A Riemannian Approach for Training Data Selection in Space-Time Adaptive Processing Applications

J-F. Degurse*, **L. Savy***, **J-Ph. Molinié***, **S. Marcos****

*Office National d'Etudes et de Recherches Aerospatiales (ONERA)
Chemin de la Hunière, Palaiseau, FRANCE
email: jean-francois.degurse@onera.fr

**Laboratoire des Signaux et des Systèmes (CNRS-Univ-Paris-Sud-Supelec)
Rue Joliot-Curie, 3, Gif-sur-Yvette, FRANCE
email: sylvie.marcos@lss.supelec.fr

Abstract: *Heterogeneous situations are a serious problem for Space-Time Adaptive Processing (STAP) in an airborne radar context. Indeed, STAP detectors need secondary training data that have to be homogeneous with the tested data, otherwise the performances of these detectors are severely impacted when facing heterogeneous environments. Hence, training data have to be carefully selected and this is traditionally done in Euclidean geometry. We introduce a new criterion for data selection. We show that it can be viewed as an approximation of the metric distance in Riemannian geometry.*

1. Introduction

STAP performs two-dimensional space and time adaptive filtering where different space channels are combined at different times [1]. In the context of radar signal processing, the aim of STAP is to remove ground clutter returns, in order to enhance slow moving target detection. Filter's weights are adaptively estimated from training data in the neighborhood of the range cell of interest, called cell under test (CUT). The estimation of these weights is always deduced, more or less directly, from an estimation of the covariance matrices of the received signal, which is the key quantity in the process of adaptation [2].
Consider a radar antenna made of N sensors that acquires M_p pulse snapshots for each l range gate for which the received data have been arranged into an $NM \times K_t$ matrix $\mathbf{X}_l$ with M the number of pulses of the spatio-temporal vector and $K_t = M_p - M + 1$ being the number of training data snapshots in the Doppler dimension. A covariance matrix $\mathbf{R}_l$ is estimated from $\mathbf{X}_l$ at each range gate, and the matrix used to filter the range CUT is estimated using covariance matrix obtained from p adjacent cells :

$$\mathbf{R}_l = \frac{\mathbf{1}}{K_t}\mathbf{X}_l\mathbf{X}_l^{\mathbf{H}} \Longrightarrow \mathbf{R}_{CUT} = \frac{\mathbf{1}}{p}\sum_{l=\mathbf{1}}^{p}\mathbf{R}_l \tag{1}$$

With $\mathrm{s_s}$ being the spatio-temporal steering vector (length NM), the STAP weights are then:

$$\mathbf{w}^{\mathbf{H}} = \mathbf{s}_{\mathbf{s}}^{\mathbf{H}}\mathbf{R}_{CUT}^{-\mathbf{1}} \tag{2}$$

This STAP method is usually referred to as the sample matrix inversion (SMI). One main consideration goes into the choice of the p training covariance matrices: how many and which matrices share the same statistics with the data sample to which the weights are to be applied. On one hand, the statistics of the clutter often change very quickly and, on the other hand, we want to use as many matrices as possible to obtain a good estimate of the covariance matrix that minimizes the estimation loss.

2. Search for an optimal training covariance matrix selection

The traditional way to answer these questions is to use a power selection criterion [3]. The power selection criterion is an application of the Euclidean distance d_1 between two matrices, which consists of the minimum Frobenius norm of the difference between one secondary data covariance matrix $\mathbf{R}_l$ and the covariance matrix of the tested data $\mathbf{R}_0$:

$$d_1^2(\mathbf{R}_l, \mathbf{R}_0) = ||\mathbf{R}_l - \mathbf{R}_0||_F^2 = trace[(\mathbf{R}_l - \mathbf{R}_0)(\mathbf{R}_l - \mathbf{R}_0)^H] \quad (3)$$

One can clearly see that this method, which works only on the signal power of the covariance matrix, doesn't take advantage of the structure of the covariance matrix.

