Tesi di Dottorato

Università degli Studi di Napoli "Federico II"

Dipartimento di Ingegneria Elettrica
e delle Tecnologie dell'Informazione

Dottorato di Ricerca in
Ingegneria Elettronica e delle Telecomunicazioni

Covariance Matrix Estimation for Radar Applications

Luca Pallotta

Il Coordinatore del Corso di Dottorato
Ch.mo Prof. Niccoló Rinaldi

Il Tutore
Ch.mo Prof. Antonio De Maio

XXVI Ciclo

Acknowledgment

I thank Selex ES and SESM for supporting my PhD's scholarship. Also, I express my gratitude to Dr. Alfonso Farina for his technical support during my research activities, the continuous assistance, encouragement and kindness demonstrated during these three years.

Contents

List of Figures

List of Tables

List of Abbreviations

AR	Autoregressive
CFAR	Constant False Alarm Rate
CNR	Clutter to Noise power Ratio
CPI	Coherent Processing Interval
DOA	Direction of Arrival
e.m.	electromagnetic
ESM	Electronic Support Measures
GIP	Generalized Inner Product
GLRT	Generalized Likelihood Ratio Test
HRR	High Resolution Radar
iid	independent and identically distributed
LRT	Likelihood Ratio Test
LMI	Linear Matrix Inequality
m.f.	matched filter
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MGLRT	Modified GLRT
MTI	Moving Target Indicator
NLCD	National Land Cover Data
ORR	Output Response Ratio
pdf	probability density function
PSD	Power Spectral Density
PSMI	Pseudo Sample Matrix Inverse
RCS	Radar Cross Section
SDP	Semidefinite Programming
STAP	Space-Time Adaptive Processing
SINR	Signal to Interference plus Noise Ratio
UMP	Uniformly Most Powerful
w.f.	whitening filter

Notations

$\boldsymbol{a}$	column vector;
$\boldsymbol{A}$	matrix;
$(\cdot)^T$	transpose operator;
$(\cdot)^\dagger$	transpose conjugate operator;
$\mathrm{tr}\,(\cdot)$	trace of the square matrix argument;
$\det(\cdot)$	determinant of the square matrix argument;
$\lambda_{min}(\cdot)$	minimum eigenvalue of the square matrix argument;
$\lambda_{max}(\cdot)$	maximum eigenvalue of the square matrix argument;
$\boldsymbol{I}$	identity matrix;
$\mathbf{0}$	matrix with zero entries;
$\mathbf{diag}(\boldsymbol{a})$	diagonal matrix whose i-th diagonal element is the i-th entry of $\boldsymbol{a}$;
$\mathbb{R}^N$	set of N-dimensional vectors of real numbers;
$\mathbb{C}^N$	set of N-dimensional vectors of complex numbers;
$\mathbb{H}^N$	$N \times N$ Hermitian matrices;
$\mathbb{C}^{N,K}$	$N \times K$ matrices of complex numbers;
$\lvert \cdot \rvert$	modulus of a complex number;
$\lVert \cdot \rVert$	Euclidean norm of a complex vector or Frobenius norm of a complex matrix;
$\boldsymbol{\lambda}(\boldsymbol{A})$	$\boldsymbol{\lambda}(\boldsymbol{A}) = [\lambda_1(\boldsymbol{A}), \lambda_2(\boldsymbol{A}), \ldots, \lambda_N(\boldsymbol{A})]$, with $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \geq \ldots \geq \lambda_N(\boldsymbol{A})$ the vector containing the eigenvalues of $\boldsymbol{A} \in \mathbb{H}^N$, arranged in decreasing order;
$f(\boldsymbol{A})$	for any $\boldsymbol{A} = \boldsymbol{U}\mathbf{diag}(\boldsymbol{\lambda})\boldsymbol{U}^\dagger \in \mathbb{H}^N$, with $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N]$ the vectors of its eigenvalues, and $\boldsymbol{U}$ the unitary matrix containing the corresponding eigenvectors, it denotes the Hermitian matrix $f(\boldsymbol{A}) = \boldsymbol{U}\mathbf{diag}(f(\boldsymbol{\lambda}))\boldsymbol{U}^\dagger$, where $f(\boldsymbol{\lambda}) = [f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_N)]$;
j	imaginary unit (i.e., $j = \sqrt{-1}$).

$\succeq$	generalized inequality: $\boldsymbol{A} \succeq \boldsymbol{0}$ means that $\boldsymbol{A}$ is an Hermitian positive semi-definite matrix;
$\succ$	generalized inequality: $\boldsymbol{A} \succ \boldsymbol{0}$ means that $\boldsymbol{A}$ is an Hermitian positive definite matrix;
$\lceil \cdot \rceil$	smallest integer greater than or equal to the argument;
$\mathbb{E}[\cdot]$	statistical expectation;
$\mathrm{var}(\cdot)$	variance;
$v(\mathcal{P})$	optimal value of the optimization problem $\mathcal{P}$;
$\mathrm{vec}(\boldsymbol{A})$	given $\boldsymbol{A} \in \mathbb{H}^N$, $\mathrm{vec}(\boldsymbol{A}) = [\boldsymbol{A}_1^T, \boldsymbol{A}_2^T, \ldots, \boldsymbol{A}_N^T]^T \in \mathbb{C}^{N^2}$, where $\boldsymbol{A}_i$ is the i-th column of the matrix $\boldsymbol{A}$.

Introduction

A radar system is an electromagnetic (e.m.) device that transmits and receives radiations, and whose main purpose is the detection of possible targets within the illuminated scene. Once the surveilled area has been probed with a suitable signal, the radar elaborates the resulting signal, which is a superposition of several contributions from different objects. Specifically, the received signal is made up of both echoes from object of tactical importance (namely, the targets), as well as unwanted contributions due to land, sea, vegetation (called clutter), thermal noise and intentional disturbance signals (namely, jammers). For a typical radar system, the power of the useful component is a small percentage of the overall disturbance power (clutter, jamming and noise); consequently, the detection problem is quite difficult. To detect the target embedded in interference, a powerful solution is the Space-Time Adaptive Processing (STAP) [1], [2]. STAP refers to a processor that simultaneously combines the signals received on multiple elements of an antenna array (the spatial domain) and from multiple pulse repetition periods (the temporal domain) of a Coherent Processing Interval (CPI), [2]. This processing is of paramount importance because, even if a target, that is located at a specific angle and with a specific Doppler frequency, cannot be distinguished from the interference component in the direction domain, may be clearly identified in the Doppler domain and/or viceversa, [1]. In Fig. 1, a pictorial representation of the interference environment that justify this last sentence is reported. The figure shows the presence of noise jamming that is localized in angle and distributed over all Doppler frequencies, whereas the clutter echo from a single cell has a Doppler frequency that depends on its aspect with respect to the radar. Finally, a point-like target is present in the scene with a specific Doppler frequency and a specific angle. A space-time adaptive processor may be seen as a two-dimensional filter that represents combined receive

beamforming and target Doppler filtering [2].

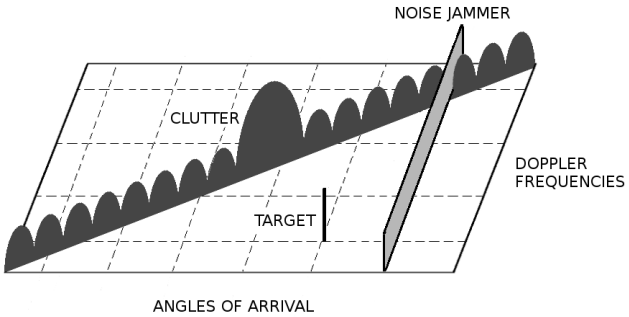


Figure 1: Example of a typical angle-Doppler interference scenario that justifies the use of a STAP.

