Riemannian medians and means with applications to radar signal processing

Marc Arnaudon, Frédéric Barbaresco and Le Yang

***Abstract*— We develop a new geometric approach for high resolution Doppler processing based on the Riemannian geometry of Toeplitz covariance matrices and the notion of Riemannian p-means. This paper summarizes briefly our recent work in this direction. First of all, we introduce radar data and the problem of target detection. Then we show how to transform the original radar data into Toeplitz covariance matrices. After that, we give our results on the Riemannian geometry of Toeplitz covariance matrices. In order to compute p-means in practical cases, we propose deterministic and stochastic algorithms, of which the convergence results are given, as well as the rate of convergence and error estimates. Finally, we propose a new detector based on Riemannian median and show its advantage over the existing processing methods.**

***Index Terms*— median, mean, Toeplitz covariance matrices, radar signal processing, Riemannian manifold**

I. Introduction

In recent years, it becomes more and more important to improve the detection performance of PD (Pulsed Doppler) radar in perturbed environment and with smaller bunch of pulses. However, the classical FFT (Fast Fourier Transform) based CFAR (Constant False Alarm Rate) detection procedures (see e.g. [23]) are not very satisfactory due to their low resolution issues. In order to overcome these drawbacks we propose in this paper a new CFAR detection procedure based on nonlinear statistics of radar Toeplitz covariance matrices.

Before explaining our main idea, we briefly introduce the radar data which we intend to analyze and the problem of target detection. For simplicity, we only consider one fixed direction in which a radar sends radio waves and we subdivide this direction into a number of cells. The radar sends each time a gust of radio waves in this direction and then receive echoes. For each echo we measure its amplitude r and phase φ, so that it can be represented by a complex number $re^{i\varphi}$. As a result, the original radar observation value of each cell is a complex vector $z = (z_0, \ldots, z_{n-1})$, where n is the number of radio waves emitted in each gust. Intuitively, a target is an object whose behavior on reflectivity or speed is very different from its environment. Now the aim of target detection is to know, according to the above observation values, whether there are targets at the locations of some cells in the fixed direction.

The fundamental difference between our detection method and the classical FFT-CFAR method (see e.g. [23]) is that, instead of using directly the original observation value z of each cell, we regard it as a realization of a centered stationary Gaussian process and identify it with its covariance matrix $R_n = \mathbf{E}[z\bar{z}]$. In other words, the new observation value for each cell is a covariance matrix, which is also Toeplitz due to the stationarity of the process. Then our new detection procedure can be formulated as follows (see Fig.1): for each cell under test, we compute the distance between the covariance matrix of the cell and the average matrix of the covariance matrices of the reference cells around the cell, if this distance is greater than some threshold, then we can conclude that there is a target at the location of the cell under test.

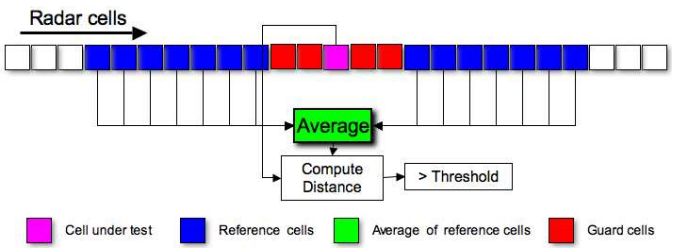

Fig. 1. New detection procedure.

In order to achieve this new detection method, there are three fundamental issues should be addressed. The first one is how to construct Toeplitz covariance matrices from the original radar observation values, the second one is how to measure the distance between two Toeplitz

Marc Arnaudon and Le Yang are with the Laboratoire de Mathématiques et Applications, CNRS : UMR 7348, Université de Poitiers, 86962 Chasseneuil Cedex, France (e-mail: Marc.Arnaudon@math.univ-poitiers.fr, Le.Yang@math.univ-poitiers.fr).

Frédéric Barbaresco is with Thales Air Systems, Thales Air Systems, Surface Radar, Technical Directorate, Advanced Developments Dept. F-91470 Limours, France (e-mail: frederic.barbaresco@thalesgroup.com).

covariance matrices and the third one is how to define the average matrix of Toeplitz covariance matrices and how to compute it. The first question will be answered in Section II through autoregressive models, the second one will be answered in Section III by giving Toeplitz covariance matrices a Riemannian structure and last one will be answered in Section IV by studying the p-means of probability measures on Riemannian manifolds. Finally, it will be shown in Section V that, as far as radar target detection is concerned, the median matrix (i.e. $p = 1$) is the most advisable choice for the average matrix of reference cells and this choice leads to a new OS-HDR-CFAR detector, whose advantage will be illustrated through numerical experiments.

II. Toeplitz covariance matrices as radar observation values

In this section, we show how to transform the original radar data into Toeplitz covariance matrices via reflection coefficients, for which an estimation method is also given. More details can be found in [12], [13] and [27].

A. Toeplitz covariance matrices parameterized by reflection coefficients via autoregressive model

As stated in Section I, the original observation value of each radar cell is a complex vector $z = (z_0, \ldots, z_{n-1})$ which is assumed to be a realization of a centered stationary Gaussian process $Z = (Z_0, \ldots, Z_{n-1})$. Moreover, we assume that the covariance matrix

$$R_n = \mathbf{E}[ZZ^*] = \begin{bmatrix} r_0 & \overline{r}_1 & \ldots & \overline{r}_{n-1} \\ r_1 & r_0 & \ldots & \overline{r}_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ r_{n-1} & \ldots & r_1 & r_0 \end{bmatrix}$$

is strictly positive definite, where $r_k = \mathbf{E}[\, Z_0 \bar{Z}_k \,]$. Now for $1 \leq k \leq l \leq n-1$, the k-th order autoregressive estimate of Z_l is given by $\hat{Z}_l = -\sum_{j=1}^{k} a_j^{(k)} Z_{l-j}$, where the autoregressive coefficients $a_1^{(k)}, \ldots, a_k^{(k)}$ are chosen such that the mean squared error $P_k = \mathbf{E}|Z_l - \hat{Z}_l|^2$ is minimized. The last autoregressive coefficient $a_k^{(k)}$ is called the k-th *reflection coefficient* and is denoted by μ_k. Observe that the classical Levinson's recursion (see e.g. [24]) yields $|\mu_k| < 1$, hence we obtain a map between two open submanifolds of $\mathbf{R}^{2n-1}$:

$$\varphi : \quad \mathcal{T}_n \longrightarrow \mathbf{R}_+^* \times \mathbf{D}^{n-1}, \quad R_n \longmapsto (r_0, \mu_1, \ldots, \mu_{n-1}),$$

where $\mathcal{T}_n$ is the manifold of Toeplitz Hermitian positive definite matrices of order n and $\mathbf{D} = \{\zeta \in \mathbf{C} : |\zeta| < 1\}$ is the open unit disk of the complex plane.

