July 2013

Reading Material

Frédéric Barbaresco
THALES LAND & AIR SYSTEMS
Voie Pierre-Gilles de Gennes, F-91470 Limours
France

Information Geometry Manifold of Toeplitz Hermitian Positive Definite Covariance Matrices: Mostow/Berger Fibration and Berezin Quantization of Cartan-Siegel Domains

Barbaresco F. , *Emeritus Member SEE, SEE Ampere Medal 2007, member IEEE*

Abstract: **This paper deals with geometry of covariance matrices to define new advanced Radar Doppler Processing based on Metric Space tools. Information Geometry has been introduced by C.R.Rao, and axiomatized by N. Chentsov, to define a distance between statistical distributions that is invariant to non-singular parameterization transformations. For Doppler/Array/STAP Radar Processing, Information Geometry Approach will give key role to Homogenous Symmetric bounded domains geometry. For Radar, we will observe that Information Geometry metric could be related to Kähler metric, given by Hessian of Kähler potential (Entropy of Radar Signal given by –log[det(R)]). To take into account Toeplitz structure of Time/Space Covariance Matrix or Toeplitz-Block-Toeplitz structure of Space-Time Covariance matrix, Parameterization known as Partial Iwasawa Decomposition could be applied through Complex Autoregressive Model or Multi-channel Autoregressive Model. Then, Hyperbolic Geometry of Poincaré Unit Disk or Symplectic Geometry of Siegel Unit Disk will be used as natural space to compute "p-mean" (p=2 for "mean", p=1 for "median") of covariance matrices via Karcher Flow derived from Weiszfeld algorithm extension on Cartan-Hadamard Manifold. This new mathematical framework will allow development of OS (Ordered Sta-tistic) concept for Hermitian Positive Definite Covariance Space/Time Toeplitz matrices or for Space-Time Toeplitz-Block-Toeplitz matrices. We will define OS-HDR-CFAR (Ordered Statistic High Doppler Resolution CFAR) and OS-STAP (Ordered Statistic Space-Time Adaptive Processing).**

Index Terms: **Median, p-mean, Center of mass, Cartan-Hadamard Manifold, Cartan Symmetric spaces, Homogenous bounded domains, Siegel Upper-half plane, Poincaré disk, Siegel disk, Fréchet metric space, Berger fibration, Mostow decomposition, Radar, CFAR, STAP**

Manuscript received September 15, 2012. This work was supported in part by US Air Force Office of Scientific Research (AFOSR) through European Office of Aerospace Research and Development (EOARD) Grant n° FA8655-11-1-3031, and by THALES Key Technology Domain Processing Control & Cognition (KTD PCC) funding.

Frédéric BARBARESCO is Senior Scientist and Advanced Studies Manager in THALES AIR SYSTEMS, Surface Radar Domain, Technical Directorate, Advanced Developments Department, Voie Pierre-Gilles de Gennes, F-91470 Limours FRANCE (corresponding author to provide phone: +33 (0)630071419 ; fax: +33 (0)164916766; e-mail: frederic.barbaresco@thalesgroup.com).

I. Introduction

INFORMATION Geometry has been introduced by C.R.Rao, and axiomatized by N. Chentsov, to define a distance between statistical distributions that is invariant to non-singular parameterization transformations. For Doppler/Array/STAP Radar Processing, Information Geometry Approach will give key role to Symmetric spaces and Homogenous bounded domains geometry. For Radar, we will observe that Information Geometry metric could be identified to Kähler metric, given by Hessian of Kähler potential (Entropy of Radar Signal given by –log[det(R)]). To take into account Toeplitz structure of Time/Space Covariance Matrix or Toeplitz-Block-Toeplitz structure of Space-Time Covariance matrix, Parameterization known as Partial Iwasawa Decomposition could be applied through Complex Autoregressive Model or Multi-channel Autoregressive Model. Then, Hyperbolic Geometry of Poincaré Unit Disk or Symplectic Geometry of Siegel Unit Disk will be used as natural space to compute "p-mean" (p=2 for "mean", p=1 for "median") of covariance matrices via Karcher Flow for Weiszfeld algorithm extension on Manifold.

This new mathematical framework will allow development of (Ordered Statistic) concept [3] for Hermitian Positive Definite Covariance Space/Time Toeplitz matrices or for Space-Time Toeplitz-Block-Toeplitz matrices. We will define OS-HDR-CFAR (Ordered Statistic High Doppler Resolution CFAR) and OS-STAP (Ordered Statistic Space-Time Adaptive Processing) algorithms for radar detection.

II. Information Geometry Manifold of Covariance Hermitian Positive Definite Matrices

In 1945, Rao has introduced Information Geometry for parameterized density of probability $p(./\theta)$ with the metric given by the formula $ds^2 = K\left[p(./\theta), p(./\theta + d\theta)\right] = d\theta^+ I(\theta) d\theta$ where $I(\theta) = \left[g_{ij}(\theta)\right]$ is the Fisher Information matrix. If we model Signal by complex circular multivariate Gaussian distribution of zero mean :

$$p(X_n / R_n) = (\pi)^{-n} . \left|R_n\right|^{-1} . e^{-(X_n - m_n)^+ R_n^{-1} . (X_n - m_n)} \quad (1)$$

then for $m_n = 0$:

$$ds^2 = d\theta^+ I(\theta) d\theta = Tr\left[\left(R_n^{-1} dR_n\right)^2\right] = \left\|R_n^{-1/2} dR_n R_n^{-1/2}\right\|_F^2 \quad (2)$$

This metric has been integrated by Siegel and the distance is :

$$dist^2(R_1, R_2) = \left\|\log\left(R_1^{-1/2}.R_2.R_1^{-1/2}\right)\right\|_F^2 = \sum_{k=1}^{n} \log^2(\lambda_k) \quad (3)$$

with $\det(R_2 - \lambda_k R_1) = 0$

This is a complete simply connected metric space of negative curvature with geodesic between R_1 and R_2 :

$$\gamma(t) = R_1^{1/2} e^{t.\log\left(R_1^{-1/2} R_2 R_1^{-1/2}\right)} R_1^{1/2} = R_1^{1/2}\left(R_1^{-1/2} R_2 R_1^{-1/2}\right)^t R_1^{1/2} \quad (4)$$

with $0 \le t \le 1$

III. From Cartan Center of Mass to Fréchet Median

Robust Covariance Matrix Mean Estimation can be addressed by Riemannian center of mass. Elie Cartan has proved back in the 1920's existence and uniqueness of center of mass for simply connected complete manifolds of negative curvature for any compact subset. In Euclidean space, the center of mass is defined for finite set of points $\{x_i\}_{i=1,...,M}$ by arithmetic mean: $x_{center} = \frac{1}{M}\sum_{i=1}^{M} x_i$ or by $\sum_{i=1}^{M} \overrightarrow{x_{center} x_i} = 0$. This point also minimizes: $x_{center} = \arg\underset{x}{Min} \sum_{i=1}^{M} d^2(x, x_i)$. This has been extended to general Riemannian manifolds by Elie Cartan that has proved that the function $f : m \in \mathrm{M} \mapsto \frac{1}{2}\int_A d^2(m,a) da$ is strictly convex (its restriction to any geodesic is strictly convex as a function of one variable) and achieves a unique minimum at a point called the center of mass of *A* for the distribution *da*. Moreover, this point is characterized by being the unique zero of the gradient vector field: $\nabla f = -\int_A \exp_m^{-1}(a) da$, where $\exp(.)$ is the "exponential map" and $\exp_m^{-1}(a)$ is the tangent vector at $m \in \mathrm{M}$ of the geodesic from m to a : $\exp_m^{-1}(a) \in T_m\mathrm{M}$. Hermann Karcher has then introduced a gradient flow intrinsic on the Manifold M that converges to the center of mass, called Karcher Barycenter:

$$m_{n+1} = \gamma_n(t_n) = \exp_{m_n}\left(-t_n.\nabla f(m_n)\right) \text{ with } \dot{\gamma}_n(0) = -\nabla f(m_n) \quad (5)$$

In the discrete case, the center of mass for set of points is:

$$m_{n+1} = \exp_{m_n}\left(t_n.\sum_{i=1}^{M} \exp_{m_n}^{-1}(x_i)\right) \quad (6)$$

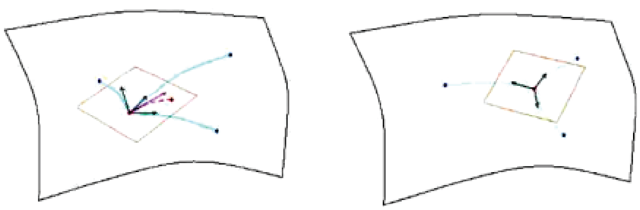

Figure 1. Fréchet-Karcher Flow on Cartan-Hadamard Manifold

But, center of mass is not useful for robust statistic. Replacing L^2 square geodesic distance by L^1 geodesic distance, we can extend this approach to estimate a <u>Median</u> in metric space (called Fermat-Weber's point in Physic). Fréchet studied Median statistic using $m_{median} = \underset{m}{Min}\, E\left[\|x - m\|\right]$ compared to $m_{mean} = \underset{m}{Min}\, E\left[\|x - m\|^2\right]$. Classically, in Euclidean space, Median point minimizes $x_{median} = \arg\underset{x}{Min}\sum_{i=1}^{M} d(x, x_i)$, equivalently $\sum_{i=1}^{M} \overrightarrow{x_{median} x_i} / \left\|\overrightarrow{x_{median} x_i}\right\| = 0$, with its Riemannian extension

$$h : m \in \mathrm{M} \mapsto \frac{1}{2}\int_A d(m,a) da \underset{Min}{\Rightarrow} \nabla h = -\int_A \frac{\exp_m^{-1}(a)}{\left\|\exp_m^{-1}(a)\right\|} da \quad (7)$$

We cannot directly extend Karcher Flow to median computation in the discrete case:

$$m_{n+1} = \exp_{m_n}\left(t_n.\sum_{k=1}^{M} \frac{\exp_{m_n}^{-1}(x_k)}{\left\|\exp_{m_n}^{-1}(x_k)\right\|}\right) \quad (8)$$

because $\left\|\exp_{m_n}^{-1}(x_k)\right\|$ could vanish if $m_n = x_k$. The uniqueness of the geometric median of a probability measure on a complete Riemannian manifold [4] has been investigated. By regarding the Weiszfeld algorithm as a sub-gradient procedure, a sub-gradient algorithm has been introduced [1] to estimate the median and to prove that this algorithm always converges :

$$m_{n+1} = \exp_{m_n}\left(t_n.\sum_{k \in G_{m_n}} \frac{\exp_{m_n}^{-1}(x_k)}{\left\|\exp_{m_n}^{-1}(x_k)\right\|}\right) \text{ with } G_{m_n} = \{k / x_k \neq m_n\} \quad (9)$$

Then, the median *A* of the *N* matrices B_k can be computed by sub-gradient Karcher flow :

$$A_{n+1} = A_n^{1/2} e^{t_n\left(\sum_{k \in G_{A_n}} \frac{\log\left(A_n^{-1/2} B_k A_n^{-1/2}\right)}{\left\|\log\left(A_n^{-1/2} B_k A_n^{-1/2}\right)\right\|_F}\right)} A_n^{1/2} \quad (10)$$

with $G_{A_n} = \{k / B_k \neq A_n\}$

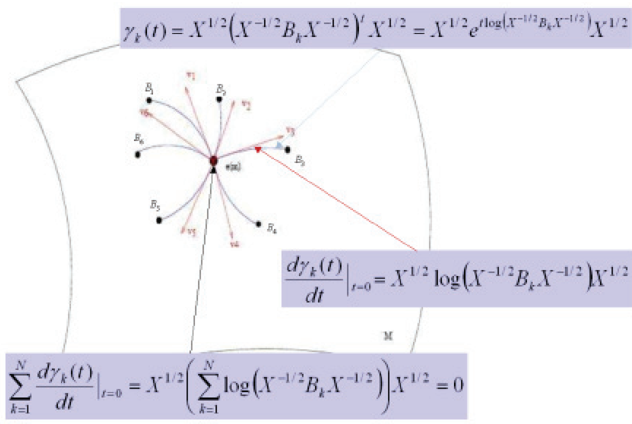

Figure 2. Fréchet-Karcher Flow on space HPD Matrices

IV. Fourier Heat Flow Equation on 1D graph of Hermitian Positive Definite matrices

We can replace Median computation by anisotropic diffusion. In normed vector space in *1D*, if we note $\hat{u}_n = (u_{n+1} + u_{n-1})/2$, Fourier diffusion Equation is given by :

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \Rightarrow$$

$$u_{n,t+1} = u_{n,t} + \frac{2\nabla t}{\nabla x^2}\left[\hat{u}_{n,t} - u_{n,t}\right] = (1-\rho).u_{n,t} + \rho.\hat{u}_{n,t} = u_{n,t} \circ_\rho \hat{u}_{n,t} \quad (11)$$

By analogy, we can define diffusion equation on a *1D* graph of *HPD(n)* by:

$$A_{n,t+1} = A_{n,t}^{1/2} e^{\frac{2\nabla t}{\nabla x^2}\log\left(A_{n,t}^{-1/2}\hat{A}_{n,t}A_{n,t}^{-1/2}\right)} A_{n,t}^{1/2} = A_{n,t}^{1/2}\left(A_{n,t}^{-1/2}\hat{A}_{n,t}A_{n,t}^{-1/2}\right)^{\rho} A_{n,t}^{1/2} = A_{n,t} \circ_\rho \hat{A}_{n,t}$$

with $\rho = \frac{2\nabla t}{\nabla x^2}$ and $\hat{A}_{n,t} = A_{n+1,t}^{1/2}\left(A_{n+1,t}^{-1/2}A_{n-1,t}A_{n+1,t}^{-1/2}\right)^{1/2}A_{n+1,t}^{1/2}$

(12)

Obiously, we can introduce anisotropy by making adaptive the parameter ρ.

From each Time covariance matrix, we can compute Doppler Spectrum. In the following exemple, we give image with range on X axis and Doppler frequency on Y axis. Fourier heat Diffusion is applied on covariance matrices and then, we draw associated Doppler spectrum of results :

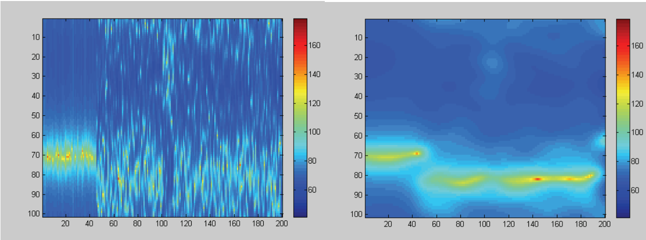

Figure 3. Fourier Heat Equation on a 1D graph of covariance matrices: isotropic diffusion

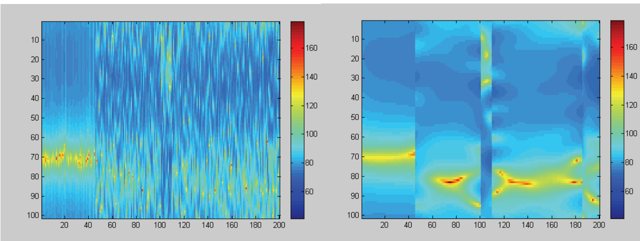

Figure 4. Fourier Heat Equation on a *1D* graph of covariance matrices: anisotropic diffusion

V. Toeplitz-Constrained Flow for Hermitian Positive Definite Covariance Matrices of Stationary Signal

Previous flow has a major drawback, it will not constrained Toeplitz structure of covariance matrices M for stationary signal $r_{n,n-k} = r_k = E\left[z_n z_{n-k}^*\right]$ but only the following structure: $JMJ = M$ with J an anti-diagonal matrix. $E = \{M \in HPD_n(C) / JMJ = M\}$ is a closed sub-manifold of $HPD_n(C)$. To take into account this constraint, Partial Iwasawa decomposition should be considered. This is equivalent for time or space signal to Complex AutoRegressive (CAR) Model decomposition (see Trench Theorem [5]) :

$$\Omega_n = (\alpha_n.R_n)^{-1} = W_n.W_n^+ = \left(1-|\mu_n|^2\right)\begin{bmatrix} 1 & A_{n-1}^+ \\ A_{n-1} & \Omega_{n-1} + A_{n-1}.A_{n-1}^+ \end{bmatrix} \quad (13)$$

$$W_n = \sqrt{1-|\mu_n|^2}\begin{bmatrix} 1 & 0 \\ A_{n-1} & \Omega_{n-1}^{1/2} \end{bmatrix} \text{ with } \Omega_{n-1} = \Omega_{n-1}^{1/2}.\Omega_{n-1}^{1/2+} \quad (14)$$

where $\alpha_n^{-1} = \left[1-|\mu_n|^2\right]\alpha_{n-1}^{-1}$, $A_n = \begin{bmatrix} A_{n-1} \\ 0 \end{bmatrix} + \mu_n \begin{bmatrix} A_{n-1}^{(-)} \\ 1 \end{bmatrix}$ and $V^{(-)} = J.V^*$

In the framework of Information Geometry, Information metric could be introduced as Kählerian metric where Kähler potential is given by the process Entropy $\tilde{\Phi}(R_n, P_0)$:

$$\tilde{\Phi}(R_n, P_0) = \log\left(\det R_n^{-1}\right) - \log(\pi.e) = -\sum_{k=1}^{n-1}(n-k).\log\left[1-|\mu_k|^2\right] - n.\log[\pi.e.P_0]$$

Information metric is then given by hessian of Entropy

$$g_{ij} \equiv \frac{\partial^2 \tilde{\Phi}}{\partial\theta_i^{(n)}\partial\theta_j^{(n)*}} \text{ where } \theta^{(n)} = \left[P_0 \quad \mu_1 \quad \cdots \quad \mu_{n-1}\right]^T \quad (15)$$

with $\{\mu_k\}_{k=1}^{n-1}$ Regularized [2] Burg reflection coefficient [6,7] and $P_0 = \alpha_0^{-1}$ mean signal Power. Kählerian metric is finally :

$$ds_n^2 = d\theta^{(n)+}\left[g_{ij}\right]d\theta^{(n)} = n\left(\frac{dP_0}{P_0}\right)^2 + \sum_{i=1}^{n-1}(n-i)\frac{|d\mu_i|^2}{\left(1-|\mu_i|^2\right)^2} \quad (16)$$

For Median autoregressive model computation, Karcher flow could be trivial. For P_0, we use classical median on real value. For $\{\mu_k\}_{k=1}^{n-1}$, we use homeomorphism of Poincaré's unit disk $\mu_{k,n+1} = \frac{\mu_{k,n} - w_n}{1-\mu_{k,n}.w_n^*}$, to fix the point under action of karcher flow at the origin from where all geodesics are radials and where the space is locally Euclidean. Dual Karcher Flow is:

$$w_n = \gamma_n \sum_{k \in G_0} \frac{\mu_{k,n}}{|\mu_{k,n}|} \text{ with } G_0 = \left\{k / |\mu_{k,n}| \neq 0\right\} \quad (17)$$

Median is deduced from each w_n at each iteration:

$$\mu_{median,n+1} = \frac{\mu_{median,n} + w_n}{1 + \mu_{median,n} w_n^*} \quad (18)$$

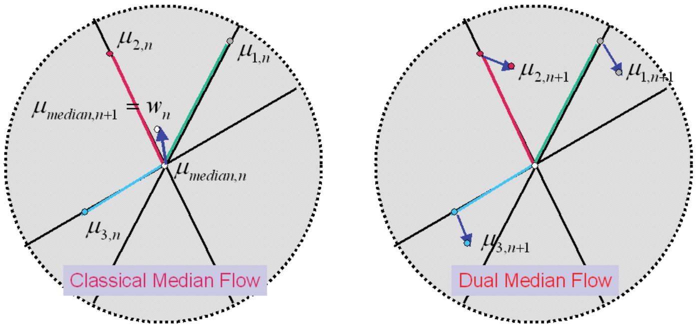

Figure 5. Modified Karcher Flow in Poincaré disk by homeomorphism

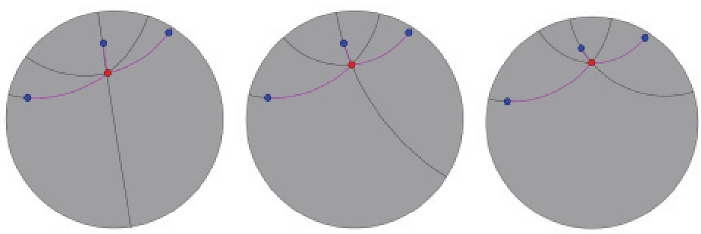

Figure 6. Classical Karcher flow in Poincaré's unit disk

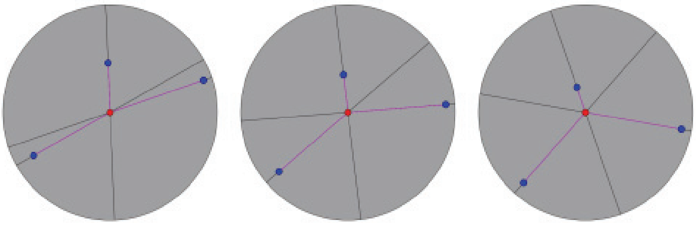

Figure 7. Dual Karcher Flow centered at the origin

VI. BLOCK-TOEPLITZ CONSTRAINED KARCHER FLOW

We will extend previous approach to space-time covariance matrix for STAP. Based on generalization of Trench Algorithm [5], if we consider Toeplitz-block-Toeplitz Hermitian Positive Definite matrix:

$$R_{p,n+1}=\begin{bmatrix} R_0 & R_1 & \cdots & R_n \\ R_1^+ & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_n^+ & \cdots & R_1^+ & R_0 \end{bmatrix}=\begin{bmatrix} R_{p,n} & \widetilde{R}_n \\ \widetilde{R}_n^+ & R_0 \end{bmatrix} \quad (19)$$

with $\widetilde{R}_n=V\begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix}^*$ and $V=\begin{bmatrix} 0 & \cdots & 0 & J_p \\ \vdots & \iddots & \iddots & 0 \\ 0 & J_p & \iddots & \vdots \\ J_p & 0 & \cdots & 0 \end{bmatrix}$ (20)

From Burg-like parameterization [11], we can deduced this inversion of Toeplitz-Block-Toeplitz matrix :

$$R_{p,n+1}^{-1}=\begin{bmatrix} \alpha_n & \alpha_n.\widehat{A}_n^+ \\ \alpha_n.\widehat{A}_n & R_{p,n}^{-1}+\alpha_n.\widehat{A}_n.\widehat{A}_n^+ \end{bmatrix} \quad (21)$$

and $R_{p,n+1}=\begin{bmatrix} \alpha_n^{-1}+\widehat{A}_n^+.R_{p,n}.\widehat{A}_n & -\widehat{A}_n^+.R_{p,n} \\ -R_{p,n}.\widehat{A}_n & R_{p,n} \end{bmatrix}$ (22)

with $\alpha_n^{-1}=\left[1-A_n^nA_n^{n+}\right]\alpha_{n-1}^{-1},\ \alpha_0^{-1}=R_0$

and $\widehat{A}_n=\begin{bmatrix} A_1^1 \\ \vdots \\ A_n^n \end{bmatrix}=\begin{bmatrix} \widehat{A}_{n-1} \\ 0_p \end{bmatrix}+A_n^n.\begin{bmatrix} J_pA_{n-1}^{n-1*}J_p \\ \vdots \\ J_pA_1^{n-1*}J_p \\ I_p \end{bmatrix}$ (23)

Where we have the following Burg-like generalized forward and backward linear prediction :

$$\begin{cases} \varepsilon_{n+1}^f(k)=\sum_{l=0}^{n+1}A_l^{n+1}(k)Z(k-l)=\varepsilon_n^f(k)+A_{n+1}^{n+1}\varepsilon_n^b(k-1) \\ \varepsilon_{n+1}^b(k)=\sum_{l=0}^{n}JA_l^{n+1}(k)^*JZ(k-n+l)=\varepsilon_n^b(k-1)+JA_{n+1}^{n+1*}J\varepsilon_n^f(k) \end{cases}$$

with $\begin{cases} \varepsilon_0^f(k)=\varepsilon_0^b(k)=Z(k) \\ A_0^{n+1}=I_p \end{cases}$

$$A_{n+1}^{n+1}=-2\left[\sum_{k=1}^{N+n}\varepsilon_n^f(k)\varepsilon_n^b(k-1)^+\right]\left[\sum_{k=1}^{N+n}\varepsilon_n^f(k)\varepsilon_n^f(k)^++\sum_{k=1}^{N+n}\varepsilon_n^b(k)\varepsilon_n^b(k)^+\right]^{-1}$$

(24)

Using Schwarz's inequality, it is easily to prove that A_{n+1}^{n+1} Burg-Like reflection coefficient matrix lies in Siegel Disk $A_{n+1}^{n+1}\in SD_p$.

VII. CARTAN-SIEGEL HOMOGENEOUS DOMAINS : SIEGEL DISK

To solve median computation of Toeplitz-Block-Toeplitz matrices, Karcher-Fréchet Flow has to be extended in Siegel Disk. Siegel Disk has been introduced by Carl Ludwig Siegel [8] through Symplectic Group $Sp_{2n}R$ that is one possible generalization of the group $SL_2R=Sp_2R$ (group of invertible matrices with determinant 1) to higher dimensions. This generalization goes further; since they act on a symmetric homogeneous space, the Siegel upper half plane, and this action has quite a few similarities with the action of SL_2R on the Poincaré's hyperbolic plane. Let F be either the real or the complex field, the Symplectic Group is the group of all matrices $M\in GL_{2n}F$ satisfying :

$$Sp(n,F)\equiv\left\{M\in GL(2n,F)\,/\,M^TJM=J\right\}$$
$$J=\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}\in SL(2n,R) \quad (25)$$

or $M=\begin{pmatrix} A & B \\ C & D \end{pmatrix}\in Sp(n,F)\Leftrightarrow A^TC$ and B^TD symmetric (26)

and $A^TD-C^TB=I_n$

The Siegel upper half plane is the set of all complex symmetric *nxn* matrices with positive definite imaginary part:

$$SH_n=\left\{Z=X+iY\in Sym(n,C)\,/\,\mathrm{Im}(Z)=Y>0\right\} \quad (27)$$

The action of the Symplectic Group on the Siegel upper half plane is transitive. The group $PSp(n,R)\equiv Sp(n,R)/\{\pm I_{2n}\}$ is group of SH_n biholomorphisms via generalized Möbius transformations:

$$M=\begin{pmatrix} A & B \\ C & D \end{pmatrix}\Rightarrow M(Z)=(AZ+B)(CZ+D)^{-1} \quad (28)$$

$PSp(n,R)$ acts as a sub-group of isometries. Siegel has proved that Symplectic transformations are isometries for the Siegel metric in SH_n. It can be defined on SH_n using the distance element at the point $Z=X+iY$, as defined by:

$$ds_{Siegel}^2=Tr\left(Y^{-1}(dZ)Y^{-1}(dZ^+)\right) \text{ with } Z=X+iY \quad (29)$$

with associated volume form : $\Omega=Tr\left(Y^{-1}dZ\wedge Y^{-1}dZ^+\right)$

C.L. Siegel has proved that distance in Siegel Upper-Half Plane is given by :

$$d_{Siegel}^2(Z_1,Z_2)=\left(\sum_{k=1}^{n}\log^2\left(\frac{1+\sqrt{r_k}}{1-\sqrt{r_k}}\right)\right) \text{ with } Z_1,Z_2\in SH_n \quad (30)$$

and r_k eigenvalues of the cross-ratio :

$$R(Z_1,Z_2)=(Z_1-Z_2)(Z_1-Z_2^+)^{-1}(Z_1^+-Z_2^+)(Z_1^+-Z_2)^{-1}. \quad (31)$$

This is deduced from the 2nd derivative of $Z\rightarrow R(Z_1,Z)$ in $Z_1=Z$ given by :

$$D^2R=2dZ(Z-Z^+)^{-1}dZ^+(Z^+-Z)^{-1}=(1/2).dZY^{-1}dZ^+Y^{-1} \quad (32)$$

and $ds^2=Tr\left(Y^{-1}dZY^{-1}dZ^+\right)=2.Tr\left(D^2R\right)$ (33)

In parallel, in China in 1945, Hua Lookeng has given the equations of geodesic in Siegel upper-half plane :

$$\frac{d^2Z}{ds^2}+i\frac{dZ}{ds}Y^{-1}\frac{dZ}{ds}=0 \quad (34)$$

Using generalized Cayley transform $W=(Z-iI_n)(Z+iI_n)^{-1}$, Siegel Upper-half Plane SH_n is transformed in unit Siegel disk $SD_n=\{W/WW^+<I_n\}$ where the metric in Siegel Disk is given by :

$$ds^2=Tr\left[(I_n-WW^+)^{-1}dW(I_n-W^+W)^{-1}dW^+\right] \qquad (35)$$

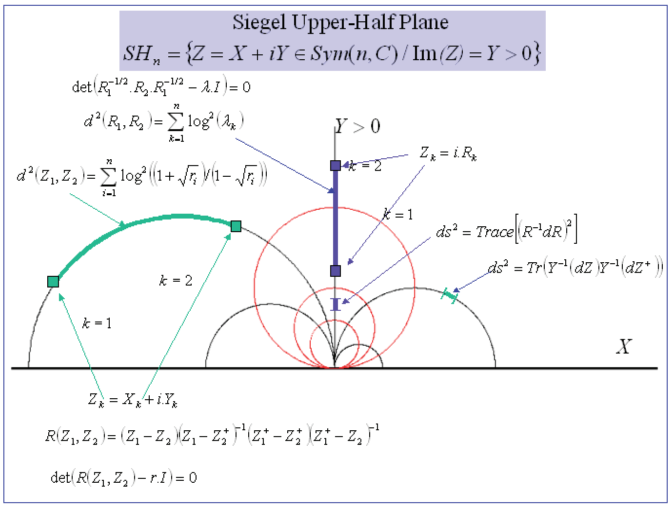

Figure 8. Geometry of Siegel Upper Half-Plane

Contour of Siegel Disk is called its Shilov boundary $\partial SD_n=\{W/WW^+-I_n=0_n\}$. We can also defined horosphere. Let $U\in\partial SD_n$ and $k\in R_*^+$, the following set is called horosphere in siegel disk :

$$H(k,U)=\{Z/0<k(I-Z^+Z)-(I-Z^+U)(I-U^+Z)\}=\left\{Z/\left\|Z-\frac{1}{k+1}U\right\|<\frac{k}{k+1}\right\}$$

(36)

Hua Lookeng and Siegel have proved [8] that the previous positive definite quadratic differential is invariant under the group of automorphisms of the Siegel Disk. Considering $M=\begin{bmatrix}A & B\\ C & D\end{bmatrix}$ such that $M^*\begin{bmatrix}I_n & 0\\ 0 & -I_n\end{bmatrix}M=\begin{bmatrix}I_n & 0\\ 0 & -I_n\end{bmatrix}$:

$$V=M(W)=(AZ+B)(CZ+D)^{-1}\Rightarrow$$
$$(I_n-VV^+)^{-1}dV(I_n-V^+V)^{-1}dV^+= \qquad (36)$$
$$(BW^++A)(I_n-WW^+)^{-1}dW(I_n-W^+W)^{-1}dW^+(BW^++A)^{-1}$$
$$\Rightarrow ds_V^2=ds_W^2$$

Complementary, Hua Lookeng has also proved that, let V,W be complex-valued matrices, if $I-VV^+>0$ and $I-WW^+>0$, then the following identity holds :

$$\det(I-VV^+)\det(I-WW^+)\le\left|\det(I-VW^+)\right|^2 \qquad (37)$$

$\det(I-A^+A)\det(I-B^+B)\le\left|\det(I-A^+B)\right|^2$ is based on Hua's matrix identity :

$$(I-B^+B)+(A-B)^+(I-AA^+)^{-1}(A-B)=(I-B^*A)(I-A^*A)^{-1}(I-A^+B)$$

(38)

using the intermediate equalities:

$$(I-B^+A)(I-A^+A)^{-1}(I-A^+B)-(I-B^+B)=(B-A)^+(I-AA^+)(B-A)$$

(39)

$$(I-A^+A)^{-1}=I+A^+(I-AA^+)^{-1}A$$

and $(I-A^+A)^{-1}A^+=A^+(I-AA^+)^{-1}$ (40)

Same kind of inequality is true for the trace :

$$Tr(I-A^+A)Tr(I-B^+B)\le\left|Tr(I-A^+B)\right|^2 \qquad (41)$$

To go further to study Siegel Disk, we need now to define what are the automorphisms of Siegel Disk SD_n. They are all defined by:

$$\forall\Psi\in Aut(SD_n),\exists U\in U(n,C)/\Psi(Z)=U\Phi_{Z_0}(Z)U^t \qquad (42)$$

with

$$\Sigma=\Phi_{Z_0}(Z)=(I-Z_0Z_0^+)^{-1/2}(Z-Z_0)(I-Z_0^+Z)^{-1}(I-Z_0^+Z_0)^{1/2} \qquad (43)$$

and its inverse :

$$G=(I-Z_0Z_0^+)^{1/2}\Sigma(I-Z_0^+Z_0)^{-1/2}=(Z-Z_0)(I-Z_0^+Z)^{-1}$$
$$\Rightarrow\begin{cases}Z=\Phi_{Z_0}^{-1}(\Sigma)=(GZ_0^++I)^{-1}(G+Z_0)\\ \text{with } G=(I-Z_0Z_0^+)^{1/2}\Sigma(I-Z_0^+Z_0)^{-1/2}\end{cases} \qquad (44)$$

By analogy with Poincaré's unit Disk, C.L. Siegel has deduced geodesic distance in SD_n :

$$\forall Z,W\in SD_n, d(Z,W)=\frac{1}{2}\log\left(\frac{1+\|\Phi_Z(W)\|}{1+\|\Phi_Z(W)\|}\right) \qquad (45)$$

VIII. MOSTOW/BERGER'S FIBRATION OF SIEGEL DISK

Information metric will be introduced as a Kähler potential defined by Hessian of multi-channel entropy $\tilde{\Phi}(R_{p,n+1})$:

$$\tilde{\Phi}(R_{p,n})=\log(\det R_{p,n}^{-1})+cte=-Tr(\log R_{p,n})+cste\Rightarrow g_{ij}=Hess[\phi(R_{p,n})]$$

Using partitioned matrix structure of Toeplitz-Block-Toeplitz matrix $R_{p,n+1}$, recursively parametrized by Burg-Like reflection coefficients matrix $\{A_k^k\}_{k=1}^{n-1}$ with $A_k^k\in SD_n$, we can give Multi-variate entropy [3], matrix extension of previous Entropy :

$$\tilde{\Phi}(R_{p,n})=-\sum_{k=1}^{n-1}(n-k).\log\det\left[1-A_k^kA_k^{k+}\right]-n.\log[\pi.e.\det R_0] \qquad (46)$$

Paul Malliavin [3] has proved that this form is a Kähler Potential of an invariant Kähler metric (Information Geometry metric in our case) that is given by matrix extenstion of (14):

$$ds^2=nTr\left[(R_0^{-1}dR_0)^2\right]+\sum_{k=1}^{n-1}(n-k)Tr\left[(I_n-A_k^kA_k^{k+})^{-1}dA_k^k(I_n-A_k^{k+}A_k^k)^{-1}dA_k^{k+}\right]$$

(47)

As we have defined a metric space, we can extend Karcher/Frechet flow in Unit Siegel Disk to compute the Median of N Toeplitz-Block-Toeplitz Hermitian Positive Definite matrices. These matrices are parameterized by Burg-Like generalized Reflection matrices $\{A_k^k\}_{k=1}^{n-1}$ with $A_k^k\in SD_n$ and Karcher/Frechet Flow in Siegel Disk will be solved by analogy of our scheme used in Poincaré unit Disk, by mean of Mostow Decomposition Theorem : every matrix M of $GL(n,C)$ can be decomposed in $M=Ue^{iA}e^S$ where S is symmetric $S=\frac{1}{2}\log\left(P^{1/2}(P^{-1/2}P^*P^{-1/2})^{1/2}P^{1/2}\right)$ with $P=M^+M$, A is antisymmetric $A=\frac{1}{2i}\log(e^{-S}Pe^{-S})$ and finally, $U=Me^{-S}e^{-iA}$ is unitary. Median in Siegel disk could be then obtained by analogy with scheme developed for median in Poincaré's disk:

Initialisation : $W_{median,0} = 0$ *and* $\{W_{1,0},...,W_{m,0}\} = \{W_1,...,W_m\}$

Iterate on n until $\|G_n\|_F < \varepsilon$

$$W_{k,n} = U_{k,n} e^{iA_{k,n}} e^{S_{k,n}} \Rightarrow H_{k,n} = U_{k,n} e^{iA_{k,n}} = W_{k,n} e^{-S_{k,n}} = e^{-\frac{S_{k,n}}{2}} W_{k,n} e^{-\frac{S_{k,n}}{2}}$$

$$\text{with } S_{k,n} = 1/2.\log\left(P_{k,n}^{1/2}\left(P_{k,n}^{-1/2} P_{k,n}^{*} P_{k,n}^{-1/2}\right)^{1/2} P_{k,n}^{1/2}\right) \text{with } P_{k,n} = W_{k,n}^{+} W_{k,n}$$

$$G_n = \gamma_n \sum_{\substack{k=1 \\ k \neq l}}^{m} H_{k,n} \quad \text{with} \quad \left\{l / \|H_{l,n}\|_F < \varepsilon\right\}$$

$$\text{For } k = 1,...,m \quad \text{then} \quad W_{k,n+1} = \Phi_{G_n}(W_{k,n})$$

$$W_{k,n+1} = \left(I - G_n G_n^{+}\right)^{-1/2} \left(W_{k,n} - G_n\right)\left(I - G_n^{+} W_{k,n}\right)^{-1} \left(I - G_n^{+} G_n\right)^{1/2}$$

$$W_{median,n+1} = \left(G G_n^{+} + I\right)^{-1} \left(G + G_n\right)$$

$$\text{with } G = \left(I - G_n G_n^{+}\right)^{1/2} W_{median,n} \left(I - G_n^{+} G_n\right)^{-1/2}$$

IX. Berezin Quantization of Cartan-Siegel domains

Symmetric Bounded Domains of C^n are key spaces for all these approaches and are particular symmetric spaces of non-compact type. Elie Cartan [4] has proved that there are only 6 types:

- 2 exceptionnal types (E6 et E7)
- 4 Classical Symmetric bounded Domains (extension of Poincaré Unit disk):

Z : Complex Rectangular Matrix

$ZZ^+ < I$ ($^+$: *transposed − conjugate*)

Type I : $\Omega^{I}_{p,q}$ complex matrices with p lines and q rows (48)

Type II : Ω^{II}_{p} complex symmetric matrices of order p

Type III : Ω^{III}_{p} complex skew symmetric matrices of order p

Type IV : Ω^{IV}_{n} complex matrices with n rows and 1 line :

$\left|ZZ^{t}\right| < 1, 1 + \left|ZZ^{t}\right|^2 - 2ZZ^{+} > 0$

kernel function for all these domains were established by Lookeng Hua:

$$K(Z,W^*) = \frac{1}{\mu(\Omega)} \det\left(I - ZW^{+}\right)^{-\nu} \quad \text{for} \begin{cases} \text{Type I} : \Omega^{I}_{p,q}, \nu = p+q \\ \text{Type II} : \Omega^{II}_{p}, \nu = p+1 \\ \text{Type III} : \Omega^{III}_{p}, \nu = p-1 \end{cases}$$

$$K(Z,W^*) = \frac{1}{\mu(\Omega)} \left(1 + ZZ^{t} W^{*} W^{+} - 2ZW^{*}\right)^{\nu} \text{ for Type IV} : \Omega^{IV}_{n}, \nu = n$$

where $\mu(\Omega)$ is euclidean volume of the domain.