There are various aspects which can be considered, when dealing with radar processing, designing radar waveforms and/or radar filters. When it comes to the capability of the radar to properly detect a target in the illuminated scene, as well as its capability to distinguish among the useful and the interference components within the received signal, the Signal to Interference plus Noise Ratio (SINR) is one of the most commonly used figure of merit. Consider now a radar comprised of an array of M antenna elements, transmitting a coherent train of L pulses, the processor that maximizes the output SINR is the coherent, linear, transversal filter [3] depicted in Fig. 2. Such a filter is com-

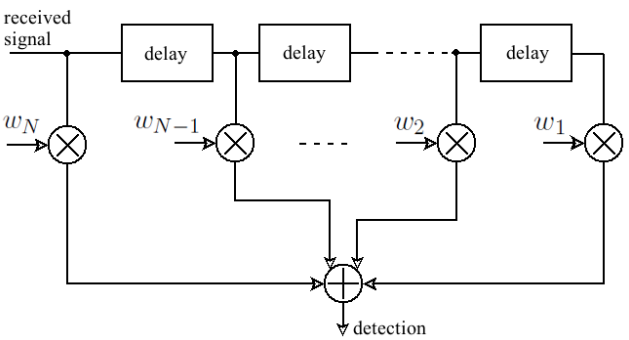


Figure 2: Optimum receiver.

pletely described by $N = LM$ complex weights $(w_1, w_2, \ldots, w_N)$, i.e. by the complex N-dimensional weight vector $\boldsymbol{w}$, that depends both on

the interference statistics as well as on the target signal model. Thus, denoting by

$$\boldsymbol{r} = \boldsymbol{p} + \boldsymbol{n} \tag{1}$$

the N-dimensional vector associated to the received signal of the cell under test, where $\boldsymbol{p}$ is the steering vector of the useful signal (assumed known) and $\boldsymbol{n}$ is the zero mean vector associated to the disturbance components, the optimum filter output is given by the inner product between the weight vector and the useful signal

$$\boldsymbol{s} = \boldsymbol{w}^{\dagger}\boldsymbol{r}. \tag{2}$$

Moreover, denoting by $\mathbb{E}[\boldsymbol{n}\boldsymbol{n}^{\dagger}] = \boldsymbol{\Sigma}$ the disturbance covariance matrix, the SINR at the output of the above mentioned filter is

$$\begin{aligned}
\mathrm{SINR}_{out} &= \frac{\mathbb{E}[|\boldsymbol{w}^{\dagger}\boldsymbol{p}|^2]}{\mathbb{E}[(\boldsymbol{w}^{\dagger}\boldsymbol{n})(\boldsymbol{w}^{\dagger}\boldsymbol{n})^{\dagger}]} = \frac{\left|\boldsymbol{w}^{\dagger}\boldsymbol{p}\right|^2}{\boldsymbol{w}^{\dagger}\boldsymbol{\Sigma}\boldsymbol{w}} \\
&= \frac{\left|\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)^{\dagger}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{p}\right)\right|^2}{\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)^{\dagger}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)} \\
&\leq \frac{\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)^{\dagger}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{p}\right)^{\dagger}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{p}\right)}{\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)^{\dagger}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}\right)} \\
&= \boldsymbol{p}^{\dagger}\boldsymbol{\Sigma}\boldsymbol{p}.
\end{aligned} \tag{3}$$

where the inequality in (3) is a consequence of the Schwarz inequality. The latter quantity attains its maximum when the equality holds, i.e. when the two terms in the product at the numerator are proportional, $\boldsymbol{\Sigma}^{1/2}\boldsymbol{w} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{p}$. Consequently, the optimum receiver is

$$\boxed{\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{p}.} \tag{4}$$

This filter can be seen as a whitening filter (w.f.) followed by a matched filter (m.f.)

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{p} = \overbrace{\left(\boldsymbol{\Sigma}^{-1/2}\right)}^{w.f.}\overbrace{\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{p}\right)}^{m.f.}, \tag{5}$$

consequently, the filter is tuned to the Doppler of the target (in a time processing) or to the angle of arrival (in a space processing) or to both of them in a space-time processing. Moreover, it properly exploits the information about the interference statistics (through the disturbance covariance matrix) in order to reduce the interference effects.

As shown in equation (4), the optimum filter to be applied on the received signal requires the exact knowledge of the true disturbance covariance matrix. However, in real radar systems this requirement cannot be satisfied and an estimate of the covariance matrix must be introduced, leading to the so-called adaptive radars [4], [5], [6]. Notice also that, accurate estimation of the disturbance covariance matrix is of paramount importance not only for adaptive receive weight vector computation [7], but also for several advanced radar signal processing algorithms, such as secondary data selection [8] and robust steering vector estimation [9].

Conventional adaptive radar receivers [4], [5], [6], are often based on the assumption that the environment remains stationary and homogeneous during the adaptation process. Precisely, they exploit an estimate of the disturbance covariance matrix resorting to a secondary data set collected from range gates spatially close to the one under test and sharing the same spectral properties [10], [11]. A classic estimate is the sample covariance matrix, which is the Maximum Likelihood (ML) estimator based on K independent and identically distributed (iid) N-dimensional zero-mean complex circular Gaussian vectors. The existence of the ML solution fails when the matrix dimension is greater than the sample support ($N > K$), whereas the sample covariance matrix achieves good performance when $K \geq 2N$ [4]. This homogeneity represents an important limitation since in real environments the number of data in which the clutter is homogeneous (often referred to as sample support) is very limited. Poor training data selection, in such adaptive detectors, can result in a remarkable degradation of the adaptive radar performance especially in regions which include varying ground surfaces such as coastal regions connecting land and sea, where the strength of the clutter may exhibit strong fluctuations. Some discussions of real-world effects and their impacts on the performance of Doppler processors and STAP detectors can be found respectively in [12] and [13]. A possible strategy to circumvent the lack of a sufficient number of homogeneous secondary data (required for achieving a satisfactory performance) is to exploit some a-priori information about the scene illuminated by the

radar, namely to perform a knowledge-based processing. Actually there are two fundamental ways to exploit the available a-priori knowledge ([14], [15], and references therein). The former is the indirect approach and uses knowledge sources to select the secondary data for the covariance estimation process [14], [15], [16]. The latter is the direct method and relies on the use of the a-priori knowledge directly in the receiver design process [14], [15], [17], [18], [19]. In both cases, it is of interest to devise procedures which exploit jointly the a-priori knowledge available about the operating environment and the training data in order to confer upon the estimator a robust adaptive behavior. The final goal is to obtain a reliable estimate of the covariance matrix, which must be well conditioned, since the computation of the weight vector, used in adaptive radar processing, involves the inverse of the estimated covariance.

As already claimed, in real scenarios the homogeneity assumptions could not hold, because secondary data may be contaminated by clutter discretes, outliers, and/or power variations. Consequently, a statistical characterization of the whole environment can be very difficult to obtain, and estimators, whose design do not rely on the multivariate probability distribution of the data, are of interest. Signal processing algorithms derived from geometric considerations on the space of the parameters to be estimated, and which do not account for the statistical characterization of the data, are available in the open literature. For instance, the least square estimator is the most natural choice [20, Ch. 8, p. 219]. In [21], an extension of the ordered statistic approach, to define a new STAP technique, based on the Riemannian p-mean computation of Toeplitz-Block-Toeplitz space-time covariance matrix is presented. Moreover, in [22], an algorithm for radar target detection is introduced, based on the Riemannian p-mean of covariance matrices computed in a neighborhood of the considered cell. For a detailed overview of this research activity see also [23] and references therein. Finally, the geometric approach is used also in other signal processing contexts; for instance in [24], the barycenter of a set of diffusion tensors is used in diffusion tensor imaging applications.

The adaptive receivers previously described refer to point-like targets, namely to targets that are contained within a single range cell. However, it is necessary also to account for radar receivers that operate in presence of targets extended in range. In fact, detection of distributed targets has gathered extensive attention among radar community during

the last three decades. This is motivated by the fact that when using High Resolution Radars (HRR's), targets can be resolved into a number of scattering centers appearing into different range cells [25, 26, 27]. Furthermore, in many practical scenarios, wherein a low/medium resolution radar is employed, the point-target model may fail: for instance, when a coastal radar is faced with detection of a large ship or when an air defense radar is detecting a cluster of point-targets flying at the same velocity in close spatial proximity. In Fig. 3, an example of a data collection from a target extended in range is given.