Using the Cramer's rule and the method of Schur complement we can show that φ is a diffeomorphism. More precisely, let R_k be the k-th order principle submatrix of R_n, then the explicit expression of φ is given by

$$\mu_k = (-1)^k \frac{\det S_k}{\det R_k}, \text{ where } S_k = R_{k+1}\begin{pmatrix} 2, \ldots, k+1 \\ 1, \ldots, k \end{pmatrix}$$

is the submatrix of R_{k+1} obtained by deleting the first row and the last column. On the other hand, if $(P_0, \mu_1, \ldots, \mu_{n-1}) \in \mathbf{R}_+^* \times \mathbf{D}^{n-1}$, then its inverse image R_n under φ can be calculated by the following algorithm:

$$r_0 = P_0, \quad r_1 = -P_0\mu_1,$$

$$r_k = -\mu_k P_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}, \quad 2 \leq k \leq n-1,$$

where

$$\alpha_{k-1} = \begin{bmatrix} r_1 \\ \vdots \\ r_{k-1} \end{bmatrix}, \quad J_{k-1} = \begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ & \ldots & & \\ 1 & \ldots & 0 & 0 \end{bmatrix},$$

and

$$P_{k-1} = P_0 \prod_{i=1}^{k-1} (1 - |\mu_i|^2). \tag{1}$$

Since φ is a diffeomorphism, the covariance matrix can be parameterized by reflection coefficients, which has a crucial advantage that, under this reparametrization, the Riemannian metric that we will give to $\mathcal{T}_n$ is diagonal. This means that the Riemannian distances and geodesics admit simple closed form formulae, which make our algorithms for computing p-means applicable. Consequently, instead of estimating the covariance matrix R_n, it suffices to estimate the reflection coefficients $(r_0, \mu_1, \ldots, \mu_{n-1})$ according to the original radar observation value $(z_0, \ldots, z_{n-1})$. This can be achieved by using the regularized Burg algorithm, to which we now turn.

B. Reflection coefficients estimation by regularized Burg algorithm

The regularized Burg algorithm (see [10] and [11]) is an alternative Bayesian composite model approach to spectral estimation. The reflection coefficients, defined in the classical Burg algorithm are estimated through a regularized method, based on a Bayesian adaptive spectrum estimation technique, proposed by Kitagawa and 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 infidelity to the data and infidelity to the frequency domain smoothness constraint. This algorithm conserves lattice structure advantages, and could be brought in widespread use with a multi-segment 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 sensitivity to coefficient value perturbations.

We briefly summarize the regularized Burg algorithm as follows:

Initialization:

$$f_0^{(k)} = b_0^{(k)} = z_k, \quad k = 0, \ldots, n-1;$$
$$P_0 = \frac{1}{n}\sum_{k=0}^{n-1} |z_k|^2 \quad \text{and} \quad a_0^{(0)} = 1.$$

Iteration: for $i = 1, \ldots, n-1$,

$$\mu_i = -\left(\frac{2}{n-i}\sum_{k=i+1}^{n} f_{i-1}^{(k)}\bar{b}_{i-1}^{(k-1)} + 2\sum_{k=1}^{i-1}\beta_k^{(i)} a_k^{(i-1)} a_{i-k}^{(i-1)}\right)$$
$$\Big/\left(\frac{1}{n-i}\sum_{k=i+1}^{n} |f_{i-1}^{(k)}|^2 + |b_{i-1}^{(k-1)}|^2 + 2\sum_{k=0}^{i-1}\beta_k^{(i)}|a_k^{(i-1)}|^2\right).$$

where

$$\beta_k^{(i)} = \gamma_1 (2\pi)^2 (k-i)^2,$$
$$\begin{cases} a_0^{(i)} = 1, \\ a_k^{(i)} = a_k^{(i-1)} + \mu_i \bar{a}_{i-k}^{(i-1)}, \ k = 1, \ldots, i-1, \\ a_i^{(i)} = \mu_i, \end{cases}$$

and

$$\begin{cases} f_i^{(k)} = f_{i-1}^{(k)} + \mu_i b_{i-1}^{k-1}, \\ b_i^{(k)} = b_{i-1}^{(k-1)} + \bar{\mu}_i f_{i-1}^{k}. \end{cases}$$

Thanks to the above regularized Burg algorithm, we can easily estimate the reflection coefficients of every radar cell according to the original radar observation values.

C. Visualization of autoregressive spectra

As another important application of the regularized Burg algorithm, we proceed to show how to apply it to visualize the autoregressive spectra, which are closely related to the speed of targets. According previous analysis, we have in fact for each radar cell an autoregressive model of order $n-1$ and the regularized Burg algorithm allows us to determine not only the reflection coefficients $(P_0, \mu_1, \ldots, \mu_{n-1})$, but also the autoregressive coefficients $(a_1^{(n-1)}, \ldots, a_{n-1}^{(n-1)})$, and the last mean squared error P_{n-1} follows easily by (1). Hence the power spectral density function of the autoregressive model is explicitly given by

$$S(f) = \frac{P_{n-1}}{\left|1 - \sum_{k=1}^{n-1} a_k^{(n-1)} e^{-2i\pi k f}\right|^2}, \quad f \in [-\frac{1}{2}, \frac{1}{2}]. \tag{2}$$

For later usage, we continue to give an example of spectra visualization. To this end, we fix a direction and subdivide its range into 200 cells, in which we insert two targets. For each cell, we simulate an autoregressive process of order 7 and regard it as the echo of the cell. As a result, the original radar observation value of each cell is a complex vector $(z_0, \ldots, z_7)$. Then according to previous analysis, for each cell we have a spectral function given by (2), all these functions are visualized in Fig. 2, where the x axis represents the cells and the y axis represents the frequency f in (2), the spectra function $S(\cdot)$ of each cell corresponds to a colored vertical line and the value $S(f)$ for a specified frequency is indicated by the colorimetric on the right.

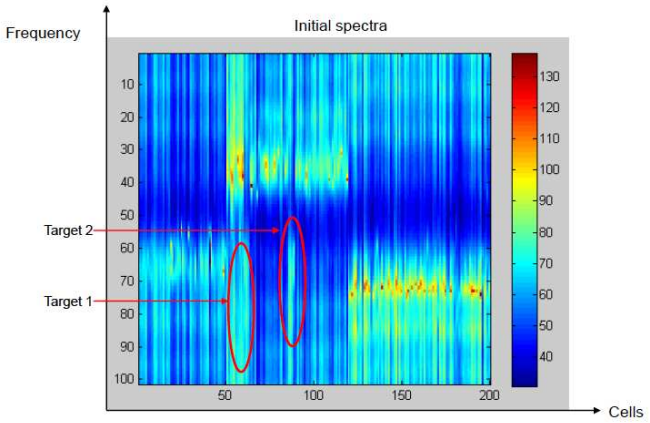

Fig. 2. Initial spectra of simulated data.