(49)

For the case (*p=q=n=1*), all these domains are reduced to the classical Poincaré unit disk :

$$\Omega^{I}_{1,1} = \Omega^{II}_{1} = \Omega^{III}_{1} = \Omega^{IV}_{1} = \left\{z \in C / zz^* < 1\right\}, K(z,w^*) = \frac{1}{\left(1 - zw^*\right)^2} \quad (50)$$

Groups of analytic automorphisms of these domains are locally isomorphic to the group of matrices which preserve following forms:

$$\text{Type I} : \Omega^{I}_{p,q}, AHA^* = H, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, \det A = 1$$

$$\text{Type II} : \Omega^{II}_{p}, AHA^* = H, AKA^t = K, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, K = \begin{pmatrix} 0 & I_p \\ -I_p & 0 \end{pmatrix}$$

$$\text{Type III} : \Omega^{III}_{p}, AHA^* = H, ALA^t = L, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, L = \begin{pmatrix} 0 & I_p \\ I_p & 0 \end{pmatrix}$$

$$\text{Type IV} : \Omega^{IV}_{n}, AHA^* = H, AHA^t = H, H = \begin{pmatrix} -I_2 & 0 \\ 0 & I_n \end{pmatrix}$$

(51)

All classical domains are circular and considered in the general framework of Elie Cartan Theory, where the origin is a distinguished point for the potential:

$$\Phi(Z,Z^*) = \log\left[\frac{K(Z,Z^*)}{K(0,0)}\right] = \log\det\left(I - ZZ^{+}\right)^{-\nu} \quad (52)$$

F.A. Berezin [3] has introduced on these Cartan-Siegel domains the concept of quantization based on construction of Hilbert spaces of analytical functions:

$$\langle f,g \rangle = c(h) \int f(Z) g(Z) \left[\frac{K(Z,Z^*)}{K(0,0)}\right]^{-1/h} d\mu(Z,Z^*)$$

$$c(h)^{-1} = \int \left[\frac{K(Z,Z^*)}{K(0,0)}\right]^{-1/h} d\mu(Z,Z^*) \quad (53)$$

$$K(gZ, gZ^*) j(g,z) j(g,Z)^* = K(Z,Z^*) \text{ with } j(g,Z) = \frac{\partial gZ}{\partial Z}$$

One example is given in dimension 1 for Poincaré unit disk $D = \{z \in C / |z| < 1\} = SU(1,1)/S^1$ with volume element $1/2i.(1 - |z|^2)^{-2} dz \wedge dz^*$:

$$g \in SU(1,1) \text{ with } g = \begin{bmatrix} a & b \\ b^* & a^* \end{bmatrix} \text{ where } |a|^2 - |b|^2 = 1$$

$$\text{with Kähler potential} : F(z) = -\log\left(1 - |z|^2\right) \quad (54)$$

$$\Rightarrow F(gz) = 2\operatorname{Re}\log\left(b^* z + a^*\right) + F(z)$$

$$\Rightarrow ds^2 = \frac{\partial^2 F(gz)}{\partial z \partial z^*} = \frac{\partial^2 F(z)}{\partial z \partial z^*}$$

It results from the last equation that the Kählerian metric is invariant under the action $g \in G$ (automorphisms of unit disk). The transform of the base point $z = 0$ of the disk by $g \in G$ is given by $g(0) = b(a^*)^{-1}$. It defines a lifting that allows to associate to all paths in disk a lift in G. In the same way, we can define a geometric lift of potential *K* in *G*:

$$g(0) = b\left(a^*\right)^{-1}$$

$$\Rightarrow F(g(0)) = -\log\left(1 - \left|b\left(a^*\right)^{-1}\right|^2\right) \underset{|a|^2 - |b|^2 = 1}{=} \log\left(1 + |b|^2\right) \quad (55)$$

$$g^{-1} = \begin{pmatrix} a^* & -b \\ -b^* & a \end{pmatrix} \Rightarrow F(g^{-1}) = F(g)$$

Obviously, all these lifts could be extended to Cartan-Siegel Domains $SD_n = \left\{Z / ZZ^{+} < I\right\}$:

Let $g=\begin{bmatrix} A & B \\ B^* & B^* \end{bmatrix}$ and $g^t J g = J$ with $J=\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$

A and B that verify $\begin{cases} A^+A - B^t B^* = I \\ B^+A - A^t B^* = 0 \end{cases}$

$$g(Z) = (AZ+B)(B^*Z+A^*)^{-1} \quad (56)$$

of Kähler Potential :

$$F(z) = -\log\det(I-ZZ^+) = -Tr[\log(I-ZZ^+)]$$

$$F(g(Z)) = F(Z) + 2\,\mathrm{Re}(Tr(\log(A^*+B^*Z)))$$

$$\Rightarrow \partial\partial^* F(g(Z)) = \partial\partial^* F(Z)$$

Geometric Lift in Cartan-Siegel domain is then given by the following:

$$g(0) = B(A^*)^{-1} \quad (57)$$

$$\Rightarrow F(g(0)) = \log\det(I+BB^+) = Tr[\log(I+BB^+)]$$

F.A. Berezin [3] has proved that for every symmetric Riemannian space, there exist a dual space being compact. The isometry groups of all the compact symmetric spaces are described by block matrices (the action of the group in terms of special coordinates is described by the same formula as the action of the group of motions of the dual domain).

$$\Gamma = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \Rightarrow \Gamma(W) = (A_{11}W + A_{12})(A_{21}W + A_{22})^{-1} \quad (58)$$

$$\text{Isometry}: \Gamma = C\Gamma C^{-1} \text{ with } C = \frac{1}{\sqrt{2}}\begin{pmatrix} I & iI \\ iI & I \end{pmatrix}$$

Berezin coordinates for Siegel domain are given by

$$\Gamma = \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix},\ \Gamma^{-1} = \begin{pmatrix} A^+ & B^t \\ B^+ & A^t \end{pmatrix} \quad (59)$$

$$\text{or equivalently}: \Gamma\Gamma^+ = I\ ,\ \Gamma L \Gamma^t = L \text{ with } L = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$$

$$g(0) = B(A^*)^{-1} \Rightarrow F(g(0)) = \ln\det(I+BB^+) = trace\ln(I+BB^+) \quad (60)$$

For this dual space, the volume and the metric are invariant:

$$d\mu(W,W^*) = H(W,W^*)\frac{d\mu_L(W,W^*)}{\pi^n}$$

$$ds^2 = \sum_{\alpha,\beta} g_{\alpha,\beta} dW^\alpha dW^{\beta *} \text{ with } g_{\alpha\beta} = -\frac{\partial^2 \log H(W,W^*)}{\partial W^\alpha \partial W^{\beta *}} \quad (61)$$

where $H(W,W^*) = \det(I+WW^+)^{-\nu}$

For arbitrary Kählerian homogeneous space, the logarithm of the density for the invariant measure is the potential of the metric.

X. Median and Conformal Busemann Barycenters

The Fréchet-median barycentre, in Poincaré/Siegel unit disk is equivalent to conformal Douady/Busemann barycenters [12]. The principle is to assign to every probability measure μ on $S^1 = \{|z|=1\}$ a point $B(\mu) \in D = \{|z|<1\}$ so that the map $\mu \mapsto B(\mu)$ is conform and satisfies: $B(\mu) = 0 \Leftrightarrow \int_{S^1} \zeta . d\mu(\zeta) = 0$

There is a unique conformally way to assign to each probability measure μ on S^1 a vector field ξ_μ on D such that: $\xi_\mu(0) = \int_{S^1} \zeta . d\mu(\zeta) = 0$. For general w in D, if we set $g_w(z) = \frac{z-w}{1-w^*z}$, the assignment is given by $\mu \mapsto \xi_\mu$:

$$\xi_\mu(w) = \frac{1}{(g_w)'(w)}\xi_{(g_w)^*(\mu)}(0) = (1-|w|^2)\int_{S^1}\left(\frac{\zeta - w}{1-w^*\zeta}\right)d\mu(\zeta) = 0 \quad (62)$$

Douady-Earle definition of conformal barycenter $B(\mu)$ of μ is the unique zero of ξ_μ in D. $\xi_\mu(z)$ can be written according to $\xi_\zeta(z)$ that is the unit tangent vector of geodesic at $z \in D$ pointing toward: $\zeta \in S^1$: $\xi_\mu(z) = \int_{S^1} \xi_\zeta(z) d\mu(\zeta)$

(63)

In Poincaré geometry of unit disk, the vector field ξ_ζ is the gradient of a function h_ζ whose level lines are the horocycles tangent to S^1 at $\zeta \in S^1$:

$$\xi_\mu = \nabla h_\mu \text{ with } h_\mu : z \mapsto \int_{S^1} h_\zeta(z) d\mu(\zeta)$$

$$h_\mu(z) == \int_{S^1} \frac{1}{2}\log\left(\frac{1-|z|^2}{|z-\zeta|^2}\right)d\mu(\zeta) = \int_{S^1} \underset{r\to 1^-}{Lim}[d(0,r) - d(z,r\zeta)]d\mu(\zeta) \quad (64)$$

with $d(z,w)$ the Poincaré distance from z to w in D. We can then observe that the Fréchet median of N points in D, $\{w_1, w_2, w_3, ..., w_N\}$, is the conformal barycenter of associated push forward N points $\{\zeta_1, \zeta_2, \zeta_3, ..., \zeta_N\}$ on S^1 :

$$Median_{Fréchet}\{w_1, w_2, w_3, ..., w_N\} = Conformal\_Barycenter\{\zeta_1, \zeta_2, \zeta_3, ..., \zeta_N\}$$

$$\zeta_i = \underset{t\to+\infty}{Lim}\,\gamma_{\xi_{\zeta_1}(z)}(t) = \underset{t\to+\infty}{Lim}\,\gamma_z^{w_i}(t)$$

where $\zeta_i = \underset{t\to+\infty}{Lim}\,\gamma_{\xi_{\zeta_1}(z)}(t) = \underset{t\to+\infty}{Lim}\,\gamma_z^{w_i}(t)$ the "push forward" association is given by the limit point when t tends to infinity along the geodesic from the barycenter toward w_i.

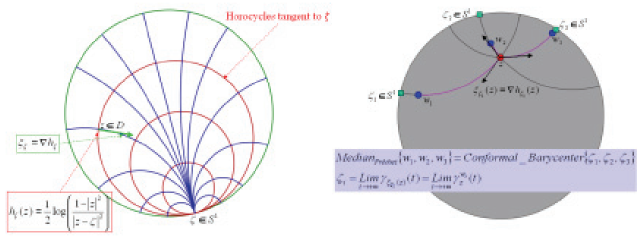

Figure 9. ***Frechet Median Barycenter = Busemann Barycenter***

All these results could be extended for Busemann barycenter on Shilov boundary of Cartan Hadamard manifolds. Following functional associated to median barycenter $x \mapsto \int_{H^n} d(x,z) d\nu(z)$ is simply connected and reached its minimum. When z goes to infinity and converges to a point θ of ∂H^n, distance function to z normalized converges to Busemann function: $B_\theta(x) = \lim_{z\to\theta} d(x,z) - d(O,z)$. Fixing the origin $O \in H^n$, let ν finite measure on ∂H^n. The following strictly convex function reaches its minimum at a unique point, independent of origin and called Busemann barycenter: $\beta_\nu(x) = \int_{\partial H^n} B_\theta d\nu(\theta)$ (65)

XI. Radar Applications for Robust Ordered-Statistic Processing: OS-HDR-CFAR and OS-STAP

In the following, we will apply previous tools to built Robust Ordered-Statistic (OS) processing. Ordered-Statistic is a very useful tool used in Radar for a long time to be robust againt outliers on scalar data from secondary data. We will define an OS-HDR-CFAR (Ordered-Statistic High Doppler Resolution Constant False Alarm Rate) algorithm jointly taking into account robustness of “matrices median” and high Doppler resolution of regularized Complex Auto-Regressive model. We will define also an OS-STAP (Ordered Statistic Space-Time Adaptive Processing), based on median computation of secondary data space-time covariance matrix with Mostow/Berger fibration applied on Multichannel Autoregressive Model.

This paper will not address Polarimetric Data processing, but obviously these tools could be extended to compact manifold to define Ordered statistic for Polarimetric covariance matrices.

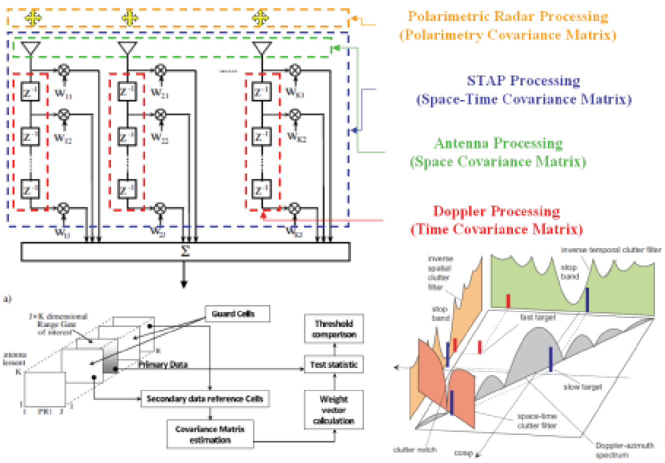

Figure 10. Doppler Processing, Antenna Processing, Space-Time Processing & Polar processing

XII. Robust Doppler Processing : OS-HDR-CFAR

The regularized Burg algorithm [2] is an alternative Bayesian composite model approach to spectral estimation. The reflection coefficients, defined in classical Burg algorithm are estimated through a regularized method, based on a Bayesian adaptative spectrum estimation technique, proposed by Kitagawa & Gersch, who use normal prior distributions expressing a smoothness priors on the solution. With these priors, autoregressive spectrum analysis is reduced to a constrained least squares problem, minimized for fixed tradeoff parameters, using Levinson recursion between autoregressive parameters. Then, a reflection coefficient is calculated, for each autoregressive model order, by minimizing the sum of the mean-squared values of the forward and backward prediction errors, with spectral smoothness constraints . Tradeoff parameters balance estimate of the autoregressive coefficients between infidility to the data and infidility to the frequency domain smoothness constraint. This algorithm conserves lattice structure advantages, and could be brought in widespread use with a multisegment regularized reflection coefficient version. The regularized Burg algorithm lattice structure offers implementation advantages over tapped delay line filters because they suffer from less round-off noise and less sensivity to coefficient value perturbations.

. Initialisation :

$f_0(k) = b_0(k) = z(k)$, $k=1,...,N$ (N : nb. pulses per burst)

$P_0 = \frac{1}{N}.\sum_{k=1}^{N} |z(k)|^2$ and $a_0^{(0)} = 1$

. Iteration (n): for $n = 1$ to M

$$\mu_n = -\frac{\frac{2}{N-n}\sum_{k=n+1}^{N} f_{n-1}(k).b_{n-1}^*(k-1) + 2.\sum_{k=1}^{n-1}\beta_k^{(n)}.a_k^{(n-1)}.a_{n-k}^{(n-1)}}{\frac{1}{N-n}\sum_{k=n+1}^{N}\left|f_{n-1}(k)\right|^2 + \left|b_{n-1}(k-1)\right|^2 + 2.\sum_{k=0}^{n-1}\beta_k^{(n)}\left|a_k^{(n-1)}\right|^2}$$

with $\beta_k^{(n)} = \gamma_1.(2\pi)^2.(k-n)^2$

$$\begin{cases} a_0^{(n)} = 1 \\ a_k^{(n)} = a_k^{(n-1)} + \mu_n.a_{n-k}^{(n-1)*} \quad , \quad k=1,...,n\text{-}1 \\ a_n^{(n)} = \mu_n \end{cases}$$

and $\begin{cases} f_n(k) = f_{n-1}(k) + \mu_n.b_{n-1}(k-1) \\ b_n(k) = b_{n-1}(k-1) + \mu_n^*.f_{n-1}(k) \end{cases}$ (66)

In the following figures, regularization property is illustrated with deletion of spurious peaks. We select the AR model of maximum order (number of pulses minus one).

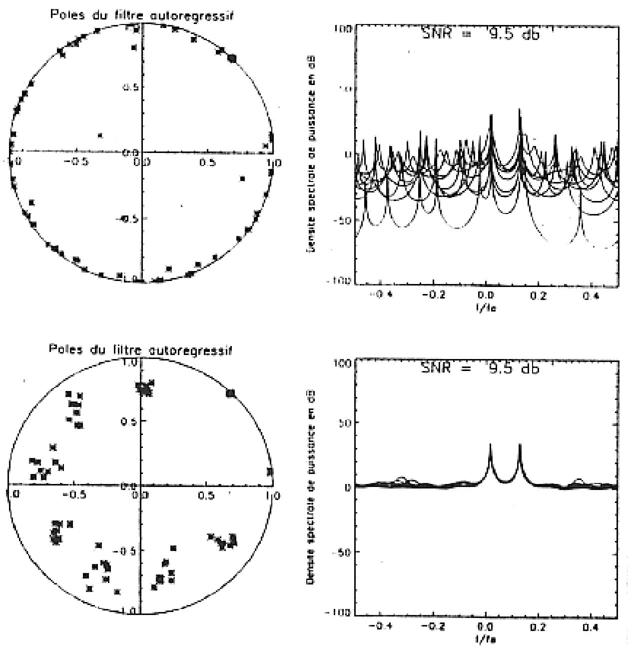

Figure 11. (up) Non-regularized & (bottom) Regularized Doppler AR Spectrum

We conserve the sliding window structure of classical CFAR : we compare the AR model under test by computing its Information Geometry distance with median AR model of secondary data in the neighborhood. Median Autoregressive model is computed by :

For $P_{median,0}$, we use classical median on real values $P_{0,k}$

For $\{\mu_k\}_{k=1}^{n-1}$:

$$w_n = \gamma_n \sum_{k \in G_0} \frac{\mu_{k,n}}{|\mu_{k,n}|} \text{ with } G_0 = \{k / |\mu_{k,n}| \neq 0\} \quad (67)$$

$$\mu_{k,n+1} = \frac{\mu_{k,n} - w_n}{1 - \mu_{k,n}.w_n^*} \quad (68)$$

$$\mu_{median,n+1} = \frac{\mu_{median,n} + w_n}{1 + \mu_{median,n} w_n^*} \quad (69)$$

The detection test is finally based on computation of the robust Information Geometry distance:

$$d^2\left[\left(P_{0,k} \quad \mu_{1,k} \quad \cdots \quad \mu_{N-1,k}\right),\left(P_{median,0} \quad \mu_{1,median} \quad \cdots \quad \mu_{N-1,median}\right)\right] =$$
$$n \log^2\left(\frac{P_{median,0}}{P_{0,k}}\right) + \sum_{i=1}^{N-1}(N-k)\left(\frac{1}{2}\log\left(\frac{1+\delta_i}{1-\delta_i}\right)\right)^2 \quad \text{with} \quad \delta_i = \left|\frac{\mu_{i,k} - \mu_{i,median}}{1 - \mu_{i,k}\mu_{i,median}^*}\right|$$

(70)

In the following figure, we compare the classical processing chain with new OS-HDR-CFAR

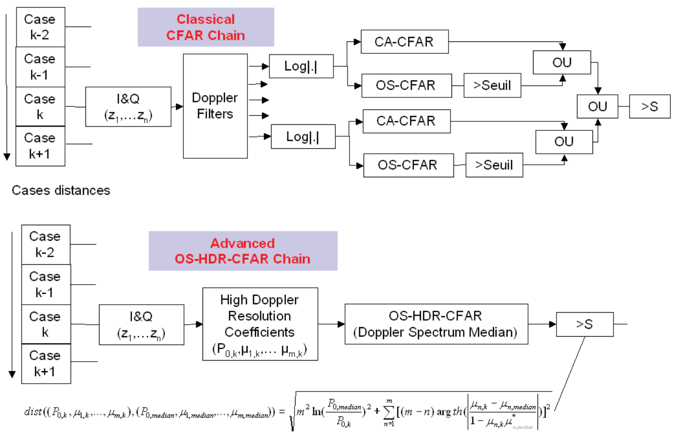

Figure 12. (top figure) Classical OS-CFAR after filter banks, (bottom figure) OS-HDR-CFAR

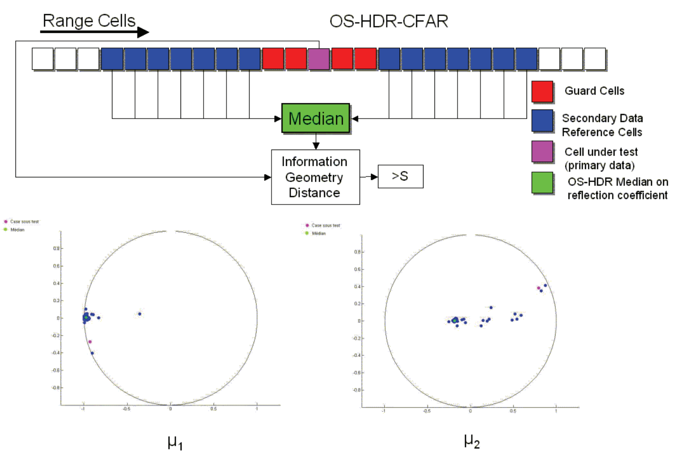

Figure 13. OS-HDR-CFAR Algorithm with illustration of two first reflection coefficients

We have tested OS-HDR-CFAR on real recorded ground Radar clutter with ingestion of synthetic slow targets.

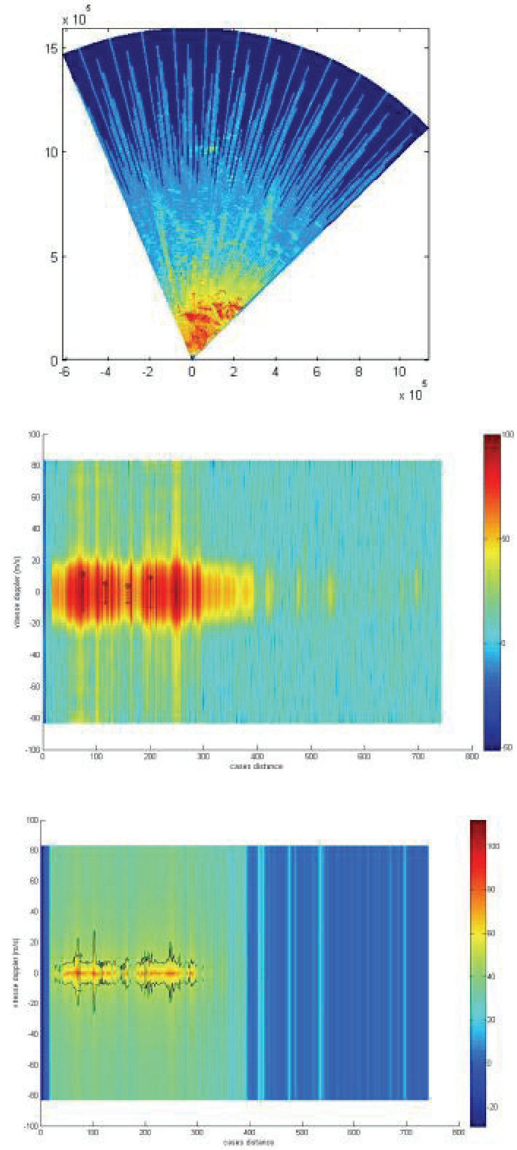

Figure 14. Reflectivity of ground clutter in recorded sector (top),Comparison of FFT Doppler spectrum (middle) and High Resolution Regularized Doppler spectrum (down)

In the following figures, we give ROC curves with Probability of detection versus probability of false alarm. We observe that OS-HDR-CFAR is better (Pd = 0.8) than OS-CFAR/Doppler-Filters (Pd = 0.65) for arbitrary fixed Pfa. We could also observe that Information Geometry approach provides better results than Optimal Transport Theory approach (based on Wasserstein distance/barycenter : black curve).

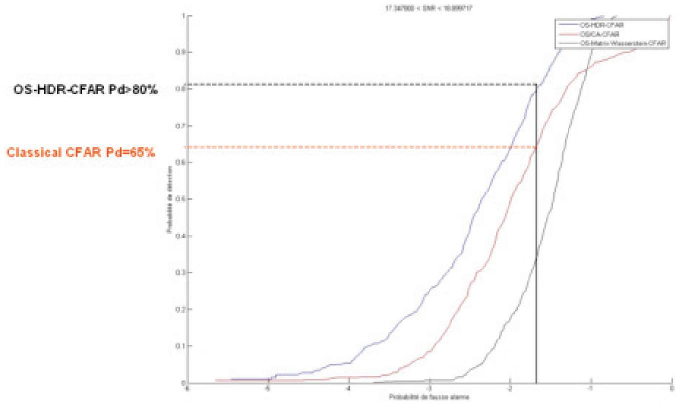

Figure 15. ROC curves for 3 approaches : OS-HDR-CFAR, OS-CFAR/Doppler-Filters, & method based on Wasserstein barycenter/distance (optimal transport theory)

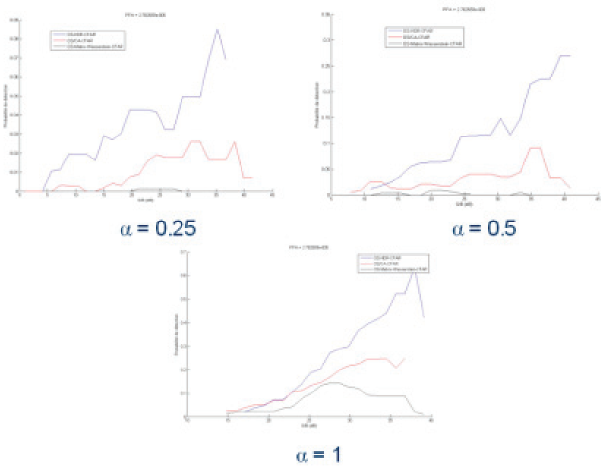

Figure 16. Probability of detection versus SNR for fixed Pfa=10^{-5}, with α relative position in Doppler of the target normalized by Doppler Clutter Spectrum Width (α=1 means that the target is positioned on the edge in Doppler of the ground clutter)

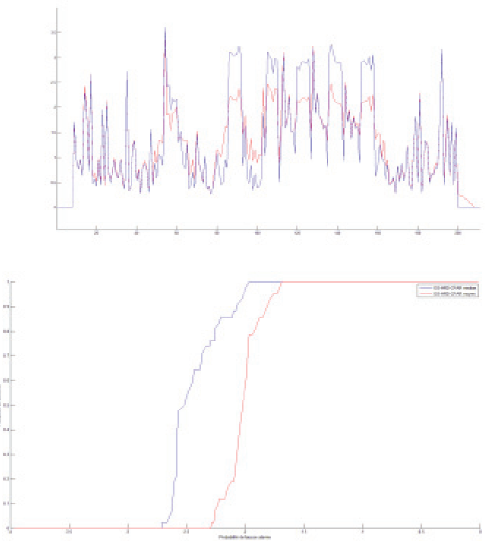

Figure 17. Comparison of “Mean” in red and “Median” in blue for clutter level estimation (top) and through Pd/Pfa curves (down).

XIII. Conclusion

Fréchet Median with Information Geometry and Geometry of of *HPD(n)* matrices is a new tool for Radar Signal Processing that could improve drastically performance and robustness of classical methods, in Doppler processing and in STAP. Obviously, these approaches could be extended to Array Processing and Polar Data Processing in the same way on respectively spatial covariance matrix and Polar covariance matrix. Future works will be dedicated to deepen close relations of Information Geometry with Lagrange Symplectic Geometry and Geometric Quantization.

Acknowledgment

We would like to thank Leon Brillouin Seminar members for fruitful discussions on “Geometric Sciences of Information”, hosted by Arshia Cont at IRCAM in Paris since 2009. We would more specifically thank Yang Le, PhD student of Poitiers University, supervised by Prof. Marc Arnaudon and myself, that has proved convergence of all flows introduced in this paper. Yang Le has win the Thales PhD Award 2012 for this work, with support of Cédric Villani (Fields Medal 2010, Institut Henri Poincaré Director) Referee Letter.

References

[1] L. Yang, « Médianes de mesures de probabilité dans les variétés riemanniennes et applications à la détection de cibles radar », Thèse de l’Université de Poitiers, tel-00664188, 2011, Thales PhD Award 2012

[2] F. Barbaresco, «Algorithme de Burg Régularisé FSDS. Comparaison avec l’algorithme de Burg MFE », GRETSI conference , pp.29–32 ,1995

[3] F. Barbaresco,”Information Geometry of Covariance Matrix”, Matrix Information Geometry Book, F. Nielsen & R. Bhatia Ed., Springer, 2012

[4] M. Émery, G. Mokobodzki, «Sur le barycentre d'une probabilité dans une variété», Séminaire de probabilité Strasbourg, n°25,p.220-233, 1991

[5] W.F. Trench,, ”An algorithm for the inversion of finite Toeplitz matrices”, J. Soc. Indust. Appl. Math. 12, pp.515–522, 1964

[6] N.H. Bingham,”Szegö’s Theorem and its Probabilistic Descendants”, 2012, http://arxiv.org/abs/1108.0368v2

[7] H.J. Landau, “Maximum Entropy and The Moment Problem”, Bulletin of the American Mathematical Society, pp.47-77, vol.16,n°1, Jan. 1987

[8] C. L. Siegel,“Symplectic geometry”,Amer. J. Math., (65), pp.1-86, 1943

[9] J.L. Koszul, «Exposés sur les espaces homogènes symétriques », Publicacao da Sociedade de Matematica de Sao Paulo, 1959

[10] P. Delsarte, Y.V. Genin,”Orthogonal Polynomial Matrices on the Unit Circle”, IEEE Trans on CS., vol.25, n°3, pp.149-160, March 1978

[11] R. Kanhouche, “A Modified Burg Algorithm Equivalent In Results to Levinson Algorithm”, http://hal.archives-ouvertes.fr/ccsd-00000624

[12] C.J. Douady, C.J. Earle,”Conformally natural extension of homeomorphisms of circle”, Acta Math. 157, pp.23-48, 1986

Frédéric Barbaresco (graduated from French High Engineering School SUPELEC in 1991) became a Member (M) of SEE in 1991, a Senior Member (SM) in 2000, and an Emeritus Member (EM) in 2011. The author become a Member (M) of IEEE in 2012. F. Barbaresco was awarded by SEE Ampere medal in 2007 and NATO SET Lecture Award in 2012 for his NATO Lecture from 2008 to 2011 in Europe and North America on “Waveform Diversity and Design for Advanced Radar Systems”.

F. Barbaresco has worked for THALES group (www.thalesgroup.com) since 1991. He is Senior Scientist and Advanced Studies Manager in Advanced Developments Department of Technical Directorate in Surface Radar Domain of THALES AIR SYSTEMS.

Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median

Frédéric Barbaresco*

*Thales Air Systems, Surface Radar Domain, Techn. Directorate, Adv. Developments Dept.
Voie Pierre-Gilles de Gennes, 91470 Limours, FRANCE
frederic.barbaresco@thalesgroup.com

Abstract: Information Geometry has been introduced by C.R.Rao, and axiomatized by N. Chentsov, to define a distance between statistical distributions that is invariant to non-singular parameterization transformations. For Doppler/Array/STAP Radar Processing, Information Geometry Approach will give key role to Homogenous Symmetric bounded domains geometry. For Radar, we will observe that Information Geometry metric could be related to Kähler metric, given by Hessian of Kähler potential (Entropy of Radar Signal given by –*log[det(R)]*). To take into account Toeplitz structure of Time/Space Covariance Matrix or Toeplitz-Block-Toeplitz structure of Space-Time Covariance matrix, Parameterization known as Partial Iwasawa Decomposition could be applied through Complex Autoregressive Model or Multi-channel Autoregressive Model. Then, Hyperbolic Geometry of Poincaré Unit Disk or Symplectic Geometry of Siegel Unit Disk will be used as natural space to compute "*p-mean*" (*p=2* for "mean", *p=1* for "median") of covariance matrices via Karcher Flow derived from Weiszfeld algorithm extension on Cartan-Hadamard Manifold. This new mathematical framework will allow development of OS (Ordered Statistic) concept for Hermitian Positive Definite Covariance Space/Time Toeplitz matrices or for Space-Time Toeplitz-Block-Toeplitz matrices. We will define OS-HDR-CFAR (Ordered Statistic High Doppler Resolution CFAR) and OS-STAP (Ordered Statistic Space-Time Adaptive Processing).

Keywords: Median, p-mean, Center of mass, Cartan-Hadamard Manifold, Cartan Symmetric spaces, Homogenous bounded domains, Siegel Upper-half plane, Poincaré disk, Siegel disk, Fréchet metric space, Berger fibration, Mostow decomposition, Radar, CFAR, STAP

1 Historical Preamble

Information Geometry has been introduced by Indian scientist Calyampudi Radhakrishna Rao [15] (http://www.crraoaimscs.org/) PhD student of R.A. Fisher, and axiomatized by N. N. Chentsov [5], to define a distance between statistical distributions that is invariant to nonsingular parameterization transformations. C.R. Rao introduced this geometry in his 1945 seminal paper on Cramer-Rao Bound. This bound was discovered in parallel by Maurice René Fréchet [6] in 1939 (in his Lecture of "Institut Henri Poincaré" of Winter 1939, in Paris) and extended later to multivariate case by Georges Darmois. We will see that this geometry could be considered in the framework of Positive Definite Matrices Geometry for Complex circular Multivariate Laplace-Gauss Law. Same kind of geometry has been also introduced by Functional Analysis approach by Prof. R. Bhatia in his recent Book [2]. For Doppler / Array / STAP Radar Processing, Information Geometry Approach will give key role to Homogeneous Symmetric bounded domains geometry. For Radar, we will propose Information Geometry metric as Kähler metric, given by Hessian of Kähler potential (Entropy of Radar Signal given by *–log det(R)* where *R* is the covariance matrix*),* which is also a Bergman metric. To take into account Toeplitz structure of Time/Space Covariance Matrix or Toeplitz-Block-Toeplitz structure of Space-Time Covariance matrix for stationary signal, Parameterization known as Partial Iwasawa Decomposition [9] could be applied through Complex Autoregressive Model or Multi-channel Autoregressive Model. Then, Hyperbolic Geometry of Poincaré Unit Disk [14] or Symplectic Geometry of Siegel Unit Disk [17, 18] will be used as natural metric space to compute p-mean (p=2 for mean, p=1 for median) of covariance matrices via Fréchet/Karcher Flow [7, 10] derived from Weiszfeld algorithm [108] extension on Cartan-Hadamard Manifold and on Fréchet Metric spaces. This new mathematical framework will allow developing concept of OS (Ordered Statistic) for Hermitian Positive Definite Covariance Space/Time Toeplitz matrices or for Space-Time Toeplitz-Block-Toeplitz matrices. We will then define OS-HDR-CFAR (Ordered Statistic High Doppler Resolution CFAR) and OS-STAP (Ordered Statistic Space-Time Adaptive Processing). This approach is based on the existence of a center of mass in the large for manifolds with non-positive curvature that was proven and used by Elie Cartan back in the 1920s [3]. The general case was employed by Calabi in an unpublished note. In 1977, Hermann Karcher [10] has proposed intrinsic flow to compute this barycenter, that we adapt for covariance matrices. This geometric foundation of Radar Signal Processing is based on general concept of Cartan-Siegel domains [4, 17, 18]. We will then give a brief history of Siegel domains studies in Europe, Russia and China. In 1935, Elie Cartan [4] proved that irreductible homogeneous bounded symmetric domains could be reduced to six types, included two exceptional ones. Four non-exceptionnal Cartans domains are now called classical models, and their extension by Siegel are considered as the higher dimensional analogues of the Poincaré unit disk [14] in the complex plane. After these seminal work of Elie Cartan, in the framework of Sympletic Geometry [17, 18], Carl Ludwig Siegel has introduced first explicit descriptions of symmetric domains, where the realization of bounded domains as unbounded domains played fundamental role (for an important class of them, these unbounded domains are Siegel domains of the first kind, with important particular case of Siegel Upper Half Plane). In 1953, Loo-Keng Hua [8] obtained the orthonormal system and the Bergman/Cauchy/Poisson kernel functions for each of the four classical domains using group representation theory.

Elie Cartan proved that all bounded homogeneous complex domains in dimension 2 and 3 are symmetric and conjectured that is true for dimension greater than 3. Ilya

Piatetski-Shapiro [13], after Hua works, has extended Siegel description to other symmetric domains and has disproved the Elie Cartan conjecture that all transitive domains are symmetric with a counter example. In parallel, A. Borel showed that if in a bounded homogeneous region a semi-simple Lie group operates transitively, then that region is symmetric. These results were strengthened by Hano and obtained by Jean-Louis Koszul [11, 12, 113] who also studied affinely homogeneous regions that are fundamental for Information Geometry and real Hessian or complex Kählerian geometries (see in book [16], Koszul's references inside). Piatetski-Shapiro introduced affinally general definition of a Siegel domain of the second kind (all symmetric domains allow a generalization of Siegel tube domains), and has proved in 1963 with S.G. Gindikin and E. Vinberg that any bounded homogeneous domain has a realization as a Siegel domain of the second kind with transitive action of linear transformation. In parallel, E. Vinberg [19] worked on the theory of homogeneous convex cones, as fundamental construction of Siegels domains (he introduced a special class of generalized matrix T-algebras), and S.G. Gindikin worked on analytic aspects of Siegels domains. More recently, classical complex Symmetric spaces have been studied by F. Berezin [1] in the framework of quantization. With Karpelevitch [114], Piatetski-Shapiro explored underlying geometry of these complex homogeneous domains manifolds, and more especially, the fibration of domains over components of the boundary. Let a bounded domain, he constructed a fibration by looking at all the geodesic that end in each boundary component and associating the end point to every point on the geodesic. This fibration is important to understand Satake compactifications, and will be studied in the framework of Mostow/Fibration for our application. For our Radar STAP and Toeplitz-Block-Toeplitz covariances matrices, we have used Berger Fibration in Unit Siegel Disk based on the theorem that all symmetric spaces are fibered on a compact symmetric space (Mostow decomposition [57,58,115]).

At the end of the paper, we will underline close relations between Fréchet median in Poincaré Unit Disk with Conformal barycenter on its boundary introduced by Douady and Earle.

Structure of the paper is explained in the following image:

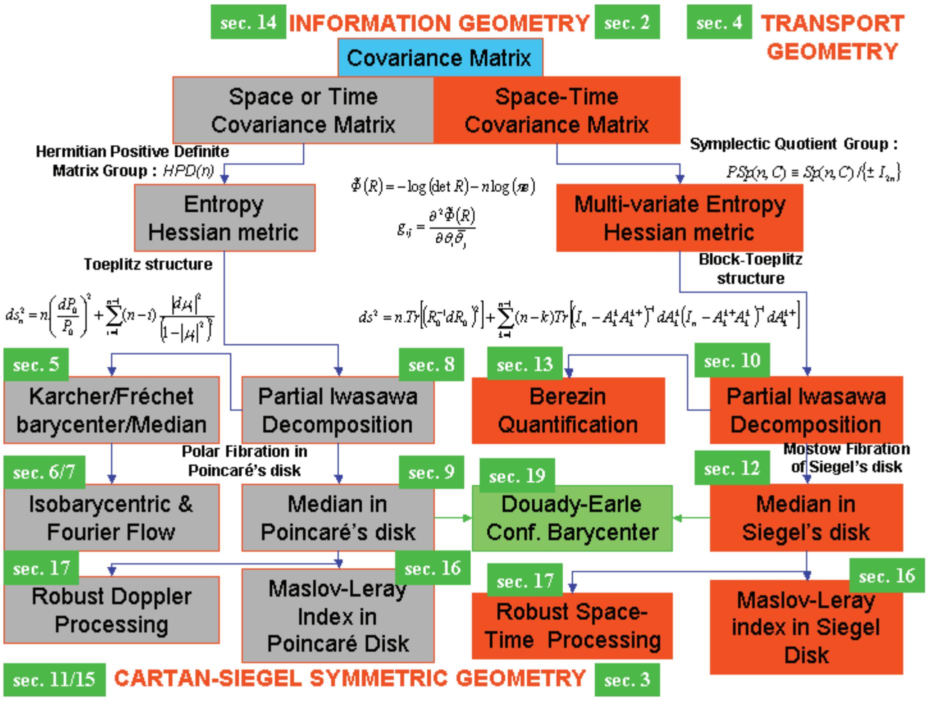

Fig. 1. paper structure with associated sections

2 Information Geometry Foundation

Foundation of Information geometry can be deduced from consideration on Kullback-Leibler Divergence. Kulback Divergence can be naturally introduced by combinatorial approach and stiling formula. Let multinomial Law of N elements spread on M levels $\{n_i\}$:

$$P_M(n_1,n_2,...,n_M / q_1,...,q_M) = N!\prod_{i=1}^{M}\frac{q_i^n}{n_i!} \quad \text{with } q_i \text{ priors, } \sum_{i=1}^{M} n_i = N \text{ and } p_i = \frac{n_i}{N} \tag{1}$$

Stirling formula gives $n! \approx n^n.e^{-n}.\sqrt{2\pi.n}$ when $n \to +\infty$. We could then observe that it converges to discret version of Kullback-Leibler:

$$\lim_{N\to+\infty}\frac{1}{N}\log[P_M] = \sum_{i=1}^{M} p_i \log\left[\frac{p_i}{q_i}\right] = K(p,q) \tag{2}$$

Based on variational approach, Donsker & Varadhan gave variational definition of Kullback divergence:

$$K(p,q) = Sup\left[E_p(\phi) - \log E_q(e^{\phi})\right] \tag{3}$$

$$\text{Consider}: \phi(\omega) = \log\left(\frac{p(\omega)}{q(\omega)}\right)$$

$$\Rightarrow E_p(\phi) - \log\left(E_q(e^{\phi})\right) = \sum_{\omega} p(\omega)\log\left(\frac{p(\omega)}{q(\omega)}\right) - \log\left[\sum_{\omega} q(\omega)\frac{p(\omega)}{q(\omega)}\right] = K(p,q) - \log(1) = K(p,q)$$

This proves that the supremum over all ϕ is no smaller than the divergence.

$$E_p(\phi) - \log\left(E_q(e^{\phi})\right) = E_p\left[\log\left(\frac{e^{\phi}}{E_q(e^{\phi})}\right)\right] = \sum_{\omega} p(\omega)\left(\log\left[\frac{q^{\phi}(\omega)}{q(\omega)}\right]\right)$$

$$\text{with} \quad q^{\phi}(\omega) = \frac{q(\omega)e^{\phi(\omega)}}{\sum_{\theta} q(\theta)e^{\phi(\theta)}} \Rightarrow K(p,q) - \left[E_p(\phi) - \log\left(E_q(e^{\phi})\right)\right] = \sum_{\omega} p(\omega)\left[\log\left(\frac{p(\omega)}{q^{\phi}(\omega)}\right)\right] \geq 0$$

using the divergence inequality.