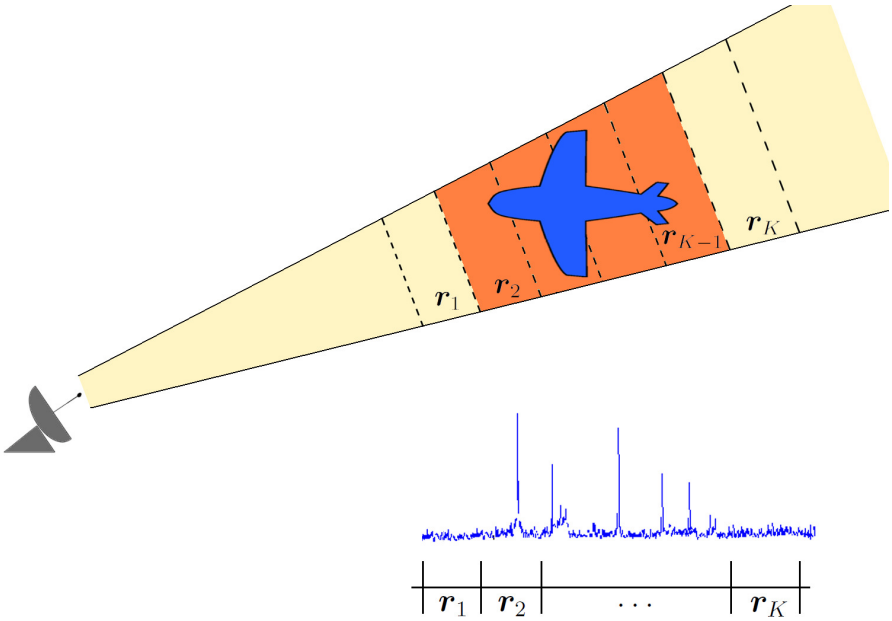


Figure 3: Data collection of a target extended in range.

Many papers have addressed detection and imaging with HRR's [28, 29, 30]. In particular, radar detection of distributed targets in white Gaussian noise of known spectral level has been considered in [31]. Therein, it is shown that properly designed HRR's allow for a significant enhancement of the detection performance, because increasing the range resolution of the radar reduces the amount of energy per cell backscattered by distributed clutter, and resolved scatterers introduce less fluctuation than an unresolved point-target.

CFAR (Constant False Alarm Rate) detection of distributed tar-

gets in Gaussian noise with unknown covariance matrix, based upon the Generalized Likelihood Ratio Test (GLRT) criterion, is addressed in [32, 33, 34, 35, 36, 37] and [38] assuming several (often different) models for the useful target echo. The disturbance returns from different range cells are modelled as independent, identically distributed, Gaussian vectors with unknown covariance matrix; moreover, a set of secondary data, free of useful signal components, is exploited to estimate the spectral properties of the disturbance. Some adaptive schemes for detecting extended targets, assuming the availability of a wide enough, although unknown, portion of secondary data free of the useful signal, are proposed in [39] and [40]. A Modified GLRT (MGLRT) which does not resort to secondary data is developed in [41]; it does not share the CFAR property but can be made bounded CFAR, thus being a viable technique to adaptively detect range-spread targets embedded in highly non-stationary environment. The modified GLRT approach is also applied in [42] to develop an adaptive algorithm with orthogonal rejection capabilities. A GLRT for the adaptive detection of Doppler-shifted, range-distributed targets embedded in noise with unknown, but structured, covariance matrix has been studied in [43]. Such a detector has been shown to be bounded CFAR via simulation. A heuristic, although effective, strategy for detecting range-spread targets in white Gaussian noise, using multiple consecutive high-resolution range profiles collected by a HHR, is proposed [44]. A generalized parametric Rao test is developed in [45] modeling the disturbance as a multi-channel auto-regressive process. By doing so the authors extend to distributed targets the interesting parametric approach developed in [46] for a point-like target.

All the above considerations highly justify the interest of the research herein conducted, whose main aim is to define new covariance matrix estimation techniques, based on both statistical argumentations and geometric considerations, exploiting advanced mathematics as for instance the convex optimization theory. These new estimates are utilized to define new adaptive radar receivers and to design new secondary data selection schemes, respectively. Moreover, applications to the problem of detecting extended targets have been considered, enforcing several structures to the disturbance covariance matrix to estimate.

The present thesis has been organized as follows:

- In Chapter 1, the problem of estimating the disturbance covariance matrix for radar signal processing applications, in the presence of

a limited number of training data, is addressed. In particular, the ML estimator of the covariance matrix is determined starting from a set of secondary data, assuming a special covariance structure (i.e. the sum of a positive semi-definite matrix plus a term proportional to the identity), and a condition number upper-bound constraint. The formulated constrained optimization problem falls within the class of MAXDET problems and an efficient procedure for its solution in closed form is developed. Remarkably, the computational complexity of the algorithm is of the same order as the eigenvalue decomposition of the sample covariance matrix.

- In Chapter 2, the problem of covariance matrix estimation for radar signal processing applications is assessed in the presence of heterogeneous secondary data. In particular, two classes of estimators, which do not require any knowledge about the probability distribution of the sample support and exploit the characteristics of the positive definite matrix space, are proposed and analyzed. Any estimator of each class is associated with a suitable distance in the considered space and is defined respectively as the geometric barycenter and the median matrix of some basic covariance matrix estimates obtained from the available secondary data set. Then, the new devised estimators are applied to the problem of secondary data selection.

- In Chapter 3, the problem of detecting an extended target embedded in homogeneous Gaussian interference with unknown but structured covariance matrix is addressed. The possible target echo, from each range bin under test, is modeled as a deterministic signal with an unknown scaling factor accounting for the target response. At the design level, some a-priori knowledge about the operating environment are exploited, enforcing the inverse interference plus noise covariance matrix to belong to a set described via unitary invariant continuous functions. Hence, the constrained ML estimates of the unknown parameters are derived, under both the H_0 and H_1 hypotheses, and the GLRT for the considered decision problem is designed.

Finally, some conclusions and hints for possible future research tracks are given.

Chapter 1

Structured Covariance Matrix Estimation with a Condition Number Constraint

In this chapter, the ML covariance matrix estimator which exploits the adaptivity provided by the training data, a special covariance structure, and a condition number upper-bound constraint (whose value can be obtained from some a-priori information, can be estimated from the available samples, or can be set according to numerical stability arguments) is devised. Specifically, the covariance matrix is modeled as the sum of two matrices, an unknown positive semi-definite matrix, describing colored interference and clutter, and a (partially known) matrix proportional to the identity one, accounting for the white disturbance term. Additionally, the estimated matrix has to comply with an upper bound on its condition number. Notice that, the ML covariance estimation with the only structural constraint has been considered in [47], while the ML estimation of a covariance matrix with only a condition number upper bound constraint and without any assumptions on its structure has been studied in [48]. Hence, the novelty of this work is to jointly account for both a structural and a condition number constraint at the design level. The core of this work is to show that the proposed constrained structured ML estimation problem can be formulated in terms of a MAXDET optimization problem [49], [11], and to design a proce-

dure for provide its analytically solution in closed form. Notice that, the proposed algorithm requires the computation of the eigenvalue decomposition of the sample covariance matrix and the solution of a scalar convex optimization problem, whose complexity is linear with respect to the number of sample eigenvalues greater than one.

At the analysis stage, the performance of the new estimator is assessed in terms of achievable SINR versus the number of available samples, both for a spatial and a Doppler processing. The results highlight that interesting SINR improvements, with respect to the estimators [47] and [48] can be achieved.

Thus, the present chapter is organized as follows. In Section 1.1, the system model is described and the main issues arising in a limited sample support covariance estimation problem are presented. In Section 1.2, the constrained structured ML estimation problem is formulated, showing that it is equivalent to a MAXDET convex optimization problem, and the procedure for its closed form solution is derived. In Section 1.3, the performance of the proposed ML estimate is assessed.