Our approach to the problem is to take a physical point of view and look for the minimum distance that fits to the minimization problem of the STAP filter. Let $\mathbf{w}_l^{\mathbf{H}} = \mathbf{s}_\mathbf{s}^{\mathbf{H}}\mathbf{R}_l{}^{\mathbf{-1}}$ be the filter weights obtained with $\mathbf{R}_l$ and applied to the tested data $\mathbf{X}_0$:

$$\mathbf{z} = \mathbf{w}_l^{\mathbf{H}}\mathbf{X}_0 \quad (4)$$

$$z = \mathbf{s}_\mathbf{s}^{\mathbf{H}}\mathbf{R}_l{}^{\mathbf{-1/2}}\mathbf{R}_l{}^{\mathbf{-1/2}}\mathbf{X}_0 \quad (5)$$

$$\mathbf{z} = \mathbf{s}_\mathbf{s}^{\mathbf{H}}\mathbf{R}_l{}^{\mathbf{-1/2}}\mathbf{y} \quad (6)$$

The term $\mathbf{y} = \mathbf{R}_l{}^{-1/2}\mathbf{X}_0$ is the whitened signal whereas $\mathbf{s}_\mathbf{s}^{\mathbf{H}}\mathbf{R}_l{}^{-1/2}$ represents the matched filter. If $\mathbf{w}_l^{\mathbf{H}}$ is the optimal weights vector, $\mathbf{y}$ is whitened hence we want $E\{\mathbf{y}\mathbf{y}^{\mathbf{H}}\} = \mathbf{I}$, i.e:

$$E\{\mathbf{R}_l^{\mathbf{-1/2}}\mathbf{X}_0\mathbf{X}_0^{\mathbf{H}}\mathbf{R}_l^{\mathbf{-1/2}}\} = \mathbf{I} \quad (7)$$

$$\mathbf{R}_l^{\mathbf{-1/2}}E\{\mathbf{X}_0\mathbf{X}_0^{\mathbf{H}}\}\mathbf{R}_l^{\mathbf{-1/2}} = \mathbf{I} \quad (8)$$

$$\mathbf{R}_l^{\mathbf{-1/2}}\mathbf{R}_0\mathbf{R}_l^{\mathbf{-1/2}} - \mathbf{I} = \mathbf{0} \quad (9)$$

This leads to a physical "distance" that we want to minimize:

$$d_2^2(\mathbf{R}_\mathbf{l}, \mathbf{R}_\mathbf{0}) = ||\mathbf{R}_l^{\mathbf{-1/2}}\mathbf{R}_0\mathbf{R}_l^{\mathbf{-1/2}} - \mathbf{I}||_\mathbf{F}^\mathbf{2} \quad (10)$$

To filter the data $\mathbf{X}_0$ with $\mathbf{R}_0$ as in (1), we will look for the adjacent covariance matrices $\mathbf{R}_l$ that have the minimal distance d_2^2 to $\mathbf{R}_0$.

3. Link to the Riemannian matrix geometry

When working on Hermitian positive-definite matrices, it is natural and desirable to work with the information geometry metric on the symmetric cone [4]. The Riemannian metric distance between two matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ is defined by [5]:

$$d_3^2(\mathbf{R_1}, \mathbf{R_2}) = ||\log(\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}})||_F^2 = \sum_{k=1}^{N} \log^2(\lambda_k) \tag{11}$$

where the λ_k are the eigenvalues of the matrix $\mathbf{R_1^{-1} R_2}$ that is to say the solution of $\det(\mathbf{R_2} - \lambda \mathbf{R_1}) = 0$.
As stated in (10), the term $\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}}$ is close to the identity matrix $\mathbf{I}$. Consequently, we can approximate (11) with the first order development of the following Taylor series:

$$\log(\mathbf{A}) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(\mathbf{A} - \mathbf{I})^n}{n} \approx \mathbf{A} - \mathbf{I} \tag{12}$$