In the same context, link with « Large Deviation Theory » & Fenchel-Legendre transform which gives that logarithm of generating function are dual to Kullback Divergence. This relation is given by:

$$\begin{aligned} &\log\left[\int e^{V(x)}q(x)dx\right] = \underset{p}{Sup}\left[\int V(x)p(x)dx - K(p,q)\right] \\ &\Leftrightarrow K(p,q) = \underset{V(.)}{Sup}\left[\int V(x)p(x)dx - \log\left[\int e^{V(x)}q(x)dx\right]\right] \\ &\Leftrightarrow K(p,q) = \underset{V(.)}{Sup}\left[E_p(V) - \log E_q\left[e^{V(x)}\right]\right] \end{aligned} \tag{4}$$

Chentsov was the first to introduce the Fisher information matrix as a Riemannian metric on the parameter space, considered as a differentiable manifold. Chentsov was led by decision theory when he considered a category whose objects are probability spaces and whose morphisms are Markov Kernels. Chentsov's great achievement was that up to a constant factor the Fisher information yields the only monotone family of Riemannian metrics on the class of finite probability simplexes. In parallel, Burbea and Rao have introduced a family of distance measures, based on the so-called α-order entropy metric, generalizing the Fisher Information metric that corresponds to the Shannon entropy. Such a choice of the matrix for the quadratic differential metric was shown to have attractive properties through the concepts of discrimination and divergence measures between probability distribution. As is well known from differential geometry, the Fisher information matrix is a covariant symmetric tensor of the second order, and hence, the associate metric is invariant under the admissible transformations of the parameters. The information geometry considers probability distributions as differentiable manifolds, while the random variables and their expectation appear as vectors and inner products in tangent spaces to these manifolds.

Chentsov has introduced a distance between parametric families of probability distributions $G_{\Theta} = \{p(./\theta) : \theta \in \Theta\}$ with Θ the space of parameters, by considering, to the first order, the difference between the log-density functions. Its variance defines a positive definite quadratic differential form based on the elements of the Fisher matrix and a Taylor expansion to the 2nd order of the Kullback divergence gives a Riemannian metric:

$$K(\theta,\tilde{\theta})\Big|_{\tilde{\theta}=\theta+d\theta} \cong K(\theta,\theta)+\left(\frac{\partial K(\theta,\tilde{\theta})}{\partial\tilde{\theta}}\right)^{+}_{\tilde{\theta}=\theta}\left(\tilde{\theta}-\theta\right)+\frac{1}{2}\left(\tilde{\theta}-\theta\right)^{+}\left(\frac{\partial^{2}K(\theta,\tilde{\theta})}{\partial\tilde{\theta}\partial\tilde{\theta}^{*}}\right)\left(\tilde{\theta}-\theta\right) \quad (5)$$

$$K\left[p(./\theta),p(./\theta+d\theta)\right]=\frac{1}{2!}\sum_{i,j}g_{ij}(\theta).d\theta_{i}.d\theta_{j}^{*}+O\left(\left|d\theta\right|^{3}\right) \quad (6)$$

$$I(\theta)=\left[g_{ij}(\theta)\right] \text{ and } \mathrm{g}_{\mathrm{ij}}(\theta)=E\left[\frac{\partial\log p(x/\theta)}{\partial\theta_{i}}\frac{\partial\log p(x/\theta)}{\partial\theta_{j}^{*}}\right]=-E\left[\frac{\partial^{2}\log p(x/\theta)}{\partial\theta_{i}\partial\theta_{j}^{*}}\right] \quad (7)$$

If we model Signal by complex circular multivariate Gaussian distribution of zero mean :

$$p(X_{n}/R_{n})=(\pi)^{-n}.\left|R_{n}\right|^{-1}.e^{-Tr\left[\hat{R}_{n}.R_{n}^{-1}\right]}$$
$$\text{with } \hat{R}_{n}=\left(X_{n}-m_{n}\right)\left(X_{n}-m_{n}\right)^{+} \text{ and } E\left[\hat{R}_{n}\right]=R_{n} \quad (8)$$

We conclude that Information metric can be written:

$$ds^{2}=d\theta^{+}I(\theta)d\theta=Tr\left[\left(R_{n}^{-1}dR_{n}\right)^{2}\right]=Tr\left[\left(d\ln R_{n}\right)^{2}\right]=\left\|R_{n}^{-1/2}dR_{n}R_{n}^{-1/2}\right\|_{F}^{2} \quad (9)$$

This metric is invariant under the action of the Linear matrix group $\left(GL_{n}(C),.\right)$:

$$R_{n}\rightarrow W_{n}.R_{n}.W_{n}^{+}\ ,\ \mathrm{W_{n}}\in GL_{n}(C) \quad (10)$$

An intrinsic metric could be also introduced by geometric study of HPD(n) Lie Group (Group of Hermitian Positive Definite Matrices nxn), that will be presented in the following as a particular case of Siegel upper half plane.
Both approaches provide the same metric:

$$ds^{2}=Tr\left[\left(R^{-1}dR\right)^{2}\right]=\left\|R^{-1/2}.dR.R^{-1/2}\right\|_{F}^{2} \quad (11)$$

This metric can be easily integrated and give the distance:

$$D^{2}\left(R_{1},R_{2}\right)=\left\|\log\left(R_{1}^{-1/2}.R_{2}.R_{1}^{-1/2}\right)\right\|^{2}=\sum_{k=1}^{n}\log^{2}\left(\lambda_{k}\right) \quad (12)$$

where $\{\lambda_{i}\}_{i=1}^{n}$ are the extended eigenvalues between R_1 and R_2:

$$\det\left(R_{1}^{-1/2}.R_{2}.R_{1}^{-1/2}-\lambda.I\right)=\det\left(R_{2}-\lambda R_{1}\right)=0\cdot \quad (13)$$

We can observe that geodesic projection on a sub-manifold M defined by $\Pi_{M}(P)=\arg\min_{S\in M} dist\left(P,S\right)$, is contractive :

$$dist\left(\Pi_{M}(P),\Pi_{M}(Q)\right)\leq dist\left(P,Q\right) \quad (14)$$

In the same way, if R and Q are two points on sub-manifold *M*, we can define distance from P to the geodesic [Q,R] by:

$$\sigma_s(t) = \sigma(s,t) = P^{1/2}\left[P^{-1/2}Q^{1/2}\left(Q^{-1/2}RQ^{-1/2}\right)^s Q^{1/2}P^{-1/2}\right]^t P^{1/2} \quad (15)$$

3 Cartan's symmetric spaces

For this differential geometry, we can also define the unique geodesic joining 2 matrices A and B. If $t \to \gamma(t)$ is the geodesic between *A* and *B*, where $t \in [0,1]$ is such that $d(A,\gamma(t)) = t.d(A,B)$, then the mean of *A* and *B* is the matrix $A \circ B = \gamma(1/2)$. The geodesic parameterized by the length as previously is given by:

$$\gamma(t) = A^{1/2}e^{t\log\left(A^{-1/2}BA^{-1/2}\right)}A^{1/2} = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2} \text{ with } 0 \le t \le 1 \quad (16)$$
$$\gamma(0) = A \;,\; \gamma(1) = B \text{ and } \gamma(1/2) = A \circ B$$

This space is a Riemannian Symmetric space of negative curvature, where for each couple of matrices *(A,B),* there exist a bijective isometry $G_{(A,B)}$ that verifies $G_{(A,B)}A = B$ and $G_{(A,B)}B = A$. This isometry has one fixed point *Z* barycenter of *(A,B),* given by $d\left(G_{(A,B)}X, X\right) = 2d(X,Z)$:

$$G_{(A,B)}X = (A \circ B)\mathrm{X}^{-1}(A \circ B) \text{ with } A \circ B = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}. \quad (17)$$

By analogy, we can compare with classical Euclidean space :

$$G_{(A,B)}X = (A \bullet B) - X + (A \bullet B) \text{ with } A \bullet B = \frac{A+B}{2} \text{ and } \|A-B\|_F \quad (18)$$

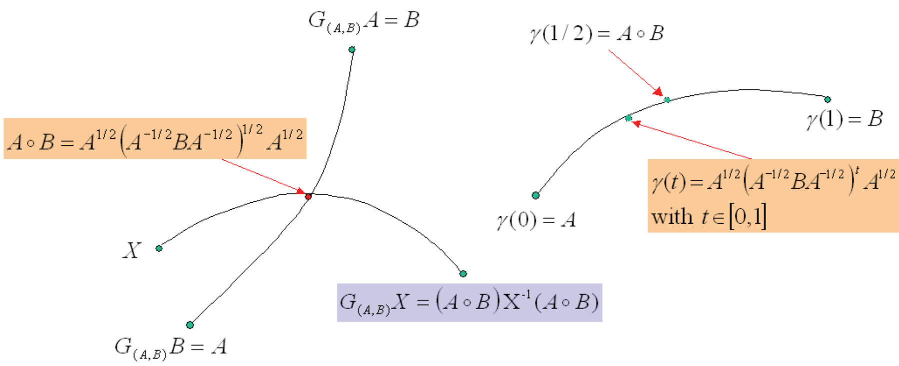

Fig. 2. Cartan-Berger Symmetric Space of Hermitian positive Definite Matrices

This space is a Cartan-Hadamard Manifold, complete, simply connected with negative sectional curvature Manifold. But this is also a Bruhat-Tits space where the distance verify the semi-parallelogram inequality:

$$\forall x_1, x_2 \quad \exists z \text{ such that } \forall x$$
$$d(x_1,x_2)^2 + 4d(x,z)^2 \leq 2d(x,x_1)^2 + 2d(x,x_2)^2 \quad \forall x \in X \tag{19}$$

Space of *HPD(n)* matrices is a metric space and a Riemannian Hermitian space of negative curvature, where we can then defined a Karcher/Fréchet Barycenter to compute mean of *N* covariance matrices.

4 Information Geometry versus Optimal Transport Theory

For a long time, main problem in statistics has been to define metrics in the space of probability measures or metrics between random variables. Recent progresses on this critical subject have been achieved in the framework of an other theory than Information Geometry, named “Optimal Transport Theory” by use of Wasserstein metrics for probability distribution and especially for multivariate Gaussian laws.

First work on probability metric was originally done by G. Monge in 1781 [91], considering the following metric in the distribution function space:

$$d(P,Q) = \underset{X,Y}{Inf}\, E\left[\left|X - Y\right|\right] \tag{20}$$

This was generalized by Appel: $d(P,Q) = \underset{X,Y}{Inf}\, E\left[f\left(\left|X - Y\right|\right)\right]$ (21)

where the infinum is taken over all joint distributions of pairs *(X,Y)* with fixed marginal distribution functions *P* and *Q*. Global survey of Optimal Transport theory is given in 2010 Field Medal Cédric Villani Book [96].

Then in 20th century, Maurice Fréchet [92,102] has proposed a metric on the space of probability distributions given by:

$$d^2(P,Q) = \underset{X,Y}{Inf}\, E\left[\left|X - Y\right|^2\right] \tag{22}$$

where the minimization is taken over all random variables *X* and *Y* having distribution *P* and *Q* respectively. Fréchet distance can be computed for n-dimensional distributions family that are closed with respect to linear transformations of the ramdom vector, and especially for Multivariate Gaussian laws. Others distance were studied by Paul Levy [93] & R. Fortét [94].

In the following table, we will compare geometry of complex circular multivariate gaussian law in two different frameworks : information geometry and geometry of Optimal transport theory. For Optimal Transport theory, I invite you to read papers of Takatsu [99], that has presented her work in Leon Brillouin Seminar in 2011 :
http://www.informationgeometry.org/Seminar/seminarBrillouin.html

Complex Gaussian Circular Law of zero mean : $p(Z/R)=\frac{1}{\pi^n \det(R)}e^{-Tr(\hat{R}R^{-1})}$ with $\begin{cases}\hat{R}=ZZ^+\\ E[\hat{R}]=R\end{cases}$	
Information Geometry	**Optimal Transport Theory**
Distance between random variables: $ds^2=d\theta^+ I(\theta)d\theta=Tr\left(\left(R^{-1}dR\right)^2\right)=\left\|R^{-1/2}dRR^{-1/2}\right\|_F^2$ $I(\theta)=\left[g_{i\bar{j}}\right]_{i,j}, g_{i\bar{j}}=-E\left[\frac{\partial^2 \log p(X/\theta)}{\partial\theta_i\partial\theta_j^*}\right]$	**Distance between random variables:** $d^2(P,Q)=\underset{X,Y}{Inf}\, E\left[\|X-Y\|^2\right]$
Metric: $g_P(R_X,R_Y)=Tr\left(P^{-1}.R_X.P^{-1}.R_Y\right)$	**Metric:** $g_P(R_X,R_Y)=Tr\left(R_X.P.R_Y\right)$
Tangent space and Exponential map: $\exp_X\left(\nu_X^Y,t\right)=X^{1/2}e^{t\log\left(X^{-1/2}YX^{-1/2}\right)}X^{1/2}$ $\begin{cases}\nu_X^Y=grad_X^Y(V)=-\exp_X^{-1}(V)\\ \nu_X^Y=-X^{1/2}\log\left(X^{-1/2}YX^{-1/2}\right)X^{1/2}\\ \exp_X\left(\nu_X^Y,t\right)=X^{1/2}e^{-t\left(X^{-1/2}.\nu_X^Y.X^{-1/2}\right)}X^{1/2}\end{cases}$	**Tangent space and Exponential map:** $\exp_{N(0,R_Y)}(t.X)=N\left(0,R_{R_Y,(t)}\right)$ $R_{R_Y,(t)}=\left((1-t)I_k+t.X\right)R_Y\left((1-t)I_k+t.X\right)$
Distance between covariance matrices: $d^2(R_X,R_Y)=\left\|\log\left(R_X^{-1/2}.R_Y.R_X^{-1/2}\right)\right\|_F^2$ $d^2(R_X,R_Y)=\sum_{k=1}^{n}\log^2(\lambda_k)$ $\det\left(R_X^{-1/2}.R_Y.R_X^{-1/2}-\lambda.I\right)=\det\left(R_Y-\lambda R_X\right)=0$	**Distance between covariance matrices:** $d^2(R_X,R_Y)=Tr[R_X]+Tr[R_Y]-2.Tr\left[\left(R_X^{1/2}R_YR_X^{1/2}\right)^{1/2}\right]$
Geodesic between covariance matrices: $\gamma(t)=R_X^{1/2}e^{t\log\left(R_X^{-1/2}R_YR_X^{-1/2}\right)}R_X^{1/2}$ $\gamma(t)=R_X^{1/2}\left(R_X^{-1/2}R_YR_X^{-1/2}\right)^t R_X^{1/2}$ $\gamma(0)=R_X$, $\gamma(1)=R_Y$ and $\gamma(1/2)=R_X\circ R_Y$	**Geodesic between covariance matrices:** $\gamma(t)=\left((1-t)I_k+t.D_{X,Y}\right)R_Y\left((1-t)I_k+t.D_{X,Y}\right)$ with $D_{X,Y}=R_X^{1/2}\left(R_X^{1/2}R_YR_X^{1/2}\right)^{-1/2}R_X^{1/2}=R_X\circ R_Y^{-1}$ $d\left(\gamma(s),\gamma(t)\right)\leq(t-s).d\left(\gamma(0),\gamma(1)\right)$
Cartan Symmetric Space: $G_{(R_X,R_Y)}R_Z=\left(R_X\circ R_Y\right)R_Z^{-1}(R_X\circ R_Y)$ with $R_X\circ R_Y=R_X^{1/2}\left(R_X^{-1/2}R_YR_X^{-1/2}\right)^{1/2}R_X^{1/2}$	**Space associated to Optimal transport:** $(x-m_x)^+R_X^{-1}(x-m_X)=\left[R_X^{-1/2}(x-m_X)\right]^+\left[R_X^{-1/2}(x-m_X)\right]$ $=(y-m_Y)^+R_Y^{-1}(y-m_Y)$ $x=\nabla\psi(y)=D_{X,Y}(y-m_Y)+m_X$ $\psi(v)=\frac{1}{2}(v-m_Y)^+D_{X,Y}(v-m_Y)+(m_X-v)$
Bruhat-Tits Space (semi-parallelogram inequality): $\forall x_1,x_2\ \ \exists z$ such that $\forall x\in X$ $d(x_1,x_2)^2+4d(x,z)^2\leq 2d(x,x_1)^2+2d(x,x_2)^2$	**Alexandrov Space:** $d(\alpha,\gamma(t))^2\geq$ $(1-t).d(\alpha,\gamma(0))^2+t.d(\alpha,\gamma(t))^2-t(1-t).d(\gamma(0),\gamma(1))^2$
Cartan-Hadamard Space: Complete, simply connected with negative sectional curvature Manifold	**Wasserstein Space:** Non-negative sectional curvature Manifold

Sectional curvature:	Sectional curvature:
$K=-Tr\left(\left(I-ZZ^{+}\right)^{-1}M\left(I-ZZ^{+}\right)^{-1}M^{+}\right)$ with $\left(I-ZZ^{+}\right)^{-1}=PP^{+}$ when $I-ZZ^{+}>0$ $K=-Tr\left(TT^{+}\right)<0$ where $T=P^{+}MP$	$K_{N(V)}(X,Y)=(3/4).Tr\left(\left([Y,X-S]\right)V\left([Y,X]-S\right)\right)^{T}$
Barycenter of N covariances matrices: $\sum_{k=1}^{N}\log\left(R^{-1/2}R_kR^{-1/2}\right)=0$ $R_{(n+1)}=R_{(n)}^{1/2}e^{\varepsilon\left(\sum_{k=1}^{N}\log\left(R_{(n)}^{-1/2}R_kR_{(n)}^{-1/2}\right)\right)}R_{(n)}^{1/2}$	**Barycenter of N covariances matrices:** $\sum_{k=1}^{N}\left(R^{1/2}R_kR^{1/2}\right)^{1/2}=R$ $K_{(n+1)}=\left(\sum_{k=1}^{N}\left(K_{(n)}K_k^2K_{(n)}\right)^{1/2}\right)^{1/2}$ with $K_i=R_i^{1/2}$

5 Cartan's center of Mass and Emery's Exponential Barycenter

We will then explained Robust Covariance Matrix Estimation based on Riemannian Geometry with center of mass. This center of mass exists only locally, except in special cases. Elie Cartan has proved back in the 1920's existence and uniqueness of center of mass for simply connected complete manifolds of negative curvature for any compact subset. This holds because a symmetric space of non-positive curvature is nothing but the quotient of a non-compact Lie group by one of its maximal compact subgroups. All these irreducible symmetric spaces have been classified by E. Cartan & M. Berger. In Euclidean space, the center of mass is defined for finite set of points $\{x_i\}_{i=1,\dots,M}$ by arithmetic mean: $x_{center}=\frac{1}{M}\sum_{i=1}^{M}x_i$ or by $x_{center}=\arg\underset{x}{Min}\sum_{i=1}^{M}\overrightarrow{xx_i}$. Appolonius of Perga has discovered that this point also minimizes the function of distances: $x_{center}=\arg\underset{x}{Min}\sum_{i=1}^{M}d^2(x,x_i)$. This extends to general Riemannian manifolds. Elie Cartan [3,4] has proved that the function:

$$f:m\in M\mapsto\frac{1}{2}\int_A d^2(m,a)da \tag{23}$$

is strictly convex (its restriction to any geodesic is strictly convex as a function of one variable), achieves a unique minimum at a point called the center of mass of A for the distribution da. Morever, this point is characterized by being the unique zero of the gradient vector field:

$$\nabla f=-\int_A \exp_m^{-1}(a)da \tag{24}$$

where $\exp(.)$ is the "exponential map" and $\exp_m^{-1}(a)$ is the tangent vector at $m\in M$ of the geodesic from m to a:

$$\exp_m^{-1}(a)\in T_mM \tag{25}$$

Herman Karcher has introduced a gradient flow intrinsic on the Manifold M that converges to the center of mass, called Karcher Barycenter :

$$m_{n+1} = \gamma_n(t) = \exp_{m_n}\left(-t.\nabla f(m_n)\right) \text{ with } \dot{\gamma}_n(0) = -\nabla f(m_n) \tag{26}$$

In the discrete case, the center of mass for finite set of points is given by:

$$m_{n+1} = \exp_{m_n}\left(t.\sum_{i=1}^{M}\exp_{m_n}^{-1}(x_i)\right) \tag{27}$$

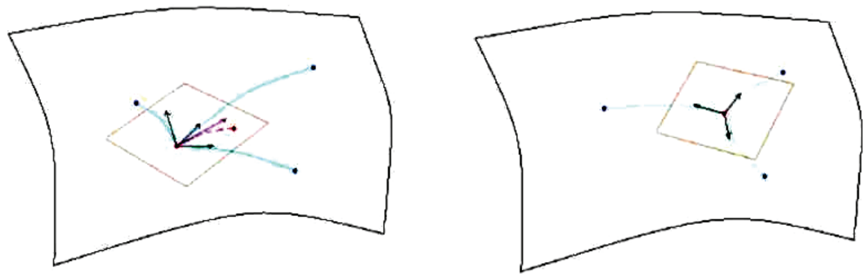

Fig. 3. Fréchet-Karcher Flow on Cartan-Hadamard Manifold

Maurice René Fréchet, inventor of Cramer-Rao bound in 1939 (published in Institut Henri Poincaré Lecture of Winter 1939 on statistics) [7], has also introduced the entire concept of Metric Spaces Geometry and functional theory on this space (any normed vector space is a metric space by defining $d(x,y) = \|y-x\|$ but not the contrary). On this base, Fréchet has then extended probability in abstract spaces.

A different point of view on center of mass or barycenter has been followed by Emery [21]. He has defined the expectation $E[x]$ as the set of all x such that:

$$\psi(x) \leq E[\psi(x)] \tag{28}$$

for all continuous convex functions. A related point of view was used by Doss and Herer who define $E[x]$ to be the set of all x such that:

$$d(z,x) \leq E[d(z,X)] \tag{29}$$

In this framework, expectation $b = E[g(x)]$ of an abstract probabilistic variable $g(x)$ where x lies on a manifold is introduced by Emery [21] as an exponential barycenter :

$$\int_M \exp_b^{-1}(g(x))P(dx) = 0 \tag{30}$$

In Classical Euclidean space, we recover classical definition of Expectation E[.] :

$$p,q \in R^n \Rightarrow \exp_p^{-1}(q) = q - p \Rightarrow E[g(x)] = \int_{R^n} g(x)P(dx) = \int_{R^n} g(x)p_X(x)dx \tag{31}$$

Statistics on manifolds is a critical aspect of different fields of applied mathematic. But Center of Mass is not useful for robust statistic. Replacing L^2 square geodesic distance by L^1 geodesic distance, we can extend this approach to estimate a _Median_ in metric space (called Fermat-Weber 's point in Physic). Fréchet studied Median statistic using Laplace's results :

$$m_{median} = \underset{m}{Min}\, E\left[\left|x - m\right|\right] \quad \text{compared to} \quad m_{mean} = \underset{m}{Min}\, E\left[\left|x - m\right|^2\right] \tag{32}$$

I would like to note that N.K. Sung and G. Stagenhaus [48] have defined an analogue Cramer-Rao-Fréchet-Darmois lower bound for median-unbiaised estimators. Based on a measure of dispersion proposed by Alamo, Stangenhaus and David, an analogue of the classical Cramér-Rao lower bound for median-unbiased estimators has been developed for absolutely continuous distributions with a single parameter, in which mean-unbiasedness, the Fisher information, and the variance are replaced by median-unbiasedness, the first absolute moment of the sample score, and the reciprocal of twice the median-unbiased estimator's density height evaluated at its median point. Could we conjecture that this result could be written by using the L^1 analogue J of L^2 Fisher information I:

$$E\left[\left|\theta - \hat{\theta}\right|\right] \geq \left[J(\theta)\right]^{-1} \quad \text{with} \quad J(\theta) = E\left[\left|\frac{\partial \log p(X/\theta)}{\partial \theta}\right|\right] \tag{33}$$

to compare with classical Cramer-Rao-Fréchet-Darmois Lower bound with Fisher Information Matrix I:

$$E\left[\left(\theta - \hat{\theta}\right)^2\right] \geq I(\theta)^{-1} \quad \text{with} \quad \left[I(\theta)\right]_{i,j} = E\left[\frac{\partial \log p(X/\theta)}{\partial \theta_i}\frac{\partial \log p(X/\theta)}{\partial \theta_j}\right] \tag{34}$$

Classically, in Euclidean space, Median point minimizes:

$$x_{median} = \arg \underset{x}{Min} \sum_{i=1}^{M} d(x, x_i) \quad \text{or equivalently} \quad x_{median} = \arg \underset{x}{Min} \sum_{i=1}^{M} \vec{xx_i} / \left\|\vec{xx_i}\right\| \tag{35}$$

This minimization could be extended for Riemannian manifold :

$$h : m \in M \mapsto \frac{1}{2}\int_A d(m,a)da \underset{Min}{\Rightarrow} \nabla h = -\int_A \frac{\exp_m^{-1}(a)}{\left\|\exp_m^{-1}(a)\right\|} da \tag{36}$$

We cannot directly extend the Karcher Flow to median computation in the discret case:

$$m_{n+1} = \exp_{m_n}\left(t.\sum_{k=1}^{M} \frac{\exp_{m_n}^{-1}(x_k)}{\left\|\exp_{m_n}^{-1}(x_k)\right\|}\right) \tag{37}$$

because $\left\|\exp_{m_n}^{-1}(x_k)\right\|$ could vanish if $m_n = x_k$.

We have investigated [85,86,87] the geometric median of a probability measure on a complete Riemannian manifold and prove the uniqueness. By regarding the Weiszfeld algorithm [108] as a sub-gradient procedure, we have introduced a sub-gradient algorithm to estimate the median and prove that this algorithm always converges :

$$m_{n+1} = \exp_{m_n}\left(t.\sum_{k\in G_{m_n}} \frac{\exp_{m_n}^{-1}(x_k)}{\left\|\exp_{m_n}^{-1}(x_k)\right\|} \right) \text{ with } G_{m_n} = \{k / x_k \neq m_n\} \tag{38}$$

Then, the median A of the N matrices B_k can be computed by sub-gradient Karcher flow :

$$A_{n+1} = A_n^{1/2} e^{\varepsilon\left(\sum_{k\in G_{A_n}} \frac{\log\left(A_n^{-1/2} B_k A_n^{-1/2}\right)}{\left\|\log\left(A_n^{-1/2} B_k A_n^{-1/2}\right)\right\|_F} \right)} A_n^{1/2} \text{ with } G_{A_n} = \{k / B_k \neq A_n\} \tag{39}$$

to compare with the mean A of the N matrices B_k that can be computed by gradient Karcher flow :

$$A_{n+1} = A_n^{1/2} e^{\varepsilon\left(\sum_{k=1}^{N} \log\left(A_n^{-1/2} B_k A_n^{-1/2}\right) \right)} A_n^{1/2} \tag{40}$$

In this last case, we have the property that for $\underset{n\to\infty}{Lim}\ A_n = A$:

$$\det(A) = \left(\prod_{k=1}^{N} \det(B_k) \right)^{1/N} \tag{41}$$

with consequence on Entropy:

$$-\log\left[\det(A)\right] = -\frac{1}{N}\sum_{k=1}^{N} \log\left[\det(B_k)\right] \Rightarrow \Phi(A) = \frac{1}{N}\sum_{k=1}^{N} \Phi(B_k) \tag{42}$$

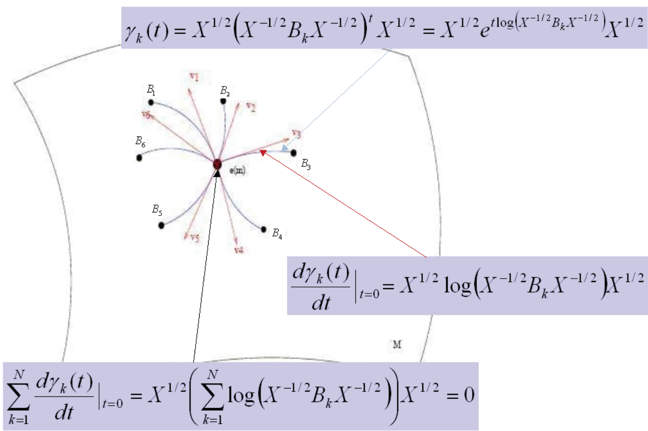

Fig. 4. Fréchet-Karcher Flow on space of Hermitian Positive Definite Matrices

6 Isobarycentric Flow

In this chapter, we propose an alternative flow to classical Frechet-Karcher Flow. To compute the barycenter in a Cartan-Hadamard Space, in this new approach that we call isobarycentric flow, the main idea is to define a flow that drives evolution of each of N points that we are looking for the barycenter. We first illustrate the isobarycentric flow in Euclidean space. For barycenter M of N points X_k in Euclidean space : $\sum_{k=1}^{N} \overrightarrow{MX_k} = 0$. We are looking for a flow $F(X_i)$ acting on each of N points X_i simultaneously at each step, so that the barycenter M is not modified such that $\sum_{k=1}^{N} \overrightarrow{MF}(X_k) = 0$.
This isobarycentric flow on X_i is given by Karcher flow induced by other *(N-1)* points $\{X_k\}_{k \neq i}$:

$$\overrightarrow{X_i F}(X_i) = \alpha \sum_{k \neq i}^{N} \overrightarrow{X_i X_k} \Rightarrow F(X_i) = X_i + \alpha \sum_{k \neq i}^{N} (X_k - X_i) \tag{43}$$

We can then easily proved that this flow doesn't change the barycenter:

$$\begin{aligned} &\sum_{i=1}^{N} (F(X_i) - M) = \sum_{i=1}^{N} (X_i - M) + \alpha \sum_{i=1}^{N} \sum_{k \neq i}^{N} \overrightarrow{X_i X_k} \\ &\Rightarrow \sum_{i=1}^{N} \overrightarrow{MF}(X_i) = \sum_{i=1}^{N} \overrightarrow{MX_i} + \alpha \sum_{i=1}^{N} \sum_{k \neq i}^{N} \overrightarrow{X_i X_k} = 0 \end{aligned} \tag{44}$$

Obviously, this isobarycentric flow could not be directly extended to Cartan-Hadamard Manifold and more especially to the space of Hermitian Positive Definite Matrices:

$$R_{i,(n+1)} = F\left[R_{i,(n)}\right] \tag{45}$$

that should verify:

$$\sum_{i=1}^{n} \log\left(R^{-1/2} F\left[R_{i,(n)}\right] R^{-1/2}\right) = 0 \quad \text{with} \quad \sum_{i=1}^{N} \log\left(R^{-1/2} R_{i,(n)} R^{-1/2}\right) = 0 \tag{46}$$

The isobarycentric flow could be approximated on *HPDD(n)* manifold when all matrices are closed to each other by using following approximation:

$$R_{i,n+1} = F\left[R_{i,(n)}\right] = R_{i,n}^{1/2} e^{\varepsilon\left(\sum_{k \neq i}^{N} \log\left(R_{i,n}^{-1/2} R_{k,n} R_{i,n}^{-1/2}\right)\right)} R_{i,n}^{1/2} \tag{47}$$

If we note $R_{i,n}^{-1/2} R_{k,n} R_{i,n}^{-1/2} = I + R_{i,n}^{-1/2}\left(R_{k,n} - R_{i,n}\right) R_{i,n}^{-1/2} = I + T_{i,k}$ (48)

By using approximation of *log(.)* and *exp(.):*

$$\log\left(I+G\right)=G-\frac{1}{2}G^{2}+\frac{1}{3}G^{3}-\ldots \quad \text{and} \quad \exp\left(H\right)=I+H+\frac{1}{2}H^{2}+\frac{1}{3!}H^{3}+\ldots$$

We are then in the same case than in Euclidean space:

$$R_{i,n+1}=R_{i,n}^{1/2}\left[I+\varepsilon\left(\sum_{k\neq i}^{N}R_{i,n}^{-1/2}\left(R_{k,n}-R_{i,n}\right)R_{i,n}^{-1/2}\right)\right]R_{i,n}^{1/2} \tag{49}$$

$$R_{i,n+1}=R_{i,n}+\varepsilon\sum_{k\neq i}^{N}\left(R_{k,n}-R_{i,n}\right)$$

In this case, barycentric flow convergence is obvious. To study convergence of barycentric flow when all matrices are not closed to each other, consideration on curvature should be studied in the more general framework of potential theory.

7 Fourier Heat Equation Flow on 1D graph of HPD(n) matrices

We can replace Median computation by anisotropic diffusion. In normed vector space in *1D*, if we note $\hat{u}_n=\left(u_{n+1}+u_{n-1}\right)/2$, Fourier diffusion Equation is given by :

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}\Rightarrow u_{n,t+1}=u_{n,t}+\frac{2\nabla t}{\nabla x^2}\left[\hat{u}_{n,t}-u_{n,t}\right]=(1-\rho).u_{n,t}+\rho.\hat{u}_{n,t}=u_{n,t}\circ_{\rho}\hat{u}_{n,t} \tag{50}$$

By analogy, we can define diffusion equation on a *1D* graph of *HPD(n)* by:

$$A_{n,t+1}=A_{n,t}^{1/2}e^{\frac{2\nabla t}{\nabla x^2}\log\left(A_{n,t}^{-1/2}\hat{A}_{n,t}A_{n,t}^{-1/2}\right)}A_{n,t}^{1/2}=A_{n,t}^{1/2}\left(A_{n,t}^{-1/2}\hat{A}_{n,t}A_{n,t}^{-1/2}\right)^{\rho}A_{n,t}^{1/2}=A_{n,t}\circ_{\rho}\hat{A}_{n,t} \tag{51}$$

$$\text{with } \rho=\frac{2\nabla \mathrm{t}}{\nabla \mathrm{x}^2} \text{ and } \hat{A}_{n,t}=A_{n+1,t}^{1/2}\left(A_{n+1,t}^{-1/2}A_{n-1,t}A_{n+1,t}^{-1/2}\right)^{1/2}A_{n+1,t}^{1/2}$$

Obiously, we can introduce anisotropy by making adaptive the parameter ρ.

From each Time covariance matrix, we can compute Doppler Spectrum. In the following exemple, we give image with range on X axis and Doppler frequency on Y axis. Fourier heat Diffusion is applied on covariance matrices and then, we draw associated Doppler spectrum of results :

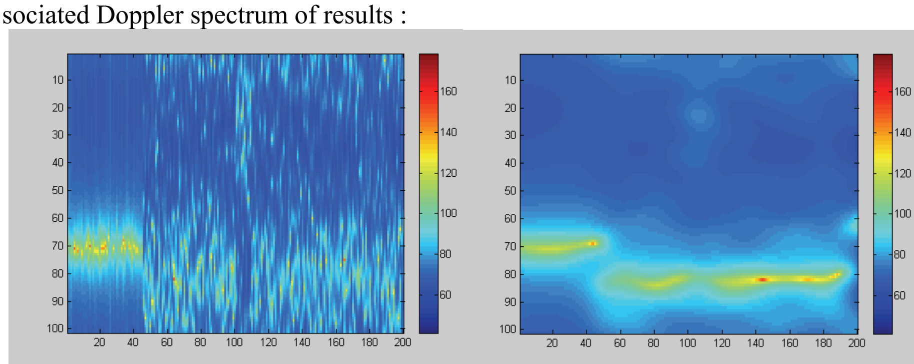

Fig. 5. Fourier Heat Equation on a 1D graph of covariance matrices: isotropic diffusion

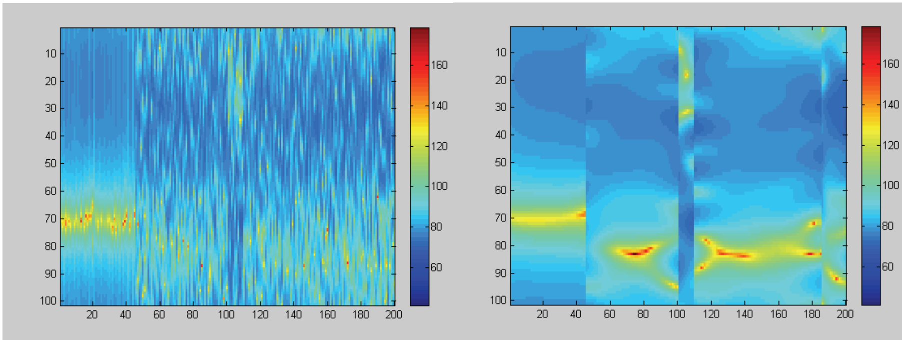

Fig. 6. Fourier Heat Equation on a *1D* graph of covariance matrices: anisotropic diffusion

8 Flow preserving Covariance matrix Toeplitz structure

All previous approaches don't take into account Toeplitz structure of covariance matrices, in case where the signal is stationary $E\left[z_n z_{n-k}^*\right] = r_{n,n-k} = r_k$ and $r_{-k} = r_k^*$. To take into account this constraint, we have used Partial Iwasawa decomposition, that is equivalent for time or space signal to complex autoregressive model decomposition (link with Gohberg-Semencul inverse covariance matrix computation):

$$\Omega_n = \left(\alpha_n.R_n\right)^{-1} = W_n.W_n^+ = \left(1-\left|\mu_n\right|^2\right)\begin{bmatrix} 1 & A_{n-1}^+ \\ A_{n-1} & \Omega_{n-1} + A_{n-1}.A_{n-1}^+ \end{bmatrix} \tag{52}$$

$$\text{with } W_n = \sqrt{1-\left|\mu_n\right|^2}\begin{bmatrix} 1 & 0 \\ A_{n-1} & \Omega_{n-1}^{1/2} \end{bmatrix} \text{ and } \Omega_{n-1} = \Omega_{n-1}^{1/2}.\Omega_{n-1}^{1/2+} \tag{53}$$

with $A_n = \left[a_1^n \quad \cdots \quad a_n^n\right]^T$ and $\mu_n = a_n^n$, respectively complex autoregressive vector and reflection coefficient (see section 17.1 for more details) .
In the framework of Information Geometry (66), we consider Information metric defined as Kählerian metric where the Kähler potential is given by the Entropy of the process $\widetilde{\Phi}(R_n)$ (called Ruppeiner metric in Physics). We describe link with Rao metric in section 14 :

$$\widetilde{\Phi}(R_n) = \log(\det R_n^{-1}) - n\log(\pi.e) = \sum_{k=1}^{n-1}(n-k).\ln\left[1-\left|\mu_k\right|^2\right] + n.\ln\left[\pi.e.P_0\right] \tag{54}$$

Information metric is given by hessian of Entropy :

$$g_{ij} \equiv \frac{\partial^2 \widetilde{\Phi}}{\partial \theta_i^{(n)} \partial \theta_j^{(n)*}} \text{ where } \theta^{(n)} = \left[P_0 \quad \mu_1 \quad \cdots \quad \mu_{n-1}\right]^T \tag{55}$$

with $\{\mu_k\}_{k=1}^{n-1}$ regularized Burg's reflection coefficient and P_0 mean Power. Kählerian metric is finally :

$$ds_n^2 = d\theta^{(n)+}\left[g_{ij}\right]d\theta^{(n)} = n.\left(\frac{dP_0}{P_0}\right)^2 + \sum_{i=1}^{n-1}(n-i)\frac{|d\mu_i|^2}{\left(1-|\mu_i|^2\right)^2} \tag{56}$$

This is linked with general result on Bergman Manifold and theory of homogeneous complex manifolds. For complex manifold, where:

$$\Omega = i^n K(z) dz_1 \wedge \ldots \wedge dz_n \wedge dz_1^* \wedge \ldots \wedge dz_n^* \tag{57}$$

is the given exterior differential form, the Hermitian differential form:

$$ds^2 = \sum_{i,j}\frac{\partial^2 \log K(z)}{\partial z_i \partial z_j^*} dz_i dz_j^* \tag{58}$$

is independant of the choice of the coordinate system. Here, parameterization is conformal:

$$ds^2 = \frac{\partial^2\left[-n.\log P_0\right]}{\partial^2 P_0} dP_0 dP_0 + \sum_{i=1}^{n-1}\frac{\partial^2\left[-(n-i).\log\left(1-|\mu_i|^2\right)\right]}{\partial\mu_i\partial\mu_i^*} d\mu_i d\mu_i^* \quad \text{where} \quad \tilde{\Phi} = \log K \tag{59}$$

9 Median by Fibration of Conformal Poincaré's unit disk

For Median autoregressive model, Karcher flow could be very simple. For P_0, we use classical median on real value. For $\{\mu_k\}_{k=1}^{n-1}$, we use homeomorphism of Poincaré's unit disk $\mu_{k,n+1} = \frac{\mu_{k,n} - w_n}{1 - \mu_{k,n}.w_n^*}$, to fixe the point under action of karcher flow at the origin where all geodesics are radials and space is quasi-euclidean. Equation of Dual Karcher Flow, in this new coordinate system, is then given by polar decomposition:

$$w_n = \gamma_n \sum_{k\in G_0}\frac{\mu_{k,n}}{\left|\mu_{k,n}\right|} \quad \text{with} \quad G_0 = \left\{k / \left|\mu_{k,n}\right| \neq 0\right\}. \tag{60}$$

Median is deduced taking into account each step w_n:

$$\mu_{median,n+1} = \frac{\mu_{median,n} + w_n}{1 + \mu_{median,n} w_n^*} \tag{61}$$

In the following this polar decomposition will be replace by Mostow decomposition in Siegel Disk. This fibration is not available for Klein model of unit disk [111].