1.1 Problem Formulation

In this section, the problem of ML estimating the positive definite covariance matrix $\mathbf{\Sigma}$ is formulated. Specifically, it is considered the availability of K secondary data $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K$, modeled as N-dimensional independent zero-mean complex circular Gaussian vectors[1], which shares the same covariance matrix

$$\mathbb{E}\left[\boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}\right] = \mathbf{\Sigma}, \qquad i = 1, \ldots, K.$$

To this end, it is necessary to specify the joint probability density function (pdf) of $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K$, i.e.

$$f(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K | \mathbf{\Sigma}) = \frac{1}{\pi^{NK} \left[\det(\mathbf{\Sigma})\right]^K} \exp\left[-\mathrm{tr}\left(K \mathbf{\Sigma}^{-1} \boldsymbol{S}_1\right)\right], \qquad (1.1)$$

[1]The proposed framework assumes Gaussian disturbance. However, there are situations, such as sea clutter at low grazing angles, where the Gaussian assumption can be no longer met and the compound-Gaussian model proves very effective to model the radar returns. In this context, alternative covariance matrix estimation strategies such as those in [50] and [10] can be conceived.

where $\boldsymbol{S}_1 = \frac{1}{K}\sum_{i=1}^{K} \boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}$ is the sample covariance matrix. Notice that, owing to the invariance principle [20, Theorem 7.2, p. 176], the ML estimate of $\boldsymbol{\Sigma}$ ($\widehat{\boldsymbol{\Sigma}}$ in the following) can be obtained from the ML estimate of $\boldsymbol{X} = \boldsymbol{\Sigma}^{-1}$ through a matrix inversion. In particular, without any structure and constraints on the covariance matrix, i.e. without any a-priori knowledge, the ML estimate of $\boldsymbol{X}$ is an optimal solution to the optimization problem

$$\widehat{\boldsymbol{X}} = \arg \min_{\boldsymbol{X} \succ \boldsymbol{0}} \left[\mathrm{tr}\, (\boldsymbol{S}_1 \boldsymbol{X}) + \log \det \left(\boldsymbol{X}^{-1} \right) \right]. \tag{1.2}$$

If $K \geq N$, (1.2) admits a well known closed form solution $\widehat{\boldsymbol{X}} = \boldsymbol{S}_1^{-1}$, namely the inverse of the sample covariance matrix; otherwise, the minimizer does not exist. This estimate is usually exploited in many adaptive radar receivers [5, 6, 51, and references therein] and, in particular, for the adaptive implementation of the optimum Doppler, spatial, and STAP processors [2, 4]. The expected SINR loss, relative to the ideal known covariance case, is kept within 3 dB if the sample support K is greater than $2N$. Unfortunately, in practical radar scenarios, such an assumption is not always verified [13]. More specifically, the size of the training set is often limited, because large swaths of homogeneous clutter/interference necessary for estimating $\boldsymbol{\Sigma}$ may not be available. Moreover, the presence of the target within the secondary data could reduce the degree of their homogeneity. In addition, the analysis of several adaptive algorithms, mostly derived assuming homogeneity of the secondary data, has shown that non-homogeneities magnify the loss between the adaptive implementation and optimum conditions [52, 53].

To reduce the sample support requirement, several solutions have been proposed in open literature:

1. to exploit structural information about $\boldsymbol{\Sigma}$, as for instance persymmetry [54], Toeplitz property [55], [56], [57], circulant structure [58], multichannel autoregressive models [59], [60], special structures imposed by the sensor and the environment [47], physical constraints [61];

2. to resort to Bayesian covariance matrix estimators [62, 63, 64, 65, 66];

3. to use knowledge-based covariance models [67, 68];

4. to consider shrinkage estimation methods [69, 70, 19, 48].

The idea followed here is to devise a covariance matrix estimator which exploits both the adaptivity provided by the training data (even if very limited) and some a-priori structural information. Namely, a covariance matrix estimator which accounts for a special covariance structure and a condition number upper-bound constraint is designed. Regarding the structure, the covariance matrix $\mathbf{\Sigma}$ is modeled as the sum of an unknown positive semidefinite matrix, describing colored interference and clutter contributes, and a matrix proportional to the identity, accounting for the white disturbance term. Furthermore, as to the condition number constraint, its upper bound value can be obtained (see Subsection 1.2.1) from some a-priori information available at the radar platform about the electromagnetic environment, in an adaptive fashion resorting to the samples $\boldsymbol{r}_i$, $i = 1, \ldots, K$, or enforcing a specific value in order to control the numerical stability. In fact, signal processors work with finite precision arithmetic and it is extremely important to account for the numerical stability of algorithms exploiting the estimated covariance matrix, or its inverse (for instance, adaptive receive weight vector computation [7], robust steering vector estimation [9], robust beamforming [71], Direction of Arrival (DOA) estimation, Autoregressive (AR) coefficient estimation [72]). Indeed, the effect of the estimated covariance roundoff error is controlled by the covariance condition number, in the sense that stable algorithms can be obtained if the estimated matrix is well conditioned with respect to the machine precision [73], [7]. Thus, through the proposed estimator, the idea is to exploit not only the structural information on the covariance matrix, but also to force an upper bound to the condition number compliant with the desired digital stability.

1.2 Derivation of the Constrained Structured Estimator

Starting from the secondary data $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K$, the problem of finding the ML estimate of the matrix $\mathbf{\Sigma}$ is considered under the following

constraints

$$\begin{aligned}&\boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I} + \boldsymbol{R},\\ &\boldsymbol{R} \succeq \mathbf{0},\\ &\sigma_n^2 \geq \sigma^2,\\ &\frac{\lambda_{max}(\boldsymbol{\Sigma})}{\lambda_{min}(\boldsymbol{\Sigma})} \leq K_{max},\end{aligned}$$

where $\boldsymbol{R}$ accounts for colored interference and clutter, whereas σ_n^2 for the power of the white disturbance term; the parameters $\sigma^2 > 0$ and K_{max} are respectively the lower bound on the white disturbance power and the upper bound on the condition number. Otherwise stated, the constrained structured covariance matrix estimator is an optimal solution to the optimization problem

$$\mathcal{P}\begin{cases}\underset{\boldsymbol{\Sigma},\boldsymbol{R},\sigma_n^2}{\min} & \operatorname{tr}\left(\boldsymbol{S}_1\boldsymbol{\Sigma}^{-1}\right) - \log\det\left(\boldsymbol{\Sigma}^{-1}\right)\\ \text{s.t.} & \frac{\lambda_{max}(\boldsymbol{\Sigma})}{\lambda_{min}(\boldsymbol{\Sigma})} \leq K_{max}\\ & \sigma_n^2\boldsymbol{I} + \boldsymbol{R} = \boldsymbol{\Sigma}\\ & \boldsymbol{R} \succeq \mathbf{0}\\ & \sigma_n^2 \geq \sigma^2\end{cases}, \tag{1.3}$$

where $\boldsymbol{\Sigma} \succ \mathbf{0}$, $\boldsymbol{R} \succeq \mathbf{0}$, and $\sigma_n^2 \in \mathbb{R}^+$ are the optimization variables. In problem $\mathcal{P}$, given in (1.3), it is assumed that $K_{max} > 1$; in fact, for $K_{max} < 1$ the problem is infeasible, while for $K_{max} = 1$ the ML estimate is trivially given by $\boldsymbol{\Sigma} = \max\left(\sigma^2, \operatorname{tr}\left(\boldsymbol{S}_1\right)/N\right)\boldsymbol{I}$. Problem $\mathcal{P}$ is a non-convex optimization problem since the objective function is a non-convex function of $\boldsymbol{\Sigma}$. However, problem $\mathcal{P}$ admits an optimal solution since the following proposition holds true

Proposition 1.2.1. *To find an optimal solution to $\mathcal{P}$, it is sufficient to solve $\mathcal{P}_1$*

$$\mathcal{P}_1\begin{cases}\underset{\boldsymbol{X},u}{\min} & \operatorname{tr}\left(\boldsymbol{S}\boldsymbol{X}\right) - \log\det\left(\boldsymbol{X}\right)\\ \textit{s.t.} & u\boldsymbol{I} \preceq \boldsymbol{X} \preceq uK_{max}\boldsymbol{I}\\ & \boldsymbol{X} \preceq \boldsymbol{I}\\ & 0 < u \leq 1\end{cases}, \tag{1.4}$$

where $\boldsymbol{S} = \boldsymbol{S}_1/\sigma^2$. Precisely, given an optimal solution $(\boldsymbol{X}^\star, u^\star)$ to the solvable problem[2] $\mathcal{P}_1$,

$$\left(\sigma^2\boldsymbol{X}^{\star-1}, \sigma^2\boldsymbol{X}^{\star-1} - \sigma^2\boldsymbol{I}, \sigma^2\right)$$

[2]By "Solvable", it is meant that the problem is feasible, bounded below, and the optimal value is attained, see [74].

is an optimal solution to $\mathcal{P}$.