III. Riemannian Geometry of Toeplitz Covariance Matrices

Thanks to the reflection coefficients, we can regard $\mathcal{T}_n$ as a Riemannian manifold whose metric, which is introduced in [13] through the Hessian of the Kähler potential

$$\Phi(R_n) = -\ln(\det R_n),$$

is given by

$$ds^2 = n\frac{dP_0^2}{P_0^2} + \sum_{k=1}^{n-1}(n-k)\frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}, \tag{3}$$

where $(P_0, \mu_1, \ldots, \mu_{n-1}) = \varphi(R_n)$.

This metric is a Bergman type metric and it is shown in [27] that this metric is not equal to the Fisher information metric of $\mathcal{T}_n$. But J. Burbea and C. R. Rao have proved in [20, Theorem 2] that the Bergman metric and the Fisher information metric do coincide for some probability density functions of particular forms. A similar potential function was used by S. Amari in [3] to derive the Riemannian metric of multi-variate Gaussian distributions by means of divergence functions.

With the metric (3) the space $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is just the product of the Riemannian manifolds $(\mathbf{R}_+^*, ds_0^2)$ and $(\mathbf{D}, ds_k^2)_{1\leq k\leq n-1}$, where

$$ds_0^2 = n\frac{dP_0^2}{P_0^2} \quad \text{and} \quad ds_k^2 = (n-k)\frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}.$$

The latter is just $n-k$ times the classical Poincaré metric of $\mathbf{D}$. Hence $(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$ is a Cartan-Hadamard manifold whose sectional curvatures K verify $-4 \leq K \leq 0$. The Riemannian distance between two different points x and y in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is given by

$$d(x,y) = \left(n\sigma(P,Q)^2 + \sum_{k=1}^{n-1}(n-k)\tau(\mu_k,\nu_k)^2 \right)^{1/2},$$

where $x = (P, \mu_1, \ldots, \mu_{n-1})$, $y = (Q, \nu_1, \ldots, \nu_{n-1})$,

$$\sigma(P,Q) = |\ln(\frac{Q}{P})| \text{ and } \tau(\mu_k,\nu_k) = \frac{1}{2}\ln\frac{1+|\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}|}{1-|\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}|}.$$

The geodesic from x to y in $\mathcal{T}_n$ parameterized by arc length is given by

$$\gamma(s,x,y) = \Big(\gamma_0(\frac{\sigma(P,Q)}{d(x,y)}s), \gamma_1(\frac{\tau(\mu_1,\nu_1)}{d(x,y)}s), \ldots, \gamma_{n-1}(\frac{\tau(\mu_{n-1},\nu_{n-1})}{d(x,y)}s) \Big),$$

where $\gamma_0(t) = Pe^{t\,\mathrm{sign}(Q-P)}$ and for $1 \leq k \leq n-1$,

$$\gamma_k(t) = \frac{(\mu_k + e^{i\theta_k})e^{2t} + (\mu_k - e^{i\theta_k})}{(1+\bar{\mu}_k e^{i\theta_k})e^{2t} + (1-\bar{\mu}_k e^{i\theta_k})},$$

with $\theta_k = \arg\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}$.

IV. Deterministic and stochastic algorithms for computing Riemannian p-means

This section is devoted to introducing the notion of Riemannian p-means. The existence and uniqueness results are given. Deterministic and stochastic algorithms for computing p-means are developed and are tested by examples. The rate of convergence and error estimates are also obtained. More details can be found in [5] and [27].

A. Riemannian p-means

Let M be a Riemannian manifold whose sectional curvatures $K(\sigma)$ verify $-\beta^2 \leq K(\sigma) \leq \alpha^2$, where α, β are positive numbers. Denote by ρ the Riemannian distance on M. Let $B(a,r)$ be an open geodesic ball in M and μ be a probability measure with support included in $B(a,r)$. Fix $p \in [1,\infty)$. We will always make the following assumptions on (r,p,μ): the support of μ is not reduced to one point. Either $p > 1$ or the support of μ is not contained in a line. The radius r satisfies

$$r < r_{\alpha,p} = \begin{cases} \frac{1}{2}\min\left\{\mathrm{inj}(M), \frac{\pi}{2\alpha}\right\}, & \text{if } 1 \leq p < 2; \\ \frac{1}{2}\min\left\{\mathrm{inj}(M), \frac{\pi}{\alpha}\right\}, & \text{if } p \geq 2, \end{cases},$$

where $\mathrm{inj}(M)$ is the injectivity radius of M.

It has been proved in [1, Theorem 2.1] that the function

$$\begin{aligned} H_p : M &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto \int_M \rho^p(x,y)\mu(dy) \end{aligned}$$

has a unique global minimizer e_p in M, the *p-mean* of μ, and moreover $e_p \in B(a,r)$. Particularly, e_1 and e_2 are the *median* and the *mean* of μ, respectively.

Remark: When μ is not necessarily compactly supported, the almost sure uniqueness of p-means when $p \geq 1$ is proved in [9]. Moreover, to our knowledge, the most precise result on the existence of a Fréchet mean on the circle is given in [21].

Remark: The existence and uniqueness of p-means in Finsler geometry are proved in [7], where algorithms for computing p-means in Finslerian case are also developed.

Remark: We do not consider the circum-center e_∞ of μ in this paper. Nevertheless, interested readers can find in [6] a simple deterministic algorithm for computing the circum-centers of probability measures in Riemannian manifolds.

In order to give the characterizations of p-means, for every $x \in B(a,r)$, we define

$$G_p(x) = \begin{cases} \displaystyle\int_{M\setminus\{x\}} \frac{-\exp_x^{-1} y}{\rho(x,y)}\mu(dy), & \text{if } p = 1; \\ \displaystyle p\int_M \rho^{p-1}(x,y)\frac{-\exp_x^{-1} y}{\rho(x,y)}\mu(dy), & \text{if } p > 1. \end{cases}$$

It is easily seen that if $p > 1$, then $G_p(x)$ is simply the *gradient* vector of H_p at x and e_p is the unique point $x \in B(a,r)$ such that $G_p(x) = 0_x$. The case $p = 1$ deserves a little more explanation. In this case, if $\mu\{x\} > 0$, then the gradient of $y \mapsto \rho(x,y)$ is not well defined at x, so that we eliminate x from the integration domain of H_1

and then take the gradient at x. This yields the vector $G_1(x)$, which is a *subgradient* of H_1 at x. If $\mu\{x\} = 0$, then the gradient of H_1 is well defined at x and coincides with $G_1(x)$ and this holds particularly when x does not belong to the support of μ. It is shown in [25] that e_1 is the unique point $x \in B(a, r)$ such that $|G_1(x)| \leq \mu\{x\}$.