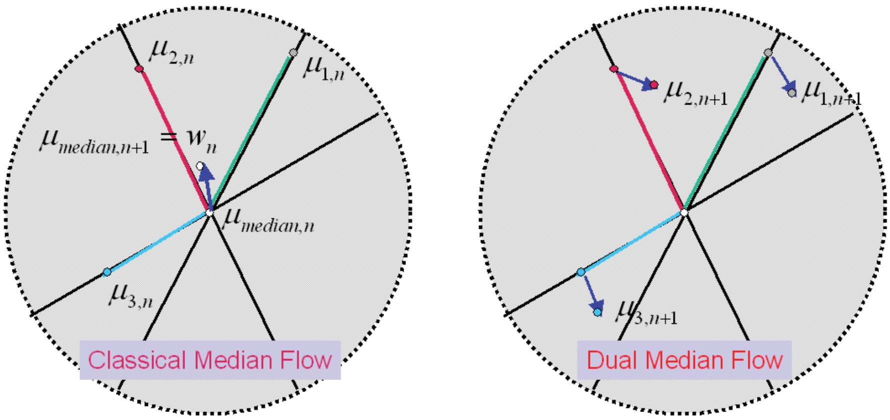

Fig. 7. : Modified Karcher Flow in Poincaré disk by homeomorphism

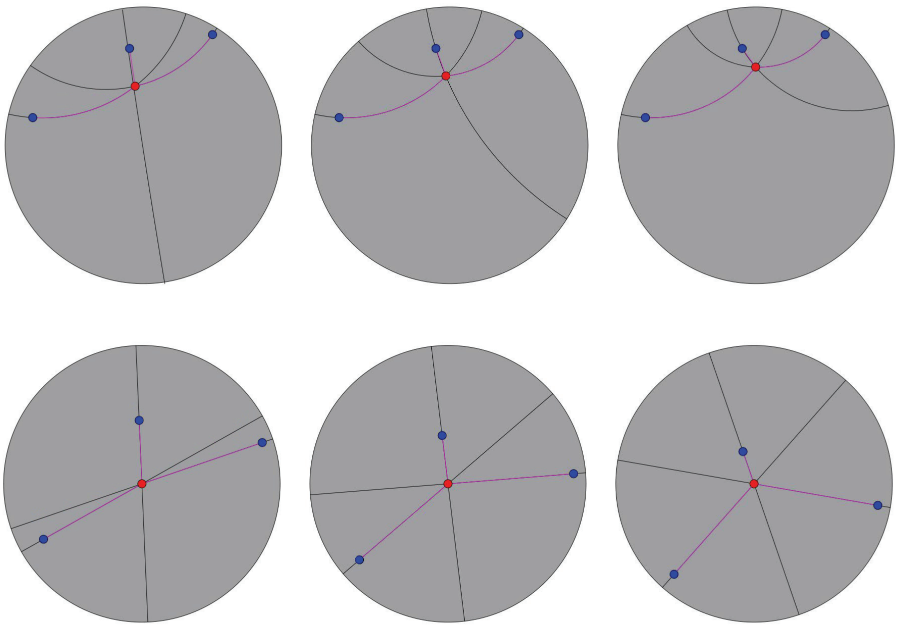

Fig. 8. Classical Karcher flow (on the top), Dual Karcher Flow (on the bottom)

As hyperbolic Poincaré's Model is conform model, angles are preserved and median will be characterized by equality between angles of tangent vectors.

More recently, Marc Arnaudon [84] has proposed a stochastic Karcher flow that converges alsmost surely to p-mean. For p=1, this stochastic flow is given by :

$$m_{n+1} = \exp_{m_n}\left(t_n . \frac{\exp^{-1}_{m_n}\left(x_{rand(n)}\right)}{\left\| \exp^{-1}_{m_n}(x_{rand(n)}) \right\|} \right) \qquad (62)$$

where for each iteration n, index of one point is selected randomly $x_{rand(n)}$. Then, m_{n+1}, driven by the flow, moves along the geodesic between m_n and $x_{rand(n)}$. Finally, in Poincaré unit disk, index $rand(n)$ is selected randomly in set $G_0 = \left\{ k / \left| \mu_{k,n} \right| \neq 0 \right\}$,at each step, displacement is given by:

$$w_n = \gamma_n . (\mu_{rand(n),n} / \left| \mu_{rand(n),n} \right|) \cdot \qquad (63)$$

Extension for Riemannian 1-Center has been applied in [112].

10 Geometry of Space-Time Covariance matrix

We will extend previous works developed for time or space covariance matrix, to space-time covariance matrix, structured as Toeplitz-block-Toeplitz Hermitian Positive Definite matrices. The problem will be considered for complex multi-channel or multi-variate data processing in the framework of Information Geometry.

Based on generalization of Trench Algorithm, if we consider Toeplitz-block-Toeplitz Hermitian Positive Definite matrix [89] :

$$R_{p,n+1} = \begin{bmatrix} R_0 & R_1 & \cdots & R_n \\ R_1^+ & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_n^+ & \cdots & R_1^+ & R_0 \end{bmatrix} = \begin{bmatrix} R_{p,n} & \widetilde{R}_n \\ \widetilde{R}_n^+ & R_0 \end{bmatrix} \qquad (64)$$

with $\widetilde{R}_n = V \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix}^*$ where $V = \begin{bmatrix} 0 & \cdots & 0 & J_p \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & 0 \\ 0 & J_p & \cdot^{\cdot^{\cdot}} & \vdots \\ J_p & 0 & \cdots & 0 \end{bmatrix}$ and J_p : anti - diagonale matrix (65)

We can apply the well-known inversion rule for a partitioned matrix, associated with adapted parameterization by mean of Block-structured Partial Iwasawa Decomposition, deduced from Burg-like generalized forward and backward linear prediction.

From Burg-like parameterization [90], we can deduced this inversion of Toeplitz-Block-Toeplitz matrix :

$$R^{-1}_{p,n+1} = \begin{bmatrix} \alpha_n & \alpha_n . \widehat{A}_n^+ \\ \alpha_n . \widehat{A}_n & R^{-1}_{p,n} + \alpha_n . \widehat{A}_n . \widehat{A}_n^+ \end{bmatrix} \quad \text{and} \quad R_{p,n+1} = \begin{bmatrix} \alpha_n^{-1} + \widehat{A}_n^+ . R_{p,n} . \widehat{A}_n & -\widehat{A}_n^+ . R_{p,n} \\ -R_{p,n} . \widehat{A}_n & R_{p,n} \end{bmatrix} \qquad (66)$$

$$\text{with}\quad \alpha_n^{-1} = \left[1 - A_n^n A_n^{n+}\right]\alpha_{n-1}^{-1},\ \ \alpha_0^{-1} = R_0 \quad \text{and} \quad \widehat{A}_n = \begin{bmatrix} A_1^1 \\ \vdots \\ A_n^n \end{bmatrix} = \begin{bmatrix} \widehat{A}_{n-1} \\ 0_p \end{bmatrix} + A_n^n . \begin{bmatrix} J_p A_{n-1}^{n-1*} J_p \\ \vdots \\ J_p A_1^{n-1*} J_p \\ I_p \end{bmatrix} \tag{67}$$

Where we have the following Burg-like generalized forward and backward linear prediction :

$$\begin{cases} \varepsilon_{n+1}^f(k) = \sum_{l=0}^{n+1} A_l^{n+1}(k) Z(k-l) = \varepsilon_n^f(k) + A_{n+1}^{n+1}\varepsilon_n^b(k-1) \\ \varepsilon_{n+1}^b(k) = \sum_{l=0}^{n} J A_l^{n+1}(k)^* J Z(k-n+l) = \varepsilon_n^b(k-1) + J A_{n+1}^{n+1*} J \varepsilon_n^f(k) \end{cases}$$

$$\text{with}\quad \begin{cases} \varepsilon_0^f(k) = \varepsilon_0^b(k) = Z(k) \\ A_0^{n+1} = I_p \end{cases} \tag{68}$$

$$A_{n+1}^{n+1} = -2\left[\sum_{k=1}^{N+n} \varepsilon_n^f(k)\varepsilon_n^b(k-1)^+\right]\left[\sum_{k=1}^{N+n} \varepsilon_n^f(k)\varepsilon_n^f(k)^+ + \sum_{k=1}^{N+n} \varepsilon_n^b(k)\varepsilon_n^b(k)^+\right]^{-1}$$

Using Schwarz's inequality, it is easily to prove that A_{n+1}^{n+1} Burg-Like reflection coefficient matrix lies in Siegel Disk $A_{n+1}^{n+1} \in SD_p$.

11 Cartan-Siegel Homogeneous Domains : Siegel Disk

To solve median computation of Toeplitz-Block-Toeplitz matrices, Karcher-Fréchet Flow has to be extended in Siegel Disk. Siegel Disk has been introduced by Carl Ludwig Siegel [17,18] through Symplectic Group $Sp_{2n}R$ that is one possible generalization of the group $SL_2R = Sp_2R$ (group of invertible matrices with determinant 1) to higher dimensions. This generalization goes further; since they act on a symmetric homogeneous space, the Siegel upper half plane, and this action has quite a few similarities with the action of SL_2R on the Poincaré's hyperbolic plane. Let F be either the real or the complex field, the Symplectic Group is the group of all matrices $M \in GL_{2n}F$ satisfying :

$$Sp(n,F) \equiv \left\{M \in GL(2n,F) / M^T J M = J\right\}, J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \in SL(2n,R) \tag{69}$$

$$\text{or}\ \ M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp(n,F) \Leftrightarrow A^T C \text{ and } B^T D \quad \text{symmetric and} \quad A^T D - C^T B = I_n \tag{70}$$

The Siegel upper half plane is the set of all complex symmetric *nxn* matrices with positive definite imaginary part:

$$SH_n = \left\{Z = X + iY \in Sym(n,C) / \operatorname{Im}(Z) = Y > 0\right\} \tag{71}$$

The action of the Symplectic Group on the Siegel upper half plane is transitive. The group $PSp(n,R) \equiv Sp(n,R)/\{\pm I_{2n}\}$ is group of SH_n biholomorphisms via generalized Möbius transformations:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \Rightarrow M(Z) = (AZ+B)(CZ+D)^{-1} \tag{72}$$

$PSp(n,R)$ acts as a sub-group of isometries. Siegel has proved that Symplectic transformations are isometries for the Siegel metric in SH_n. It can be defined on SH_n using the distance element at the point $Z = X + iY$, as defined by:

$$ds^2_{Siegel} = Tr\left(Y^{-1}(dZ)Y^{-1}(dZ^+)\right) \text{ with } Z = X + iY \tag{73}$$

with associated volume form : $\Omega = Tr\left(Y^{-1}dZ \wedge Y^{-1}dZ^+\right)$
C.L. Siegel has proved that distance in Siegel Upper-Half Plane is given by :

$$d^2_{Siegel}(Z_1,Z_2) = \left(\sum_{k=1}^{n} \log^2\left(\frac{1+\sqrt{r_k}}{1-\sqrt{r_k}}\right)\right) \text{ with } Z_1, Z_2 \in SH_n \tag{74}$$

and r_k eigenvalues of the cross-ratio :

$$R(Z_1,Z_2) = (Z_1 - Z_2)(Z_1 - Z_2^+)^{-1}(Z_1^+ - Z_2^+)(Z_1^+ - Z_2)^{-1}. \tag{75}$$

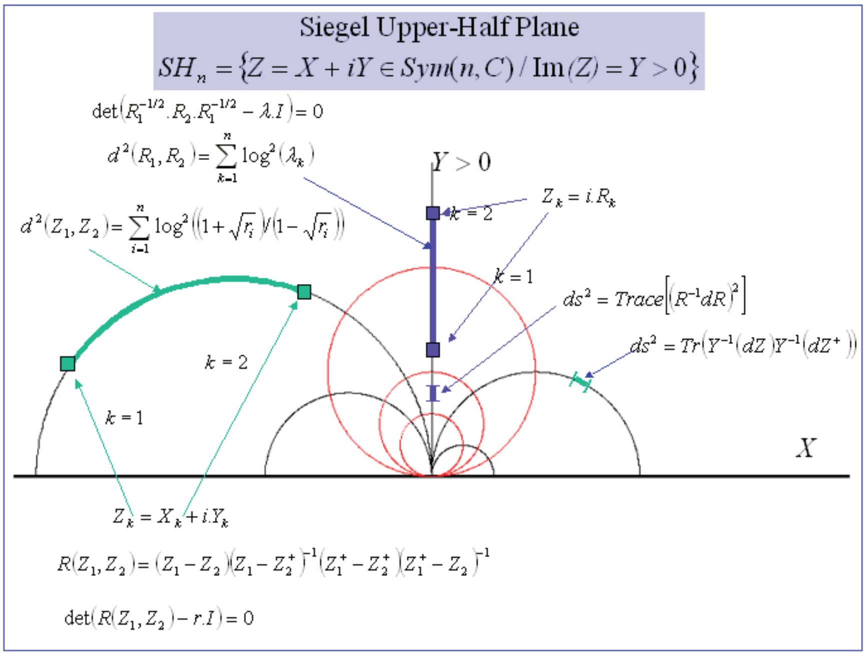

Fig. 9. Geometry of Siegel Upper Half-Plane

This is deduced from the 2nd derivative of $Z \to R(Z_1, Z)$ in $Z_1 = Z$ given by :

$$D^2 R = 2dZ\left(Z - Z^+\right)^{-1} dZ^+ (Z^+ - Z)^{-1} = (1/2).dZY^{-1}dZ^+Y^{-1} \tag{76}$$

and $ds^2 = Tr\left(Y^{-1}dZY^{-1}dZ^+\right) = 2.Tr\left(D^2R\right)$ (77)

In parallel, in China in 1945, Hua Lookeng has given the equations of geodesic in Siegel upper-half plane [8]:

$$\frac{d^2Z}{ds^2} + i\frac{dZ}{ds}Y^{-1}\frac{dZ}{ds} = 0 \tag{78}$$

Using generalized Cayley transform $W = (Z - iI_n)(Z + iI_n)^{-1}$, Siegel Upper-half Plane SH_n is transformed in unit Siegel disk $SD_n = \{W / WW^+ < I_n\}$ where the metric in Siegel Disk is given by :

$$ds^2 = Tr\left[\left(I_n - WW^+\right)^{-1} dW \left(I_n - W^+W\right)^{-1} dW^+\right] \tag{79}$$

Contour of Siegel Disk is called its Shilov boundary $\partial SD_n = \{W / WW^+ - I_n = 0_n\}$. We can also defined horosphere. Let $U \in \partial SD_n$ and $k \in R_*^+$, the following set is called horosphere in siegel disk :

$$H(k,U) = \left\{Z / 0 < k\left(I - Z^+Z\right) - \left(I - Z^+U\right)\left(I - U^+Z\right)\right\} = \left\{Z / \left\| Z - \frac{1}{k+1}U \right\|\right\} < \frac{k}{k+1} \tag{80}$$

Hua Lookeng [8] has proved that the previous positive definite quadratic differential is invariant under the group of automorphisms of the Siegel Disk.

Considering $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ such that $M^* \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} M = \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix}$:

$$\begin{aligned} & V = M(W) = (AZ + B)(CZ + D)^{-1} \Rightarrow \\ & \left(I_n - VV^+\right)^{-1} dV \left(I_n - V^+V\right)^{-1} dV^+ = \\ & \left(BW^+ + A\right)\left(I_n - WW^+\right)^{-1} dW \left(I_n - W^+W\right)^{-1} dW^+ \left(BW^+ + A\right)^{-1} \\ & \Rightarrow ds_V^2 = ds_W^2 \end{aligned} \tag{81}$$

Complementary, Hua Lookeng [20] has also proved that, let V, W be complex-valued matrices, if $I - VV^+ > 0$ and $I - WW^+ > 0$, then the following identity holds :

$$\det\left(I - VV^+\right)\det\left(I - WW^+\right) \leq \left|\det\left(I - VW^+\right)\right|^2 \tag{82}$$

$\det\left(I - A^+A\right)\det\left(I - B^+B\right) \leq \left|\det\left(I - A^+B\right)\right|^2$ is based on Hua's matrix identity :

$$\left(I - B^+B\right) + (A - B)^+\left(I - AA^+\right)^{-1}(A - B) = \left(I - B^*A\right)\left(I - A^*A\right)^{-1}\left(I - A^+B\right) \tag{83}$$

using the intermediate equalities:

$$\left(I - B^{+}A\right)\left(I - A^{+}A\right)^{-1}\left(I - A^{+}B\right) - \left(I - B^{+}B\right) = \left(B - A\right)^{+}\left(I - AA^{+}\right)\left(B - A\right) \tag{84}$$

$$\left(I - A^{+}A\right)^{-1} = I + A^{+}\left(I - AA^{+}\right)^{-1}A \quad \text{and} \quad \left(I - A^{+}A\right)^{-1}A^{+} = A^{+}\left(I - AA^{+}\right)^{-1} \tag{85}$$

Same kind of inequality is true for the trace :

$$Tr\left(I - A^{+}A\right)Tr\left(I - B^{+}B\right) \leq \left|Tr\left(I - A^{+}B\right)\right|^{2} \tag{86}$$

To go further to study Siegel Disk, we need now to define what are the automorphisms of Siegel Disk SD_n . They are all defined by:

$$\forall \Psi \in Aut(SD_n), \exists U \in U(n, C) / \Psi(Z) = U\Phi_{Z_0}(Z)U^{t} \tag{87}$$

$$\text{with } \Sigma = \Phi_{Z_0}(Z) = \left(I - Z_0 Z_0^{+}\right)^{-1/2}\left(Z - Z_0\right)\left(I - Z_0^{+}Z\right)^{-1}\left(I - Z_0^{+}Z_0\right)^{1/2} \tag{88}$$

and its inverse :

$$\begin{aligned} & G = \left(I - Z_0 Z_0^{+}\right)^{1/2}\Sigma\left(I - Z_0^{+}Z_0\right)^{-1/2} = \left(Z - Z_0\right)\left(I - Z_0^{+}Z\right)^{-1} \\ & \Rightarrow \begin{cases} Z = \Phi_{Z_0}^{-1}(\Sigma) = (GZ_0^{+} + I)^{-1}\left(G + Z_0\right) \\ \text{with } G = \left(I - Z_0 Z_0^{+}\right)^{1/2}\Sigma\left(I - Z_0^{+}Z_0\right)^{-1/2} \end{cases} \end{aligned} \tag{89}$$

By analogy with Poincaré's unit Disk, C.L. Siegel has deduced geodesic distance in SD_n [109] :

$$\forall Z, W \in SD_n, d(Z, W) = \frac{1}{2}\log\left(\frac{1 + \left\|\Phi_Z(W)\right\|}{1 + \left\|\Phi_Z(W)\right\|}\right) \tag{90}$$

12 Mostow/Berger's Fibration of Siegel Disk

As in previous case, Information metric will be introduced as a Kähler potential defined by Hessian of multi-channel/Multi-variate entropy $\widetilde{\Phi}\left(R_{p,n+1}\right)$, from (64) :

$$\widetilde{\Phi}\left(R_{p,n}\right) = -\log\left(\det R_{p,n}\right) + cste = -Tr\left(\log R_{p,n}\right)\mu + cste \Rightarrow g_{i\bar{j}} = Hess\left[\widetilde{\phi}\left(R_{p,n}\right)\right] \tag{91}$$

Using partitioned matrix structure of Toeplitz-Block-Toeplitz matrix $R_{p,n+1}$, recursively parametrized by Burg-Like reflection coefficients matrix $\left\{A_k^k\right\}_{k=1}^{n-1}$ with $A_k^k \in SD_n$, we can give a new expression of the Multi-variate entropy from (66)(67):

$$\widetilde{\Phi}(R_{p,n}) = -\sum_{k=1}^{n-1}(n-k).\log\det\left[1 - A_k^k A_k^{k+}\right] - n.\log\left[\pi.e.\det R_0\right] \tag{92}$$

Paul Malliavin [52] has proved that this form is a Kähler Potential of an invariant Kähler metric that is given by :

$$ds^2 = n.Tr\left[\left(R_0^{-1}dR_0\right)^2\right] + \sum_{k=1}^{n-1}(n-k)Tr\left[\left(I_n - A_k^k A_k^{k+}\right)^{-1} dA_k^k \left(I_n - A_k^{k+} A_k^k\right)^{-1} dA_k^{k+}\right] \quad (93)$$

Median Matrix Estimation of N Radar Space-Time sample data Covariance matrices : Study of Karcher/Frechet Barycenter and median in Siegel Disk based on Mostow Decomposition

As we have defined a metric space, we can extend Karcher/Frechet flow in Unit Siegel Disk to compute the Median of N Toeplitz-Block-Toeplitz Hermitian Positive Definite matrices. These matrices are parametrized by Burg-Like generalized Reflection coefficient [90] matrices $\left\{A_k^k\right\}_{k=1}^{n-1}$ with $A_k^k \in SD_n$ and Karcher/Frechet Flow in Siegel Disk will be solved by analogy of our scheme used in Poincaré unit Disk, by mean of Mostow Decomposition Theorem. Mostow's decomposition theorem is a refinement of the polar decomposition [35,36,55,56,57,58]. This theorem is related to geometric properties of the non-positively curved space of postive definite Hermitian matrices and to a characterisation of its geodesic subspaces.

> ***Mostow Theorem:***
> Every matrix M of $GL(n,C)$ can be decomposed in:
>
> $$M = Ue^{iA}e^{S} \quad (94)$$
>
> where U is unitary, A is real antisymmetric and S is real symmetric

Mostow Theorem is deduced from following Lemma and Corollary :

> ***Lemma:***
> Let A and B two positive definite Hermitian matrices, there exist a unique positive definite Hermitian matrix X such that :
>
> $$XAX = B \quad (95)$$

$A^{1/2}$ is unique Hermitian positive definite square root of A :

$$\begin{aligned} XAX = B &\Rightarrow A^{1/2}X\left(A^{1/2}A^{1/2}\right)XA^{1/2} = A^{1/2}BA^{1/2} \Rightarrow \left(A^{1/2}XA^{1/2}\right)^2 = A^{1/2}BA^{1/2} \\ &\Rightarrow A^{1/2}XA^{1/2} = \left(A^{1/2}BA^{1/2}\right)^{1/2} \Rightarrow X = A^{-1/2}\left(A^{1/2}BA^{1/2}\right)^{1/2}A^{-1/2} \end{aligned} \quad (96)$$

We can observe that X is geodesic center of A^{-1} and B for symmetric space of Hermitian positive definite matrices.

> ***Corollary:***
> if M is Hermitian Positive Definite, there exist a unique real symmetric matrix S such that :
>
> $$M^* = e^{S}M^{-1}e^{S} \quad (97)$$

M Positive Definite Hermitian Matrix, M^* and M^{-1} with same property. From previous Lemma, there exist a unique hermitian positive definite matrix X such that:

$$M^* = XM^{-1}X \quad (98)$$

Exponential providing an homeomorphism between symmetric and positive definite symmetric spaces, it can be proved proof that X is positive definite

$$M = \left(M^*\right)^* = X^* M^{*-1} X^* \Rightarrow M^{*-1} = X^{*-1} M X^{*-1} \Rightarrow M^* = X^* M^{-1} X^* \tag{99}$$

because $M^* = X M^{-1} X \Rightarrow X^* = X$

If we come back to Mostow Theorem : $M = U e^{iA} e^S$

$$\begin{aligned} &\Rightarrow P = M^+ M = e^S e^{2iA} e^S \\ &\Rightarrow P^* = e^S e^{-2iA} e^S = e^{2S}\left(e^{-S} e^{-2iA} e^{-S}\right) e^{2S} \Rightarrow P^* = e^{2S} P^{-1} e^{2S} \end{aligned} \tag{100}$$

Lemma and corollary will induce :

$$P^* = e^{2S} P^{-1} e^{2S} \Rightarrow e^{2S} = P^{1/2}\left(P^{-1/2} P^* P^{-1/2}\right)^{1/2} P^{1/2} \tag{101}$$

And then :

$$S = 1/2.\log\left(P^{1/2}\left(P^{-1/2} P^* P^{-1/2}\right)^{1/2} P^{1/2}\right) \text{ with } P = M^+ M \tag{102}$$

Based on Exponential injectivity from $e^{2iA} = e^{-S} P e^{-S}$, we can deduce that :

$$A = \frac{1}{2i}\log\left(e^{-S} P e^{-S}\right) \text{ with } P = M^+ M \tag{103}$$

and finally, $U = M e^{-S} e^{-iA}$ (104)

Median in Siegel disk could be then obtained by analogy with numerical scheme developed for median in Poincaré's disk. Numerical scheme based on Mostow Decomposition theorem and Siegel Disk automorphism is given by :

$$\begin{aligned} &\{Z_1, ..., Z_m\} \text{ in Siegel Half-Plane} \\ &\text{For } i = 1,...,m : W_i = \left(Z_i - iI\right)\left(Z_i + iI\right)^{-1} \\ &\textit{Initialisation :} \\ &W_{median,0} = 0 \quad et \quad \{W_{1,0}, ..., W_{m,0}\} = \{W_1, ..., W_m\} \end{aligned} \tag{105}$$

Iterate on n until $\|G_n\|_F < \varepsilon$

$$
\left|
\begin{array}{l}
W_{k,n} = U_{k,n} e^{iA_{k,n}} e^{S_{k,n}} \Rightarrow H_{k,n} = U_{k,n} e^{iA_{k,n}} = W_{k,n} e^{-S_{k,n}} = e^{-\frac{S_{k,n}}{2}} W_{k,n} e^{-\frac{S_{k,n}}{2}} \quad with: \\
S_{k,n} = 1/2.\log\left(P_{k,n}^{1/2}\left(P_{k,n}^{-1/2} P_{k,n}^{*} P_{k,n}^{-1/2}\right)^{1/2} P_{k,n}^{1/2}\right) \text{ with } P_{k,n} = W_{k,n}^{+} W_{k,n} \\
G_n = \gamma_n \sum_{\substack{k=1 \\ k \neq l}}^{m} H_{k,n} \quad \text{with} \quad \left\{ l / \|H_{k,n}\|_F < \varepsilon \right\} \\
\text{For } k = 1,...,m \quad \text{then} \quad W_{k,n+1} = \Phi_{G_n}\left(W_{k,n}\right) \\
W_{k,n+1} = \left(I - G_n G_n^{+}\right)^{-1/2} \left(W_{k,n} - G_n\right)\left(I - G_n^{+} W_{k,n}\right)^{-1} \left(I - G_n^{+} G_n\right)^{1/2} \\
W_{median,n+1} = \Phi_{G_n}^{-1}\left(W_{median,n}\right) \Rightarrow \begin{cases} W_{median,n+1} = \Phi_{G_n}^{-1}(W_{median,n}) = (GG_n^{+} + I)^{-1}\left(G + G_n\right) \\ G = \left(I - G_n G_n^{+}\right)^{1/2} W_{median,n} \left(I - G_n^{+} G_n\right)^{-1/2} \end{cases}
\end{array}
\right.
\tag{106}
$$

13 Hua Kernel for Cartan-Siegel Domains, Berezin Quantization & Geometric lift

Symmetric Bounded Domains of C^n are key spaces for all these approaches and are particular symmetric spaces of non-compact type. Elie Cartan [4] has proved that there are only 6 types:

- 2 exceptionnal types (E6 et E7)
- 4 Classical Symmetric bounded Domains (extension of Poincaré Unit disk):

$$
\begin{array}{l}
Z : \text{Complex Rectangular Matrix} \\
ZZ^{+} < I \quad (^{+} : transposed - conjugate) \\
\text{Type I} : \Omega_{p,q}^{I} \text{ complex matrices with p lines and q rows} \\
\text{Type II} : \Omega_{p}^{II} \text{ complex symmetric matrices of order p} \\
\text{Type III} : \Omega_{p}^{III} \text{ complex skew symmetric matrices of order p} \\
\text{Type IV} : \Omega_{n}^{IV} \text{ complex matrices with n rows and 1 line :} \\
\left|ZZ^{t}\right| < 1, 1 + \left|ZZ^{t}\right|^{2} - 2ZZ^{+} > 0
\end{array}
\tag{107}
$$

kernel function for all these domains were established by Lookeng Hua:

$$
\begin{array}{l}
K\left(Z,W^{*}\right) = \dfrac{1}{\mu(\Omega)} \det\left(I - ZW^{+}\right)^{-\nu} \quad \text{for } \begin{cases} \text{Type I} : \Omega_{p,q}^{I}, \nu = p + q \\ \text{Type II} : \Omega_{p}^{II}, \nu = p + 1 \\ \text{Type III} : \Omega_{p}^{III}, \nu = p - 1 \end{cases} \\
K\left(Z,W^{*}\right) = \dfrac{1}{\mu(\Omega)} \left(1 + ZZ^{t} W^{*} W^{+} - 2ZW^{*}\right)^{-\nu} \text{ for Type IV} : \Omega_{n}^{IV}, \nu = n \\
\text{where } \mu(\Omega) \text{ is euclidean volume of the domain.}
\end{array}
\tag{108}
$$

For the case ($p=q=n=1$), all these domains are reduced to the classical Poincaré unit disk :

$$\Omega^{I}_{1,1} = \Omega^{II}_{1} = \Omega^{III}_{1} = \Omega^{IV}_{1} = \{z \in C / zz^{*} < 1\}, K(z,w^{*}) = \frac{1}{(1 - zw^{*})^{2}} \tag{109}$$

Groups of analytic automorphisms of these domains are locally isomorphic to the group of matrices which preserve following forms:

$$\begin{aligned}
&\text{Type I}: \Omega^{I}_{p,q}\ , AHA^{*} = H, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, \det A = 1 \\
&\text{Type II}: \Omega^{II}_{p}\ , AHA^{*} = H, AKA^{t} = K, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, K = \begin{pmatrix} 0 & I_p \\ -I_p & 0 \end{pmatrix} \\
&\text{Type III}: \Omega^{III}_{p}\ , AHA^{*} = H, ALA^{t} = L, H = \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}, L = \begin{pmatrix} 0 & I_p \\ I_p & 0 \end{pmatrix} \\
&\text{Type IV}: \Omega^{IV}_{n}\ , AHA^{*} = H, AHA^{t} = H, H = \begin{pmatrix} -I_2 & 0 \\ 0 & I_n \end{pmatrix}
\end{aligned} \tag{110}$$

All classical domains are circular and considered in the general framework of Elie Cartan Theory, where the origin is a distinguished point for the potential:

$$\Phi(Z,Z^{*}) = \log\left[\frac{K(Z,Z^{*})}{K(0,0)}\right] = \log\det(I - ZZ^{+})^{-\nu} \tag{111}$$

F.A. Berezin [1] has introduced on these Cartan-Siegel domains the concept of quantization based on construction of Hilbert spaces of analytical functions:

$$\begin{aligned}
&\langle f, g\rangle = c(h)\int f(Z)g(Z)\left[\frac{K(Z,Z^{*})}{K(0,0)}\right]^{-1/h} d\mu(Z,Z^{*}) \\
&c(h)^{-1} = \int\left[\frac{K(Z,Z^{*})}{K(0,0)}\right]^{-1/h} d\mu(Z,Z^{*}) \\
&K(gZ,gZ^{*})j(g,z)j(g,Z)^{*} = K(Z,Z^{*}) \text{ with } j(g,Z) = \frac{\partial gZ}{\partial Z}
\end{aligned} \tag{112}$$

One example is given in dimension 1 for Poincaré unit disk $D = \{z \in C / |z| < 1\} = SU(1,1)/S^{1}$ with volume element $1/2i.(1-|z|^{2})^{-2} dz \wedge dz^{*}$:

$$\begin{aligned}
&g \in SU(1,1) \text{ with } g = \begin{bmatrix} a & b \\ b^{*} & a^{*} \end{bmatrix} \text{ where } |a|^{2} - |b|^{2} = 1 \\
&\text{with Kähler potential}: F(z) = -\log(1 - |z|^{2}) \Rightarrow F(gz) = 2\,\mathrm{Re}\log(b^{*}z + a^{*}) + F(z) \\
&\Rightarrow ds^{2} = \frac{\partial^{2}F(gz)}{\partial z \partial z^{*}} = \frac{\partial^{2}F(z)}{\partial z \partial z^{*}}
\end{aligned} \tag{113}$$

It results from the last equation that the Kählerian metric is invariant under the action $g \in G$ (automorphisms of unit disk). The transform of the base point $z = 0$ of the disk by $g \in G$ is given by $g(0) = b(a^{*})^{-1}$. It defines a lifting that allows to associate to all paths in disk a lift in G. In the same way, we can define a geometric lift of potential K in G:

$$g(0)=b\left(a^*\right)^{-1}\Rightarrow F(g(0))=-\log\left(1-\left|b\left(a^*\right)^{-1}\right|^2\right)\underset{|a|^2-|b|^2=1}{=}\log\left(1+|b|^2\right)$$
$$g^{-1}=\begin{pmatrix} a^* & -b \\ -b^* & a \end{pmatrix}\Rightarrow F(g^{-1})=F(g) \tag{114}$$

Obviously, all these lifts could be extended to Cartan-Siegel Domains $SD_n=\{Z\,/\,ZZ^+<I\}$:

$$\text{Let } g=\begin{bmatrix} A & B \\ B^* & B^* \end{bmatrix} \text{ and } g^tJg=J \text{ with } J=\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$
$$\text{A and B that verify } \begin{cases} A^+A-B^tB^*=I \\ B^+A-A^tB^*=0 \end{cases}$$
$$g(Z)=(AZ+B)\left(B^*Z+A^*\right)^{-1} \tag{115}$$
$$\text{of Kähler Potential}: F(z)=-\log\det\left(I-ZZ^+\right)=-Tr\left[\log\left(I-ZZ^+\right)\right]$$
$$F(g(Z))=F(Z)+2\,\mathrm{Re}\left(Tr\left(\log\left(A^*+B^*Z\right)\right)\right)\Rightarrow \partial\partial^*F(g(Z))=\partial\partial^*F(Z)$$

Geometric Lift in Cartan-Siegel domain is then given by the following:

$$g(0)=B\left(A^*\right)^{-1}\Rightarrow F(g(0))=\log\det\left(I+BB^+\right)=Tr\left[\log\left(I+BB^+\right)\right] \tag{116}$$

F.A. Berezin [1] has proved that for every symmetric Riemannian space, there exist a dual space being compact. The isometry groups of all the compact symmetric spaces are described by block matrices (the action of the group in terms of special coordinates is described by the same formula as the action of the group of motions of the dual domain).

$$\Gamma=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\Rightarrow \Gamma(W)=\left(A_{11}W+A_{12}\right)\left(A_{21}W+A_{22}\right)^{-1}$$
$$\text{Isometry}: \Gamma=C\Gamma C^{-1} \text{ with } C=\frac{1}{\sqrt{2}}\begin{pmatrix} I & iI \\ iI & I \end{pmatrix} \tag{117}$$

Berezin coordinates for Siegel domain are given by

$$\Gamma=\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix},\ \Gamma^{-1}=\begin{pmatrix} A^+ & B^t \\ B^+ & A^t \end{pmatrix} \quad :$$
$$\text{or equivalently}: \Gamma\Gamma^+=I\ ,\ \Gamma L\Gamma^t=L \text{ with } L=\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \tag{118}$$

$$g(0)=B\left(A^*\right)^{-1}\Rightarrow F(g(0))=\ln\det\left(I+BB^+\right)=trace\ln\left(I+BB^+\right) \tag{119}$$

For this dual space, the volume and the metric are invariant:

$$d\mu\left(W,W^*\right)=H\left(W,W^*\right)\frac{d\mu_L\left(W,W^*\right)}{\pi^n}$$

$$ds^2=\sum_{\alpha,\beta}g_{\alpha,\beta}dW^{\alpha}dW^{\beta*} \text{ with } g_{\alpha\beta}=-\frac{\partial^2\log H\left(W,W^*\right)}{\partial W^{\alpha}\partial W^{\beta*}} \tag{120}$$

where $H(W,W^*)=\det\left(I+WW^+\right)^{-\nu}$

For arbitrary Kählerian homogeneous space, the logarithm of the density for the invariant measure is the potential of the metric.

14 Information metric, Entropy metric and Bergman metric

In previous chapter, we have first introduced Information Geometry metric and have used Entropy metric that is also a Bergman metric. In 1984, Jacob Burbea [49] has written a very interesting paper where he has studied relations of differential metrics in probability spaces through entropy functionals with information metric and Bergman metric. He has considered probability density functions $p(t/z)$ where $z\in D$ and D is a manifold imbedded in C^n with Hermitian Fisher information matrix :

$$g_{i\bar{j}}(z)=\int p^{-1}(t/z)\frac{\partial p(t/z)}{\partial z_i}\frac{\partial p(t/z)}{\partial z_j^*}d\mu(t)=\int p(t/z)\frac{\partial\log p(t/z)}{\partial z_i}\frac{\partial\log p(t/z)}{\partial z_j^*}d\mu(t) \tag{121}$$

with $$ds^2(z)=\sum_{i=1}^{N}g_{i\bar{j}}dz_idz_j^*=\left\|p^{1/2}\partial\log p\right\|_{\mu}^2 \tag{122}$$

that is locally invariant under holomorphic transformations of z.
Considering $f(t/z)$ function of an open subset of some Fréchet space, the tangent in the direction of $(u,v)\in C^nxC^n$ is given by :

$d_{(u,v)}f(./z)=\partial_uf(./z)+\bar{\partial}_vf(./z)$ with $\partial_uf(./z)=\sum_{k=1}^{N}\frac{\partial f(./z)}{\partial z_k}u_k$ and $\bar{\partial}_vf(./z)=\sum_{k=1}^{N}\frac{\partial f(./z)}{\partial z_k^*}v_k$

He has defined the φ_α-entropy functional, with φ_α a concave function :

$$H_{\varphi_\alpha}(p)=\int\varphi_\alpha\left[p(t/z)\right]d\mu(t) \tag{123}$$

$$\varphi_\alpha(s)=\begin{cases}(\alpha-1)^{-1}\left(s-s^{\alpha}\right) & \text{if } \alpha\neq 1\\ -s\log s & \text{if } \alpha=1\end{cases} \tag{124}$$

Complex Hessian at p in the direction of f is then defined by Fréchet Derivatives:

$$dH_{\varphi_\alpha}(p;f)=\frac{d}{ds}H_{\varphi_\alpha}\left(p+sf\right)\Big|_{s=0}=\int\varphi_\alpha'\left[p(t)\right]f(t)d\mu(t) \tag{125}$$

$$d^2H_{\varphi_\alpha}(p;f,g)=\int\varphi_\alpha''\left[p(t)\right]f(t)g(t)d\mu(t) \tag{126}$$

An Hermitian positive definite differential metric can be defined with $f = \partial_u p$:

$$ds_\alpha^2(z) = -\frac{1}{4\alpha}\Delta_{\partial p} H_{\varphi_\alpha}(p) = -\frac{1}{\alpha} d^2 H\varphi_\alpha(p;\partial_u p, \bar{\partial}_u p) = -\frac{1}{\alpha}\int \varphi_\alpha''(p(t/z))\left|\partial_u p(t/z)\right|^2 d\mu(t) \quad (127)$$

For $\alpha = 1$, we recover the classical Rao metric, based on Shannon entropy.

To make the relation with Bergman metric, Burbea has defined $p(t/z)$ has a square of the modulus of a normalized function : $p(t/z) = |\psi(t/z)|^2$ such that $\|\psi(./z)\|_\mu^2 = 1$, where a non normalized function is given by :

$$g(t/z) = \sqrt{K(z,z^*)}.\psi(t/z) \quad \text{where} \quad K(z,w^*) = \langle g(t/z), g(t/w)\rangle_\mu \quad (128)$$

where is K a sesqui-holomorphic Bergman Kernel on *DxD* by use of Hartog's theorem. To recover Bergman metric, Burbea has considered then a pseudo-distance in the form of $\lambda(z,w) = \sqrt{1 - \left|\int \psi(t/z)\psi(t/w)d\mu(t)\right|} = \sqrt{1 - \left|\langle \psi(t/z), \psi(t/w)\rangle_\mu\right|}$ with :

$$ds_{Bergman}^2 = d^2\lambda(z,w)\big|_{w=z} = \|d\psi\|_\mu^2 - \left|\langle \psi, d\psi\rangle_\mu\right|^2 = K^{-2}\left[K\partial\bar{\partial}K - |\partial K|^2\right] = \partial\bar{\partial}\log K \quad (129)$$

$$ds_{Bergman}^2 = \sum_{i,j=1}^{N} \frac{\partial \log K(z,z^*)}{\partial z_i \partial z_k} dz_i dz_k^* \quad (130)$$

Then, Burbea has showed that the projective pseudo distance is in fact the Skwarczynski pseudo-distance:

$$\lambda(z,w) = \sqrt{1 - \sqrt{\frac{|K((z,w^*)|^2}{K(z,z^*)K(w,w^*)}}} \quad (131)$$

Relation between Information metric and Bergman metric is given by this fundamental relation:

$$\begin{cases} p(t/z) = |\psi(t/z)|^2 \\ g(t/z) = \sqrt{K(z,z^*)}.\psi(t/z) \end{cases} \Rightarrow \log p(t/z) = \log g(t/z) + \log g^*(t/z) - \log K(z,z^*) \quad (132)$$

By virtue of the Cauchy-Riemann equations, the following relation is given:

$$\frac{\partial^2 \log p(t/z)}{\partial z_i \partial z_j^*} = -\frac{\partial^2 \log K(z,z^*)}{\partial z_i \partial z_j^*} \quad (133)$$

We recover equivalence of Rao-Chentsov Information metric with Bergman metric by taking expectation of previous equation:

$$g_{ij} = -E\left[\frac{\partial^2 \log p(t/z)}{\partial z_i \partial z_j^*}\right] = \frac{\partial^2 \log K(z,z^*)}{\partial z_i \partial z_j^*} \quad (134)$$

15 Complex Riccati Equation in Cartan-Siegel Domains

Recently, Russian M.I. Zelikin [61] has introduced Complex Riccati equations, revealing an intrinsic connection between these Riccati equations as a flow on Cartan-Siegel Homogeneous Domains. These Riccati equations that arise in the classical calculus of variations define a flow on the generalized Siegel upper half-plane.