Proof. See Appendix A. □

Problem $\mathcal{P}_1$ is a convex optimization problem, since the objective function is a convex function and the constraints are Linear Matrix Inequalities (LMIs). Precisely, it is a MAXDET problem, see [49]. Then, $\mathcal{P}_1$ can be efficiently solved numerically by interior-point methods. Notice that the objective function of problem $\mathcal{P}_1$ is a strictly convex function, [75, Theorem 7.6.7, p. 466], in terms of the variable $\boldsymbol{X}$. This implies that the minimizer $\boldsymbol{X}^\star$ of $\mathcal{P}_1$ is unique.
The core of this work relies on the design of a procedure for the closed form solution of $\mathcal{P}_1$, and hence of $\mathcal{P}$. To this end, denote by

$$\boldsymbol{S} = \boldsymbol{V}\,\mathbf{diag}\,(\boldsymbol{d})\,\boldsymbol{V}^\dagger$$

the eigenvalue decomposition of $\boldsymbol{S}$, where $\boldsymbol{V}$ is the unitary matrix containing the eigenvectors, and $\boldsymbol{d} \in \mathbb{R}^N$ is the vector of the corresponding eigenvalues arranged in decreasing order, i.e. $d_1 \geq d_2 \geq \ldots \geq d_N \geq 0$. The following result concerning problem $\mathcal{P}_1$ holds true

Lemma 1.2.2. *For any fixed* $\bar{u} \in]0, 1]$*, the optimal solution* $\boldsymbol{X}^\star(\bar{u})$ *to problem*

$$\mathcal{P}_1(\bar{u}) \begin{cases} \min\limits_{\boldsymbol{X}} & \text{tr}\,(\boldsymbol{S}\boldsymbol{X}) - \log\det(\boldsymbol{X}) \\ s.t. & \bar{u}\boldsymbol{I} \preceq \boldsymbol{X} \preceq \bar{u}K_{max}\boldsymbol{I} \\ & \boldsymbol{X} \preceq \boldsymbol{I} \end{cases}, \tag{1.5}$$

is given by

$$\boldsymbol{X}^\star(\bar{u}) = \boldsymbol{V}\,\mathbf{diag}\,(\boldsymbol{\lambda}^\star(\bar{u}))\,\boldsymbol{V}^\dagger, \tag{1.6}$$

where $\boldsymbol{\lambda}^\star(\bar{u}) = [\lambda_1^\star(\bar{u}), \ldots, \lambda_N^\star(\bar{u})]$*, with*

$$\lambda_i^\star(\bar{u}) = \min\left(\min\left(K_{max}\bar{u}, 1\right), \max\left(\bar{u}, \frac{1}{d_i}\right)\right), \quad i = 1, \ldots, N. \tag{1.7}$$

Proof. See Appendix B. □

Let us now define the vector function

$$\begin{aligned} \boldsymbol{\lambda}^\star(u) = \Bigg[&\min\left(\min\left(K_{max}u, 1\right), \max\left(u, \frac{1}{d_1}\right)\right), \ldots, \\ &\min\left(\min\left(K_{max}u, 1\right), \max\left(u, \frac{1}{d_N}\right)\right)\Bigg], \end{aligned} \tag{1.8}$$

which assigns to any $u \in]0,1]$ the vector of the optimal eigenvalues to problem $\mathcal{P}_1(u)$ given in (1.5).

Theorem 1.2.3. *Let $u^\star$ be an optimal solution to the following optimization problem*

$$\mathcal{P}_2 \left\{ \begin{array}{ll} \min\limits_{u} & \sum_{i=1}^{N} G_i(u) \\ s.t. & 0 < u \leq 1 \end{array} \right. , \tag{1.9}$$

where, for any $i = 1, \ldots, N$, $G_i(u) = d_i \lambda_i^\star(u) - \log \lambda_i^\star(u)$, namely

$$G_i(u) = \left\{ \begin{array}{ll} -\log K_{max} - \log u + K_{max} d_i u & \text{if} \quad 0 < u \leq \frac{1}{K_{max}} \\ d_i & \text{if} \quad \frac{1}{K_{max}} \leq u \leq 1 \end{array} \right. \tag{1.10}$$

if $d_i \leq 1$, and

$$G_i(u) = \left\{ \begin{array}{ll} -\log K_{max} - \log u + K_{max} d_i u & \text{if} \quad 0 < u \leq \frac{1}{K_{max} d_i} \\ \log d_i + 1 & \text{if} \quad \frac{1}{K_{max} d_i} \leq u \leq \frac{1}{d_i} \\ -\log u + d_i u & \text{if} \quad \frac{1}{d_i} \leq u \leq 1 \end{array} \right. \tag{1.11}$$

if $d_i > 1$. Then, an optimal solution to $\mathcal{P}_1$ is

$$(\boldsymbol{X}^\star, u^\star) = \left(\boldsymbol{V} \mathbf{diag}\left(\boldsymbol{\lambda}^\star\right) \boldsymbol{V}^\dagger, u^\star \right), \tag{1.12}$$

where $\boldsymbol{\lambda}^\star = [\lambda_1^\star, \ldots, \lambda_N^\star] = \boldsymbol{\lambda}^\star(u^\star)$, with the vector function $\boldsymbol{\lambda}^\star(u)$ defined in (1.8).

Proof. See Appendix C. □

Notice that the formulation of Theorem 1.2.3 holds even when $d_i = 0$, interpreting $\frac{1}{d_i} = +\infty$. Therefore, resorting to Theorem 1.2.3, problem $\mathcal{P}_1$ reduces, essentially, to the univariate minimization problem $\mathcal{P}_2$. Let us, now, study the properties of the optimization problem $\mathcal{P}_2$, precisely of its objective function

$$G(u) = \sum_{i=1}^{N} G_i(u), \tag{1.13}$$

with $G_i(u)$ defined in (1.10) or in (1.11), depending on the value of the corresponding d_i. Firstly, the function $G(u)$ is a continuous function over the interval $u \in]0,1]$, since it is the sum of continuous functions. Secondly, although the constraint $u \in]0,1]$ of problem $\mathcal{P}_2$ defines an open set, $\mathcal{P}_2$ is solvable as proved in the following theorem

Theorem 1.2.4. *Let $d_1 \geq d_2 \geq \ldots \geq d_N$ the eigenvalues of $\boldsymbol{S}$. The optimal value $v(\mathcal{P}_2)$ is attainable and*

- *if $d_1 \leq 1$, an optimal solution to $\mathcal{P}_2$ is $u^\star = \frac{1}{K_{max}}$;*
- *if $1 < d_1 \leq K_{max}$, an optimal solution to $\mathcal{P}_2$ is $u^\star = \frac{1}{d_1}$;*
- *if $d_1 > K_{max}$, an optimal solution to $\mathcal{P}_2$ complies with $u^\star \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$.*

Proof. See Appendix D. □

From theorem 1.2.4, to completely solve $\mathcal{P}_2$, it is needed to analyze the case $d_1 > K_{max}$. Hence, it has to be proven

Lemma 1.2.5. *Let $d_1 > K_{max}$. The function $G(u)$ has a continuous derivative over the interval $u \in \left]0, \frac{1}{K_{max}}\right]$. Moreover, $G(u)$ is a univariate convex function in the interval $u \in \left]0, \frac{1}{K_{max}}\right]$.*