Replacing $\mathbf{A}$ by $\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}}$ in (12) yields:

$$||\log(\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}})||^2 \approx ||\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}} - \mathbf{I}||^2 \tag{13}$$

We can then approximate the Riemannian metric distance from (11) by:

$$d^2(\mathbf{R_1}, \mathbf{R_2}) \approx ||\mathbf{R_1^{-1/2} R_2 R_1^{-1/2}} - \mathbf{I}||_F^2 \tag{14}$$

We deduce that the proposed physical “distance” (10) is an approximation of the Riemannian metric distance (11). It implies that this distance should be a much better criterion for the selection of training data than the Euclidean distance (3).

4. Symmetrization and simplification

We can demonstrate the following simplification of (10):

$$d_2^2(\mathbf{R}_l, \mathbf{R_0}) = ||\mathbf{R}_l^{-1/2} \mathbf{R}_0 \mathbf{R}_l^{-1/2} - \mathbf{I}||_{\mathbf{F}}^{\mathbf{2}} = ||(\mathbf{R}_l^{-1} \mathbf{R}_0 - \mathbf{I})||_{\mathbf{F}}^{\mathbf{2}} \tag{15}$$

The criterion described in (10) called physical “distance” does not meet all the properties of a mathematical distance. In particular, d_2 is not symmetric:

$$d_2(\mathbf{R}_l, \mathbf{R_0}) \neq d_2(\mathbf{R_0}, \mathbf{R}_l) \tag{16}$$

This issue could be a problem to set up a data selection strategy. Indeed, the chosen criterion is based on the whitening of the interference. Therefore, if R_l and R_0 are built from data containing the same interference but the power of the interference of the data that form the matrix R_l are less powerful than the interference of R_0, the “distance” won’t increase as the signal would be still whitened, although over-whitened.

We can however resolve this problem, by defining a symmetric physical “distance”:

$$d_4^2 = [\frac{1}{2}(d_2(\mathbf{R}_l, \mathbf{R_0}) + d_2(\mathbf{R_0}, \mathbf{R}_l))]^2 \tag{17}$$

5. Results

The performance of the distances described above are compared on realistic synthetic data, simulating a nose AMSAR antenna in a GMTI mode. Characteristics of the configuration used to simulate the data can be found in [6]. In these data, different range gate areas with different types of clutter are present. Area **A** is made of clutter seen by sidelobes. Areas **B** and **D** are Gaussian clutter which powers are σ_0 and $\sigma_0/10$ respectively and area **C** is composed of spiky clutter. The areas **A** and **C** are very heterogeneous, i.e we can't use many adjacent range gates to estimate the weights of the filter whereas in areas **B** and **D**, clutter is homogeneous, so we can use many data from adjacent range gates, as long as these data come from the respective areas.

We plot the normalized distance between the covariance of the tested range $l = 0$ gate and the range $l = 10$. The normalization is achieved by dividing both distances by the distance between two noise covariance matrices of the same size and the same number of estimates.