Let $W = X + iY$ with $\mathrm{Im}(Y) > 0$ in the Siegel Upper-half-plane, then the following complexified Riccati equation :

$$\dot{W} = (C + W)A^{-1}\left(C^T + W\right) - B \quad \text{with } A, B \text{ symmetric matrices} \tag{135}$$

$$\dot{W} - WA^{-1}W - CA^{-1}W - WA^{-1}C^T - CA^{-1}C^T + B = 0 \tag{136}$$

is a flow in this Siegel space. This Riccati equation is link with the following classical calculus of variation for the functional:

$$S = \frac{1}{2}\int_{t_0}^{t_1}\left[\left\langle A(t)\dot{h}, \dot{h}\right\rangle + 2\left\langle C(t)\dot{h}, h\right\rangle + \left\langle B(t)h, h\right\rangle\right]dt \tag{137}$$

This Riccati equation is obtained from the canonical Hamiltonian system of ordinary differential equations :

$$\begin{pmatrix} \dot{h} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} -A^{-1}C^T & A^{-1} \\ -CA^{-1}C^T + B & CA^{-1} \end{pmatrix}\begin{pmatrix} h \\ p \end{pmatrix} \tag{138}$$

where the previous block matrix belongs to the Lie algebra of the Lie group Sp(n,R) for all t (we can observe the same block structure than in equation (52)).

16 Maslov index for shilov boundary of Poincaré/Siegel Disks

Before introducing Radar applications of previous flows, we introduce a last tool that could be used for shilov boundary of Siegel Disk : Arnold-Maslov-Leray index . We will provide Jean-Louis Clerc formula for computing the Arnold-Maslov-Leray Index for Siegel dounded domains, analogue of Souriau index for Lagrangian Manifold, using the automorphy kernel of the Siegel Disk, where the shilov boundary is the manifold of Lagrangian subspaces. This Arnold-Maslov-Leray index could be studied in the future in the framework of Information Geometry for Toeplitz-Block-Toeplitz matrices.

In Poincaré's disk SD_1, ideal triangle can be considered as a limit of geodesic triangle in SD_1 when triangle points converge to shilov boundary, using aire A(T) of oriented geodesic triangle:

$$A(T) = \arg\left(\frac{1 - z_1 z_2^*}{1 - z_1^* z_2}\right) + \arg\left(\frac{1 - z_2 z_3^*}{1 - z_2^* z_3}\right) + \arg\left(\frac{1 - z_3 z_1^*}{1 - z_3^* z_1}\right) \tag{139}$$

We recover canonical automorphism kernel of SD_1 :

$$\begin{aligned}
&\forall z, w \in D, k(z,w) = 1 - zw^* \ , \ \ k(z,w) = k^*(w,z) \\
&g = \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} \in PSU(1,1) \\
&\Rightarrow k(gz, gw) = (\beta^* z + \alpha^*)^{-1} k(z,w)(\beta w^* + \alpha)^{-1}
\end{aligned} \tag{140}$$

As SD_1 is simply connected, there exist a unique computation of geodesic aire argument that is cancelled on *{z1,z2,z3}* from which we can deduce geodesic triangle aire and Maslov-Leray index :

$$\begin{aligned}
&c(z_1, z_2, z_3) = k(z_1, z_2)k(z_2, z_1)^{-1}k(z_2, z_3)k(z_3, z_2)^{-1}k(z_3, z_1)k(z_1, z_3)^{-1} \\
&i(\sigma_1, \sigma_2, \sigma_3) = \frac{1}{\pi} \lim_{\substack{z_i \to \sigma_i \\ i=1,2,3}} \arg c(z_1, z_2, z_3)
\end{aligned} \tag{141}$$

Maslov Index has been generalized in Siegel disk with Symplectic aire by J.L. Clerc, using canonical automorphism kernel of Siegel disk SD_n and oriented Symplectic aire:

$$\begin{aligned}
&D = \{Z \in V / I - ZZ^+ > 0\} \text{ and } S = \{Z \in V / Z^+ = Z^{-1}\} \\
&S_T^3 = \{(\sigma_1, \sigma_2, \sigma_3) \in S^3 / \sigma_i T \sigma_j, \forall i \neq j\} \\
&k(Z,W) = \det\left(K(Z,W)^{r/n}\right) \text{ with } K(Z,W) \ \textit{canonical automorphism kernel of } SD_n \\
&c(Z_1, Z_2, Z_3) = k(Z_1, Z_2)k(Z_2, Z_1)^{-1}k(Z_2, Z_3)k(Z_3, Z_2)^{-1}k(Z_3, Z_1)k(Z_1, Z_3)^{-1} \\
&i(\sigma_1, \sigma_2, \sigma_3) = \frac{1}{2\pi} \lim_{\substack{z_i \to \sigma_i \\ i=1,2,3}} \arg c(Z_1, Z_2, Z_3)
\end{aligned} \tag{142}$$

17 Radar Applications for Robust Ordered-Statistic Processing: OS-HDR-CFAR and OS-STAP

In the following, we will apply previous tools to built Robust Ordered-Statistic (OS) processing. Ordered-Statistic is a very useful tool used in Radar for a long time to be robust againt outliers on scalar data from secondary data. We will define an OS-HDR-CFAR (Ordered-Statistic High Doppler Resolution Constant False Alarm Rate) algorithm jointly taking into account robustness of “matrices median” and high Doppler resolution of regularized Complex Auto-Regressive model. We will define also an OS-STAP (Ordered Statistic Space-Time Adaptive Processing), based on median computation of secondary data space-time covariance matrix with Mostow/Berger fibration applied on Multichannel Autoregressive Model.

This paper will not address Polarimetric Data processing, but obviously these tools could be extended to compact manifold to define Ordered statistic for Polarimetric covariance matrices.

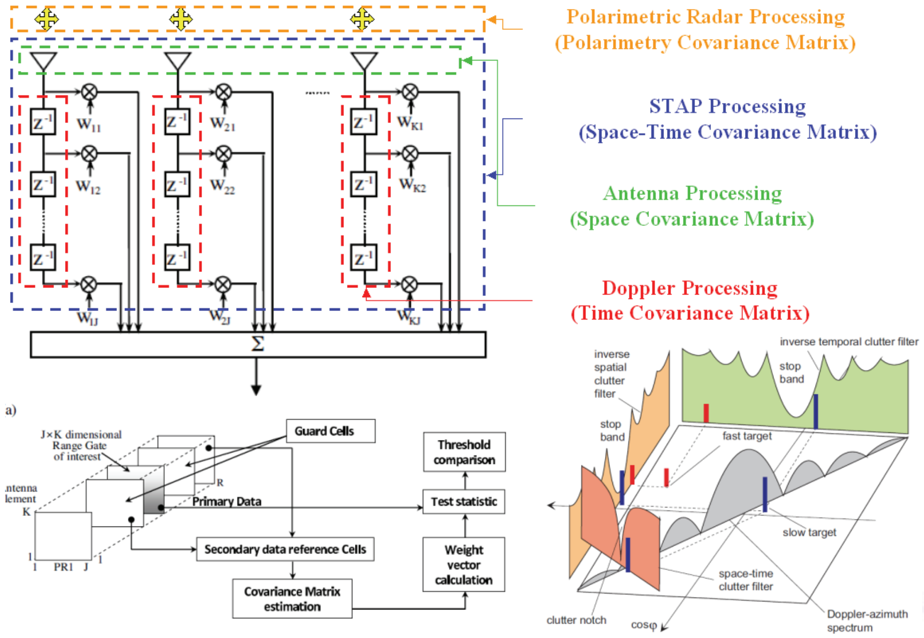

Fig. 10. Doppler Processing, Antenna Processing, Space-Time Processing & Polar processing

17.1 Robust Doppler Processing : OS-HDR-CFAR

The regularized Burg algorithm [62,63] is an alternative Bayesian composite model approach to spectral estimation. The reflection coefficients, defined in classical Burg algorithm are estimated through a regularized method, based on a Bayesian adaptative spectrum estimation technique, proposed by Kitagawa & Gersch, who use normal prior distributions expressing a smoothness priors on the solution. With these priors, autoregressive spectrum analysis is reduced to a constrained least squares problem, minimized for fixed tradeoff parameters, using Levinson recursion between autoregressive parameters. Then, a reflection coefficient is calculated, for each autoregressive model order, by minimizing the sum of the mean-squared values of the forward and backward prediction errors, with spectral smoothness constraints . Tradeoff parameters balance estimate of the autoregressive coefficients between infidility to the data and infidility to the frequency domain smoothness constraint. This algorithm conserves lattice structure advantages, and could be brought in widespread use with a multisegment regularized reflection coefficient version. The regularized Burg algorithm lattice structure offers implementation advantages over tapped delay line filters because they suffer from less round-off noise and less sensivity to coefficient value perturbations.

$$
\begin{aligned}
&.\ \textit{Initialisation :} \\
&f_0(k) = b_0(k) = z(k) \quad , \quad k=1,...,N \ \ (\text{N : nb. pulses per burst}) \\
&P_0 = \frac{1}{N}.\sum_{k=1}^{N} |z(k)|^2 \quad \text{and} \quad a_0^{(0)} = 1 \\
&.\ \textit{Iteration}\,(\text{n}):\ \text{for}\ \ n = 1\ \text{to}\ M \\
&\mu_n = -\frac{\dfrac{2}{N-n}\displaystyle\sum_{k=n+1}^{N} f_{n-1}(k).b_{n-1}^{*}(k-1) + 2.\sum_{k=1}^{n-1}\beta_k^{(n)}.a_k^{(n-1)}.a_{n-k}^{(n-1)}}{\dfrac{1}{N-n}\displaystyle\sum_{k=n+1}^{N} |f_{n-1}(k)|^2 + |b_{n-1}(k-1)|^2 + 2.\sum_{k=0}^{n-1}\beta_k^{(n)}.\left|a_k^{(n-1)}\right|^2} \quad \text{with} \quad \beta_k^{(n)} = \gamma_1.(2\pi)^2.(k-n)^2 \\
&\begin{cases} a_0^{(n)} = 1 \\ a_k^{(n)} = a_k^{(n-1)} + \mu_n.a_{n-k}^{(n-1)*} \quad , \quad k=1,...,n\text{-}1 \\ a_n^{(n)} = \mu_n \end{cases} \quad \text{and} \quad \begin{cases} f_n(k) = f_{n-1}(k) + \mu_n.b_{n-1}(k-1) \\ b_n(k) = b_{n-1}(k-1) + \mu_n^{*}.f_{n-1}(k) \end{cases}
\end{aligned}
\tag{143}
$$

In the following figures, regularization property is illustrated with deletion of spurious peaks. We select the AR model of maximum order (number of pulses minus one).

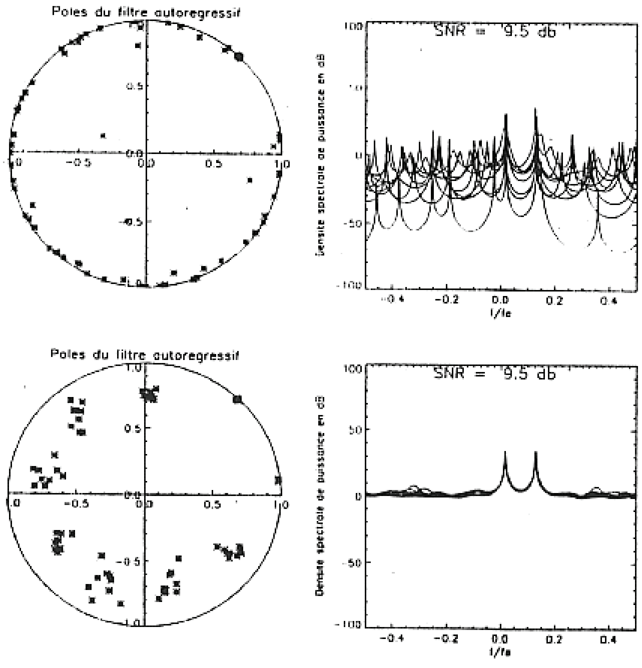

Fig. 11. (up) Non-regularized & (bottom) Regularized Doppler AR Spectrum

We conserve the sliding window structure of classical CFAR : we compare the AR model under test by computing its Information Geometry distance with median AR model of secondary data in the neighborhood. Median Autoregressive model is computed by :

For $P_{median,0}$, we use classical median on real values $P_{0,k}$

For $\{\mu_k\}_{k=1}^{n-1}$:

$$w_n = \gamma_n \sum_{k \in G_0} \frac{\mu_{k,n}}{|\mu_{k,n}|} \text{ with } G_0 = \{k / |\mu_{k,n}| \neq 0\}$$

$$\mu_{k,n+1} = \frac{\mu_{k,n} - w_n}{1 - \mu_{k,n}.w_n^*}$$

$$\mu_{median,n+1} = \frac{\mu_{median,n} + w_n}{1 + \mu_{median,n} w_n^*} \tag{144}$$

The detection test is finally based on computation of the robust Information Geometry distance:

$$d^2\left[\left(P_{0,k} \quad \mu_{1,k} \quad \cdots \quad \mu_{N-1,k}\right),\left(P_{median,0} \quad \mu_{1,median} \quad \cdots \quad \mu_{N-1,median}\right)\right] = n \log^2\left(\frac{P_{median,0}}{P_{0,k}}\right) + \sum_{i=1}^{N-1}(N-k)\left(\frac{1}{2}\log\left(\frac{1+\delta_i}{1-\delta_i}\right)\right)^2 \text{ with } \delta_i = \left|\frac{\mu_{i,k} - \mu_{i,median}^*}{1 - \mu_{i,k}\mu_{i,median}^*}\right| \tag{145}$$

In the following figure, we compare the classical processing chain with new OS-HDR-CFAR

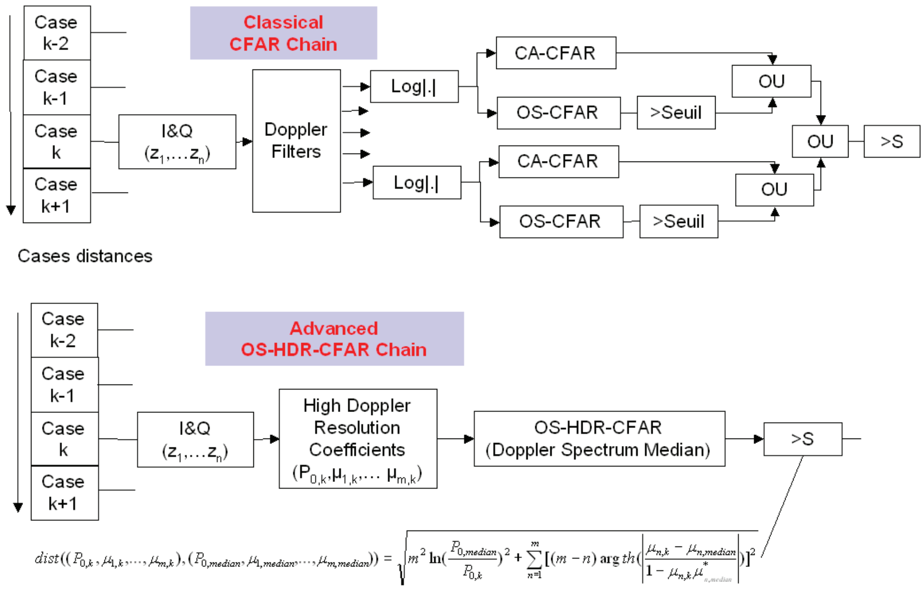

Fig. 12. (top figure) Classical OS-CFAR after filter banks, (bottom figure) OS-HDR-CFAR

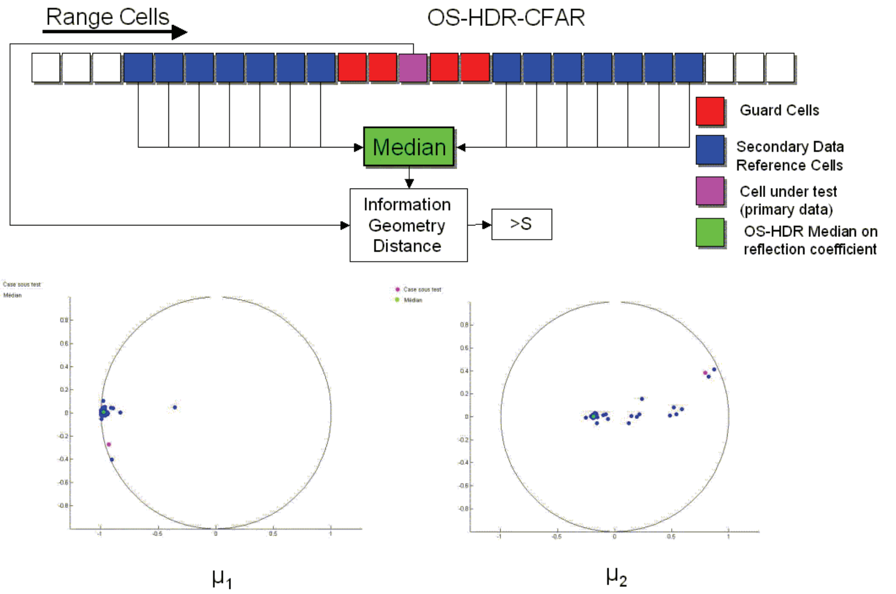

Fig. 13. OS-HDR-CFAR Algorithm with illustration of two first reflection coefficients

We have tested OS-HDR-CFAR on real recorded ground Radar clutter with ingestion of synthetic slow targets.

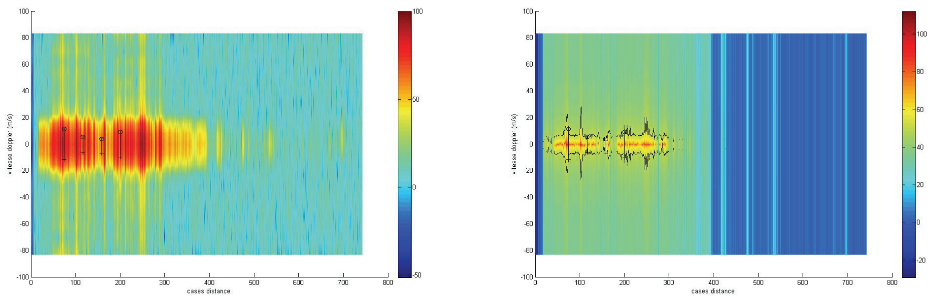

Fig. 14. Comparison of FFT Doppler spectrum (at left) and High Resolution Regularized Doppler spectrum (at right)

In the following figures, we give ROC curves with Probability of detection versus probability of false alarm. We observe that OS-HDR-CFAR is better (Pd = 0.8) than OS-CFAR/Doppler-Filters (Pd = 0.65) for arbitrary fixed Pfa. We could also observe that Information Geometry approach provides better results than Optimal Transport Theory approach (based on Wasserstein distance/barycenter : black curve).

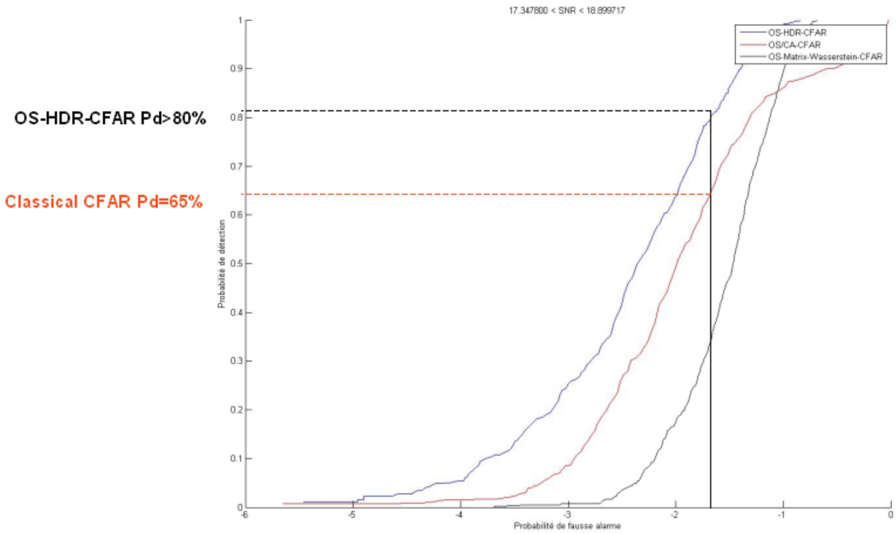

Fig. 15. ROC curves for 3 approaches : OS-HDR-CFAR, OS-CFAR/Doppler-Filters, & method based on Wasserstein barycenter/distance (optimal transport theory)

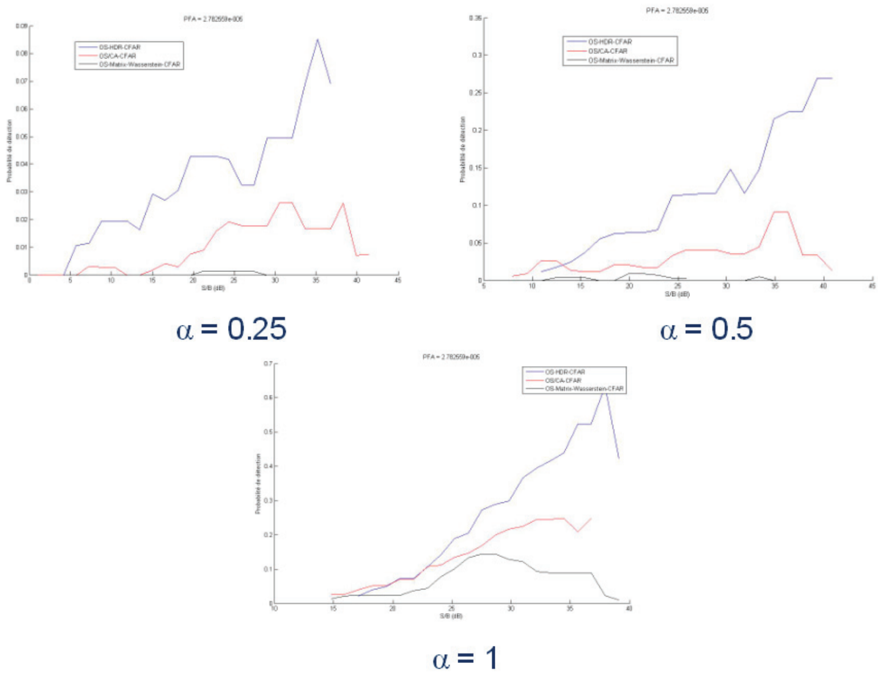

Fig. 16. Probability of detection versus SNR for fixed Pfa=10^{-5}, with α relative position in Doppler of the target normalized by Doppler Clutter Spectrum Width (α=1 means that the target is positioned on the edge in Doppler of the ground clutter)

To prove that OS-HDR-CFAR is an Ordered-Statistic CFAR, robust against outliers, we have compared the case where all targets are ingested every 33 range cells (CFAR window are limited to 32 range cells), and the case where all targets are ingested every 7 range cells. We can observe that OS-HDR-CFAR performance are not altered by targets in the secondary data window.

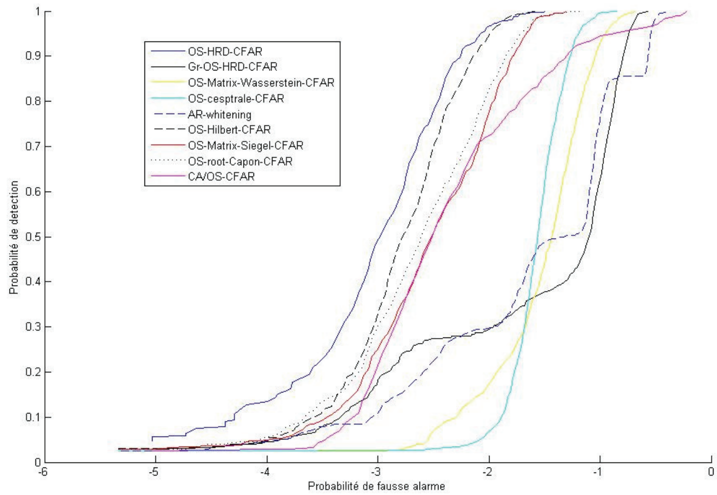

Fig. 17. α=0.25; SNRmean=17dB; target every 33 range cells

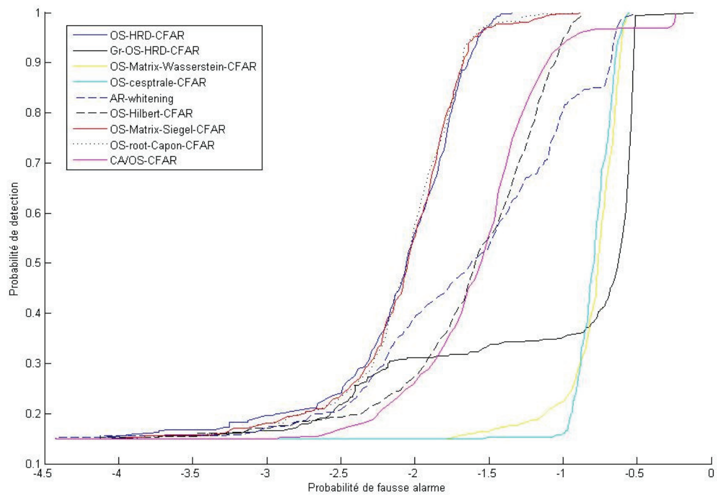

Fig. 18. α=0.25; SNRmean=17dB; 1 target every 7 range cells

17.2 Robust Space-Time Processing: OS-STAP

We propose in this second applicative part to study robust STAP (Space-Time Adaptative Processing) based on Median of Sample Covariance Matrix, with advantages to be tolerant to the presence of targets or non-homogeneities in the secondary data. First, we give some basic elements of STAP theory.

Brennan and Reed have proposed in 1973, STAP for radar target based on interference covariance matrix estimation from target-free training data, weight vector calculation and threshold on statistical test. Signal snapshot Radar data model is given by:

$$Z_t = \alpha V(\varphi,\omega) + n \tag{146}$$

with $R = E\left[(Z_t - \alpha V)(Z_t - \alpha V)^+\right]$ where $R = R_{noise} + R_{clutter} + R_{jammer}$

A single "primary" data may contain a target return with space-time steering vector $V(\varphi,\omega)$ and unknown complex amplitude α, while other independent "secondary" data vectors are available that are zero mean, and share the same space-time Covariance matrix R for noise, clutter and jamming. Snapshot for range gate l is given by :

$$Z_{t,l} = \begin{bmatrix} z_{1l} \\ z_{2l} \\ \vdots \\ z_{Ml} \end{bmatrix} \text{ with } z_{ml} = \begin{bmatrix} z_{1ml} \\ z_{2ml} \\ \vdots \\ z_{Nml} \end{bmatrix} \quad \text{where } z_{nml} \begin{cases} \text{n}^{\text{th}} \text{ element} \\ \text{m}^{\text{th}} \text{ pulse} \\ \text{l}^{\text{th}} \text{ range gate} \end{cases} \tag{147}$$

Optimal step filter can be interpreted by successive processing of a Matched filter and a Whitening Filter :

$$Z_{output} = \hat{w}^+ Z_t = \left(S^{-1/2}V\right)^+ \left(S^{-1/2}\right)^+ Z_t \tag{148}$$

In this formula, Z_t is the primary data vector snapshot, and S is a sample covariance matrix based on M secondary data vectors $Z(k)$:

$$S = M.\hat{R} = \sum_{k=1}^{M} Z(k)Z(k)^+ \tag{149}$$

The Generalized Likelihood Ratio Test (GLRT) assumes that the covariance is known, and is deduced by maximization over the unknown parameter α :

$$\Lambda_{GLRT} = \frac{\left|\hat{w}^+ Z_t\right|^2}{V^+ S^{-1} V\left[1 + \left(Z_t^+ S^{-1} Z_t\right)\right]} > \eta_0 \quad \text{with} \quad \hat{w} = S^{-1}V \tag{150}$$

Classically, sample covariance matrix of secondary data is based on N.R. Goodman's Theorem. Consider M independent identically distributed N-variate complex Gaussian random variable $Z(k)$, $k=1,...,M$ as a sample of size N from a population with PDF $p(Z/R_Z)$. Let *HPD(n)* be the set of NxN Hermitian positive definite matrices. Over the domain *HPD(n)* the maximum likelihood estimator $\hat{R}_Z$ of the covariance matrix R_Z is:

$$\hat{R}_Z = \frac{1}{M}\sum_{k=1}^{M} Z(k)Z(k)^+ \cdot \tag{151}$$

Proof is based on

$$E_{p(Z/R_Z)}\left[Z^+ HZ\right] = Tr\left(R_Z H\right) = -i\left.\frac{d\Phi_{R_Z,H}}{d\theta}\right|_{\theta=0} \tag{152}$$

deduced from characteristic function of Hermitian form Z^+HZ :

$$\Phi_{R_Z,H}(\theta) = E_{p(Z/R_Z)}\left[e^{i\theta.Z^+HZ}\right] = \det{}^{-1}\left(I - i\theta.R_Z H\right) \cdot \tag{153}$$

In [88], authors have introduced the Parametric Adaptive Matched Filter (PAMF) methodology for STAP and detection, approximating the interference spectrum with a multichannel autoregressive (AR) model of low order, attaining modeling fidelity using a small fraction of the Reed-Brennan rule training data set, and offering dramatic improvement in performance over the conventional AMF, with only a small fraction of the secondary data required by the AMF. Multi-channel parameter identification algorithms considered were the Strand—Nuttall (SN) and the least-squares (LS) algorithms for AR model identification.

Matched Filter that is deduced from LDU decomposition $R^{-1} = \left(A^{-1}\right)^+ D^{-1/2} D^{-1/2} A^{-1}$ can be written as follow :

$$\Lambda_{MF} = \frac{\left|\left(D^{-1/2}A^{-1}V\right)^+\left(D^{-1/2}A^{-1}Z_t\right)\right|}{\left(D^{-1/2}A^{-1}V\right)^+\left(D^{-1/2}A^{-1}V\right)} = \frac{\left|\left(D^{-1/2}u\right)^+\left(D^{-1/2}\varepsilon\right)\right|}{\left(D^{-1/2}u\right)^+\left(D^{-1/2}u\right)} \quad \text{with} \quad \begin{cases} u = A^{-1}V \\ \varepsilon = A^{-1}Z_t \end{cases} \tag{154}$$

Multi-channel element vectors are approximated at order P :

$$\begin{cases} \upsilon(n) = D_P^{-1/2}\varepsilon(n) = D_P^{-1/2}\sum_{k=0}^{P} A^+(k) Z_t(n-k+P) \\ \zeta(n) = D_P^{-1/2}u(n) = D_P^{-1/2}\sum_{k=0}^{P} A^+(k) V(n-k+P) \end{cases} \tag{155}$$

with $A^+(0) = I_J$ and $n = 0,1,\ldots,N-P-1$

$Z_t : \left\{Z_t(n) \in C^J \,/\, n = 0,1,\ldots,N-1\right\}$ time series of the data

$Z_t = \left[Z_t^T(0) \quad Z_t^T(1) \quad \cdots \quad Z_t^T(N-1)\right]^T$

with final PMF test given by: $$\Lambda_{PMF} = \frac{\left|\sum_{n=0}^{N-P-1}\zeta^+(n)\upsilon(n)\right|^2}{\sum_{n=0}^{N-P-1}\zeta^+(n)\zeta(n)} \tag{156}$$

PAMF is deduced from Multivariate Autoregressive Model and Identification Algorithms where for stability, all the system poles must lie inside the Siegel Disk. For STAP PAMF, we use Multivariate Burg algorithm:

$$\left[A_1^n, A_2^n, .., A_n^n\right] = \left[A_1^{n-1}, A_2^{n-1}, .., A_{n-1}^{n-1}, 0\right] + A_n^n\left[JA_{n-1}^{n-1*}J, JA_{n-1}^{n-1*}J, .., JA_{n-1}^{n-1*}J, I\right]$$

$$\begin{cases} \varepsilon_n^f(k) = \sum_{l=0}^{n} A_l^n(k) Z(k-l) \\ \varepsilon_n^b(k) = \sum_{l=0}^{n} JA_l^n(k)^* JZ(k-n+l) \end{cases} \text{ with } J = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & 1 & 0 \\ 0 & \iddots & \iddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \ A_0^n = I \quad (157)$$

$$\begin{cases} \varepsilon_{n+1}^f(k) = \varepsilon_n^f(k) + A_{n+1}^{n+1} \varepsilon_n^b(k-1) \\ \varepsilon_{n+1}^f(k) = \varepsilon_n^b(k-1) + JA_{n+1}^{n+1*} J \varepsilon_n^f(k) \end{cases} \text{ with } \ \varepsilon_0^f(k) = \varepsilon_0^b(k) = Z(k)$$

$$A_{n+1}^{n+1} = -2\left[\sum_{k=1}^{N+n} \varepsilon_n^f(k) \varepsilon_n^b(k-1)^+\right]\left[\sum_{k=1}^{N+n} \varepsilon_n^f(k) \varepsilon_n^f(k)^+ + \sum_{k=1}^{N+n} \varepsilon_n^b(k) \varepsilon_n^b(k)^+\right]^{-1}$$

Multivariate Burg coefficient is in Siegel Unit disk SD_n. We have by Schwarz inequality that $A_{n+1}^{n+1}.A_{n+1}^{n+1+} < I_{n+1}$.

For the time being, we have not yet tested computation of Median Toeplitz-Block-Toeplitz covariance matrix by Mostow/Berger Fibration and Frechet-Karcher Flow on Reflection Coefficient matrix A_{n+1}^{n+1} of multichannel Autoregressive model: In collaboration with DRDC Canada, we have planned to do that in near future. For the time being, Balaji [80] from DRDC has tested the mean matrix iteration given by Karcher Flow:

$$R_{i,n+1} = R_{i,n}^{1/2} e^{\varepsilon\left(\sum_{k \in Secondarydata}^{N} \log\left(R_{i,n}^{-1/2} R_k R_{i,n}^{-1/2}\right)\right)} R_{i,n}^{1/2} \text{ where } R_k = Z(k)Z(k)^+ \quad (158)$$

A simulated covariance matrix is used as an example. The snapshots are drawn from a clairvoyant clutter covariance matrix. The number of apertures and pulses are chosen to be 12 and no dispersive effects, such as ICM, are assumed. A side-looking array, satisfying the DPCA condition is chosen so that the clutter covariance matrix rank is given by Brennan's rule. The performance of the Riemmanian mean for a single run is shown in following figure. In this instance, improvment over the LSMI is evident in some area close to the clutter notch. If the improvement in performance is due to the Riemannian mean algorithm more closely approximating the clairvoyant covariance matrix. This confirms the conjecture that the improved performance is due to better approximation of the true covariance matrix by the proposed Riemannian mean algorithm.

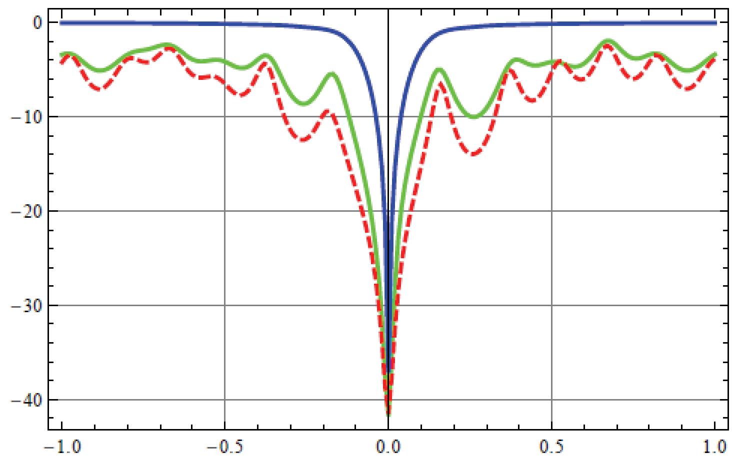

Fig. 19. LSMI STAP algorithm (blue : optimum, Red : LSMI with arithmetic mean, Green : LSMI with Riemannian mean)

Apart from inversion, an eigenvector projection algorithm using the Riemannian mean can also be investigated. Naïvely, one expects that the most important eigenvectors are those corresponding to the strongest eigenvalues. Furthermore, theseeigenvectors are better estimated using fewer samples. The EVP performance for a single run using the Riemannian mean are shown in following figure. Once agian, clear improvement in performance is evident. In fact, improvement over the inversion is also observed near the important clutternotch region .

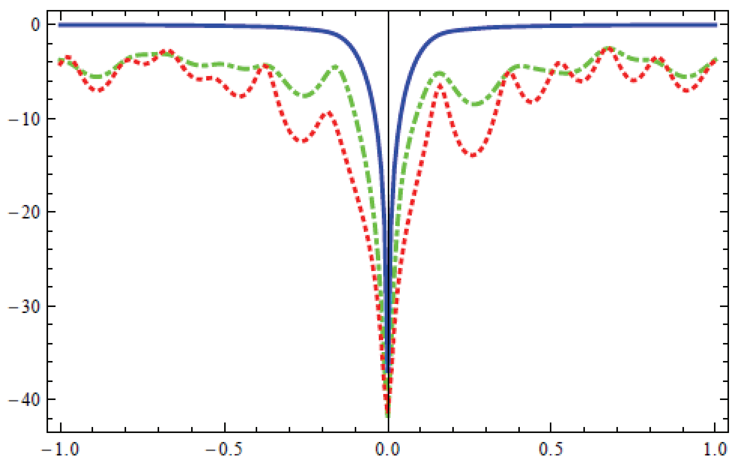

Fig. 20. EVP STAP algorithm (blue : optimum, Red : EVP with arithmetic mean, Green : EVP with Riemannian mean)

18 Miscellaneous: Shape Manifold

I would like to conclude with some remarks on “shape manifold”. In image processing, it is very useful to make statistics on shape. We can use previous approach if we can define “Shape manifold” or “shape space”, and in case of Metric space, we can extend definition of Fréchet Mean for shapes.

I will give a very simple example. If we consider a set of right triangles $\{a_i, b_i, h_i\}_{i=1}^N$, where one right triangle could be defined by one point on the surface/manifold $S = \{(a,b,h) / h^2 = a^2 + b^2\}$, then the Fréchet p-mean is defined by the minimum of:

$$\{A,B,H\} = \arg \underset{\{A,B,H\}}{Min} \sum_{i=1}^{N} d^{p}_{geodesic}\left(\{a_i, b_i, h_i\}, \{A,B,H\}\right) \tag{159}$$

where geodesic is considered on the surface S.

If we consider no longer right triangles, but triangles such that the angle between a and b is greater than 90°, $\angle(a,b) > 90°$,one such triangle is one point in the cone $C = \{(a,b,h) / h^2 > a^2 + b^2\}$. Then, if we built the following Hermitian Positive Definite matrix :

$$\begin{bmatrix} h & a - ib \\ a + ib & h \end{bmatrix} > 0 \Leftrightarrow h^2 > a^2 + b^2 \tag{160}$$

The problem is then reduce to previous study of *HPD(2)* matrices, where the coordinates system is given by (h, μ) where $\mu = \frac{a + ib}{h}$ is in Poincaré unit disk $|\mu| < 1$, and $h \in R^+$. Fréchet p-mean is computed by previous method in product space $R^+ xD$.

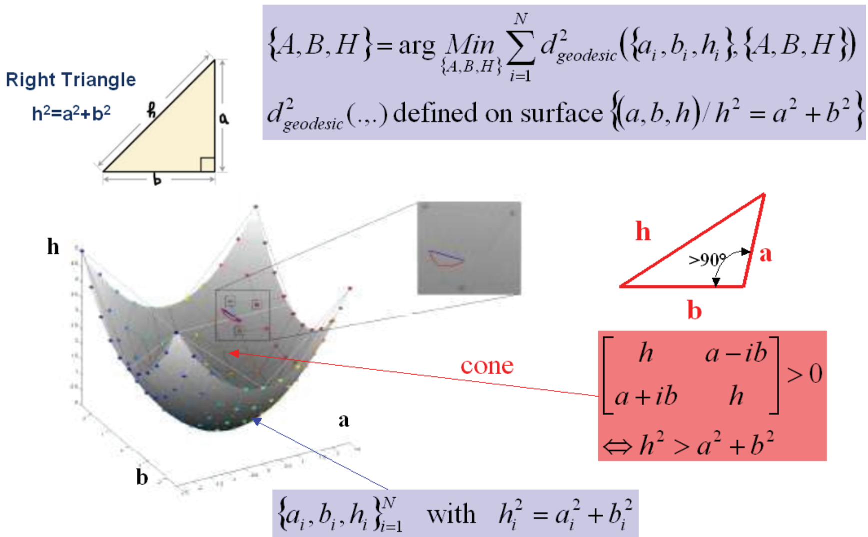

Fig. 21. Shape manifold of Right triangles and cone for triangle where $\angle(a,b) > 90°$

19 From Fréchet-median barycentre in Poincaré's disk to Douady-Earle Conformal barycentre on its boundary

I would like to conclude this paper with some remarks to compare the Fréchet-median barycentre in Poincaré unit disk with Douady-Earle Conformal barycentre on its boundary [116,117].