Proof. See Appendix E. □

Let us further investigate the characteristics of the univariate convex function $G(u)$ when $d_1 > K_{max}$; the goal is to exploit its structure in order to develop an explicit procedure to solve $\mathcal{P}_2$. To this end, let us define some auxiliary quantities. Denote by $\bar{N}$, the number of d_i's greater than 1, i.e. $d_i > 1$, $i = 1, \ldots, \bar{N}$. The vector $\boldsymbol{v}$

$$\boldsymbol{v} = [d_1, d_2, \ldots, d_{\bar{N}}, 1], \tag{1.14}$$

contains the eigenvalues greater than 1, and its last entry is equal to 1; also, $\boldsymbol{v} = [v_1, v_2, \ldots, v_{\bar{N}}, v_{\bar{N}+1}]$ contains its entries in non-increasing order. Thus, the following theorem is proved

Theorem 1.2.6. *Assuming $d_1 > K_{max}$, an optimal solution $u^\star$ to $\mathcal{P}_2$ is given by*

1) $u^\star = \frac{1}{d_1}$, *if* $\left.\frac{dG(u)}{du}\right|_{u=\frac{1}{d_1}} = 0$;

2) $u^\star = \frac{1}{K_{max}}$, *if* $\left.\frac{dG(u)}{du}\right|_{u=\frac{1}{K_{max}}} \leq 0$;

3) if 1) *and* 2) *are not satisfied,* $u^\star$ *is the optimal point (with probability one*[3]*) if and only if*

$$u^\star = \frac{N+\alpha-\beta+1}{\sum_{i=1}^{\alpha} d_i + \sum_{i=\beta}^{N} K_{max} d_i} \tag{1.15}$$

with $\alpha \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ *the largest index such that* $\frac{1}{v_\alpha} < u^\star$*, and* $\beta \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ *the smallest index such that* $\frac{1}{v_\beta K_{max}} > u^\star$*.*

Proof. See Appendix F. □

From Theorem 1.2.6, if conditions 1) and 2) are not satisfied, the search for the optimal solution $u^\star$ of $\mathcal{P}_2$, requires finding the integers α and β such that

$$u_{\alpha,\beta} = \frac{N+\alpha-\beta+1}{\sum_{i=1}^{\alpha} d_i + \sum_{i=\beta}^{N} K_{max} d_i}, \tag{1.16}$$

$$\frac{1}{v_\alpha} < u_{\alpha,\beta} \leq \frac{1}{v_{\alpha+1}}, \qquad \text{and} \qquad \frac{1}{K_{max} v_{\beta-1}} \leq u_{\alpha,\beta} < \frac{1}{K_{max} v_\beta}. \tag{1.17}$$

An efficient procedure to find the optimal point $u^\star$, exploiting conditions (1.16) and (1.17), based on the idea in [48], is now described. The trick is to check iteratively the satisfaction of the conditions (1.16) and (1.17), once efficiently fixed the values of α and β. To this end, notice that, if the intersection of intervals (1.17) is empty, $u_{\alpha,\beta}$ cannot be an optimal value. Otherwise, the intersection is one of the following subintervals $\left]\frac{1}{v_\alpha}, \frac{1}{K_{max} v_\beta}\right[, \left]\frac{1}{v_\alpha}, \frac{1}{v_{\alpha+1}}\right], \left[\frac{1}{K_{max} v_{\beta-1}}, \frac{1}{K_{max} v_\beta}\right[, \left[\frac{1}{K_{max} v_{\beta-1}}, \frac{1}{v_{\alpha+1}}\right]$, and the optimal value needs to belong to that intersection. The procedure is composed by the following steps

1. Set $\alpha = 1$, $\beta = 2$ and increase β until $\frac{1}{v_\alpha} > \frac{1}{K_{max} v_\beta}$.
2. Compute $u_{\alpha,\beta}$. If $u_{\alpha,\beta}$ belongs to the current intersection, let $u^\star = u_{\alpha,\beta}$ and exit. Otherwise go to step 3).
3. if $\frac{1}{v_{\alpha+1}} > \frac{1}{K_{max} v_\beta}$ increase β and go to step 2). Otherwise increase α and go to step 2).

[3] It is assumed that $v_i \neq v_j$ and $v_i \neq K_{max} v_j$ for any $i \neq j$ with $1 \leq i, j \leq \bar{N}+1$, which hold with probability one since the probability density function of $[d_1, d_2, \ldots, d_{\max\{K,N\}}]$ is an absolute continuous function.

The above technique is capable to grant an optimal solution to problem $\mathcal{P}_2$ with a linear computational complexity, with respect to the number of sample covariance matrix eigenvalues greater than 1, i.e. N in the worst case.
In Fig. 1.1, a schematic representation of the procedure for the computation of $\widehat{\boldsymbol{\Sigma}}$ is given. Notice that $\widehat{\boldsymbol{\Sigma}}$ is a shrinkage estimator[4] which regularizes the sample covariance matrix explicitly accounting for a condition number constraint and leading to a well-conditioned structured estimator.

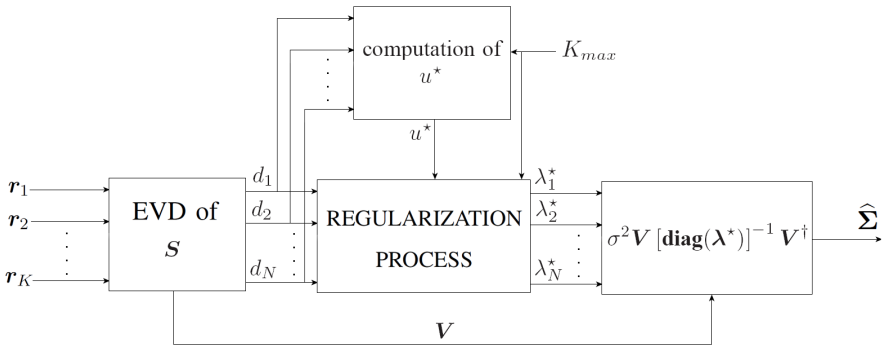


Figure 1.1: Schematic representation of the estimation procedure.

1.2.1 Selection of the Parameters K_{max} and σ^2

In this section, the selection of the parameters K_{max} and σ^2 in (1.3) is addressed. The focus is on radar applications, wherein the main disturbance contributions are due to thermal noise, jammers, and clutter [7, 2]. Thus, σ^2, which corresponds to an a-priori known lower bound on the power of the white disturbance term, can be evaluated characterizing the power level related to the isolated operation of the receiver's components. Moreover, as to the computation of the parameter K_{max}, there are three main approaches that can be considered: the knowledge-based, the numerical-stability oriented, and the adaptive method. In

[4] A shrinkage covariance estimator $\widehat{\boldsymbol{\Sigma}}$ is a matrix sharing the same eigenvectors as the sample covariance matrix $\boldsymbol{S} = \boldsymbol{V}\mathbf{diag}(\boldsymbol{d})\boldsymbol{V}^\dagger \succeq \mathbf{0}$, but transforming its eigenvalues, i.e. $\widehat{\boldsymbol{\Sigma}}=\boldsymbol{V}\mathbf{diag}\left([g_1(d_1,d_2,\ldots,d_N), g_2(d_1,d_2,\ldots,d_N),\ldots,g_N(d_1,d_2,\ldots,d_N)]\right)\boldsymbol{V}^\dagger \succeq \mathbf{0}$.

the following subsections, all of them are discussed in detail.

Knowledge-Based Selection of K_{max}

A knowledge-based selection of K_{max} resorts to a-priori information available at the radar platform about the electromagnetic environment. Precisely, exploiting Electronic Support Measures (ESM), a rough prediction of the jammer attributes can be obtained (such as their location, bandwidth, and power). Furthermore, as to the clutter contribution, it can be predicted through the interaction between a digital terrain map, such as the National Land Cover Data (NLCD), and Radar Cross Section (RCS) clutter models, see [76], [16], [77, Ch. 15, 16]. Starting from this information and some a-priori knowledge on the power of the white disturbance term, a rough estimate of the condition number of the covariance matrix can be performed.