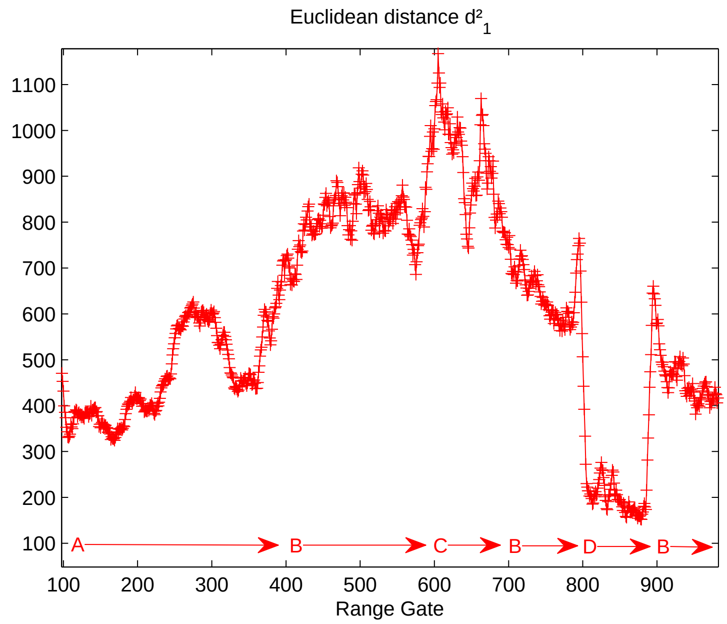

Figure 1: Euclidean distance between $\mathbf{R}_0$ and $\mathbf{R}_{10}$ for different clutter type

The result of the Euclidean distance in Fig. 1 shows that we're not able to distinguish the different types of clutter along range gates, except for, as expected, the **D** area which is a less powerful clutter area.

In Fig. 2 however, the physical "distance" is able to separate almost all the different kinds of ground echos. It only fails to detect the change between Gaussian clutter σ_0 and $\sigma_{0/10}$.

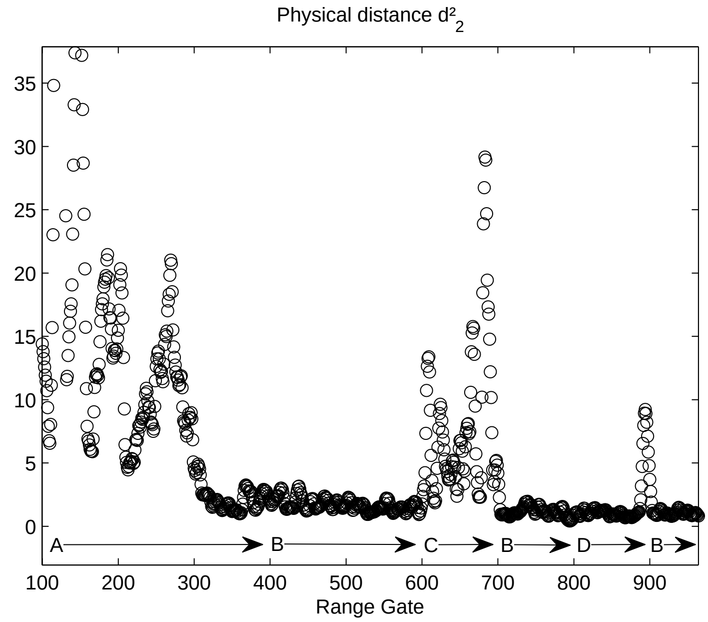

Figure 2: Physical criterion “distance” between $\mathbf{R}_{\mathit{0}}$ and $\mathbf{R}_{\mathit{10}}$ for different clutter type