Considering G the group of all conformal automorphisms of Poincaré unit disk $D = \{z \in C / |z| < 1\}$ and G_+ the subgroup, of index 2 in G, or orientation preserving maps : $z \mapsto \lambda \frac{z-a}{1-a^* z}$ with $|z| = 1$ and $|a| < 1$. The group G operates on *D* but also on the set $P(S^1)$ of probability measure on $S^1 = \{z \in C / |z| = 1\}$. The principle is to assign to every probability measure μ on S^1 a point $B(\mu) \in D$ so that the map $\mu \mapsto B(\mu)$ is conform and satisfies:

$$B(\mu) = 0 \Leftrightarrow \int_{S^1} \zeta . d\mu(\zeta) = 0 \tag{161}$$

There is a unique conformally way to assign to each probability measure μ on S^1 a vector field ξ_μ on *D* such that:

$$\xi_\mu(0) = \int_{S^1} \zeta . d\mu(\zeta) = 0 \tag{162}$$

For general w in *D*, the assignment is given by $\mu \mapsto \xi_\mu$:

$$\xi_\mu(w) = \frac{1}{(g_w)'(w)} \xi_{(g_w)^*(\mu)}(0) = \left(1 - |w|^2\right) \int_{S^1} \left(\frac{\zeta - w}{1 - w^* \zeta} \right) d\mu(\zeta) = 0 \tag{163}$$

with $g_w(z) = \frac{z-w}{1-w^* z}$.

Douady-Earle definition of conformal barycenter is then the following:

The unique zero of ξ_μ in *D* is the conformal Barycenter $B(\mu)$ of μ

Demonstration uses that the Jacobian of ξ_μ at $w = 0$ is strictly positive:

$$J\xi_\mu(0) = \left|(\xi_\mu)_w(0)\right|^2 - \left|(\xi_\mu)_{w^*}(0)\right|^2 = \iint_{S^1 x S^1} \left| z^2 - \zeta^2 \right| d\mu(\zeta) x d\mu(z) > 0 \tag{164}$$

Douady and Earle provide a second proof of the uniqueness of $B(\mu)$ that will underline the link with Fréchet Median that we can compute by deterministic flow defined in section 9 (or its stochastic version by Arnaudon [84]).

$\xi_\mu(z)$ can be written according to $\xi_\zeta(z)$ that is the unit tangent vector of geodesic at $z \in D$ pointing toward $\zeta \in S^1$:

$$\xi_\mu(z) = \int_{S^1} \xi_\zeta(z) d\mu(\zeta) \tag{165}$$

In Poincaré geometry of unit disk, the vector field ξ_ζ is the gradient of a function h_ζ whose level lines are the horocycles tangent to S^1 at $\zeta \in S^1$:

$$\begin{aligned} &\xi_\mu = \nabla h_\mu \quad \text{with} \quad h_\mu : z \mapsto \int_{S^1} h_\zeta(z) d\mu(\zeta) \\ &h_\mu(z) == \int_{S^1} \frac{1}{2}\log\left(\frac{1-|z|^2}{|z-\zeta|^2}\right) d\mu(\zeta) = \int_{S^1} \underset{r \to 1^-}{Lim}\left[d(0,r) - d(z, r\zeta)\right] d\mu(\zeta) \end{aligned} \tag{166}$$

with $d(z,w)$ the Poincaré distance from *z* to *w* in *D*.

Uniqueness of $B(\mu)$, that is a critical point of h_μ, is given by strict convexity of $-h_\mu$ restriction to Poincaré geodesics.

We can then observe that the Fréchet median of N points in *D*, $\{w_1, w_2, w_3, ..., w_N\}$, is the conformal barycenter of associated push forward N points $\{\zeta_1, \zeta_2, \zeta_3, ..., \zeta_N\}$ on S^1:

$$\begin{aligned} &Median_{Fréchet}\{w_1, w_2, w_3, ..., w_N\} = Conformal\_Barycenter\{\zeta_1, \zeta_2, \zeta_3, ..., \zeta_N\} \\ &\zeta_i = \underset{t \to +\infty}{Lim} \gamma_{\xi_{\zeta_1}(z)}(t) = \underset{t \to +\infty}{Lim} \gamma_z^{w_i}(t) \end{aligned} \tag{167}$$

where the "push forward" association $\zeta_i = \underset{t \to +\infty}{Lim} \gamma_{\xi_{\zeta_1}(z)}(t) = \underset{t \to +\infty}{Lim} \gamma_z^{w_i}(t)$ is given by the limit point when *t* tends to infinity along the geodesic from the barycenter toward w_i.

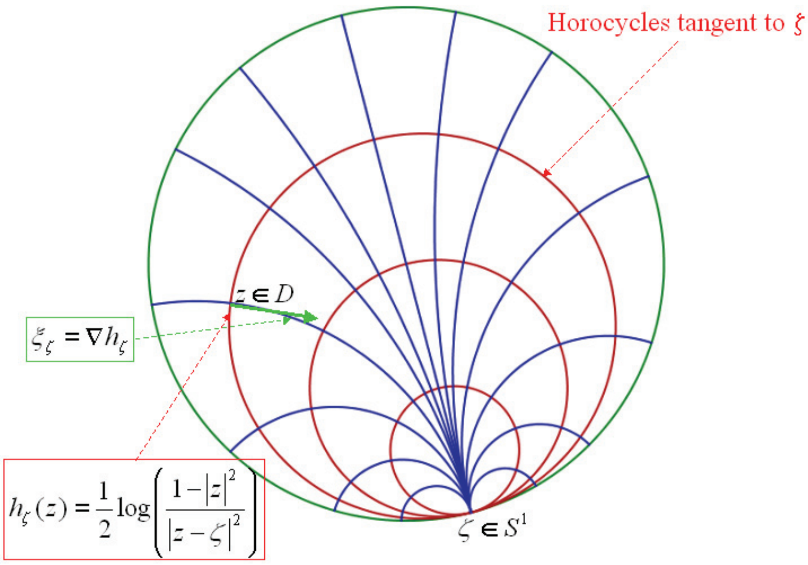

Fig. 22. Vector Field Gradient and horocycles level lines

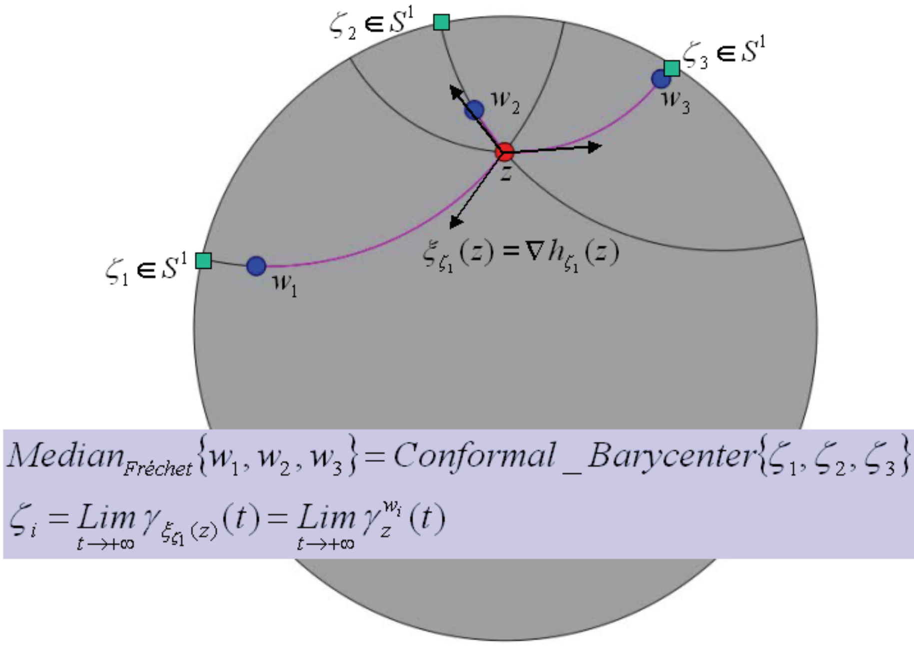

Fig. 23. Fréchet Median and Douady-Earle Conformal Barycenter

For points on Poincaré Disk boundary, the Karcher Flow of Fréchet median converges to Douady-Earle Conformal barycenter :

$$m_{n+1} = \gamma_n(t) = \exp_{m_n}\left(-t.\nabla f(m_n)\right) \text{ with } \dot{\gamma}_n(0) = -\nabla f(m_n)$$
$$f(m) = \frac{1}{2}\int_{S^1} d(m,a)da \underset{Min}{\Rightarrow} \nabla f = -\int_{S^1} \frac{\exp_m^{-1}(a)}{\left\|\exp_m^{-1}(a)\right\|} da \quad (168)$$

Frédéric Paulin [119] has constructed on the boundary of a hyperbolic group (in Gromov's sense) a natural visual measure and a natural crossratio and has defined a barycentre for every probability measure on the boundary, extending the Douady-Earle construction. Extension of Conformal structure for boundary of CAT(-1) space has also been studied by Marc Boudon in [118].

M. Itoh [120] has recently defined a map from complete Riemannian manifold of negative curvature to its boundary in term of its Poisson Kernel, as Douady and Earle map, to investigate geometry of the pull-back metric of the Fisher information metric by this map based on a paper of Friedrich [121] that studied the space of probability measure with respect to the Fisher information metric.

More recently, H. Airault, P. Malliavin and A. Thalmaier [122] have extended study of Brownian motion on the diffeomorphism group of the circle to Brownian motion on Jordan curves in C based on a Douady-Earle type conformal extension of vector fields on the circle to the disk. The aim of one of Malliavin's projects was the construction of natural measures on infinite dimensional spaces, like Brownian measures on the diffeomorphism group of the circle, on the space of univalent functions of the unit disk and on the space of Jordan curves in the complex plane. He understood that unitarizing measures for representations of Virasoro algebra can be approached as invariant measures of Brownian motion on the diffeomorphism group with a certain drift defined in terms of a Kähler potential [50,51,52,53].

20 Conclusion

Fréchet Median with Information Geometry and Geometry of of *HPD(n)* matrices is a new tool for Radar Signal Processing that could improve drastically performance and robustness of classical methods, in Doppler processing and in STAP. Obviously, these approaches could be extended to Array Processing and Polar Data Processing in the same way on respectively spatial covariance matrix and Polar covariance matrix. Future works will be dedicated to deepen close relations of Information Geometry with Lagrange Symplectic Geometry and Geometric Quantization.

I would like to to give many thanks to all member of Brillouin seminar for interesting discussion, hosted in IRCAM by Arshia Cont, since 2009. I am especially very graceful for Le YANG under supervision of Marc Arnaudon, who has proven consistency and convergence of all these algorithms with rigorous developments and generalizations [86].

In 1943 seminal Maurice Fréchet's paper [6], where he introduced for the first time what it is called nowdays "Cramer-Rao Bound", we can read at the bottom of first page "*Le contenu de ce mémoire a formé une partie de notre cours de statistique mathématique de l'Institut Henri Poincaré pendant l'hiver 1939-1940*". With help of Cédric Villani, new Intitut Henri Poincaré director, we have looked for this Fréchet Lecture in IHP without success. I have recently visited in Archive of French Academy of Science, quai Conti, the "Fonds Fréchet" that are made of 28 boxes with all original manuscripts, papers and works of Maurice Fréchet. For the time being, I have had only time to read papers of Box 16 where Fréchet study statistics of human profiles and cranes with computing resources of first IHP Computation Center. I hope to have enough time soon to explore all these "Fonds Fréchet" to find this historical Lecture. More recently, M. Emery gave me the advice to look for Fréchet's document in Pantheon-Sorbonne university Archive.

21 Appendix : Iwasawa, Cartan and Hua Coordinates

- ***Cartan Decomposition on Poincaré Unit Disk***

$D = \{z / |z| < 1\}$

$g \in SU(1,1)$ with $g = \begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix}$ and $g(z) = \dfrac{az+b}{b^* z + a^*}$ where $|a|^2 - |b|^2 = 1$

Cartan Decomposition : $g = u_\Phi d_\tau u_\Psi$

with $u_\Phi = \begin{pmatrix} e^{i\Phi} & 0 \\ 0 & e^{-i\Phi} \end{pmatrix}$ and $d_\tau = \begin{pmatrix} ch(t) & sh(t) \\ sh(t) & ch(t) \end{pmatrix}$

$$\Rightarrow \begin{cases} a = e^{i(\Phi+\Psi)} ch(\tau) \\ b = e^{i(\Phi-\Psi)} sh(\tau) \end{cases} \Rightarrow z = b\left(a^*\right)^{-1} = th(\tau) e^{i2\Phi}$$

$$ds^2 = 8\left(d\tau^2 + sh^2(2\tau) d\theta^2\right) \Rightarrow \Delta_{LB}. = \frac{\partial^2 .}{\partial \tau^2} + \coth(2\tau)\frac{\partial .}{\partial \tau} + \frac{1}{sh^2(\tau)}\frac{\partial^2 .}{\partial \Phi^2}$$

$$F(z) = -\ln\left(1 - |z|^2\right) = 2\ln ch(\tau)$$

- ***Iwasawa Decomposition on Poincaré Unit Disk Disk (Lemma of Iwasawa for radial coordinates in Poincaré Disk)***

$D = \{z / |z| < 1\}$

$g \in SU(1,1)$ with $g = \begin{pmatrix} a & b \\ b^* & a^* \end{pmatrix}$ and $g(z) = \dfrac{az+b}{b^* z + a^*}$ where $|a|^2 - |b|^2 = 1$

Iwasawa Dec.: $g = h\left(K_\theta D_\tau N_\xi\right)$ with $h(g) = CgC^{-1}$, $C = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$

$K_\theta = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$, $D_\tau = \begin{pmatrix} ch(t) & sh(t) \\ sh(t) & ch(t) \end{pmatrix}$ and $N_\xi = \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix}$

$$\Rightarrow \begin{cases} a = e^{i\theta/2}\left(ch(\tau/2) + i\dfrac{u}{2}e^{-\frac{\tau}{2}}\right) \\ b = e^{i\theta/2}\left(sh(\tau/2) - i\dfrac{u}{2}e^{-\frac{\tau}{2}}\right) \end{cases} \text{ with } u = \xi e^\tau$$

- ***Hua-Cartan Decomposition on Siegel Unit Disk (Lemma of Hua for radial coordinates in Siegel Disk)***

$\tau = [\tau_1 \quad \tau_2 \quad \cdots \quad \tau_n]$ with $0 \le \tau_n \le \tau_{n-1} \le \cdots \le \tau_1$

$A_0(\tau) = diag[ch(\tau_1) \quad ch(\tau_2) \quad \cdots \quad ch(\tau_n)]$

$B_0(\tau) = diag[sh(\tau_1) \quad sh(\tau_2) \quad \cdots \quad sh(\tau_n)]$

$$g = \begin{bmatrix} A & B \\ B^* & A^* \end{bmatrix} \in Sp(n) \ , g = \begin{bmatrix} U^t & 0 \\ 0 & U^+ \end{bmatrix} \begin{bmatrix} A_0 & B_0 \\ B_0 & A_0 \end{bmatrix} \begin{bmatrix} V^* & 0 \\ 0 & V \end{bmatrix}$$

there exist U and V unitary complex matrices of order n

$$\begin{cases} A = U^t A_0(\tau) V^* \\ B = U^t B_0(\tau) V \end{cases} \Rightarrow \begin{pmatrix} A_0(\tau) & B_0(\tau) \\ B_0(\tau) & A_0(\tau) \end{pmatrix} = \exp \begin{pmatrix} 0 & Z_2(\tau) \\ Z_2(\tau) & 0 \end{pmatrix}$$

with $Z_2(\tau) = diag[\tau_1 \quad \tau_2 \quad \cdots \quad \tau_n]$

Let $Z = B(A^*)^{-1} = U^t P U$, $P^2 = B_0^2 (A_0^{-1})^2 = \mathrm{diag}[eigen(ZZ^+)]$

$P = diag[th(\tau_1) \quad th(\tau_2) \quad \cdots \quad th(\tau_n)]$

- ***Iwasawa Decomposition on Siegel Unit Disk (Iwasawa coordinates in Siegel Disk)***

$SD_n = \{Z / ZZ^+ < I\}$ and $g(Z) = (AZ + B)(B^* Z + A^*)^{-1}$

$$g = \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix}, h(g) = CgC^{-1} \text{ with } C = \frac{1}{\sqrt{2}} \begin{pmatrix} I & -iI \\ I & iI \end{pmatrix}$$

$$K = \left\{ g / h(g) = \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix}, U \text{ unitary order n} \right\} \Rightarrow g = \frac{1}{2} \begin{pmatrix} U + U^* & -i(U - U^*) \\ i(U - U^*) & U + U^* \end{pmatrix} = C^{-1} \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} C$$

$$A = \left\{ A = \begin{pmatrix} A_0 + B_0 & 0 \\ 0 & A_0 - B_0 \end{pmatrix} = \begin{pmatrix} diag[e^{\tau_1} \quad \cdots \quad e^{\tau_n}] & 0 \\ 0 & diag[e^{-\tau_1} \quad \cdots \quad e^{-\tau_n}] \end{pmatrix} \right\}$$

$$N = \left\{ N / N = \begin{Bmatrix} I & S \\ 0 & I \end{Bmatrix}, S \text{ real matrix of order n} \right\}$$

$$h(A) = \begin{pmatrix} A_0 & B_0 \\ B_0 & A_0 \end{pmatrix}, \ h(N) = \begin{pmatrix} I + i/2.S & -i/2.S \\ i/2.S & I - i/2.S \end{pmatrix} \Rightarrow h(KAN) = \begin{pmatrix} A_1 & B_1 \\ B_1^* & A_1^* \end{pmatrix}$$

$$\text{with } \begin{cases} A_1 = U\left[A_0 + i(A_0 + B_0) \frac{1}{2} S \right] \\ B_1 = U\left[B_0 - i(A_0 + B_0) \frac{1}{2} S \right] \end{cases}$$

- **Iwasawa/Cartan Coordinates on Siegel Unit Disk (Iwasawa/Cartan coordinates relation in Siegel Disk)**

$$M_S = \begin{pmatrix} I + i/2.S & -i/2.S \\ i/2.S & I - i/2.S \end{pmatrix},\ \begin{pmatrix} A_0 & B_0 \\ B_0 & A_0 \end{pmatrix} M_S = M_{\tilde{S}} \begin{pmatrix} A_0 & B_0 \\ B_0 & A_0 \end{pmatrix}$$

with $(A_0 + B_0)S = \tilde{S}(A_0 - B_0)$

$$g = \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \Rightarrow \begin{cases} \text{Cartan}: \begin{cases} A = U^t A_0 V^* \\ B = U^t B_0 V \end{cases} \\ \text{Iwasawa}: \begin{cases} A = U_1 \left[A_0 + i(A_0 + B_0)\frac{1}{2} S \right] \\ B = U_1 \left[B_0 - i(A_0 + B_0)\frac{1}{2} S \right] \end{cases} \end{cases}$$

$$Z = B\left(A^*\right)^{-1} = U_1 H U_1^t = U^t P U$$

with $H = \left[B_0(A_0 + B_0) - \frac{i}{2}\tilde{S} \right]\left[A_0(A_0 + B_0) - \frac{i}{2}\tilde{S} \right]^{-1}$

22 References:

1. Feliks Aleksandrovich Berezin. Quantization in complex symmetric spaces. Math. USSR, n°9, pp.341-379, 1975.
2. Rajendra Bhatia. Positive Definite Matrices (Princeton Series in Applied Mathematics). Princeton University Press, illustrated edition edition, December 2006.
3. Elie Cartan. Groupes simples clos et ouverts et géométrie riemannienne. Journal de mathématiques pures et appliquées, (8), pp. 1-34, 1929
4. Elie Cartan. Sur les domaines bornés de l'espace de n variables complexes. Abh. Math. Seminar Hamburg, 1935.
5. Nikolai Nikolaevich Chentsov. Statistical Decision Rules and Optimal Inferences. Transactions of Mathematics Monograph, numero 53, 1982. Published in Russian in 1972.
6. Maurice Fréchet. Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. Review of the International Statistical Institute, 11(3/4), pp.182-205, 1943. published in IHP Lecture in 1939.
7. Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. Annales de l'institut Henri Poincaré, 10(4), pp.215-310, 1948.
8. Loo-Keng Hua. Harmonic analysis of functions of several complex variables in the classical domains. American Mathematical Society, 1958.
9. K. Iwasawa. On some types of topological groups. Ann. of Math. (2), 50, pp. 507-558, 1949.
10. Hermann Karcher. Riemannian center of mass and mollifier smoothing. Communications on Pure and Applied Mathematics, 30(5), pp.509-541, 1977.
11. Jean-Louis Koszul. Sur la forme hermitienne canonique des espaces homogènes complexes. Can. J. of. Math., (7), pp.562-576, 1955.
12. Jean-Louis Koszul. Exposés sur les espaces homogènes symétriques. Publicacao da Sociedade de Matematica de Sao Paulo, 1959.
13. Ilya Piatetskii-Shapiro. Automorphic Functions and the Geometry of Classical Domains. Gordon & Breach Science Publishers, 1969. 340 pages.

14. Henri Poincaré. Les fonctions analytiques de deux variables et la représentation conforme. Rendiconti del Circolo Matematico di Palermo (1884 - 1940), 23, pp.185-220, 1907.
15. Calyampudi Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society, 37, pp.81-89, 1945.
16. Hirohiko Shima. The geometry of Hessian structures. World Scientic, 2007.
17. C. L. Siegel. Symplectic geometry. Amer. J. Math., (65), pp.1-86, 1943.
18. C. L. Siegel. Analytic functions of several complex variables. Princeton Univ. Press, 1949.
19. Ernest Borisovich Vinberg. The theory of convex homogeneous cones. Trans. Amer. Math. Soc., (12), 1963. Trudy Moskov. Mat. Obshch.
20. L.K. Hua,. Inequalities involving determinant, Acta Math. Sinica,n°5, pp. 463-470, 1955
21. M. Émery, G. Mokobodzki,. Sur le barycentre d'une probabilité dans une variété , Sém. de proba. Strasbourg, n°25, p. 220-233, 1991
22. J. Mitchell. A Theorem in the Geometry of Matrices, Proc. AMS, vol.2, n°2, April 1951
23. J. Mitchell. Potential Theory in the Geometry of Matrices, Trans. AMS, vol.79,n°2, 1955
24. J. Mitchell. The kernel function in the geometry of matrices, Duke Math. J. vol.19, pp.575-584, 1944
25. G. Fubini. The distance in General Fuchsian Geometries, PNASpp.700-708, 1940
26. R. Hermann. Geometric Aspects of Potential Theory in the Bounded Symmetric Domains, Math. Annalen, 148, pp.349-366, 1962
27. R. Hermann. Geometric Aspects of Potential Theory in the Bounded Symmetric Domains II, Math. Annalen, 151, pp.143-149, 1963
28. R. Hermann. Geometric Aspects of Potential Theory in the Bounded Symmetric Domains III, Math. Annalen, 153, pp.384-394, 1963
29. R. Hermann. Geodesics of Bounded, Symmetric Domains, CMH, vol.35, pp.1-8, 1960
30. C.T. Sturm. Nonlinear Martingale Theory for Processes with Values in Metric spaces of Nonpositive Curvature,The Annals of Probability, vol.30,n°3, pp.1195-1222, 2002
31. J. Picard. Barycentres et martingales sur une variété, Ann. IHP, vol.30, pp.647-702, 1994
32. S. Doss. Sur la moyenne d'un élément aléatoire dans un espace distancié, Bull. Sci. Math., 73, pp.48-72, 1949
33. A. Es-Sahib & H. Heinrich. Barycentres canoniques sur un espace métrique à courbure négative, Séminaire de probabilités XXXIII, Lecture Notes in Math, 1709, pp.355-370, Springer, 1999
34. W. Herer. Espérance mathématique au sens de Doss d'une variable aléatoire dans un espace métrique, C.R. Acad. Sci. Paris Ser.I, 302, pp.131-134, 1991
35. A.B. Tumpach. Mostow's Decomposition Theorem for L*-groups and Applications to Affine coadjoint Orbits and Stable Manifolds, preprint 2011
36. A.B. Tumpach. Variétés Kählériennes et hyperkählériennes de dimension infinie, PhD, Ecole Polytechnique, Palaiseau, France, July 2005.
37. S. Bergman. The Kernel Function and Conformal Mapping, Mathematical Survey and Monographs,n°5 AMS, 2000
38. A. Weil. Introduction à l'étude des variétés Kählériennes, actualités scientifiques et industrielles 1267, Publications de l'Institut de Mathématiques de l'université de Nancago VI, Ed. Hermann Paris, 1958
39. A. Lichnerowicz. Variétés complexes et tenseur de Bergmann. Annales de l'institut Fourier, 15 no. 2, p. 345-407, 1965
40. A. Lichnerowicz. Sur les groupes d'automorphismes de certaines variétés kählériennes, C. R. Acad. Sci., Paris 239, 1344-1346, 1954

41. J.L. Clerc & K. Koufani. Primitive du cocycle de Maslov généralisé, Mathematische Annalen 337, pp.91-138, 2007
42. J. Leray. Complément à la théorie d'Arnold de l'indice de Maslov, Convegno di geometrica simplettica et fisica matematica, Instituto di Alta Matematica, Roma, 1973
43. J. Leray. Analyse lagrangienne et mécanique quantique, Publications de l'Université Louis Pasteur, Strasbourg, 1978
44. J.-M. Souriau. Construction explicite de l'indice de Maslov, Group Theoretical Methods in Physics, Lecture Notes in Physics, vol. 50, Springer-Verlag, pp. 17–148, 1975
45. R. Godement. Introduction à la théorie des groupes de Lie, Springer, 2004
46. R. Mneimné, F. Testard. Introduction à la théorie des groupes de Lie classiques, Hermann, 1986
47. J.M. Souriau. La structure Symplectique de la mécanique décrite par Lagrange en 1811, Journée d'histoire des sciences de Marseille, « Mécanique et Mathématiques », Vieille Charité, 4-5 Octobre 1985
48. N.K. Sung, G. Stangenhaus, H.T. David. A Cramer-Rao analogue for Median-Unbiaised Estimators, Trabajos De Estadistica, vol.5,n°2, pp.83-94, 1990
49. J. Burbea and C. R. Rao. Differential metrics in probability spaces, Probability and Mathematical Statistics, Vol. 3, Fasc. 2, pp. 241-258, 1984.
50. H. Airault and P. Malliavin. Regularized Brownian Motion on the Siegel Disk of Infinite Dimension, Ukrainian Mathematical Journal,vol.52, n°9, pp.1324-1333, 2000
51. P. Malliavin. The canonical diffusion above the diffeomorphism group of the circle, C.R. Acad. Sci., 329, pp.325-329, 1999
52. P. Malliavin. Invariant or quasi-invariant probability measures for infinite dimensional groups, Japan. J. Math., n°3, pp.19-47, 2008
53. H. Airault. Stochastic analysis on finite dimensional Siegel disks, approach to the finite dimensional Siegel disk and upper-half plane, Bulletin des sciences mathématiques, vol. 128, no7, pp. 605-659, 2004
54. M. Berger. Les espaces symétriques non-compact, Annales scientifques de l'E.N.S., 3ème série, tome 74,n°2, pp.85-177, 1957
55. G. Larotonda. Geodesic Convexity Symmetric Spaces and Hilbert-Schmidt operators, PhD thesis, Buenos Aires, 2004
56. S. A. Al-Bashrawi. Reductive Lie Groups and Mostow Fibration, Master, Dhahran, 2004
57. G.D. Mostow. Covariant Fiberings of Klein spaces, American Journal of Mathematics, 77, pp.247-278, 1955
58. G. D. Mostow . Covariant Fiberings of Klein Spaces, II, American Journal of Mathematics, Vol. 84, No. 3 , pp. 466-474, 1962
59. H. Maass. Lectures on Siegel's Modular Functions, Tata Institute of Fundamental Research, Bombay, 1955, noted by T.P. Srinivasan
60. C.L. Siegel. Lectures on Quadratic Forms, Tata Institute of Fundamental Research, Bombay, 1957, noted by K.G. Ramanathan
61. M.I. Zelikin. Control Theory and Optimization I (Homogeneous Spaces and the Riccati Equation in the Calculus of Variations), Springer, 2000
62. F. Barbaresco. Algorithme de Burg Régularisé FSDS, Comparaison avec l'algorithme de Burg MFE, XVème colloque GRETSI, vol.1, pp.29-32, September 1995
63. F. Barbaresco. Super Resolution Spectrum Analysis Regularization : Burg, Capon and Ago-antagonistic Algorithms, EUSIPCO-96, pp.2005-8, Trieste, Sept.1996
64. F. Barbaresco. Recursive Eigendecomposition via Autoregressive Analysis and Ago-antagonistic Regularization, Proc. ICASSP-97, Munich, Avril 1997

65. F. Barbaresco. Innovative Tools for Radar Signal Processing Based on Cartan's Geometry of SPD Matrices and Information Geometry, IEEE RadarCon Rome, May 2008
66. F. Barbaresco. Interactions between Symmetric Cone and Information Geometries : Bruhat-Tits and Siegel Spaces Models for High Resolution Autoregressive Doppler Imagery, ETVC'08 Conf., Ecole Polytechnique, Nov. 2008, published by Springer, in Lecture Notes in Computer Science, vol.5416, February 2009
67. F. Barbaresco & G. Bouyt. Espace Riemannien symétrique et géométrie des espaces de matrices de covariance : équations de diffusion et calculs de médianes, GRETSI conference, Dijon, September 2009
68. F. Barbaresco. New Foundation of Radar Doppler & Array Processing Based on German Mathematical Works, International Radar Symposium, Hamburg, Sept. 2009
69. M. Oudin, J.P. Delmas & F. Barbaresco. Siegel distance-based covariance matrix selection for Space-Time Adaptive Processing, Radar'09, Bordeaux, October 2009
70. F. Barbaresco. New Foundation of Radar Doppler Signal Processing based on Advanced Differential Geometry of Symmetric Spaces: Doppler Matrix CFAR and Radar Application, Radar'09 Conference, Bordeaux, October 2009
71. N. Charon & F. Barbaresco. Une nouvelle approche pour la détection de cibles dans les images radar basée sur des distances et moyennes dans des espaces de matrices de covariance, Revue traitement du signal, volume 26, n°4, Décembre 2009
72. F. Barbaresco. Robust Median-Based STAP in Inhomogeneous Secondary Data :Frechet Information Geometry of Covariance Matrices, 2nd French-Singaporian SONDRA Workshop on "EM Modeling, New Concepts and Signal Processing For Radar Detection and Remote Sensing, Cargese France, 25-28 May 2010
73. F. Barbaresco. Geometric Science of Information: Modern Geometric Foundation of Radar Signal Processing, 8th International IEEE Conference on Communications, Bucharest, Romania, 10th-12th June, 2010
74. C. Chaure & F. Barbaresco. New Generation Doppler Radar Processing : Ultra-fast Robust Doppler Spectrum Barycentre Computation Scheme in Poincaré's Unit Disk, EURAD'10 conference, Paris, Sept. 2010
75. Le Yang, M. Arnaudon and F. Barbaresco. Riemannian Median, Geometry of Covariance Matrices and Radar Target Detection, EURAD'10 conference, Paris, Sept. 2010
76. Le Yang, M. Arnaudon and F. Barbaresco. Geometry of Covariance Matrices and Computation of Median, MaxEnt'10 conference,France, July 2010
77. F. Barbaresco. Science géométrique de l'Information : Géométrie des matrices de covariance, espace métrique de Fréchet et domaines bornés homogènes de Siegel, Conférence GRETSI'11, Bordeaux, Sept. 2011
78. F. Barbaresco. Robust Statistical Radar Processing in Fréchet Metric Space: OS-HDR-CFAR and OS-STAP Processing in Siegel Homogeneous Bounded Domains, Proceedings of IRS'11, International Radar Conference, Leipzig, Sept. 2011
79. F. Barbaresco. Geometric Radar Processing based on Fréchet Distance : Information Geometry versus Optimal Transport Theory, Proceedings of IRS'11, International Radar Conference, Leipzig, Sept. 2011
80. Bhashyam Balaji, F. Barbaresco. Application of Riemannian Mean of Covariance Matrices to Space-Time Adaptive Processing, submitted to IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), USA 2012
81. M. Arnaudon. Espérances conditionnelles et C-martingales dans les variétés, Séminaire de Probabilités-XXVIII, Lecture Notes in Mathematics 1583. (Springer, Berlin, 1994), pp. 300-311.

82. M. Arnaudon. Barycentres convexes et approximations des martingales dans les variétés, Séminaire de Probabilités-XXIX, Lecture Notes in Mathematics 1613. (Springer, Berlin, 1995), pp. 70-85.
83. M. Arnaudon and X. M. Li. Barycenters of measures transported by stochastic flows, The Annals of probability, 33 (2005), no. 4, 1509-1543.
84. M. Arnaudon, C. Dombry, A.Phan, L.Yang. Stochastic algorithms for computing means of probability measures, Stochastic Processes and their Applications, to appear , http://hal.archives-ouvertes.fr/hal-00540623
85. Le Yang. Riemannian Median and Its Estimation, LMS Journal of Computation and Mathematics, v.13, pp 461 – 479, 6 January 2011
86. Le Yang. Medians of probability measures in Riemannian manifolds and applications to radar target detection, PhD Thesis, 15th December 2011, Poitiers
87. L. Yang, M. Arnaudon and F. Barbaresco. Geometry of covariance matrices and computation of median, AIP Conf. Proc. Volume 1305 (2011), pp. 479-486.
88. J. R. Roman & al. Parametric Adaptive Matched Filter for Airborne Radar Applications, IEEE Trans. on AES, Vol. 36, N°. 2, pp. 677-692, April 2000
89. M. Wax, T. Kailath. Efficient Inversion of Toeplitz-Block Toeplitz Matrix, IEEE Trans. ASSP, vol.31, n°5, pp.1218-1221, Oct. 1983
90. R. Kanhouche. A Modified Burg Algorithm Equivalent In Results to Levinson Algorithm, CMLA preprint, Sept. 2003, http://hal.inria.fr/docs/00/00/06/24/PDF/BurgLevRK.pdf
91. G. Monge. Mémoire sur la théorie des déblais et des remblais, Histoire de l'Académie des Sciences de Paris, p. 666-704, 1781
92. M. Fréchet. Sur la distance de deux lois de probabilité, C.R. Acad. Sci. Paris, n°244, p.689-692, 1957
93. P. Levy. Distance de deux variables aléatoires et distance de deux lois de probabilité, Généralités sur les probabilités de variables aléatoires, Ed. M. Fréchet, Gauthier-Villars, Paris, p. 286-301, 1937
94. R. Fortet and E. Mourier. Convergence de la répartition empirique vers la répartition théorique, Ann. Sci. Ecole Norm. Sup. N°70, p.267-285, 1953
95. Y. Brenier. Polar factorisation and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. , vol.44,n°4,p.375-417, 1991
96. Cédric Villani. Optimal Transport, Old and New, Springer-Verlag 2008, http://math.univ-lyon1.fr/homes-www/villani/Cedrif/B07D.StFlour.pdf
97. D.C. Dowson and B. V. Landau. The Fréchet Distance between Multivariate Normal Distributions, Journal of Multivariate Analysis, n°12, p.450-455, 1982
98. R.J. McCann. A convexity Principle for Interacting Gases, Adv. Math., vol.128, n°1, p. 153-179, 1997
99. A. Takatsu. On Wasserstein Geometry of the Space of Gaussian Measures, to appear in Osaka J. Math., 2011, http://arxiv.org/abs/0801.2250
100. M. Calvo & J.M. Oller. A Distance between Multivariate Normal Distribution Based in an Embedding into the Siegel Group, Jour. of Multivariate Analysis, vol.35,n°2,p.223-242, 1990
101. M. Agueh & G. Carlier. Barycenters in the Wasserstein Space, SIAM J.Math.Anal., 2011
102. G. Dall'Aglio. Fréchet Classes: The Beginnings, Advances in Probability Distributions with Given Marginals, Mathematics and Its Applications, Kluwer, 1991
103. M. Fréchet, Sur la précision comparée de la moyenne et de la médiane, Aktuarské Vedy, n°5, pp. 29-34, 1935
104. M. Fréchet, Comparaison des diverses mesures de la dispersion, Revue, de l'Institut International de Statistique, 8 (1), 1-12., 1940

105. A.N. Kolmogorov, The Method of the Median in the Theory of Errors,, Matermatichiskii Sbornik, 38, 47-50, 1931
106. Roger Koenker, The Median is the message: toward the Fréchet Median, Journal de la Société Française de Statistique, tome 147, n◦2, 2006
107. M. Armatte. Fréchet et la médiane : un moment dans une histoire de la robustesse, Journal de la Société Française de Statistique, tome 147,n°2, 2006
108. E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum, Tohoku Math. Journal 43: 355–386., 1937
109. G. Bassanelli, On horospheres and holomorphic endomorphisms of the Siegel disc, Rendiconti del Seminario Matematico della Università di Padova, t.70,pp.147-165, 1983
110. J.L. Koszul, Domaines bornés homogènes et orbites de groupes et transformations affines, Bulletin de la SMF , t.89, pp.515-533, 1961
111. Frank Nielsen and Richard Nock. 2010. Hyperbolic Voronoi Diagrams Made Easy. In Proceedings of the 2010 International Conference on Computational Science and Its Applications (ICCSA '10). IEEE Computer Society, Washington, DC, USA, 74-80. DOI=10.1109/ICCSA.2010.37, http://dx.doi.org/10.1109/ICCSA.2010.37
112. Marc Arnaudon, Frank Nielsen: On Approximating the Riemannian 1-Center CoRR abs/1101.4718:, 2011
113. Geometry of Homogeneous Bounded Domains, CIME Summer Schools Urbino Italy 1967, published by Springer 2011
114. Karpelevitch, F. I. , The geometry of geodesics and the eigenfunctions of the Beltrami-Laplace operator on symmetric spaces. Trudy Moskov. Mat. Obsc. 14, pp. 48-185; transl. as Trans. Moscow Math. Soc. , pp. 51-199, 1967
115. G. D. Mostow, Some new decomposition theorems for semi-simple groups, Mem. Amer. Math. Soc. (1955), p. 31-54
116. A. Douady and C.J. Earle, Conformally natural extension of homeomorphisms of circle, Acta Math. 157, pp.23-48
117. C.J. Earle, Angular derivatives of the barycentric extension, Complex variables Theory Appl. 11, p.189-195, 1986
118. M. Bourdon, Structure conforme au bord et flot géodésique d'un CAT(-1)-espace, L'enseignement mathématique, t.41, pp.63-102, 1995
119. F. Paulin, Un groupe hyperbolique est déterminé par son bord, J. London Math. Soc. 54 , pp. 50-74, 1996
120. M. Itoh and Y. Shishido, Fisher information metric and Poisson kernels, Differential Geom. Appl. , Vol 26, pp. 347-356, 2008
121. T.Friedrich, Die Fisher-Information und symplektische Strukturen, Math. Nachr. 153, pp.273-296, 1991
122. H. Airault, P. Malliavin and A. Thalmaier, Brownian measures on Jordan-Virasoro curves associated to the Weil-Petersson metric, Journal of Functional Analysis 259, pp. 3037-3079, 2010

« Il est clair que si l'on parvenait à démontrer que tous les domaines homogènes dont la forme $\Phi = \sum_{i,j} \frac{\partial^2 \log K(z,\bar{z})}{\partial z_i \partial \bar{z}_j} dz_i d\bar{z}_j$ *est définie positive sont symétriques, toute la théorie des domaines bornés homogènes serait élucidée. C'est là un problème de géométrie hermitienne certainement très intéressant »*

Last sentence in Elie Cartan, « Sur les domaines bornés de l'espace de n variables complexes », Abh. Math. Seminar Hamburg, 1935

Radar Detection for Non-Stationary Time-Doppler Signal based on Fréchet Distance of Geodesic Curves on Covariance Matrix Information Geometry Manifold

Frédéric Barbaresco[*]

[*]Thales Air Systems, Surface Radar Domain, Technical Directorate
Voie Pierre-Gilles de Gennes, 91470 Limours, France
email: frederic.barbaresco@thalesgroup.com

Abstract: *New challenge in Radar is the processing of non-stationary signal corresponding to fast time variation of Doppler Spectrum in one burst. We can observe this phenomenon for high speed or abrupt Doppler variations of clutter or target signal but also in case of target migration during the burst duration due to high range resolution. We will assume that each non-stationary target signal in one burst can be split into several short signals with less Doppler resolution but stationary, represented by time sequence of stationary covariance matrices or a geodesic polygon on covariance matrix manifold. For micro-Doppler analysis of these cases, we adapt the Fréchet distance between two curves in the plane with a natural extension to more general geodesic curves in abstract metric spaces used for covariance matrix manifold. This new approach could be used for robust detection of target with non-stationary Time-Doppler spectrum (NS-OS-HDR-CFAR: Non-Stationary Ordered Statistic High Doppler Resolution CFAR), False Alarm filtering for inhomogeneous clutter (statistics of plots Time-Doppler fluctuation), but also emerging civil applications like wind-shear, micro-burst, downdraft and wake-vortex detection.*

1. Non-stationary Time-Doppler signal in Radar: Targets and Clutters

Non stationary Doppler signal is challenging for radar processing. Main processing chains implemented in radars make the assumption of Doppler stationary signal during the burst waveform duration. This assumption is not always true, especially in case of high speed Doppler fluctuation or in case of abrupt Doppler changes as observed in fig1. For helicopter.