B. Deterministic algorithms for computing p-means

The deterministic algorithms for computing p-means that we are going to present are essentially gradient descent type procedures. To begin with, we choose a sequence of stepsizes $(t_k)_{k\geq 0}$ such that

$$0 < t_k < T_1(\varepsilon, \beta, r, p), \ \lim_{k\to\infty} t_k = 0 \text{ and } \sum_{k=0}^{\infty} t_k = \infty,$$

where the constant

$$T_1(\varepsilon, \beta, r, p) = p\varepsilon^p(\pi p^2(2r)^{2p-1}\beta \coth(2\beta r) + p\varepsilon^{p-1})^{-1},$$

with $\varepsilon = \rho(\operatorname{supp}\mu, \partial B(a, r))/2$.

After that, let $x_0 \in B(a, r)$ be an arbitrary starting point and define

$$x_{k+1} = \exp_{x_k}(-t_k G(x_k)), \quad k \geq 0.$$

Then the sequence $(x_k)_{k\geq 0}$ converges to e_p.

We proceed to give some heuristics of the above algorithm. In Fig. 3 below, μ is supported by four data points y_1, y_2, y_3 and y_4, with possibly different weights. In order to compute the p-mean of μ, we start from some point x_0 and compute the weighted sum of the four unit tangent vectors (black arrows) in $T_{x_0}M$ pointing to the data points. This gives us the tangent vector $-G_p(x_0)$. Then we go along the geodesic (blue line) stating from x_0 with velocity $-G_p(x_0)$ for a t_0 time to arrive at the next iteration point x_1. We repeat this procedure and finally we will arrive at e_p.

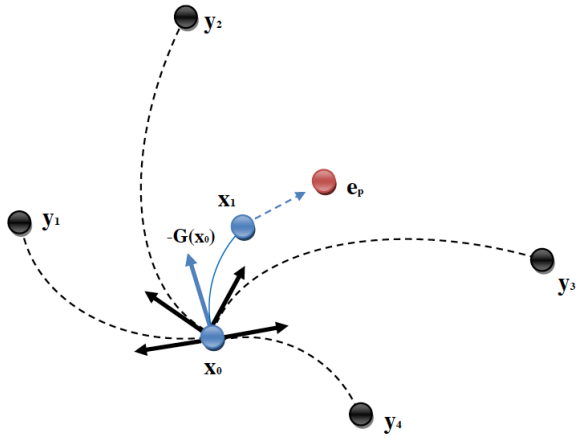

Fig. 3. Deterministic algorithms for computing p-means.

Now we give an example of the above algorithm applied to median computation in the Poincaré disk. In Fig. 4, μ is supported by the three blue points with equal weight and the red point stands for the iteration sequence x_k. The starting state is at the top left and the end state is at the bottom right. The top right and bottom left pictures are two intermediate states of the iteration process. Finally, the red point converges to the median of the three blue points.

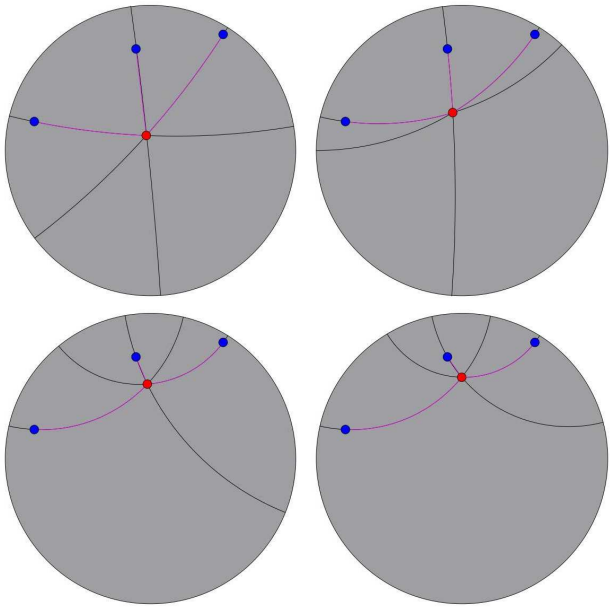

Fig. 4. Median computation by subgradient algorithm in the Poincaré disk.

Before giving the error estimates of the above deterministic algorithms, we give a little explanation on a constant which will be involved in our presentation. It is shown in [5] and [25] that there exists some constant $C_{p,\mu} > 0$ such that for every $x \in B(a, r)$,

$$H_p(x) - H_p(e_p) \geq \frac{C_{p,\mu}}{2}\rho^2(x, e_p).$$

Moreover, for the case when $1 < p \leq 2$, we can choose $C_{p,\mu} = p(2r)^{p-2}\min(p-1, 2\alpha r\cot(2\alpha r))$. But if $p = 1$ or $p > 2$, then $C_{p,\mu}$ maybe depend on the shape of $\operatorname{supp}\mu$ and one can only determine $C_{p,\mu}$ according to concrete cases.

Now the error estimates can be summarized as follows: assume $t_k < C_{p,\mu}^{-1}$ holds for every k, then the following error estimations hold:

i) if $1 \le p < 2$, then for $k \ge 1$,

$$\rho^2(x_k, e_p) \le 4r^2 \prod_{i=0}^{k-1}(1 - C_{p,\mu}t_i) + C(\beta, r, p)\left(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1}(1 - C_{p,\mu}t_i) + t_{k-1}^2\right) := b_k;$$

ii) if $p \ge 2$, then for $k \ge 1$,

$$H_p(x_k) - H_p(e_p) \le (2r)^p \prod_{i=0}^{k-1}(1 - C_{p,\mu}t_i) + C(\beta, r, p)\left(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1}(1 - C_{p,\mu}t_i) + t_{k-1}^2\right) := c_k,$$

where the constant $C(\beta, r, p)$

$$= \begin{cases} p^2(2r)^{2p-1}\beta \coth(2\beta r), & \text{if } 1 < p < 2; \\ p^3(2r)^{3p-4}\left(2\beta r \coth(2\beta r) + p - 2\right), & \text{if } p \ge 2. \end{cases}$$

Moreover, the sequences $(b_k)_k$ and $(c_k)_k$ both tend to zero when k tends to infinity.