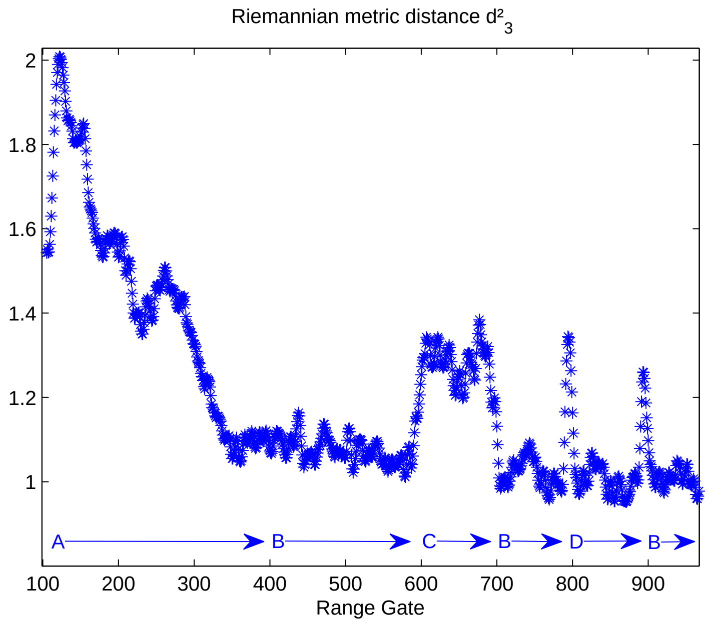

Figure 3: Riemannian metric distance between $\mathbf{R}_{\mathit{0}}$ and $\mathbf{R}_{\mathit{10}}$ for different clutter type

Finally, in Fig. 3, the metric distance in Riemannian geometry performs very well, detecting all the clutter changes. On both Fig. 2 and Fig. 3, the distances between matrices point out the heterogeneity of the areas and the fact that it is not possible to use many adjacent range gates as training data. In Fig. 4, as predicted, the symmetric physical “distance” does not fail in detecting the change of clutter at range gate number 700. Except from this point, both distances d_4 and d_2 are very close.

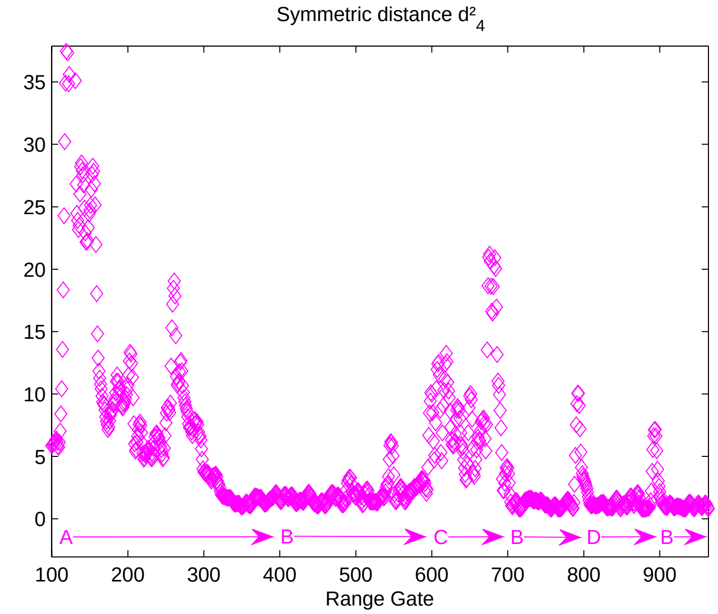

Figure 4: Symmetric physical criterion "distance" between $\mathbf{R}_0$ and $\mathbf{R}_{10}$ for different clutter type

6. Conclusion

A new approach for the selection of training data have been investigated. This new approach outperforms the classical approach in detecting heterogeneity and homogeneity of the interference in the fast time domain. With these methods, an overall processing strategy can be set up to determine how many and which training data are to be chosen to build the adaptive filters.

References

[1] Melvin, W.L., "A STAP overview", *IEEE Aerospace And Electronic Systems Magazine*, 2004, vol. 19, no. 1, pp. 19-35

[2] Klemm, R., "Principles of space-time adaptive processing", *The Institution of Electrical Engineers (IEE)*, 2002, pp. 644.

[3] Kogon, S.M.; Zatman, M.A.; , "STAP adaptive weight training using phase and power selection criteria," Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on , vol.1, no., pp.98-102 vol.1, 4-7 Nov. 2001

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

[5] Bhatia, R., "Positive Definite Matrices.", Princeton Series in Applied Mathematics. Princeton, 501 University Press, illustrated edition, Princeton (2006)

[6] Degurse, J.-F. and Savy, L. and Perenon, R. and Marcos, S., "An extended formulation of the Maximum Likelihood Estimation algorithm. Application to space-time adaptive processing", PRadar Symposium (IRS), 2011 Proceedings International,pp.763 -768, Sept. 2001