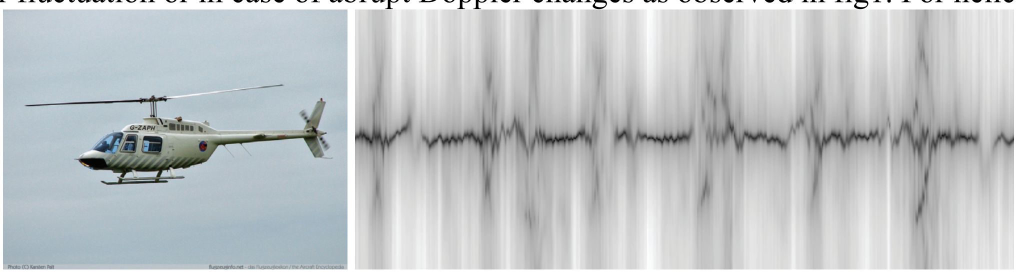

Figure 1. BELL 206 helicoper Time-Doppler Signature in one radar cell in staring mode

Non stationarity of Doppler signal is also the main cause of false alarm due to abrupt/fast Doppler variations of clutter like inhomogeneous sea clutter where spikes or breaking waves echoes will generate false detection due to these furtive events, as illustrated in fig. 2.

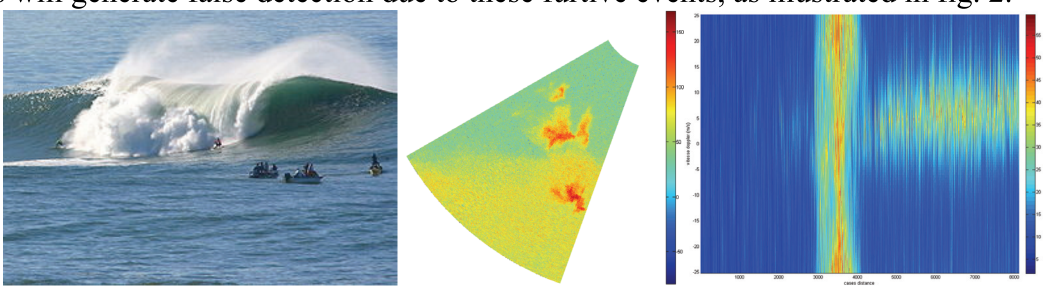

Figure 2. Non-homogenesous Sea clutter (with rain clutter) : fast Doppler spectrum fluctuation of sea clutter

Non stationary Time-Doppler signature could be also observed by civil radar in the framework of safety airport issue, for wind hazards monitoring like wind-shear/micro-burst/downdraft/wake-vortex, as illustrated in the following two images.
(see UFO: http://www.transport-research.info/web/projects/project_details.cfm?ID=45116)

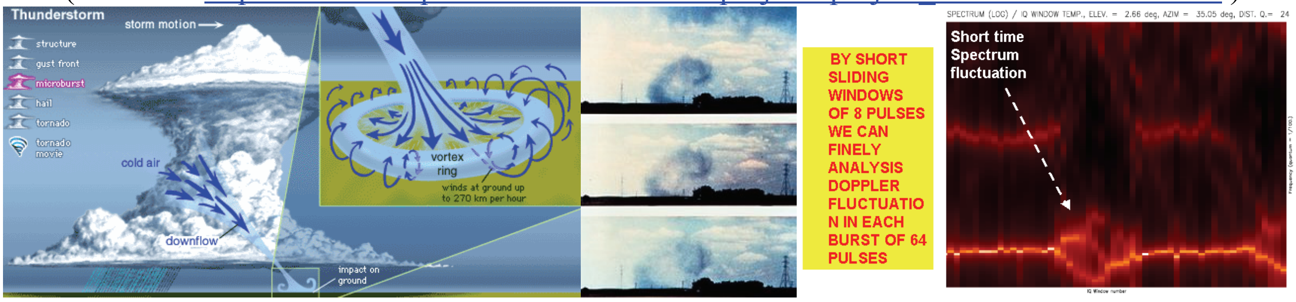

Figure 3. Time Doppler signatures of wind shear instabilities and fast downdrafts

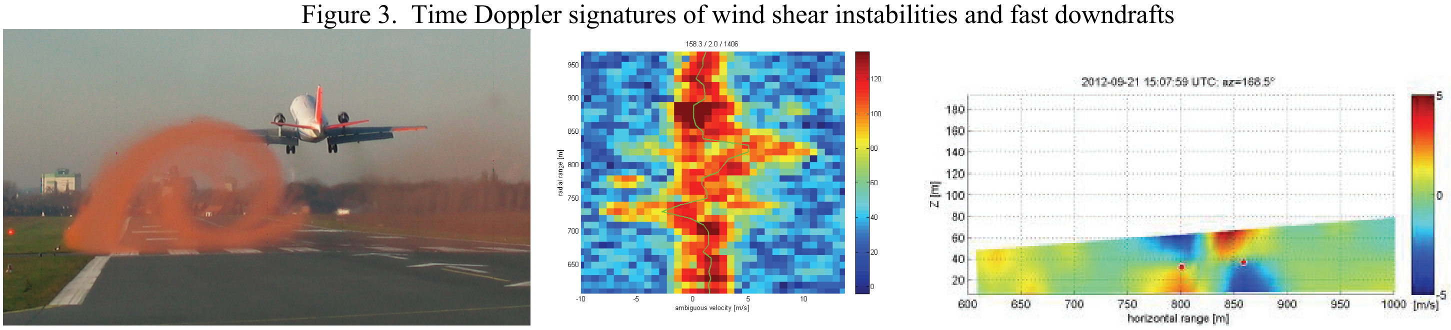

Figure 4. Doppler signature of wake-vortex in ground effect, with strength (circulation) depending of aircraft (mass/wingspan x speed) ratio at generation time and decaying according to Eddy Dissipation Rate

Wake-Vortex Time-Doppler Signatures are also observed when they are generated by helicopter blades or wind-farm blades.

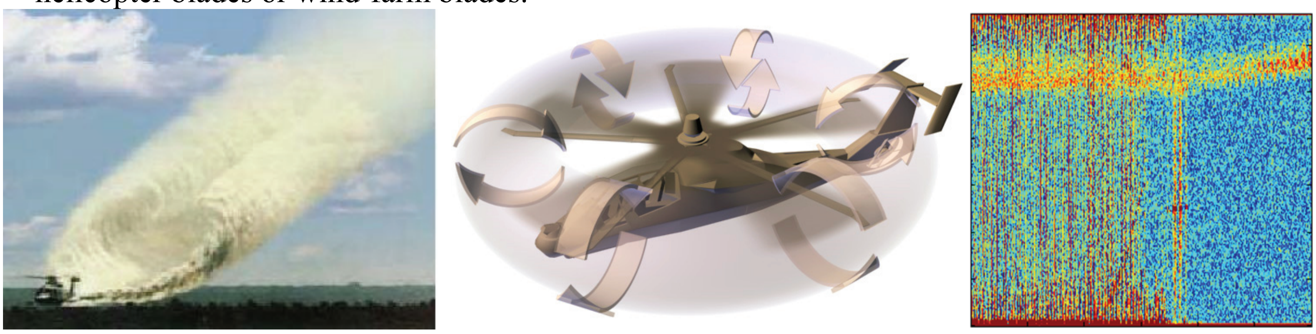

Figure 5. Doppler signature of helico blades turbulences and vortices generating during flights phases (landing/take-off/cruise) or behind a moutain crest-line before furtive helico pop-up

Last case is about radar where target is migrating during the burst duration due to high range resolution. Target could be observed only on some pulses of the burst but not on all pulses of the waveform.

In the following, we will consider Time-Doppler signature as a geodesic path on an Information Geometry manifold (of covariance matrices) as explained in chapter 2 by analyzing the time series in the burst with a short sliding window. In chapter 3, we recall classical definition of distance between paths in Euclidean Space based on Fréchet's works. In chapter 4, we extend this classical Fréchet distance between paths to be used for our Time-Doppler geodesic paths on Information Geometry Manifolds and for NS-OS-HDR-CFAR.

2. Covariance Matrix Information Geometry Manifold: Geodesic and Distance between Micro Doppler spectrums

In the burst, we can consider through a sliding window a sub-set of pulses where we could consider the signal as locally stationary. For this sub-set of pulses, we use the Trench theorem [6] proving that their THPD (Toeplitz Hermitian Positive Definite) Covariance matrix could be parameterized by Complex Auto-Regressive (CAR) model for the time series. All THPD matrices are diffeomorphic to $(r_0, \mu_1,..., \mu_n) \in R^+ x D^n$ (r_0 is a real "scale" parameter, μ_k are called reflection/Verblunsky coefficients of CAR model in D the complex unit Poincare disk, and are "shape" parameters). This result has been found previously by Samuel Verblunsky in 1936 [7]. We have observed that this CAR parameterization of the THPD matrix could be also interpreted as Partial Iwasawa decomposition of the matrix in Lie Group Theory [5]. At this step, to introduce the "natural" metric of this model, we used jointly Burbea/Rao [5] results in Information Geometry and Koszul [5] results in Hessian geometry, where the conformal metric is given by the Hessian of the Entropy of the CAR model. This metric has all good properties of invariances, and could be also recover as Information Geometry metric. Distance between 2 THPD matrices is then easily computed by distance in product space : r_0 on R^+, and μ_k in D^n Poincare unit polydisk. To regularize the CAR inverse problem on very short sub-set of Burst pulses, we have used a "regularized" Burg reflection coefficient [8] avoiding prior selection of AR model order (impossible to retrieve with Ikaike criterion).

In 1945, Rao has introduced Information Geometry for parameterized density of probability $p(./\theta)$ with the metric given by the formula $ds^2 = K[p(./\theta), p(./\theta + d\theta)] = d\theta^+ I(\theta) d\theta$ where $I(\theta) = [g_{ij}(\theta)]$ is the Fisher Information matrix. If we model Radar Signal by complex circular multivariate Gaussian distribution of zero mean :

$$p(X_n / R_n) = (\pi)^{-n} . |R_n|^{-1} . e^{-(X_n - m_n)^+ R_n^{-1} . (X_n - m_n)} \quad (1)$$

then for $m_n = 0$: $ds^2 = d\theta^+ I(\theta) d\theta = Tr\left[\left(R_n^{-1} dR_n \right)^2 \right] = \left\| R_n^{-1/2} dR_n R_n^{-1/2} \right\|_F^2$ (2)

This metric has been integrated by Siegel and the distance is given by:

$$dist^2(R_1, R_2) = \left\| \log\left(R_1^{-1/2} . R_2 . R_1^{-1/2} \right) \right\|_F^2 = \sum_{k=1}^{n} \log^2(\lambda_k) \text{ with } \det(R_2 - \lambda_k R_1) = 0 \quad (3)$$

This is a complete simply connected metric space of negative curvature with geodesic between R_1 and R_2 : $\gamma(t) = R_1^{1/2} e^{t . \log\left(R_1^{-1/2} R_2 R_1^{-1/2} \right)} R_1^{1/2} = R_1^{1/2} \left(R_1^{-1/2} R_2 R_1^{-1/2} \right)^t R_1^{1/2}$ with $0 \le t \le 1$ (4)

To take into account Toeplitz structure in case of stationary signal, Partial Iwasawa decomposition should be considered. This is equivalent for time or space signal to Complex AutoRegressive (CAR) Model decomposition (see Trench Theorem [6] or Verblunsky [7]) :

$$\Omega_n = (\alpha_n . R_n)^{-1} = W_n . W_n^+ = \left(1 - |\mu_n|^2 \right) \begin{bmatrix} 1 & A_{n-1}^+ \\ A_{n-1} & \Omega_{n-1} + A_{n-1} . A_{n-1}^+ \end{bmatrix} , \quad W_n = \sqrt{1 - |\mu_n|^2} \begin{bmatrix} 1 & 0 \\ A_{n-1} & \Omega_{n-1}^{1/2} \end{bmatrix}$$

with $\Omega_{n-1} = \Omega_{n-1}^{1/2} . \Omega_{n-1}^{1/2+}$ where $\alpha_n^{-1} = \left[1 - |\mu_n|^2 \right] \alpha_{n-1}^{-1}$, $A_n = \begin{bmatrix} A_{n-1} \\ 0 \end{bmatrix} + \mu_n \begin{bmatrix} A_{n-1}^{(-)} \\ 1 \end{bmatrix}$ and $V^{(-)} = J . V^*$

In the framework of Information Geometry, Information metric could be introduced as Kählerian metric where Kähler potential is given by the process Entropy $\tilde{\Phi}(R_n, P_0)$:

$$\tilde{\Phi}(R_n, P_0) = \log\left(\det R_n^{-1} \right) - \log(\pi . e) = -\sum_{k=1}^{n-1} (n-k) . \log\left[1 - |\mu_k|^2 \right] - n . \log[\pi . e . P_0] \quad (5)$$

Information metric is then given by hessian of Entropy $g_{ij} \equiv \frac{\partial^2 \tilde{\Phi}}{\partial \theta_i^{(n)} \partial \theta_j^{(n)*}}$ where $\theta^{(n)} = [P_0 \quad \mu_1 \quad \cdots \quad \mu_{n-1}]^T$ with $\{\mu_k\}_{k=1}^{n-1}$ Regularized [2] Burg reflection coefficient [6,7] and $P_0 = \alpha_0^{-1}$ mean signal Power. Kählerian metric is finally :

$$ds_n^2 = d\theta^{(n)+}[g_{ij}]d\theta^{(n)} = n\left(\frac{dP_0}{P_0}\right)^2 + \sum_{i=1}^{n-1}(n-i)\frac{|d\mu_i|^2}{\left(1-|\mu_i|^2\right)^2} \tag{6}$$

The robust Information Geometry distance is then given in product space by:

$$d^2\left[\left(P_{0,1},\{\mu_{i,1}\}_{i=1}^{N-1}\right),\left(P_{0,2},\{\mu_{i,2}\}_{i=1}^{N-1}\right)\right] = n\log^2\left(\frac{P_{0,2}}{P_{0,1}}\right) + \sum_{i=1}^{N-1}(N-k)\left(\frac{1}{2}\log\left(\frac{1+\delta_i}{1-\delta_i}\right)\right)^2$$

with $\delta_i = \left|\frac{\mu_{i,1}-\mu_{i,2}}{1-\mu_{i,1}\mu_{i,2}^*}\right|$ (7)

3. Fréchet Distance between curves and Geodesic extension on Manifold

In this chapter, we recall classical definition of distance between paths in Euclidean space. If we consider two curves, we can define similarity between each other. Classically, Hausdorff distance, that is the maximum distance between a point on one curve and its nearest neighbor on the other curve, is classically used but it does not take into account the flow of the curves, which is important for Time-Doppler analysis in our approach. The Fréchet distance [1,2,4] between two curves is defined as the minimum length of a leash required to connect a dog and its owner as they walk without backtracking along their respective curves from one endpoint to the other. The Fréchet metric takes the flow of the two curves into account; the pairs of points whose distance contributes to the Fréchet distance sweep continuously along their respective curves.

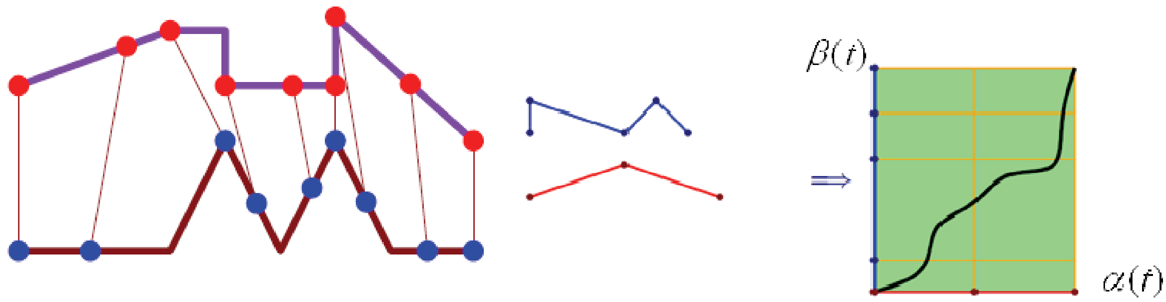

Figure 6. Fréchet Distance between two polygonal curves (α and β indexing all matching of points)

Let P and Q be two given curves, the Fréchet distance between P and Q is defined as the infimum over all reparameterizations α and β of $[0,1]$ of the maximum over all $t \in [0,1]$ of the distance in between $P(\alpha(t))$ and $Q(\beta(t))$. In mathematical notation, the Fréchet distance $d_{Fréchet}(P,Q)$ is:

$$\begin{cases} d_{Fréchet}(P,Q) = \underset{\alpha,\beta}{Inf} \underset{t\in[0,1]}{Max}\{d(P(\alpha(t)),Q(\beta(t)))\} \\ \alpha \text{ and } \beta : [0,1] \to [0,1] \text{ Nondecreasing and surjective} \end{cases} \tag{8}$$

Alt and Godau [2] have introduced a polynomial-time algorithm to compute the Fréchet distance between two polygonal curves in Euclidean space. For two polygonal curves with m and n segments, the computation time is $O(mn\log(mn))$. Alt and Godau have defined the free-space diagram between two curves for a given distance threshold ε is a two-dimensional region in the parameter space that consist of all point pairs on the two curves at distance at most ε: $D_\varepsilon(P,Q)=\{(\alpha,\beta)\in[0,1]^2 / d_{Fréchet}(P(\alpha(t)),Q(\beta(t)))\leq\varepsilon\}$. The Fréchet distance $d_{Fréchet}(P,Q)$ is at most ε if and only if the free-space diagram $D_\varepsilon(P,Q)$ contains a path which from the lower left corner to the upper right corner which is monotone both in the horizontal and in the vertical direction.

In an $n\times m$ free-space diagram, shown in following figure, the horizontal and vertical directions of the diagram correspond to the natural parametrizations of P and Q respectively. Therefore, if there is a monotone increasing curve from the lower left to the upper right corner of the diagram (corresponding to a monotone mapping), it generates a monotonic path that defines a matching between point-sets P and Q.

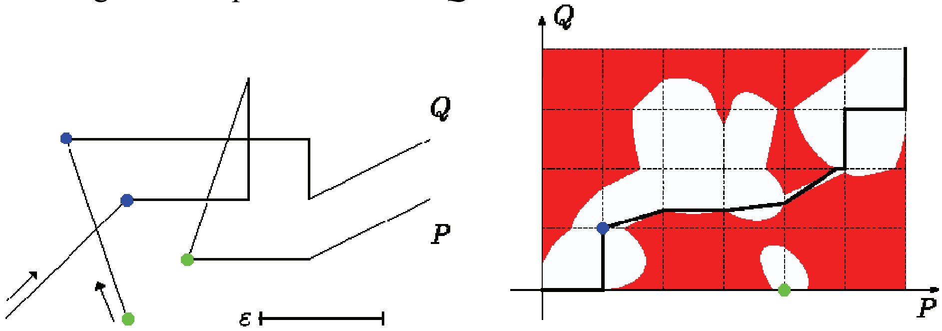

Figure 7. Fréchet free-space diagram for 2 polygonal curves P and Q with monotonicity in both directions

4. Fréchet Distance of Geodesic Curves representative of Time-Doppler Signatures on Information Geometry Manifold

We extend previous distance between paths of chapter 3 to Information Geometry Manifold of chapter 2. When the two curves are embedded in a more complex metric space, the distance between two points on the curves is most naturally defined as the geodesic length of the shortest path between them. If we consider N subsets of M Radar pulses in the burst, the Doppler burst can then be described by a poly-geodesic lines on Information Geometry Manifold. The set of N covariances matrices $\{R(t_1),R(t_2),\ldots,R(t_N)\}$ describe a discrete "polygonal" geodesic path on Information Geometry Manifold, and we can extend previous Frechet Distance but with Geodesic distance of chapter 2:

$$\begin{cases} d_{Fréchet}(R_1,R_2)=\underset{\alpha,\beta}{Inf}\ \underset{t\in[0,1]}{Max}\{d_{geo}(R_1(\alpha(t)),R_2(\beta(t)))\} \\ \text{with } d_{geo}^2(R_1(\alpha(t)),R_2(\beta(t)))=\left\|\log\left(R_1^{-1/2}(\alpha(t))R_2(\beta(t))R_1^{-1/2}(\alpha(t))\right)\right\|^2 \end{cases} \quad (9)$$

As classical Fréchet distance doesn't take into account with Inf[Max] close dependence of elements between points of time series paths, we propose to define a new distance given by:

$$d_{geo-path}(R_1,R_2)=\underset{\alpha,\beta}{Inf}\left\{\int_0^1 d_{geo}(R_1(\alpha(t)),R_2(\beta(t)))dt\right\} \quad (10)$$

We have then to find the solution for computing the geodesic minimal path on the Fréchet free-space diagram. The length of the path is not given by euclidean metric $ds^2=dt^2$ (where $L=\int_L ds$) but geodesic metric weighted by *d(.,.)* of the free-space diagram :

$$L_g=\int_L g.ds=\int_L ds_g \quad \text{with } ds_g=d(R_1(\alpha(t),R_2(\beta(t))).dt \quad (11)$$

This optimal shortest path could be computed by classical "Fast Marching method" [9].

NS-OS-HDR-CFAR will be based on distance (10) between sequence of"cell under test" covariance matrices and median sequence of auxilary data covariance matrices as in [5].

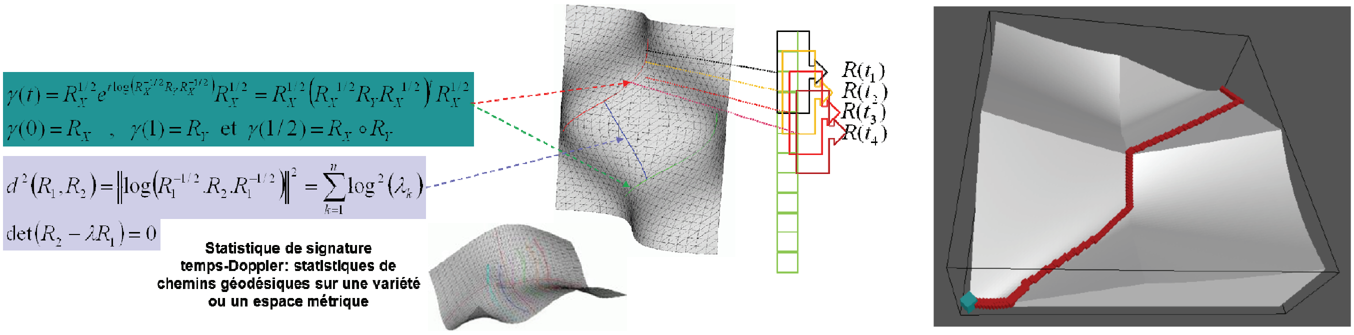

Figure 6. (on left) Geodesic Path on Information Geometry Manifold where 1 non stationary burst is decomposed on a sequence of stationary covariance matrices onTHPD matrix manifold, (on right) shortest path on free-space diagram

References:

[1] M. Fréchet, “Sur quelques points du calcul fonctionnel », Rendiconti del Circolo Mathematico di Palermo, n°22, p1-74, 1906

[2] M. Fréchet, « L’espace des courbes n’est qu’un semi-espace de Banach », General Topology and Its Relation to Modern Analysis and Algebra, pp.155-156, Prague, 1962

[3] H. Alt, and M. Godau. “Computing the Fréchet distance between two polygonal curves”. Internat. J. Comput. Geom. Appl., n°5, pp.75-91, 1995.

[4] Fréchet M.R., « Les éléments aléatoires de nature quelconque dans un espace distancié », Ann. 551 Inst. Henri Poincaré, vol.10 n°4, pp. 215–310, 1948

[5] Barbaresco F., “Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median”, In R. Bhatia & F. Nielsen Ed., “Matrix Information Geometry”, Springer Lecture Notes in Mathematics, 2012

[6] Trench W.F.,”An algorithm for the inversion of finite Toeplitz matrices”, J. Soc. Indust. Appl. Math. 12, pp.515–522, 1964

[7] Verblunsky S., “On positive harmonic functions”, Proc. London Math. Soc. 38, pp. 125-157, 1935; Proc. London Math. Soc. 40, pp. 290-320., 1936

[8] Barbaresco, F., “Super Resolution Spectrum Analysis Regularization: Burg, Capon and Ago-antagonistic Algorithms”, EUSIPCO-96, pp. 2005–2008, Trieste, Sept. 1996

[9] Barbaresco F. , “Computation of Most Threatening Radar Trajectories Areas and Corridors based on Fast-Marching & Level Sets”, IEEE CISDA Symposium, Paris, 2011

[10] P.W. Michor and D. Mumford, “An overview of the Riemannian metrics on spaces of curves using the Hamiltonnian approach”, Appl. Comput. Harm. Analysis 23 (1): 74-113, 2007

[11] A.Chouakria-Douzal and P.N. Nagabhushan, “Improved Fréchet distance for time series”, in Data Sciences and Classification, Springer, pp 13-20, 2006

[12] M. Bauer & al., “Constructing reparametrization invariant metrics on spaces of plane curves”, preprint http://arxiv.org/abs/1207.5965

[13] M. Fréchet, « L’espace dont chaque élément est une courbe n’est qu’un semi-espace de Banach », Annales scientifiques de l’ENS, 3ème série, tome 78, n°3, p.241-272, 1961

[14] M. Fréchet, « L’espace dont chaque élément est une courbe n’est qu’un semi-espace de Banach II», Annales scientifiques de l’ENS, 3ème série, tome 80, n°2, p.135-137, 1963

[15] Frédéric Chazal & al., “Gromov-Hausdorff Stable Signatures for Shapes using Persistence”, Eurographics Symposium on Geometry Processing 2009, Marc Alexa and Michael Kazhdan (Guest Editors), Volume 28, Number 5, 2009

[16] F. Cagliari, B. Di Fabio, and C. Landi, “The Natural Pseudo-Distance as a quotient pseudo metric, and applications”, Preprint, http://amsacta.unibo.it/3499/1/Forum_Submission.pdf

[17] P. Frosini and C. Landi, “No embedding of the automorphisms of a topological space into a compact metric space endows them with a composition that passes to the limit”, Appl. Math. Lett. 24, no. 10, 1654–1657, 2011

Information/Contact Geometries and Koszul Entropy

Frédéric Barbaresco

Thales Land & Air Systems, Voie Pierre-Gilles de Gennes, F91470 Limours, France
frederic.barbaresco@thalesgroup.com

Abstract. Based on Koszul theory of sharp convex cone hessian geometry, Information Geometry metric could be introduced by Koszul 2-form as hessian of Koszul-Vinberg Characteristic function logarithm (KVCFL). The front of the Legendre mapping of this KVCFL is the graph of a convex function, the Legendre transform of this KVCFL. By analogy in thermodynamic with Dual Massieu-Duhem potentials (Free Energy & Entropy), in large deviation theory with Cumulant Generating & Rate functions (Legendre duality by Laplace Principle) or in Legendre duality in Mechanics, the Legendre transform of KVCFL could be interpreted as a "Koszul Entropy". This Legendre duality is considered in more general framework of Contact Geometry, the odd-dimensional twin of symplectic geometry, with Legendre fibration & mapping.

Keywords: Characteristic Function, Koszul forms, Laplace Principle, Massieu-Duhem Potential, Legendre Duality, Contact Geometry, Information Geometry

1 Characteristic function in geometry/statistic/thermodynamic

We study use of Koszul Characteristic Function in Information Geometry, in Large Deviations Theory with Laplace Principle, in Mechanics with Contact Geometry and in Thermodynamics with Massieu-Duhem Potentials. We try to explore close inter-relations between these 4 domains. First, derivatives of the Koszul-Vinberg Characteristic Function Logarithm (KVCFL) $\log \psi_\Omega(x) = \log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi$ are invariant by the automorphisms of the convex cone, and KVCFL Hessian defines a Riemannian metric, useful in Information Geometry. Tool of "characteristic function" was first introduced by Henri Poincaré in Statistics and used systematically by Paul Levy, using property that all moments of statistical laws could be deduced from its derivatives. In Large Deviation Theory, Laplace Principle allows to introduce the (scaled cumulant) Generating Function $\phi(x) = \underset{n \to \infty}{Lim} \frac{1}{n} \log \psi_n(x)$ with $\psi_n(x) = \int e^{nx\varsigma} p(\Sigma_n = \varsigma) d\varsigma$ as the more natural element to characterize behavior of statistical variable for large deviations. In Mechanics, Legendre Duality gives the relation between the variational Euler-Lagrange and the symplectic Hamilton-Jacobi formulations of the equations of motion. Finally, in thermodynamic, the Massieu characteristic function (Duhem potential) ϕ is able to provide all body properties from their derivatives and is Legendre

adfa, p. 1, 2011.

transform of Entropy *S*: $S = \phi - \frac{1}{T} . \frac{\partial \phi}{\partial (1/T)}$. "Characteristic function" and its derivatives capture all information of random variable/system/physical model.

2 Koszul Characteristic Function/Entropy by Legendre Duality

We define Koszul-Vinberg hessian metric of convex sharp cone, and observe that. the Fisher information metric coincides with the canonical Koszul Hessian metric (Koszul 2-form). Then, by Legendre duality, we introduce Koszul-Vinberg Entropy.

2.1 Koszul-Vinberg Characteristic Function and Metric of convex sharp cone

J.L. Koszul [1] and E. Vinberg have introduced an affinely invariant Hessian metric on a sharp convex cone Ω through its characteristic function ψ. In the following, Ω is a sharp open convex cone in a vector space E of finite dimension on R (a convex cone is sharp if it does not contain any full straight line). In dual space E^* of E, Ω^* is the set of linear strictly positive forms on $\overline{\Omega} - \{0\}$. Ω^* is the dual cone of Ω and is a sharp open convex cone. If $\xi \in \Omega^*$, then the intersection $\Omega \cap \{x \in E / \langle x, \xi \rangle = 1\}$ is bounded. $G = Aut(\Omega)$ is the group of linear transform of E that preserves Ω. $G = Aut(\Omega)$ operates on Ω^* by $\forall g \in G = Aut(\Omega), \forall \xi \in E^*$ then $\tilde{g}.\xi = \xi \circ g^{-1}$.

Koszul-Vinberg Characteristic function definition: Let $d\xi$ be the Lebesgue measure on E^*, the following integral: $\psi_\Omega(x) = \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi \quad \forall x \in \Omega$ (1)

with Ω^* the dual cone is an analytic function on Ω, with $\psi_\Omega(x) \in]0, +\infty[$, called the ***Koszul-Vinberg characteristic function*** of cone Ω, with the properties:

- The Bergman kernel of $\Omega + iR^{n+1}$ is written as $K_\Omega(\mathrm{Re}(z))$ up to a constant where K_Ω is defined by the integral: $K_\Omega(x) = \int_{\Omega^*} e^{-\langle \xi, x \rangle} \psi_{\Omega^*}(\xi)^{-1} d\xi$
- ψ_Ω is analytic function defined on the interior of Ω and $\psi_\Omega(x) \to +\infty$ as $x \to \partial\Omega$
- If $g \in Aut(\Omega)$ then $\psi_\Omega(gx) = |\det g|^{-1} \psi_\Omega(x)$ and since $tI \in G = Aut(\Omega)$ for any $t > 0$, we have $\psi_\Omega(tx) = \psi_\Omega(x) / t^n$
- ψ_Ω is logarithmically strictly convex, and $\phi_\Omega(x) = \log(\psi_\Omega(x))$ is strictly convex
- ***Koszul 1-form*** : The differential 1-form $\alpha = d\phi_\Omega = d \log \psi_\Omega = d\psi_\Omega / \psi_\Omega$ (2) is invariant by all automorphisms $G = Aut(\Omega)$ of Ω. If $x \in \Omega$ and $u \in E$ then $\langle \alpha_x, u \rangle = -\int_{\Omega^*} \langle \xi, u \rangle . e^{-\langle \xi, x \rangle} d\xi$ and $\alpha_x \in -\Omega^*$ (3)
- ***Koszul 2-form*** $\boldsymbol{\beta}$: The symmetric differential 2-form $\beta = D\alpha = d^2 \log \psi_\Omega$ (4) is a positive definite symmetric bilinear form on E invariant under $G = Aut(\Omega)$. $D\alpha > 0$ (Schwarz inequality and $d^2 \log \psi_\Omega(u, v) = \int_{\Omega^*} \langle \xi, u \rangle \langle \xi, v \rangle e^{-\langle \xi, u \rangle} d\xi$)

- ***Koszul-Vinberg Metric***: $D\alpha$ defines a Riemanian structure invariant by $Aut(\Omega)$, and then the Riemanian metric is given by $g = d^2 \log\psi_\Omega$ (5)

$$d^2 \log\psi(x) = d^2\left[\log\int\psi_u du\right] = \frac{\int\psi_u d^2\log\psi_u du}{\int\psi_u du} + \frac{1}{2}\frac{\iint\psi_u\psi_v\left(d\log\psi_u - d\log\psi_v\right)^2 dudv}{\iint\psi_u\psi_v dudv}$$

A diffeomorphism is used to define dual coordinate : $x^* = -\alpha_x = -d\log\psi_\Omega(x)$ (6)

With $\langle df(x), u\rangle = D_u f(x) = \left.\frac{d}{dt}\right|_{t=0} f(x+tu)$. When the cone Ω is symmetric, the map $x^* = -\alpha_x$ is a bijection and an isometry with a unique fixed point (the manifold is a Riemannian Symmetric Space given by this isometry): $(x^*)^* = x$, $\langle x, x^*\rangle = n$ and $\psi_\Omega(x)\psi_{\Omega^*}(x^*) = cste$. x^* is characterized by $x^* = \arg\min\left\{\psi(y)/ y\in\Omega^*, \langle x, y\rangle = n\right\}$ and x^* is the center of gravity of the cross section $\left\{y\in\Omega^*, \langle x,y\rangle = n\right\}$ of Ω^*:

$$x^* = \int_{\Omega^*}\xi.e^{-\langle\xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi \text{ and } \langle -x^*, h\rangle = d_h\log\psi_\Omega(x) = -\int_{\Omega^*}\langle\xi,h\rangle e^{-\langle\xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi \quad (7)$$

From this last equation, we can deduce "***<u>Koszul Entropy</u>*" *defined as Legendre Transform of minus logarithm of Koszul-Vinberg characteristic function*** $\Phi(x)$:

$\Phi^*(x^*) = \langle x, x^*\rangle - \Phi(x)$ with $x^* = D_x\Phi$ and $x = D_{x^*}\Phi^*$ where $\Phi(x) = -\log\psi_\Omega(x)$

By (7), we can write: $-\langle x^*, x\rangle = \int_{\Omega^*}\log e^{-\langle\xi,x\rangle}.e^{-\langle\xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi$

$$\Phi^*(x^*) = -\int_{\Omega^*}\log e^{-\langle\xi,x\rangle}.e^{-\langle\xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi + \log\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi$$

and

$$\Phi^*(x^*) = \left[\left(\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi\right).\log\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi - \int_{\Omega^*}\log e^{-\langle\xi,x\rangle}.e^{-\langle\xi,x\rangle}d\xi\right] \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi$$

We can then consider this Legendre transform as an entropy, named **Koszul Entropy**:

$$\Phi^* = -\int_{\Omega^*}\frac{e^{-\langle\xi,x\rangle}}{\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi}\log\frac{e^{-\langle\xi,x\rangle}}{\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi}d\xi = -\int_{\Omega^*}p_x(\xi)\log p_x(\xi)d\xi \quad (8)$$

With $p_x(\xi) = e^{-\langle\xi,x\rangle} \Big/ \int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi = e^{-\langle x,\xi\rangle - \log\int_{\Omega^*}e^{-\langle\xi,x\rangle}d\xi} = e^{-\langle x,\xi\rangle + \Phi(x)}$ and $x^* = \int_{\Omega^*}\xi.p_x(\xi)d\xi$

J.L. Koszul [1] and J. Vey [2] have proved more, by the following theorem:

Koszul-Vey Theorem: Let M be a connected Hessian manifold with Hessian metric g. Suppose that admits a closed 1-form α such that $D\alpha = g$ and there exists a group G of affine automorphisms of M preserving α :

- If M/G is quasi-compact, then the universal covering manifold of M is affinely isomorphic to a convex domain Ω real affine space not containing any full straight line.
- If M/G is compact, then Ω is a sharp convex cone.

If we denote by S_c the level surface of ψ_Ω: $S_c = \{\psi_\Omega(x) = c\}$ which is a non-compact submanifold in Ω, and by ω_c the induced metric of $d^2\log\psi_\Omega$ on S_c, then assuming that the cone Ω is homogeneous under $G(\Omega)$, Sasaki proved that S_c is a homogeneous hyperbolic affine hypersphere and every such hyperspheres can be obtained in this way .Sasaki also remarks that ω_c is identified with the affine metric and S_c is a global Riemannian symmetric space when Ω is a self-dual cone. Let Ω be a regular convex cone and let $g = d^2 \log \psi_\Omega$ be the canonical Hessian metric, then each level surface of the characteristic function ψ_Ω is a minimal surface of the Riemannian manifold (Ω, g).

2.2 Koszul Forms and Metric for Symmetric Positive Definite Matrices

Let v be the volume element of g. We define a closed 1-form α and β a symmetric bilinear form by: $D_X v = \alpha(X) v$ and $\beta = D\alpha$. The forms α and β are called the first Koszul form and the second Koszul form for a Hessian structure *(D; g)* respectively:

$$v = \left(\det\left[g_{ij}\right]\right)^{1/2} dx^1 \wedge .. dx^n \Rightarrow \alpha_i = \frac{\partial}{\partial x^i} \log\left(\det\left[g_{ij}\right]\right)^{\frac{1}{2}} v \text{ and } \beta_{ij} = \frac{\partial \alpha_i}{\partial x^j} = \frac{1}{2} \frac{\partial^2 \log \det\left[g_{kl}\right]}{\partial x^i \partial x^j}$$

A pair *(D; g)* of a flat connection *D* and a Hessian metric g is called a Hessian structure. J.L. Koszul studied a flat manifold endowed with a closed 1-form α such that $D\alpha$ is positive definite, whereupon $D\alpha$ is a Hessian metric. A Hessian structure *(D; g)* is said to be of Koszul type, if there exists a closed 1-form α such that $g = D\alpha$. The 2nd Koszul form β plays a role similar to the Ricci tensor for Kählerian metric.

We can apply this Koszul geometry framework for Symmetric Positive Definite Matrices. Let the inner product $\langle x, y \rangle = Tr(xy), \forall x, y \in Sym_n(R)$, Ω be the set of symmetric positive definite matrices is an open convex cone and is self-dual $\Omega^* = \Omega$.

$$\psi_\Omega(x) = \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi \underset{\substack{\langle x,y\rangle = Tr(xy) \\ \Omega^* = \Omega \text{ self-dual}}}{=} \det x^{-\frac{n+1}{2}} \psi(I_n) \quad (9)$$

$$g = d^2 \log \psi_\Omega = -\frac{n+1}{2} d^2 \log \det x \text{ and } x^* = -d \log \psi_\Omega = \frac{n+1}{2} d \log \det x = \frac{n+1}{2} x^{-1} \quad (10)$$

Let be the regular convex cone consisting of all positive definite symmetric matrices of degree n. Then $(D, Dd \log \det x)$ is a Hessian structure on Ω, and each level surface of $\det x$ is a minimal surface of the Riemannian manifold $(\Omega, g = -Dd \log \det x)$.

J.L. Koszul has introduced a more general 1-form definition given by :

$$\alpha = -\frac{1}{4} d\Psi(X) \text{ with } \Psi(X) = Tr_{g/b}\left[ad(JX) - Jad(X)\right] \quad \forall X \in g \quad (11)$$

We can illustrate it for Poincaré's Upper Half Plane $V = \{z = x + iy \,/\, y > 0\}$. Let vector fields $X = y\dfrac{d}{dx}$ and $Y = y\dfrac{d}{dy}$, and J tensor of complex structure V defined by $JX = Y$. As $[X, Y] = -Y$ and $ad(Y).Z = [Y, Z]$ then

$$\begin{cases} Tr\left[ad(JX) - Jad(X)\right] = 2 \\ Tr\left[ad(JY) - Jad(Y)\right] = 0 \end{cases} \quad (12)$$

The Koszul 1-form and then the Koszul/Poincaré metric is given by:

$$\Psi(X)=2\frac{dx}{y}\Rightarrow\alpha=-\frac{1}{4}d\Psi=-\frac{1}{2}\frac{dx\wedge dy}{y^2}\Rightarrow ds^2=\frac{dx^2+dy^2}{2y^2} \quad (13)$$

This could be also applied for Siegel's Upper Half Plane $V=\{Z=X+iY\,/\,Y>0\}$

$$\begin{cases} SZ=(AZ+B)D^{-1} \\ A^TD=I,B^TD=D^TB \end{cases} \text{ with } S=\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \text{ and } J=\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

$$\Psi(dX+idY)=\frac{3p+1}{2}Tr\left(Y^{-1}dX\right)\Rightarrow\begin{cases} \alpha=-\frac{1}{4}d\Psi=\frac{3p+1}{8}Tr\left(Y^{-1}dZ\wedge Y^{-1}d\overline{Z}\right) \\ ds^2=\frac{(3p+1)}{8}Tr\left(Y^{-1}dZY^{-1}d\overline{Z}\right) \end{cases} \quad (14)$$

For covariance Hermitian Positive Definite (HPD) matrix, if we take $Z=iR$ (with $X=0$), the metric $ds^2=Tr\left[\left(R^{-1}dR\right)^2\right]$ is equivalent to Information metric for multivariate Gaussian law of covariance matrix R and zero mean.