Remark: It is shown in [27] that if the stepsizes are chosen to be a multiple of harmonic series, then the convergence rate of the above algorithm is at least sublinear. For example, if $t_k = c(k+1)^{-1}$ with $c > C_{p,\mu}^{-1}$, then for $1 \le p \le 2$ we have $\rho^2(x_k, e_p) = O(k^{-1})$.

C. Stochastic algorithms for computing p-means

We continue to introduce the stochastic gradient descent algorithms for computing p-means. As before, we start by choosing a sequence of stepsizes $(t_k)_{k\ge1}$ such that

$$0 < t_k < T_2(\varepsilon, p, r, \mu), \sum_{k=1}^{\infty} t_k = \infty \text{ and } \sum_{k=1}^{\infty} t_k^2 < \infty,$$

where the constant

$$T_2(\varepsilon, p, r, \mu) = \min(C_{p,\mu}^{-1}, \varepsilon(2r)^{1-p}p^{-1}).$$

Now let $(P_k)_{k\ge1}$ be a sequence of independent $B(a, r)$-valued random variables with the same law μ and $x_0 \in B(a, r-\varepsilon)$ be an arbitrary initial point, then we define inductively a random walk $(X_k)_{k\ge0}$ by $X_0 = x_0$ and

$$X_{k+1} = \exp_{X_k}\left(-t_{k+1}\operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})\right), \quad k \ge 0,$$

where $F_p(x, y) = \rho^p(x, y)$, with the convention $\operatorname{grad}_x F_p(\cdot, x) = 0$. Then the random walk $(X_k)_{k\ge0}$ converges in L^2 and almost surely to e_p.

As before, we give some heuristics of the above stochastic algorithm. In Fig. 5 below, μ is supported by four data points y_1, y_2, y_3 and y_4, with possibly different weights. In order to compute the p-mean of μ, we start from some point X_0 and randomly choose a point according to the law μ, assume y_2 is chosen, then after going along the geodesic joining X_0 and y_2 (dashed line) for a distance $t_1 p\rho^{p-1}(X_0, y_2)$ we arrive at X_1. Similarly, we randomly choose for the second time another point according to the law μ and assume this time y_4 is chosen, then after going along the geodesic joining X_1 and y_4 for a distance $t_2 p\rho^{p-1}(X_1, y_4)$ we arrive at X_2. By repeating this procedure, we will arrive at e_p almost surely.

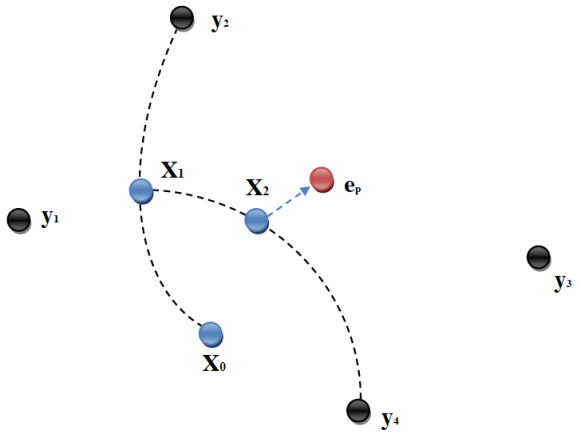

Fig. 5. Stochastic algorithms for computing p-means.

The following example focuses on the case $M = \mathbf{R}^d$ and $p = 2$, where drastic simplifications occur. Assume $M = \mathbf{R}^d$ and μ is a compactly supported probability measure on $\mathbf{R}^d$, the above stochastic gradient algorithm simplifies into $X_0 = x_0$ and for $k \ge 0$,

$$X_{k+1} = X_k - t_{k+1}\operatorname{grad}_{X_k} F_p(\cdot, P_{k+1}). \tag{4}$$

If furthermore $p = 2$, then clearly $e_2 = \mathbf{E}[P_1]$ and $\operatorname{grad}_x F_p(\cdot, y) = 2(x - y)$. As a result, (4) becomes

$$X_{k+1} = (1 - 2t_{k+1})X_k + 2t_{k+1}P_{k+1}.$$

Then easy induction yields for $k \ge 1$,

$$X_k = x_0 \prod_{j=0}^{k-1}(1-2t_{k-j}) + 2\sum_{j=0}^{k-1} P_{k-j}t_{k-j}\prod_{\ell=0}^{j-1}(1-2t_{k-\ell}).$$

Now, taking $t_k = \dfrac{1}{2k}$, we have

$$\prod_{j=0}^{k-1}(1 - 2t_{k-j}) = 0 \quad \text{and} \quad \prod_{\ell=0}^{j-1}(1 - 2t_{k-\ell}) = \frac{k-j}{k}$$

so that

$$X_k = \sum_{j=0}^{k-1} P_{k-j}\frac{1}{k} = \frac{1}{k}\sum_{j=1}^{k} P_j.$$

The stochastic gradient algorithm estimating the mean e_2 of μ is given by the empirical mean of a growing sample of independent random variables with distribution μ. In this simple case, our convergence result is nothing but the Strong Law of Large Numbers.

Now we give an example of the above stochastic algorithm applied to median computation in the Poincaré disk. In Fig. 6, μ is supported by the three blue points with equal weight. The black point stands for the state of the inhomogeneous Markov chain $(X_k)_{k\geq 0}$ and the red path represents one of its trajectory. The green point is the median of the blue points computed by the subgradient algorithm in the preceding subsection. The starting state is at the top left and the end state is at the bottom right. The top right and bottom left pictures are two intermediate states of the iteration process. Finally, the black point converges to the green point.

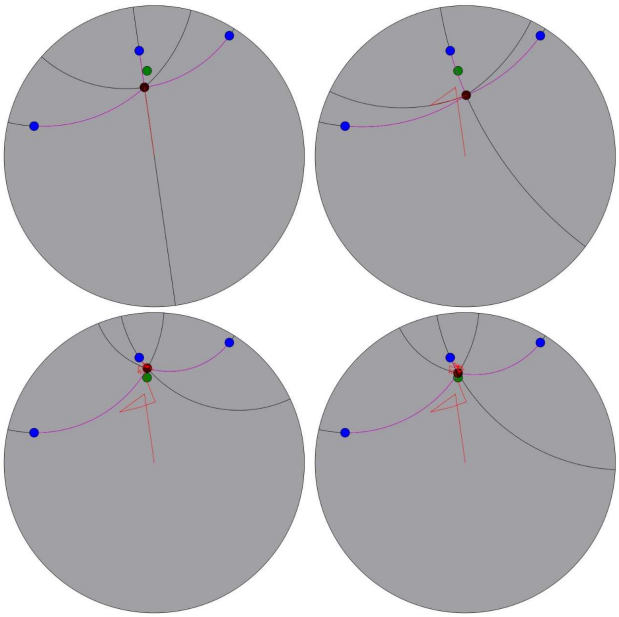

Fig. 6. Median computation by stochastic gradient algorithm in the Poincaré disk.