Numerical-Stability Oriented Selection of K_{max}

An important task, in digital processing design, is the numerical stability of the outputs from the implemented algorithms, with respect to the accuracy of the input data. Thus, it is extremely relevant to guarantee a stable computation with respect to the roundoff errors corrupting the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$. It is worth pointing out that there is a fundamental tradeoff between the number of bits available in the computer to accomplish matrix inversion and the allowable eigenvalue spread (ruled by the condition number) of the input covariance [78, pp. 312-313], [7, p. 132]. In this context, a suitable choice of K_{max} allows for a control on the algorithm stability. For instance, the adaptive receive weight vector $\boldsymbol{w}$ is given by the solution of the following linear system $\frac{\widehat{\boldsymbol{\Sigma}}}{\sigma^2}\boldsymbol{w} = \boldsymbol{p}$, where $\boldsymbol{p}$ is the steering vector. Consequently, due to a perturbation $\boldsymbol{E}$ of the matrix $\frac{\widehat{\boldsymbol{\Sigma}}}{\sigma^2}$, [7], the computed weight vector is the solution to $\left(\frac{\widehat{\boldsymbol{\Sigma}}}{\sigma^2} + \boldsymbol{E}\right)\boldsymbol{w}_p = \boldsymbol{p}$, where $|\boldsymbol{E}(h,k)| \leq \epsilon'$ with ϵ' the machine precision. Thus, from [79], the sensitivity of the weight vector to the machine precision is upper bounded[5] by $\frac{||\boldsymbol{w}-\boldsymbol{w}_p||}{||\boldsymbol{w}||} \leq \epsilon' N K_{max}$, i.e. it can be controlled through an appropriate selection of K_{max}.

[5] A similar result holds true even if $\boldsymbol{p}$ is perturbed.

Adaptive Selection of K_{max}

In this subsection, an adaptive estimator of K_{max}, based on the K secondary data $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K$, is presented. The trainer principle, exploited by the predictor, is to extract information starting from the diagonal blocks of $\boldsymbol{S}$; each block is the sample covariance matrix of the corresponding sub-vector extracted by $\frac{\boldsymbol{r}_1}{\sigma}, \ldots, \frac{\boldsymbol{r}_K}{\sigma}$. Indeed, due to the available sample support K, the estimation of the sub-covariance matrix from sub-vectors extracted by $\frac{\boldsymbol{r}_1}{\sigma}, \ldots, \frac{\boldsymbol{r}_K}{\sigma}$ might be performed more reliably than the estimation of the entire covariance matrix. Moreover, given the stationarity property (ensured by the use of either a uniform linear array or a regularly spaced pulse train) of the random vectors $\frac{\boldsymbol{r}_i}{\sigma},\ i = 1, \ldots, K$, the information provided by different diagonal blocks of $\boldsymbol{S}$ of the same size can be combined to produce an estimate.

As to the dimension of the sub-blocks, which are meaningfully to analyze, its value is related to the dimension of the subspace in which the disturbance concentrates most of its power and depends on the specific radar application. Namely, for a spatial processing, through the analysis of sub-matrices of dimension smaller than or equal to the number J of jammers, it is difficult to acquire reliable information about the condition number of the matrix $\frac{\boldsymbol{\Sigma}}{\sigma^2}$, since all the directions are almost surely completely affected by the interference power. Consequently, the knowledge of J is assumed, whose value can be obtained adaptively resorting to the ESM of the radar platform, and start to analyze sub-blocks of dimension greater than or equal to $J+1$. It is also assumed that $J < N$, which is a reasonable assumption in the radar context.

In a Doppler processing, instead, the size of the sub-block to be processed can be evaluated through an analysis of the estimated power distribution in the disturbance signal space, for instance, as the number of eigenvalues of $\boldsymbol{S}$ which corresponds to the 98% of the whole disturbance power associated to non-zero sample eigenvalues. As to the notation adopted to describe the computation of K_{max}, with $\boldsymbol{S}_i^L$ is indicated the $i-$th sub-matrix of dimension L extracted by $\boldsymbol{S}$, namely

$$\boldsymbol{S}_i^L(h,k) = \boldsymbol{S}(i+h, i+k), \quad (h,k) \in \{1, \ldots, L\}^2.$$

Based on the aforementioned guidelines, the proposed predictor $K_{max}^{(*)}$ is now described.

- $K \leq \max(J, 1)$,

$$K_{max}^{(*)} = \max(1, d(1)), \tag{1.18}$$

i.e. it is equal to its ML estimate.

- $\max(J,1) < K < 2N$,

$$K_{max}^{(*)} = \frac{\lambda_{max}^{\diamond}\left(\bar{\boldsymbol{S}}^{(c)}\right)}{\lambda_{min}^{\diamond}\left(\bar{\boldsymbol{S}}^{(c)}\right)}, \quad \text{where} \quad \bar{\boldsymbol{S}}^{(c)} = \frac{1}{N-c+1}\sum_{i=1}^{N-c+1}\boldsymbol{S}_i^{(c)}, \quad (1.19)$$

with $\lambda_{min}^{\diamond}(\cdot) = \max(1, \lambda_{min}(\cdot))$, $\lambda_{max}^{\diamond}(\cdot) = \max(1, \lambda_{max}(\cdot))$ and $c = \max\left(\min\left(\lceil K/2 \rceil, \lceil N/2 \rceil\right), J+1\right)$.

- $K \geq 2N$

$$K_{max}^{(*)} = \alpha\frac{\lambda_{max}^{\diamond}\left(\bar{\boldsymbol{S}}^{(c)}\right)}{\lambda_{min}^{\diamond}\left(\bar{\boldsymbol{S}}^{(c)}\right)} + (1-\alpha)\frac{\lambda_{max}^{\diamond}\left(\boldsymbol{S}\right)}{\lambda_{min}^{\diamond}\left(\boldsymbol{S}\right)}, \quad (1.20)$$

where $\bar{\boldsymbol{S}}^{(c)} = \frac{1}{N-c+1}\sum_{i=1}^{N-c+1}\boldsymbol{S}_i^{(c)}$, $c = \max\left(\lceil N/2 \rceil, J+1\right)$, $\alpha = (-1/(2N))(K-4N)u(-K+4N)$, and u the Heaviside step function.

An important remark is now given: when $K_{max} \geq K_{max}^{\star} = \frac{\lambda_{max}^{\diamond}(\boldsymbol{S})}{\lambda_{min}^{\diamond}(\boldsymbol{S})}$, then the FML estimate [47], is attainable for problem $\mathcal{P}$, given in (1.3), and the proposed constrained structured estimator coincides with the FML one. The last condition holds true for $K \geq 4N$.

1.3 Performance Analysis

In this section, numerical results on the performance of the proposed constrained structured ML estimate, in terms of normalized average SINR, are presented. The following two scenarios are considered:

- spatial processing in the presence of jamming and white interference;
- Doppler processing in the presence of bimodal clutter plus white noise.