2.3 Characteristic Function and Laplace Principle of Large Deviations

Let μ be a positive Borel Measure on euclidean space V. Assume that the following integral is finite for all x in an open set $\Omega\subset V$: $\psi_x(y)=\int e^{-\langle y,x\rangle}d\mu(x)$ (15)

For $x\in\Omega$, consider the probability measure : $p(y,dx)=\frac{1}{\psi_x(y)}e^{-\langle y,x\rangle}d\mu(x)$ (16)

then mean is given by : $m(y)=\int xp(y,dx)=-\nabla\log\psi_x(y)$ (17)

and covariance $\langle V(y)u,v\rangle=\int\langle x-m(y),u\rangle\langle x-m(y),v\rangle p(y,dx)=D_uD_v\log\psi_x(y)$ (18)

In his book "*Calcul des probabilités*", Poincaré introduced first the term "*fonction caractéristique*" because of Massieu work developed in his "*thermodynamique*" book.

Large deviation principle [5]***:*** Let $\{\Sigma_n\}$ be a sequence of random variables indexed by the positive integer n, and let $P(\Sigma_n\in d\varsigma)=P(\Sigma_n\in[\varsigma,\varsigma+d\varsigma])$ denote the probability measure associated with these random variables. We say that Σ_n or $P(\Sigma_n\in d\varsigma)$ satisfy a large deviation principle if the limit the rate function $I(\varsigma)=\underset{n\to\infty}{Lim}-\frac{1}{n}\log P(\Sigma_n\in d\varsigma)$ exists: $P(\Sigma_n\in d\varsigma)\approx e^{-n.I(\varsigma)}d\varsigma$ or $p(\Sigma_n=\varsigma)\approx e^{-n.I(\varsigma)}$

If we write the Legendre transform of a function $h(x)$ defined by $g(k)=\underset{x}{Sup}\{k.x-h(x)\}$, we can express the following principle and Laplace's method:

Laplace Principle[5]***:*** Let $\{(\Omega_n,F_n,P_n),n\in N\}$ be a sequence of probability spaces, Θ a complete separable metric space, $\{Y_n,n\in N\}$ a sequence of random variables such that Y_n maps Ω_n into Θ, and I a rate function on Θ. Then, Y_n satisfies the Laplace principle on Θ with rate function I if for all bounded, continuous functions f mapping Θ into R: $\underset{n\to\infty}{Lim}\frac{1}{n}\log E_{P_n}\left[e^{n.f(Y_n)}\right]=\underset{x\in\Theta}{Sup}\{f(x)-I(x)\}$ (19)

If Y_n satisfies the large deviation principle on Θ with rate function I, then $P_n(Y_n \in dx) \approx e^{-n.I(x)}dx$: $\underset{n\to\infty}{Lim}\frac{1}{n}\log E_{P_n}\left[e^{n.f(Y_n)}\right] = \underset{n\to\infty}{Lim}\frac{1}{n}\log\int_{\Omega_n} e^{nf(Y_n)}.dP_n$

$\underset{n\to\infty}{Lim}\frac{1}{n}\log\int_{\Theta} e^{n.f(x)}P_n(Y_n \in dx) \approx \underset{n\to\infty}{Lim}\frac{1}{n}\log\int_{\Theta} e^{n.f(x)}e^{-n.I(x)}.dx = \underset{n\to\infty}{Lim}\frac{1}{n}\log\int_{\Theta} e^{n.[f(x)-I(x)]}.dx$

The asymptotic behavior of the last integral is determined by the largest value of the integrand [5]: $\underset{n\to\infty}{Lim}\frac{1}{n}\log E_{P_n}\left[e^{n.f(Y_n)}\right] = \frac{1}{n}\log\left[\underset{x\in\Theta}{Sup}\left\{e^{n.[f(x)-I(x)]}\right\}\right] = \underset{x\in\Theta}{Sup}\{f(x)-I(x)\}$ (20)

The generating function of Σ_n is defined as $\psi_n(x) = \int e^{nx\varsigma}p(\Sigma_n = \varsigma)d\varsigma$. The function $\phi(x)$ defined by the limit: $\phi(x) = \underset{n\to\infty}{Lim}\frac{1}{n}\log\psi_n(x)$ is called the (scaled cumulant) generating function of Σ_n. It is also called the log-generating function or free energy function of Σ_n. In the following in thermodynamic, it will be called Massieu Potential. The existence of this limit is equivalent to writing $\psi_n(x) \approx e^{n.\phi(x)}$.

Gärtner/Varadhan Theorems: $I(\varsigma) = \underset{x}{Sup}\{x.\varsigma - \phi(x)\}$ and $\phi(x) = \underset{\varsigma}{Sup}\{x.\varsigma - I(\varsigma)\}$

2.4 Contact Geometry in Mechanics, Cartan invariant & Legendre Mapping

As described by Vladimir Arnold [6], in the general case, we can define the Hamiltonian H as the fiberwise Legendre transformation of the Lagrangian L: $H(p,q,t) = \underset{\dot{q}}{Sup}(p.\dot{q} - L(q,\dot{q},t))$. Due to strict convexity, $H(p,q,t) = p.\dot{q} - L(q,\dot{q},t)$ supremum is reached in a unique point $\dot{q}$ such that $p = \partial_{\dot{q}}L(q,\dot{q},t)$, and we have also $\dot{q} = \partial_p H(p,q,t)$. If we consider total differential of Hamiltonian:

$$\left.\begin{aligned} dH &= \dot{q}dp + pd\dot{q} - \partial_q Ldq - \partial_{\dot{q}}Ld\dot{q} - \partial_t Ldt = \dot{q}dp - \partial_q Ldq - \partial_t Ldt \\ &= \partial_p Hdp + \partial_q Hdq + \partial_t Hdt \end{aligned}\right\} \Rightarrow \begin{cases} \dot{q} = \partial_p H \\ -\partial_q L = \partial_q H \end{cases}$$

Euler-Lagrange equation $\partial_t\partial_{\dot{q}}L - \partial_q L = 0$ with $p = \partial_{\dot{q}}L$ and $-\partial_q L = \partial_q H$ provides the 2nd Hamilton equation $\dot{p} = -\partial_q H$ with $\dot{q} = \partial_p H$ in Darboux coordinates.

Considering Pfaffian form $\omega = p.dq - H.dt$ related to Poincaré-Cartan integral invariant [11], based on $\omega = \partial_{\dot{q}}L.dq - (\partial_{\dot{q}}L.\dot{q} - L)dt = L.dt + \partial_{\dot{q}}L\varpi$ with $\varpi = dq - \dot{q}.dt$, P. Dedecker [7] has observed, that the property that among all forms $\theta \equiv L.dt \bmod \varpi$ the form $\omega = p.dq - H.dt$ is the only one satisfying $d\theta \equiv 0 \bmod \varpi$, is a particular case of more general T. Lepage congruence [8] related to transversality condition.

Legendre transform and contact geometry where used in Mechanic [6] and in Thermodynamic [9,10]. Integral submanifolds of dimension n in 2n+1 dimensional contact manifold are called Legendre submanifolds. A smooth fibration of a contact manifold, all of whose are Legendre, is called a Legendre Fibration. In the neighbourhood of each point of the total space of a Legendre Fibration there exist contact Darboux coordinates *(z, q, p)* in which the fibration is given by the projection *(z, q, p) =>(z, q)*.

Indeed, the fibres *(z, q) = cst* are Legendre subspaces of the standard contact space. A Legendre mapping is a diagram consisting of an embedding of a smooth manifold as a Legendre submanifold in the total space of a Legendre fibration, and the projection of the total space of the Legendre fibration onto the base. Let us consider the two Legendre fibrations of the standard contact space R^{2n+1} of 1 -jets of functions on R^n: $(u,p,q)\mapsto(u,q)$ and $(u,p,q)\mapsto(p.q-u,p)$, the projection of the l-graph of a function $u=S(q)$ onto the base of the second fibration gives a Legendre mapping $q\mapsto\left(q\frac{\partial S}{\partial q}-S(q),\frac{\partial S}{\partial q}\right)$. If S is convex, the front of this mapping is the graph of a convex function, the Legendre transform of the function S: $\left(S^*(p),p\right)$ (21)

2.5 Massieu Characteristic Function & Duhem potentials in Thermodynamic

In 1869, François MASSIEU, French Engineer from Corps des Mines, has presented two papers to French Science Academy on « Characteristic function » in Thermodynamic. Massieu has demonstrated that some mechanical and thermal properties of physical and chemical systems could be derived from two potentials called "characteristic functions". The infinitesimal amount of heat dQ received by a body produces external work of dilatation, internal work, and an increase of body sensible heat. The last two effects could not be identified separately and are noted dE (function E accounted for the sum of mechanical and thermal effects by equivalence between heat and work). The external work $P.dV$ is thermally equivalent to $A.P.dV$ (with A the conversion factor between mechanical and thermal measures). The first principle provides $dQ=dE+A.P.dV$. For a closed reversible cycle (Joule/Carnot principles) $\int\frac{dQ}{T}=0$ that is the complete differential dS of a function S of $dS=\frac{dQ}{T}$.

If we select volume V and temperature T as independent variables:

$T.dS=dQ\Rightarrow T.dS-dE=A.P.dV\Rightarrow d(TS)-dE=S.dT+A.P.dV$ (22)

If we set $H=TS-E$, then we have $dH=S.dT+A.P.dV=\frac{\partial H}{\partial T}.dT+\frac{\partial H}{\partial V}.dV$ (23)

Massieu has called H the "characteristic function" because all body characteristics could be deduced of this function: $S=\frac{\partial H}{\partial T}$, $P=\frac{1}{A}\frac{\partial H}{\partial V}$ and $E=TS-H=T\frac{\partial H}{\partial T}-H$

If we select pression P and temperature T as independent variables:

Massieu characteristic function is then given by $H'=H-AP.V$ (24)

We have $dH'=dH-AP.dV-AV.dP=S.dT-AV.dP=\frac{\partial H'}{\partial T}.dT+\frac{\partial H'}{\partial P}.dP$ (25)

And we can deduce: $S=\frac{\partial H'}{\partial T}$ and $V=-\frac{1}{A}\frac{\partial H'}{\partial P}$ (26)

And inner energy: $E=TS-H=TS-H'-AP.V\Rightarrow E=T\frac{\partial H'}{\partial T}-H'+P.\frac{\partial H'}{\partial P}$ (27)

The most important result consists in deriving all body properties dealing with thermodynamics from Massieu characteristic function and its derivatives :***« je montre,***

***dans ce mémoire, que toutes les propriétés d'un corps peuvent se déduire d'une fonction unique, que j'appelle la fonction caractéristique de ce corps»* [3].**

In thermodynamics, the Massieu potential (previously called generating function) is the Legendre transform of the Entropy (previously called rate function), and depends on the inverse temperature $\rho = 1/kT$ **:**

$\phi(\rho) = -k\rho.F = S - E/T$ where F is the Free Energy $F = E - TS$ (28)

$$-\frac{\partial\phi(\rho)}{\partial(k\rho)} = \frac{\partial(k\rho F)}{\partial(k\rho)} = F + k\rho\frac{\partial F}{\partial(k\rho)} = F + k\rho\left(\frac{\partial F}{\partial T}\right)_V\left(\frac{\partial T}{\partial(k\rho)}\right) \quad (29)$$

$$-\frac{\partial\phi(\rho)}{\partial(k\rho)} = F + k\rho(-S)\left(-\frac{1}{(k\rho)^2}\right) = F + \frac{S}{k\rho} = (E - TS) + TS = E \quad (30)$$

With E the inner energy. The Legendre transform of the Massieu potential provides the Entropy S: $L(\phi) = k\rho.\frac{\partial\phi(\rho)}{\partial(k\rho)} - \phi(\rho) = k\rho.(-E) - k\rho.F = k\rho(F - E) = -S$ (31)

On these bases, Duhem [4] has founded Mechanics on the principles of Thermodynamics. Duhem has defined [4] a more general potential function $\Omega = G(E - TS) + W$ with G mechanic equivalent of heat. In case of constant volume, $W = 0$, the potential Ω becomes Helmoltz Free Energy and in the case of constant pressure, $W = P.V$, the potential Ω becomes Gibbs-Duhem Free Enthalpy. Duhem has written: "***Nous avons fait de la Dynamique un cas particulier de la Thermodynamique, une Science qui embrasse dans des principes communs tous les changements d'état des corps, aussi bien les changements de lieu que les changements de qualités physiques*** " [4].

3 References

1. Koszul J.L., Variétés localement plates et convexité, Osaka J. Math. , n°2, p.285-290, 1965
2. Vey J., Sur les automorphismes affines des ouverts convexes saillants, Annali della Scuola Normale Superiore di Pisa, Classe di Science, 3e série, tome 24,n°4, p.641-665, 1970
3. Massieu F., Thermodynamique: Mémoire sur les fonctions caractéristiques des divers fluides et sur la théorie des vapeurs, 92-pgs, Académie des Sciences, 1876
4. Duhem P., Sur les équations générales de la Thermodynamique, Annales Scientifiques de l'Ecole Normale Supérieure, 3e série, tome VIII, p. 231, 1891
5. Laplace P.S., Mémoire sur la probabilité des causes sur les évènements. Mémoires de Mathématique et de Physique, Tome Sixième, 1774
6. Arnold, V.I., Givental, A.G., Symplectic geometry, Encyclopedia of mathematical science, vol. 4., Springer Verlag (translated from Russian)
7. Dedecker P., A property of differential forms in the calculus of variations, Pacific J. Math. Volume 7, Number 4,p. 1545-1549, 1957
8. Lepage T., Sur les champs géodésiques du calcul des variations, Bull. Acad. Roy. Belg., CL. Sci.27, p.716-729, pp. 1036-1046, 1936
9. Mrugala R., On contact and metric structures on thermodynamic spaces, RIMS Kokyuroku, 1142 . pp. 167-181, 2000
10. Arnold V.I., Contact geometry: the geometrical method of Gibbs's thermodynamics, Proceedings of the Gibbs Symposium, p.163–179, Amer. Math. Soc., Providence, RI, 1990
11. Cartan E., Leçons sur les invariants intégraux, Hermann, Paris, 1922

Geometric Radar Processing based on Fréchet Distance : Information Geometry versus Optimal Transport Theory

Frédéric Barbaresco*

*Thales Air Systems, Surface Radar Domain, Technical Directorate, Advanced Developments Dept.
Voie Pierre-Gilles de Gennes, 91470 Limours, France
email: frederic.barbaresco@thalesgroup.com

***Abstract**: In the framework of Optimal Transport Theory, Fréchet-Wasserstein distance could be used to define distance between multivariate Gaussian laws with positive curvature geometry. We compare this approach with Information geometry for Covariance Radar Matrices Processing, where Fisher metric and Siegel-Rao distance provides geometry of negative curvature.*

1. Optimal Transport Theory and Fréchet-Wasserstein distance

For a long time, one important problem in statistics is to define metrics in the space of probability measures or metrics between random variables. Recent progresses on this critical subject have been achieved in the framework of Optimal Transport Theory by use of Wasserstein metrics for probability distribution and especially for multivariate Gaussian laws. First work on probability metric was originally done by G. Monge in 1781 [1], considering the following metric in the distribution function space : $d(P,Q)=\underset{X,Y}{Inf}\,E\left[\left|X-Y\right|\right]$ (1)

This was generalized by Appel for : $d(P,Q)=\underset{X,Y}{Inf}\,E\left[f\left(\left|X-Y\right|\right)\right]$ (2)

where the infinum is taken over all joint distributions of pairs *(X,Y)* with fixed marginal distribution functions *P* and *Q*. Global survey of Optimal Transport theory is given in 2010 Field Medal Cédric Villani Book [6].

Then during 20th century, Maurice Fréchet [2] has proposed a metric on the space of probability distributions given by : $d^2(P,Q)=\underset{X,Y}{Inf}\,E\left[\left|X-Y\right|^2\right]$ (3)

where the minimization is taken over all random variables *X* and *Y* having distribution *P* and *Q* respectively. Fréchet distance can be computed for n-dimensional distributions family that are closed with respect to linear transformations of the ramdom vector, and especially for Multivariate Gaussian laws. Others distance were studied by Paul Levy [3] & R. Fortét [4].

Fréchet distance could be considered as Wasserstein distance $W_n(\mu,\nu)$ of order *2 (n=2),* when Wasserstein distance of order *n* is defined by [6] :

Let $(\wp,d)$ be a polish metric space, and let $n\in[1,+\infty)$. For any two probability measure μ,ν on $\wp$, the Wasserstein distance of order *n* between μ and ν is defined by the formula :

$$W_n(\mu,\nu)=\left(\underset{\pi\in\Pi(\mu,\nu)}{Inf}\int_{\wp}d(x,y)^n\,d\pi(x,y)\right)^{1/n}=Inf\left\{\left(E\left[d(X,Y)^n\right]\right)^{1/n},law(X)=\mu,law(Y)=\nu\right\}$$

Probability measure $\pi\in\wp(X,Y)$ is called a coupling of μ and ν with its projections coincides with μ and ν. We call π an optimal coupling if it attains the infinum of previous wasserstain distance. The metric space $(\wp,W_n)$ is called L^n-*Wasserstein* space.

Monge distance could be considered as Wasserstein distance of order *n=1*, $W_1(\mu,\nu)$, also called Kantorovich-Rubinstein distance. If *X*, *Y* are square integrable random variables with *X* has distribution *P* and *Y* has distribution *Q*, then :

$$E\left[\left|X-Y\right|^2\right]=E\left[Tr\left((X-Y)^+(X-Y)\right)\right]=\left|E(X)-E(Y)\right|^2+Tr\left[R_X\right]+Tr\left[R_Y\right]-2Tr\left[R_{X,Y}\right] \qquad (4)$$

with $R_X = Cov(X)$, $R_Y = Cov(Y)$ and $R_{X,Y} = Cov(X,Y)$. Then the problem is equivalent to [7] : If $K = \left\{ \Psi \in \Re^{kxk} / \begin{pmatrix} R_X & \Psi \\ \Psi^+ & R_Y \end{pmatrix} \geq 0 \right\}$, then $Cov(P,Q) \subset K$, solution is given by

$Sup\{Tr(\Psi) / \Psi \in Cov(P,Q)\}$ where $Cov(P,Q) = \{\Psi \in \Re^{kxk} / \exists X \sim P, Y \sim Q, \Psi = Cov(X,Y)\}$

For Multivariate Gaussian measures P and Q on R^k, with means m_X and m_Y and non-singular covariance matrices R_X and R_Y, The Fréchet-Wasserstein distance is given by :

$$d^2(N(m_X, R_X), N(m_Y, R_Y)) = |m_X - m_Y|^2 + Tr[R_X] + Tr[R_Y] - 2.Tr\left[\left(R_X^{1/2} R_Y R_X^{1/2}\right)^{1/2}\right] \quad (5)$$

The proof could be given by considering the random vector $W = \begin{bmatrix} X \\ Y \end{bmatrix}$ and the covariance matrix of W: $R_W = \begin{bmatrix} R_X & \Psi \\ \Psi^+ & R_Y \end{bmatrix}$. If we consider the case of zero mean :

We have directly that : $E\left[\|X - Y\|^2\right] = Tr\left[R_X + R_Y - \Psi - \Psi^+\right] = Tr[R_X] + Tr[R_Y] - Tr\left[\Psi + \Psi^+\right]$

Solution is given by extreme value of $Tr\left[\Psi + \Psi^+\right]$ subject to the constraint that R_W is a covariance matrix. Considering the Hermitian form :

$$G = \begin{bmatrix} x \\ y \end{bmatrix}^+ R_W \begin{bmatrix} x \\ y \end{bmatrix} = x^+ R_X x + y^+ R_Y y + x^+ \Psi y + y^+ \Psi^+ x \underset{G \geq 0}{\Rightarrow} G = \sum_{r=1}^{m} \left| s_r^+ x + t_r^+ y \right|^2$$

$$\Rightarrow R_X = \sum_{r=1}^{m} s_r s_r^+ ; R_Y = \sum_{r=1}^{m} t_r t_r^+ ; \Psi = \sum_{r=1}^{m} s_t t_r^+$$

Problem is then reduce to :

$Tr\left[\sum_{r=1}^{m} s_r t_r^+ + \sum_{r=1}^{m} t_r s_t^+\right]$ with the contraints that : $R_X = \sum_{r=1}^{m} s_r s_r^+$ and $R_Y = \sum_{r=1}^{m} t_r t_r^+$

This problem is solved by introducing Lagrange Multipliers M and N :

$$L = Tr\left[\sum_{r=1}^{m} s_r t_r^+ + \sum_{r=1}^{m} t_r s_t^+\right] + Tr\left(\sum_{r=1}^{m} s_r s_r^+\right) M + Tr\left(\sum_{r=1}^{m} t_r t_r^+\right) N$$

Optimization provides : $\underset{\{s_r, t_r\}}{Max}\, L \Rightarrow \begin{cases} t_r = M.s_r \\ s_r = N.t_r \end{cases} r = 1, \ldots, m$

As $\Psi = \sum_{r=1}^{m} s_r t_r^+$ then $\sum_{r=1}^{m} t_r t_r^+ = R_Y \Rightarrow M.R_X.M = R_Y$

From which we deduce : $R_X^{1/2}.M.R_X.M.R_X^{1/2} = R_X^{1/2}.R_Y.R_X^{1/2} \Rightarrow \left(R_X^{1/2}.M.R_X^{1/2}\right)^2 = R_X^{1/2}.R_Y.R_X^{1/2}$

And then finally $R_X^{1/2}.M.R_X^{1/2} = \left(R_X^{1/2}.R_Y.R_X^{1/2}\right)^{1/2} \Rightarrow M = R_X^{-1/2}\left(R_X^{1/2}.R_Y.R_X^{1/2}\right)^{1/2} R_X^{-1/2}$

But as $\Psi = \sum_{r=1}^{m} s_r t_r^+ \Rightarrow \Psi = \sum_{r=1}^{m} s_r (M.s_r)^+ = R_X.M$, we conclude that :

$\Psi = R_X.M = R_X^{1/2}\left(R_X^{1/2}.R_Y.R_X^{1/2}\right)^{1/2} R_X^{-1/2}$ with properties of trace,

$Tr\left(\Psi + \Psi^+\right) = 2.Tr\left[\left(R_X^{1/2}.R_Y.R_X^{1/2}\right)^{1/2}\right]$, we can also prove that

$$Y = R_X^{1/2}\left(R_X^{1/2}.R_Y.R_X^{1/2}\right)^{-1/2} R_X^{1/2}.X \quad (6)$$

McCann [8] has defined the optimal transport plans between Gaussian measures on R^k and has proved that the displacement interpolation between any two Gaussian measures is also a

Gaussian measure. For two Gaussian measures $N(m_Y, R_Y)$ and $N(m_X, R_X)$, we can define a positive definite Hermitian matrix : $D_{X,Y} = R_X^{1/2}\left(R_X^{1/2} R_Y R_X^{1/2}\right)^{-1/2} R_X^{1/2} = R_X \circ R_Y^{-1}$ (7)

The geodesic between $N(m_X, R_X)$ and $N(m_Y, R_Y)$ is then given by $N(m_{(t)}, R_{(t)})$ with [9] :

$$\begin{cases} m_{(t)} = (1-t)m_Y + t.m_X \\ R_{(t)} = \left((1-t)I_k + t.D_{X,Y}\right)R_Y\left((1-t)I_k + t.D_{X,Y}\right) \end{cases} \tag{8}$$

It could be proved that : $W_2\left[\left(m_{(s)}, R_{(s)}\right), \left(m_{(t)}, R_{(t)}\right)\right] \leq (t-s).W_2\left[\left(m_{(0)}, R_{(0)}\right), \left(m_{(1)}, R_{(1)}\right)\right]$

with $\left(m_{(0)}, R_{(0)}\right) = \left(m_Y, R_Y\right)$ and $\left(m_{(1)}, R_{(1)}\right) = \left(m_X, R_X\right)$

If we define the following function : $\psi(v) = \frac{1}{2}\left(v - m_Y\right)^+ D_{X,Y}\left(v - m_Y\right) + \left(m_X - v\right)$ (9)

We could observe that if $x = \nabla\psi(y) = D_{X,Y}\left(y - m_Y\right) + m_X$

$$\left(x - m_x\right)^+ R_X^{-1}\left(x - m_X\right) = \left[R_X^{-1/2}\left(x - m_X\right)\right]^+\left[R_X^{-1/2}\left(x - m_X\right)\right] = \left(y - m_Y\right)^+ R_Y^{-1}\left(y - m_Y\right)$$

$$\Rightarrow \left(x - m_x\right)^+ R_X^{-1}\left(x - m_X\right) = \left[R_X^{-1/2} D_{X,Y}\left(y - m_Y\right)\right]^+\left[R_X^{-1/2} D_{X,Y}\left(y - m_Y\right)\right]$$

Proof is given by : $= \left[\left(R_X^{1/2} R_Y R_X^{1/2}\right)^{-1/2} R_X^{1/2}\left(y - m_Y\right)\right]^+\left[\left(R_X^{1/2} R_Y R_X^{1/2}\right)^{-1/2} R_X^{1/2}\left(y - m_Y\right)\right]$

$$= \left(y - m_Y\right)^+ R_X^{1/2}\left(R_X^{1/2} R_Y R_X^{1/2}\right)^{-1} R_X^{1/2}\left(y - m_Y\right) = \left(y - m_Y\right)^+ R_Y^{-1}\left(y - m_Y\right)$$

Using previous result and that $\det D_{X,Y} = \left(\det R_X / \det R_Y\right)^{1/2}$, then $\left(I_k, \nabla\psi\right)_{\#} N\left(m_Y, R_Y\right)$ is the optimal transport between $N(m_Y, R_Y)$ and $N(m_X, R_X)$. In the case of Gaussian measures with zero mean, , the Wasserstein metric is given by [9] : $g_{N(0,R_Y)}(X,Y) = Tr\left(X.R_Y.Y\right)$ (10)

if we identify the tangent space at $N(0, R_Y)$ via the exponential map :

$\exp_{N(0,R_Y)}(t.X) = N\left(0, R_{R_Y,(t)}\right)$ with $R_{R_Y,(t)} = \left((1-t)I_k + t.X\right)R_Y\left((1-t)I_k + t.X\right)$ (11)

We could observe that this space forms a Riemannian length space metrized by the Wasserstein distance, geodesically convex and simply connected (uniqueness of geodesic). This Riemannian manifold is of nonnegative curvature and is flat [6,9]. This space is also an Alexandrov space because of the following property of Wasserstein distance :

$$W_2\left(\alpha, \gamma(t)\right)^2 \geq (1-t).W_2\left(\alpha, \gamma(0)\right)^2 + t.W_2\left(\alpha, \gamma(t)\right)^2 - t(1-t).W_2\left(\gamma(0), \gamma(1)\right)^2$$

Wasserstein barycenter [11] is given by $\underset{\mu}{Inf} \sum_{k=1}^{N} W_2^2(\mu, \nu_k)$. Wasserstein barycenter R of N multivariate Gaussian laws $\left\{N\left(0, R_k\right)\right\}_{k=1}^{N}$ should verify: $\sum_{k=1}^{N}\left(R^{1/2} R_k R^{1/2}\right)^{1/2} = R$. Wasserstein barycenter is given by iterative process : $K^{(n+1)} = \left(\sum_{k=1}^{N}\left(K^{(n)} K_k^2 K^{(n)}\right)^{1/2}\right)^{1/2}$ with $K_i = R_i^{1/2}$

2. Comparison of Fréchet-Wasserstein Distance with Siegel/Rao Distance from Information Geometry

In Information Geometry, we equip $N(m_Y, R_Y)$ with the Fisher information metric which is different from the Wasserstein metric W_2. Fisher metric provides a negative constant sectional curvature while Wasserstein metric is flat [6,9]. Foundation of Information geometry is based on Kullback-Leibler Divergence. Kullback Divergence can be naturally introduced by combinatorial elements and stirling formula. Let multinomial Law of N elements spread on M levels $\{n_i\}$:

$$P_M\left(n_1,n_2,...,n_M / q_1,...,q_M\right)=N!\prod_{i=1}^{M}\frac{q_i^{n_i}}{n_i!}$$

with q_i priors, $\sum_{i=1}^{M} n_i = N$ and $p_i = \frac{n_i}{N}$. Stirling formula gives $n! \approx n^n.e^{-n}.\sqrt{2\pi.n}$ when $n \to +\infty$. We could then observe that it converges to discret version of Kullback-Leibler :

$$\underset{N\to+\infty}{Lim}\frac{1}{N}\log\left[P_M\right]=\sum_{i=1}^{M} p_i.\log\left[\frac{p_i}{q_i}\right]=K(p,q)$$

Donsker and Varadhan have proposed a variational definition of Kullback divergence :

$$K(p,q)=\underset{\phi}{Sup}\left[E_p(\phi)-\log E_q(e^{\phi})\right]$$

$$\text{Consider}:\phi(\omega)=\ln\left(\frac{p(\omega)}{q(\omega)}\right)$$

$$\Rightarrow E_p(\phi)-\ln\left(E_q(e^{\phi})\right)=\sum_{\omega} p(\omega)\ln\left(\frac{p(\omega)}{q(\omega)}\right)-\ln\left[\sum_{\omega} q(\omega)\frac{p(\omega)}{q(\omega)}\right]=K(p,q)-\ln(1)=K(p,q)$$

This proves that the supremum over all ϕ is no smaller than the divergence.

$$E_p(\phi)-\ln\left(E_q(e^{\phi})\right)=E_p\left[\ln\left(\frac{e^{\phi}}{E_q(e^{\phi})}\right)\right]=\sum_{\omega} p(\omega)\left(\ln\left[\frac{q^{\phi}(\omega)}{q(\omega)}\right]\right)$$

with $q^{\phi}(\omega)=\frac{q(\omega)e^{\phi(\omega)}}{\sum_{\theta} q(\theta)e^{\phi(\theta)}} \Rightarrow K(p,q)-\left[E_p(\phi)-\ln\left(E_q(e^{\phi})\right)\right]=\sum_{\omega} p(\omega)\left[\ln\left(\frac{p(\omega)}{q^{\phi}(\omega)}\right)\right]\geq 0$

using the divergence inequality. Link with « Large Deviation Theory » & Fenchel-Legendre transform which gives that logarithm of generating function are dual to Kullback Divergence is given by :

$$\log\left[\int e^{V(x)}q(x)dx\right]=\underset{p}{Sup}\left[\int V(x)p(x)dx-K(p,q)\right]$$

$$\Leftrightarrow K(p,q)=\underset{V(.)}{Sup}\left[\int V(x)p(x)dx-\log\left[\int e^{V(x)}q(x)dx\right]\right]$$

$$\Leftrightarrow K(p,q)=\underset{V(.)}{Sup}\left[E_p(V)-\log E_q\left[e^{V(x)}\right]\right]$$

Chentsov was the first to introduce the Fisher information matrix as a Riemannian metric on the parameter space, considered as a differentiable manifold. Chentsov was led by decision theory when he considered a category whose objects are probability spaces and whose morphisms are Markov Kernels. Chentsov's great achievement was that up to a constant factor the Fisher information yields the only monotone family of Riemannian metrics on the class of finite probability simplexes. In parallel, Burbea and Rao have introduced a family of distance measures, based on the so-called α-order entropy metric, generalizing the Fisher Information metric that corresponds to the Shannon entropy. Such a choice of the matrix for the quadratic differential metric was shown to have attractive properties through the concepts of discrimination and divergence measures between probability distribution. As is well known from differential geometry, the Fisher information matrix is a covariant symmetric tensor of the second order, and hence, the associate metric is invariant under the admissible transformations of the parameters. The information geometry considers probability distributions as differentiable manifolds, while the random variables and their expectation appear as vectors and inner products in tangent spaces to these manifolds.

Chentsov has introduced a distance between parametric families of probability distributions $G_\Theta = \{p(./\theta) : \theta \in \Theta\}$ with Θ the space of parameters, by considering, to the first order, the difference between the log-density functions. Its variance defines a positive definite quadratic differential form based on the elements of the Fisher matrix and a Taylor expansion to the 2nd order of the Kullback divergence gives a Riemannian metric:

$$K(\theta,\tilde{\theta})\Big|_{\tilde{\theta}=\theta+d\theta} \cong K(\theta,\theta) + \left(\frac{\partial K(\theta,\tilde{\theta})}{\partial\tilde{\theta}}\right)^+_{\tilde{\theta}=\theta}\left(\tilde{\theta}-\theta\right) + \frac{1}{2}\left(\tilde{\theta}-\theta\right)^+\left(\frac{\partial^2 K(\theta,\tilde{\theta})}{\partial\tilde{\theta}\partial\tilde{\theta}^*}\right)\left(\tilde{\theta}-\theta\right)$$

$$K\left[p(./\theta), p(./\theta+d\theta)\right] = \frac{1}{2!}\sum_{i,j} g_{ij}(\theta).d\theta_i.d\theta_j^* + O\left(\left|d\theta\right|^3\right) \qquad (12)$$

$$\text{where } I(\theta) = \left[g_{ij}(\theta)\right] \text{ and } \mathrm{g_{ij}}(\theta) = E\left[\frac{\partial \ln p(x/\theta)}{\partial\theta_i}.\frac{\partial \ln p(x/\theta)}{\partial\theta_j^*}\right] \qquad (13)$$

Previous theory of information geometry can be developed for multivariate Gaussian laws. In this section, we will illustrate information geometry on complex circular multivariate Gaussian distribution of zero mean that classically models Radar data and given by :

$$p(X_n/R_n) = (\pi)^{-n}.\left|R_n\right|^{-1}.e^{-Tr\left[\hat{R}_n.R_n^{-1}\right]} \text{ with } \hat{R}_n = \left(X_n - m_n\right).\left(X_n - m_n\right)^+ \text{ and } E\left[\hat{R}_n\right] = R_n \qquad (14)$$

Fisher matrix elements are provided by derivatives of first moments:

$$g_{ij}(\theta) = -E\left[\frac{\partial \ln p(X_n/\theta_n)}{\partial\theta_i.\partial\theta_j}\right] = -Tr\left[\partial_i R_n.\partial_j R_n^{-1}\right] + \partial_i m_n^+.R_n^{-1}.\partial_j m_n \qquad (15)$$

If we assume zero-mean process $m_n = 0$, we deduce from $R_n.R_n^{-1} = I_n \Rightarrow \partial R_n = -R_n.\partial R_n^{-1}.R_n$ that : $g_{ij}(\theta) = Tr\left[\left(R_n.\partial_i R_n^{-1}\right).\left(R_n.\partial_j R_n^{-1}\right)\right]$. Then, we obtain an extended expression of the metric: $ds^2(\theta) = Tr\left[R_n.\left(\sum_i \partial_i R_n^{-1}.d\theta_i\right).R_n.\left(\sum_j \partial_j R_n^{-1}.d\theta_j\right)\right]$ that can be simplified by using that $dR_n^{-1} = \sum_i \partial_i R_n^{-1} d\theta_i$ .We conclude that Information metric can be written:

$$ds^2 = Tr\left[\left(R_n dR_n^{-1}\right)^2\right] = Tr\left[\left(R_n^{-1} dR_n^{-1}\right)^2\right] = Tr\left[\left(d\ln R_n\right)^2\right] \qquad (16)$$

We write this metric synthetically by mean of Frobenius Norm :

$$ds^2 = \left\|R_n^{-1/2} dR_n R_n^{-1/2}\right\|^2 \text{ with } \left\|A\right\|^2 = \langle A, A\rangle \text{ and } \langle A, B\rangle = Tr\left(AB^T\right) \qquad (17)$$

This metric is invariant under the action of the Linear matrix group$\left(GL_n(C),.\right)$: $R_n \to W_n.R_n.W_n^+$, $\mathrm{W_n} \in GL_n(C)$. We can observe that this metric is also invariant by inversion : As $R^{-1}R = I \Rightarrow dR^{-1}R = -R^{-1}dR \Rightarrow ds_R^2 = ds_{R^{-1}}^2$

We have considered the case of zero mean $m = 0$, but for general case, we have :

$ds^2 = dm^+ R^{-1} dm + Tr\left(\left(R^{-1}dR\right)^2\right)$ associated with the following isometries :

$$\begin{aligned} &\left(m, R\right) \to \left(m', R'\right) = \left(A^+ m + a, A^+ R A\right) \\ &ds^2 \mapsto ds'^2 = ds^2 \quad \text{with } \left(a, A\right) \in C^n \mathrm{x} GL(n, C) \end{aligned} \qquad (18)$$

For the time being, there is no explicit expression of Information Geometry distance, when means are not equal to zero $m \neq 0$ [10]. This problem is open in Information Geometry !

By integration, the distance between 2 Radar covariance matrices, R_X and R_Y, is deduced from their extended eigen-values $\{\lambda_k\}_{k=1}^n$:

$$d^2(R_X,R_Y)=\left\|\log\left(R_X^{-1/2}.R_Y.R_X^{-1/2}\right)\right\|^2=\sum_{k=1}^{n}\log^2(\lambda_k) \text{ with } \det\left(R_X^{-1/2}.R_Y.R_X^{-1/2}-\lambda.I\right)=\det(R_Y-\lambda R_X)=0$$

We can also defined the unique geodesic joining 2 matrices , R_X and R_Y. If $t\to\gamma(t)$ is the geodesic between , R_X and R_Y, where $t\in[0,1]$ is such that $d(R_X,\gamma(t))=t.d(R_X,R_Y)$, then the mean of R_X and R_Y is the matrix $R_X\circ R_Y=\gamma(1/2)$. The geodesic parameterized by the length as previously is given by :

$$\gamma(t)=R_X^{1/2}e^{t\log\left(R_X^{-1/2}R_YR_X^{-1/2}\right)}R_X^{1/2}=R_X^{1/2}\left(R_X^{-1/2}R_YR_X^{-1/2}\right)^tR_X^{1/2} \text{ with } 0\le t\le 1$$
$$\gamma(0)=R_X \quad , \quad \gamma(1)=R_Y \text{ and } \gamma(1/2)=R_X\circ R_Y \tag{19}$$

This space is a Bruhat-Tits space. A Bruhat-Tits space is a space with complete metric that satisfies the semi-parallelogram law:

$$\forall x_1,x_2\in X \quad \exists z\in X \text{ such that } d(x_1,x_2)^2+4d(x,z)^2\le 2d(x,x_1)^2+2d(x,x_2)^2 \quad \forall x\in X \tag{20}$$

This space is also a symmetric space as defined by Elie Cartan, such that for every pair *(A,B)* of points of *M* there exists a bijective isometry $G_{(A,B)}$ from *M* to itself with following properties : $G_{(A,B)}A=B$ and $G_{(A,B)}B=A$ (21)

$G_{(A,B)}$ has a unique fixed point $A\circ B$ and $d\left(G_{(A,B)}X,X\right)=2d(X,A\circ B)$ (22)

For space of symmetric definite positive matrices, bijective isometry is given by :

$$G_{(A,B)}X=(A\circ B)X^{-1}(A\circ B) \text{ with } A\circ B=A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2} \tag{23}$$

Barycenter R of N multivariate Gaussian laws $\{N(0,R_k)\}_{k=1}^{N}$ verify: $\sum_{k=1}^{N}\log\left(R^{-1/2}R_kR^{-1/2}\right)=0$.

Barycenter is given by iterative process : $R_{(n+1)}=R_{(n)}^{1/2}e^{\varepsilon\left(\sum_{k=1}^{N}\log\left(R_{(n)}^{-1/2}R_kR_{(n)}^{-1/2}\right)\right)}R_{(n)}^{1/2}$

References:

[1] G. Monge, “Mémoire sur la théorie des déblais et des remblais“, Histoire de l’Académie des Sciences de Paris, p. 666-704, 1781

[2] M. Fréchet, “Sur la distance de deux lois de probabilité“, C.R. Acad. Sci. Paris, n°244, p.689-692, 1957

[3] P. Levy, “Distance de deux variables aléatoires et distance de deux lois de probabilité”, Généralités sur les probabilités de variables aléatoires, Ed. M. Fréchet, Gauthier-Villars, Paris, p. 286-301, 1937

[4] R. Fortet and E. Mourier, “Convergence de la répartition empirique vers la répartition théorique”, Ann. Sci. Ecole Norm. Sup. N°70, p.267-285, 1953

[5] Y. Brenier, “Polar factorisation and monotone rearrangement of vector-valued functions”, Comm. Pure Appl. Math. , vol.44,n°4,p.375-417, 1991

[6] Cédric Villani, “Optimal Transport, Old and New”, Springer-Verlag 2008, http://math.univ-lyon1.fr/homes-www/villani/Cedrif/B07D.StFlour.pdf

[7] D.C. Dowson and B. V. Landau, “The Fréchet Distance between Multivariate Normal Distributions”, Journal of Multivariate Analysis, n°12, p.450-455, 1982

[8] R.J. McCann, “A convexity Principle for Interacting Gases”, Adv. Math., vol.128, n°1, p. 153-179, 1997

[9] A. Takatsu, “On Wasserstein geometry of the space of Gaussian measures”, arXiv:0801.2250v3, 11 February 2009, to appear in Osaka J. Math.

[10] M. Calvo & J.M. Oller, “A Distance between Multivariate Normal Distribution Based in an Embedding into the Siegel Group, Jour. of Multivariate Analysis, vol.35,n°2,p.223-242, 1990

[11] M. Agueh & G. Carlier, “Barycenters in the Wasserstein Space”, SIAM J.Math.Anal., 2011