We end this section by showing another example of the stochastic gradient algorithm applied to median computation. In Fig. 8, M is the Euclidean plane $\mathbf{R}^2$ and μ is an absolutely continuous law with density given in Fig. 7. The red path represents one trajectory of the inhomogeneous Markov chain $(X_k)_{k\geq 0}$ corresponding to $p = 1$, with linear interpolation between the different steps. The red point is e_1. White blobs represent the values of $(P_k)_{k\geq 1}$.

It is well known that the convergence rate of the law of large numbers is given by the central limit theorem. This is also the case for our stochastic gradient descent algorithm. Now we turn to a central limit theorem which

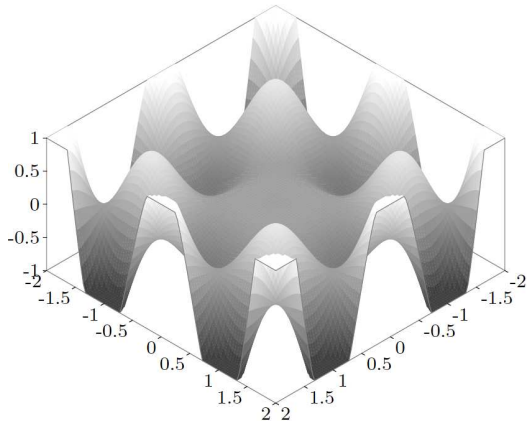

Fig. 7. Density function on the square $[-2,2]\times[-2,2]$.

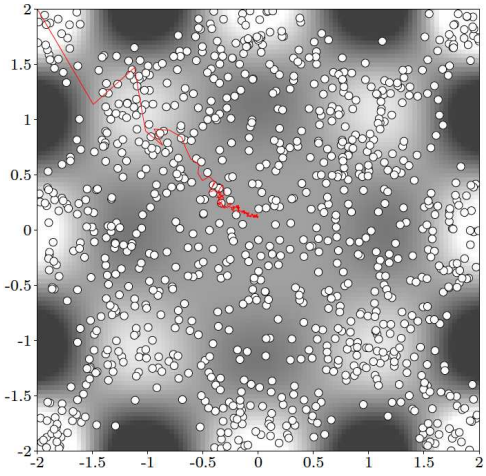

Fig. 8. Median computation by stochastic gradient algorithm on the square $[-2,2]\times[-2,2]$

gives the convergence speed of our stochastic gradient descent algorithm.

For every $k \geq 1$, let $t_k = \min(\delta/k, T_2(\varepsilon, p, r, \mu))$ for some $\delta > C_{p,\mu}^{-1}$. Since the exponential map at e_p is a diffeomorphism onto $B(a,r)$, we can define for every $n \geq 1$ a Markov chain $(Y_k^n)_{k\geq 0}$ in $T_{e_p}M$ by

$$Y_k^n = \frac{k}{\sqrt{n}} \exp_{e_p}^{-1} X_k.$$

Assume H_p is C^2 in a neighborhood of e_p, then the sequence of processes $\left(Y_{[nt]}^n\right)_{t\geq 0}$ converges weakly in $\mathbb{D}((0,\infty), T_{e_p}M)$ to a diffusion process given by

$$y_\delta(t) = \sum_{i=1}^{d} t^{1-\delta\lambda_i} \int_0^t s^{\delta\lambda_i - 1} \langle \delta\sigma \, dB_s, e_i\rangle e_i, \quad t \geq 0,$$

where B_t is the standard Brownian motion in $T_{e_p}M$ and $\sigma \in \mathrm{End}(T_{e_p}M)$ satisfying

$$\sigma\sigma^* = \mathbb{E}\left[\mathrm{grad}_{e_p} F_p(\cdot, P_1) \otimes \mathrm{grad}_{e_p} F_p(\cdot, P_1)\right],$$

$(e_i)_{1\leq i\leq d}$ is an orthonormal basis diagonalizing the symmetric bilinear form $\nabla dH_p(e_p)$ and $(\lambda_i)_{1\leq i\leq d}$ are the associated eigenvalues. Moreover, the process y_δ satisfies the following SDE:

$$dy_\delta(t) = \frac{1}{t}[y_\delta(t) - \delta\nabla dH_p(y_\delta(t),\cdot)^\sharp]dt + \delta\sigma dB_t,$$

where $\nabla dH_p(y,\cdot)^\sharp$ is the dual vector of $\nabla dH_p(y,\cdot)$.

V. Applications: A new OS-HDR-CFAR detector

With the notion of p-means, we could choose in principle the "average matrix" stated in Section I to be the p-mean of reflection coefficients. But considering the heavily perturbed radar environment, the most appropriate choice should be the *median*, which is a prominent ordered statistic estimator due to its robustness. In fact, ordered statistic is a very useful tool used in Radar for a long time to be robust against outliers on scalar data from secondary data.

Now we are to apply the tools developed in the previous sections to build an Ordered-Statistic High Doppler Resolution Constant False Alarm Rate (OS-HDR-CFAR) detector (see [16], [17], [18] and [19]) jointly taking into account the robustness of median matrix and high Doppler resolution of regularized complex autoregressive model. This new detection procedure is shown in Fig. 9. To be more precise, firstly, we use the regularized Burg algorithm to compute the reflection coefficients of each cell in the fixed direction. After that, for each cell under test, we use the subgradient algorithm and the Riemannian structure of Toeplitz covariance matrices to compute the median of the reflection coefficients of the reference cells around the cell under test. Finally, we compute the Riemannian distance between the reflection coefficients of the cell under test and the median reflection coefficients just obtained, if this distance is greater than some threshold, then we can conclude that there is a target at the location of the cell under test.

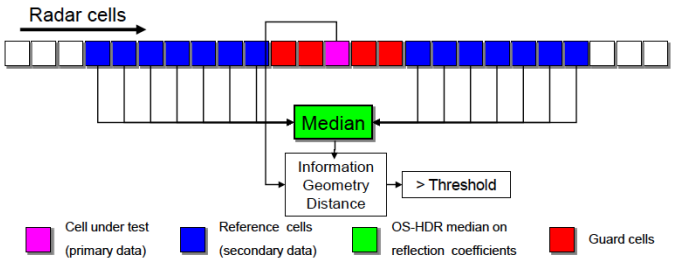

Fig. 9. OS-HDR-CFAR detection procedure.

The comparison of the classical CFAR detector and the new OS-HDR-CFAR one is shown in Fig. 10.

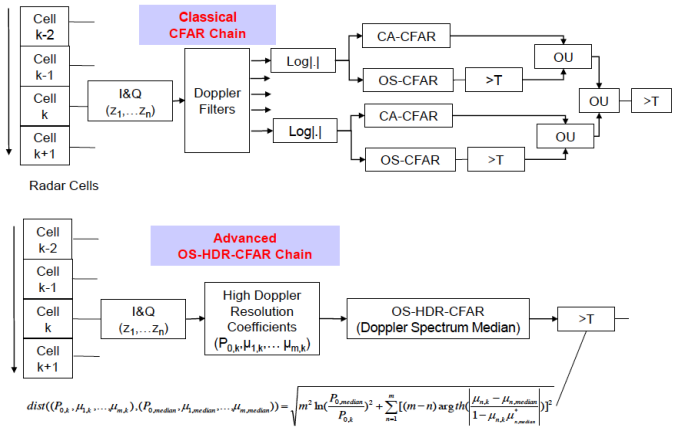

Fig. 10. Comparison of classical CFAR detector and the new OS-HDR-CFAR detector.

A. Test of OS-HDR-CFAR on simulated data

We proceed to show the advantage of the new detector over the classical ones. For this purpose, we come back to the example considered in Section II-C and work within this context.

Since the autoregressive spectra are closely related to the speed of targets, we first investigate the spectral performance of the new detector. Our results on the median spectra and the mean spectra are shown in Fig. 11. To be more precise, in order to obtain the median spectra, for each cell ranging from the 8-th to the 193-th, we use the subgradient algorithm in Section IV-B to compute the median of the reflection coefficients of the window (cf. Fig. 9) centered on this cell and consisting of 15 reference cells around it without 4 guard cells. After that, we apply (2) to each of the median reflection coefficients to obtain a spectral function $S(\cdot)$. As in Section II-C, all these functions are visualized in the picture at the top of Fig. 11, where the x axis represents the cells and the y axis represents the frequency f in (2), the spectra function $S(\cdot)$ of each cell corresponds to a colored vertical line and the value $S(f)$ for a specified frequency is indicated by the colorimetric on the right. Replacing the median by the mean in the above procedure yields the mean spectra which is shown at the bottom of Fig. 11.

Now we are ready to analyze these spectra. On the one hand, it is easily seen that the median spectra are robust against the second target but the mean spectra are strongly affected by it as there is an evident distortion at the location where the second target appears in the picture of mean spectra. This phenomena well explains the fact that the median is much more robust than the mean when outliers appear (see e.g. [22] and [26]). As the robustness is a desired property for detectors, the

OS-HDR-CFAR detector is advantageous over the mean-HDR-CFAR detector as far as robustness is concerned. On the other hand, the three plateau in Fig. 2 are simulated to represent transitions of clutter. After filtered by median, the edges of clutter transitions are much better marked and, in contrast, they are not well marked after the mean filtering. This is another advantage of the OS-HDR-CFAR detector over the mean-HDR-CFAR detector when the targets are near the transitions of clutter.

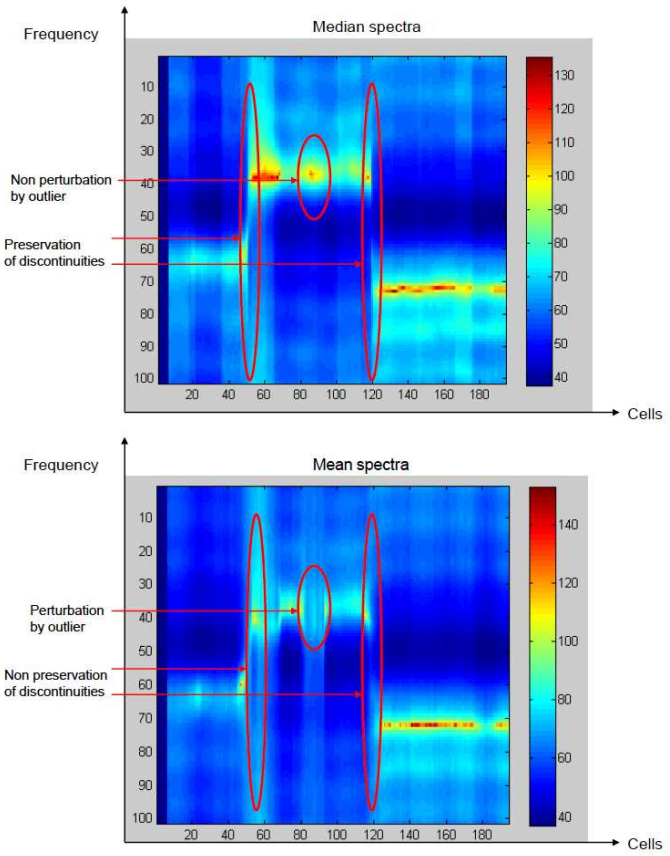

Fig. 11. Comparison of the median spectra and the mean spectra.

We finish our test on simulated data by showing the detection performance of the three detectors: OS-HDR-CFAR, Mean-HDR-CFAR and FFT-CFAR, the results are shown in Fig. 12, in which the x axis stands for the cells and the y axis stands for the Riemannian distance between the original reflection coefficients of each cell and the median reflection coefficients of the cell computed when producing the median spectra. It is easily seen that the two inserted targets are much better detected by the OS-HDR-CFAR detector than by the other two detectors. One can also observe that the performance of the classical FFT-CFAR detector is very poor.

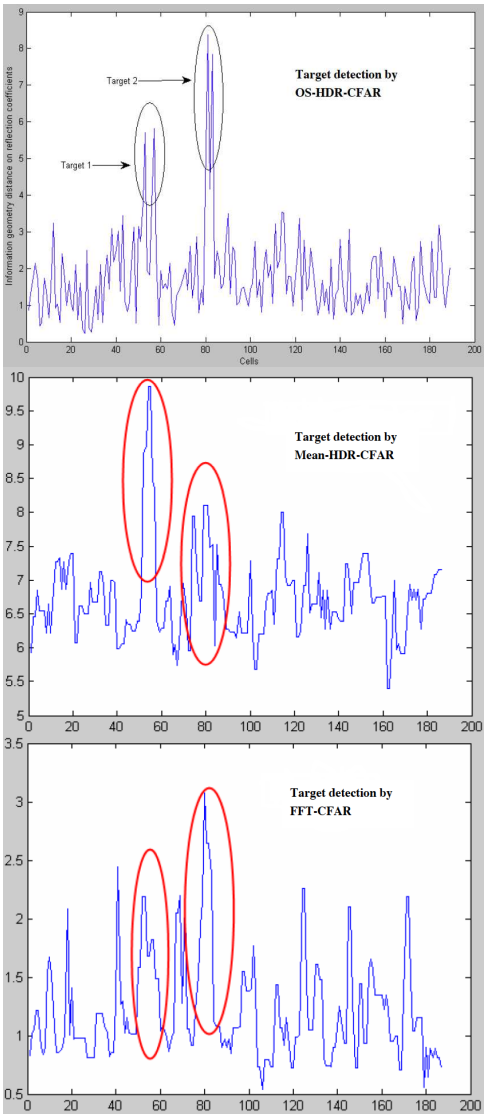

Fig. 12. Comparison of the detection quality. From top to bottom: OS-HDR-CFAR, Mean-HDR-CFAR and FFT-CFAR.

B. Test of OS-HDR-CFAR on real data

We have also tested OS-HDR-CFAR on real recorded ground Radar clutter with ingestion of synthetic slow targets. In Fig. 13, we give ROC curves (see e.g. [23]) with Probability of detection versus Probability of false alarm. We observe that OS-HDR-CFAR is better ($P_d = 0.89$) than OS-CFAR/Doppler-filters ($P_d = 0.65$) and Mean-HDR-CFAR for arbitrarily fixed P_{fa}.

VI. Conclusions

A new approach for high resolution Doppler processing is developed. In order to achieve this method, we have studied the Riemannian geometry of Toeplitz covariance matrices and the p-means of probability measure on Riemannian manifolds. Deterministic and

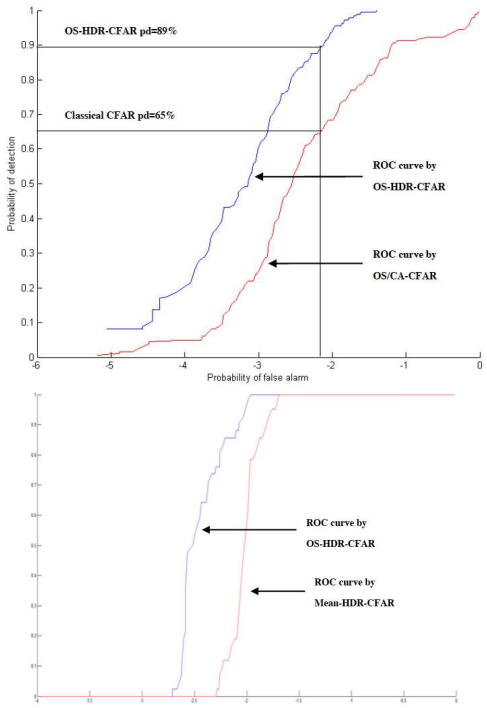

Fig. 13. Comparison of ROC curves.

stochastic algorithms for computing p-means are given, as well as their error estimates. Finally, it is shown that this new processing method possesses many advantages over the existing ones, especially for the cases when targets are near and move slowly.

REFERENCES

[1] B. Afsari, Riemannian L^p center of mass : existence, uniqueness, and convexity, Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.

[2] B. Afsari et al., On The Convergence of Gradient Descent for Finding the Riemannian Center of Mass, arXiv:1201.0925v1, (2011).

[3] S. Amari and A. Cichocki, Information geometry of divergence functions, Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 58, No. 1, 2010.

[4] M. Arnaudon and X. M. Li, Barycenters of measures transported by stochastic flows, The Annals of probability, 33 (2005), no. 4, 1509-1543.

[5] M. Arnaudon, C. Dombry, A. Phan and L. Yang, Stochastic algorithms for computing means of probability measures, Stochastic Processes and their Applications, Vol. 122, No. 4, pp. 1437-1455 (2012)

[6] M. Arnaudon and F. Nielsen, On approximating the Riemannian 1-center, Computational Geometry, Vol 46, No.1, pp. 93-104 (2013).

[7] M. Arnaudon and F. Nielsen, Medians and means in Finsler geometry, LMS J. Comput. Math. vol 15 (2012), pp. 23-37.

[8] M. Arnaudon, F. Barbaresco and L. Yang, Medians and means in Riemannian geometry: existence, uniqueness and computation, Matrix Information Geometry (R. Bhatia and F. Nielsen, Eds.), pp. 169-198, Springer (2012)

[9] M. Arnaudon and L. Miclo, Means in complete manifolds: uniqueness and approximation, hal-00717677-version 1 (2012), to appear in ESAIM P&S.

[10] F. Barbaresco, Algorithme de Burg Régularisé FSDS. Comparaison avec lalgorithme de Burg MFE, XVème colloque GRETSI 1, pp. 29-32 (1995)

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

[12] F. Barbaresco, Innovative Tools for Radar Signal Processing Based on Cartan's Geometry of SPD Matrices and Information Geometry, IEEE International Radar Conference (2008).

[13] F. Barbaresco, Interactions between Symmetric Cone and Information Geometries, ETVC'08, Springer Lecture Notes in Computer Science (2009), pp. 124-163.

[14] F. Barbaresco and G. Bouyt, Espace Riemannien symétrique et géométrie des espaces de matrices de covariance : équations de diffusion et calculs de médianes, GRETSI'09 conference, Dijon, September 2009

[15] 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

[16] 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

[17] 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

[18] 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

[19] F. Barbaresco, Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median, Matrix Information Geometry (R. Bhatia and F. Nielsen, Eds.), pp. 199-256, Springer (2012)

[20] J. Burbea and C. R. Rao, Differntial metrics in probability spaces, Probability and Mathematical Statistics, Vol. 3, Fasc. 2, pp. 241-258, 1984.

[21] B. Charlier, Necessary and sufficient condition for the existence of a Fréchet mean on the circle, hal-00620965, version 1, (2011).

[22] P. T. Fletcher et al., The geometric median on Riemannian manifolds with application to robust atlas estimation, NeuroImage, 45 (2009), S143-S152.

[23] M. A. Richards, Fundamentals of Radar Signal Processing, McGraw-Hill, New York, 2005.

[24] P. P. Vaidyanathan The Theory of Linear Prediction, Morgan and Claypool. 2008

[25] L. Yang, Riemannian median and its estimation, LMS J. Comput. Math. vol 13 (2010), pp. 461-479.

[26] L. Yang, Some properties of Fréchet medians in Riemannian manifolds, preprint hal-00633203, version 2 (2011).

[27] L. Yang, Medians of probability measures in Riemannian manifolds and applications to radar target detection. PhD thesis, Université de Poitiers, tel-00664188, version 1 (2011).