1.3.1 Spatial Processing in the Presence of Jamming and White Interference

The considered radar system is equipped with a uniform linear array of $N = 20$ elements, with a spacing between the antennas equal to

$d = \lambda_0/2$, where λ_0 is the radar operating wavelength, that points in the boresight direction. The overall disturbance is composed of jammers and white interference. Hence, a structured covariance [47] is assumed, that can be expressed as $\boldsymbol{\Sigma} = \boldsymbol{R} + \sigma_a^2 \boldsymbol{I}$, where σ_a^2 is the actual power level of the white disturbance term, whereas $\boldsymbol{R}$ is the covariance matrix associated to J (narrowband or wideband) jammers, defined by

$$\boldsymbol{R}(n,m) = \sum_{i=1}^{J} \sigma_i^2 \mathrm{sinc}\left[0.5 B_f (n-m)\phi_i\right] e^{j(n-m)\phi_i} \quad (n,m) \in \{1,\ldots,N\} \tag{1.21}$$

where $B_f = B/f_0$ is the fractional bandwidth, B is the instantaneous bandwidth of the desired signal (coinciding with the jammer's bandwidth), σ_i^2 is the power associated with the i-th jammer, and ϕ_i is the jammer phase angle with respect to the antenna phase center. Precisely, $\phi = 2\pi d(\sin\theta)/\lambda_0$, where θ is the angle off-boresight of the jammer.
To assess the performance of the proposed estimator, the normalized average SINR[6] is considered as figure of merit, which is defined as

$$\mathrm{SINR}_{av} = \frac{1}{MC} \sum_{i=1}^{MC} \frac{|\boldsymbol{w}_i^\dagger \boldsymbol{p}|^2}{\left(\boldsymbol{w}_i^\dagger \boldsymbol{\Sigma} \boldsymbol{w}_i\right) \mathrm{SINR}_{opt}} \tag{1.24}$$

where SINR_{opt} is the optimal value of the SINR given by $\mathrm{SINR}_{opt} = \boldsymbol{p}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{p}$, achieved by the optimal weight vector $\boldsymbol{\Sigma}^{-1}\boldsymbol{p}$, corresponding

[6]The statistical expectation of the normalized SINR is computed resorting to the arithmetic mean of a number MC of Monte Carlo trials. As to the Monte Carlo simulation, a bound on the variance of the normalized average SINR is

$$\begin{aligned} \mathrm{var}(\mathrm{SINR}_{av}) &= \mathbb{E}[(\mathrm{SINR}_{av} - \mathbb{E}[\mathrm{SINR}_{av}])^2] = \mathrm{var}\left[\frac{\sum_{i=1}^{MC} \mathrm{SINR}_i}{MC}\right] \\ &= \frac{1}{(MC)^2} \sum_{i=1}^{MC} \mathrm{var}(\mathrm{SINR}_i) = \frac{1}{MC} \mathrm{var}(\mathrm{SINR}_i) \leq \frac{1}{MC}, \end{aligned} \tag{1.22}$$

where SINR_i is the SINR of the i-th Monte Carlo trial. The last inequality stems from the fact that SINR is less than 1. Then, the standard deviation of the estimation error is upper bounded by

$$\sqrt{\mathrm{var}(\mathrm{SINR}_{av})} \leq \frac{1}{\sqrt{MC}}. \tag{1.23}$$

Thus, using equation (1.23), the number of Monte Carlo trials can be set in order to guarantee the required accuracy.

to the steering vector $\boldsymbol{p} = (1, 1, \ldots, 1)^T$. The adaptive estimate of the weight vector is $\boldsymbol{w}_i = \widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{p}$, where $\widehat{\boldsymbol{\Sigma}}_i$ is the data-dependent estimate of $\boldsymbol{\Sigma}$ at the i-th trial. Finally, MC is the number of Monte Carlo trials[7], used to estimate SINR_{av}.

In the following analysis, the focus is on 3 different scenarios. The first accounts for only 1 jammer, whose power is $\sigma_1^2 = 30$ dB, whereas its phase is $\phi_1 = 20$ deg. The second includes $J = 3$ jammers. They share the same power $\sigma_i^2 = 30$ dB, $i = (1, 2, 3)$, and phases $\phi_1 = 20$ deg, $\phi_2 = 40$ deg, and $\phi_3 = 60$ deg (i.e. they are concerning to 3 different angles of arrival). In these two situations, it is considered first the narrowband ($B_f = 0$) and then the wideband ($B_f = 0.3$) environment. Finally, in the third simulation, it has been accounted for $J = 3$ jammers, with different powers, $\sigma_1^2 = 10$ dB, $\sigma_2^2 = 20$ dB, and $\sigma_3^2 = 30$ dB, phases, $\phi_1 = 20$ deg, $\phi_2 = 40$ deg, and $\phi_3 = 60$ deg, and fractional bandwidths, $B_{f_1} = 0.2$, $B_{f_2} = 0$, $B_{f_3} = 0.3$. For all these cases, the considered simulation setting assumes three different values for the actual power level of the white interference, $\sigma_a^2 = 0$, 5, and 10 dB, respectively, whereas the nominal lower bound to the aforementioned power level is $\sigma^2 = 0$ dB.

As to K_{max}, both the a-priori knowledge of the true condition number, i.e. $K_{max} = \lambda_{max}(\boldsymbol{\Sigma})/\lambda_{min}(\boldsymbol{\Sigma})$, and the predictor $K_{max}^{(*)}$, proposed in Subsection 1.2.1, have been considered.

In Fig. 1.2, the normalized average SINR is plotted versus the number of independent snapshots (secondary data), for the proposed algorithm, the one with only a condition number constraint [48], and the FML [47]. The sub-plots refer to the different cases analyzed in the first scenario. Moreover, Fig. 1.3 refers to the second scenario, whereas Fig. 1.4 refers to the last. The curves highlight that the constrained structured estimator can achieve, for the considered values of the parameters, an higher (or comparable) normalized average SINR than the FML algorithm (green dashed curve) and the technique proposed in [48] (orange curve), both with the true condition number (blue curve with dots) and the proposed predictor (red curve with crosses). Specifically, the curves show that in the presence of a perfect knowledge of the white disturbance power, i.e. if the parameter $\sigma_a^2 = 0$ dB, the performance coincides with that of the FML, whereas a SINR gain is present with respect to the estimation technique of [48] (see Tables 1.4-1.6). On the contrary, if $\sigma_a^2 > 0$ dB, an interesting SINR gain is present with respect

[7] In the numerical results, it has been considered $MC = 500$.

to the FML (as shown in Tables 1.1-1.3)[8]. In particular, the proposed estimator with the proposed predictor of K_{max} exhibits a SINR gain of 1.8 dB with respect to the FML, in the presence of 3 wideband jammers and with 10 dB power level of the white interference. Furthermore, the proposed estimator exhibits a SINR gain of 0.9 dB with respect to the algorithm of [48], in the presence of 3 narrowband jammers and with 0 dB power level of the white interference. Notice that, for comparison purpose, the PSMI (Pseudo Sample Matrix Inverse) is also considered in the simulations of Figs. 1.2, 1.3, and 1.4 (black dotted line with circles). In particular, the PSMI [4] computes the inverse of the sample covariance matrix when the condition $K \geq N$ holds true; conversely, it utilizes the pseudo inverse of the sample matrix when $K < N$. As expected, the curves show a severe performance degradation of the PSMI with respect to all the other analyzed algorithms; however, it can be seen that as N increases (in particular for $K >> N$) the performance of the PSMI tends to reach those of the other algorithms herein considered.

Table 1.1: Maximum SINR gain (in dB) of the proposed estimator, both with the true condition number and its proposed predictor, with respect to the FML. The values refer to the simulations of Fig. 1.2.

case	1.2a	1.2b	1.2c	1.2d	1.2e	1.2f
true K_{max}	0	0	0	0	1.7	1.6
$K_{max}^{(*)}$	0	0	0.3	0.3	2	1.9

In Fig. 1.5, the effect of K_{max} on the jammer cancellation and its connection with the required processor wordlength, [7], is shown for some values of K, assuming a power level of the white disturbance $\sigma_a^2 = 0$ dB, and the presence of 1 narrowband jammer with power $\sigma_1^2 = 30$ dB and a phase angle $\phi = 25$ deg (100 Monte Carlo independent trials have been considered). Therein, the average Output Response Ratio (ORR) is plotted, i.e. the average ratio between the squared modulus of the

[8]Notice that, if $d_N \geq 1$, then the proposed estimator coincides with the estimator which accounts only for a condition number constraint. Moreover, since $\sigma_a^2 + \lambda_{min}(\boldsymbol{R}) = \lambda_{min}(\mathbb{E}[\boldsymbol{S}]) \geq \mathbb{E}[\lambda_{min}(\boldsymbol{S})] = \mathbb{E}[d_{min}]$, where the upper bound becomes tighter and tighter as K increases, it is expected that the probability that the minimum eigenvalue is less than 1 increases as long as $\sigma_a^2 + \lambda_{min}(\boldsymbol{R})$ is close to 1. This explains the results obtained in the simulations, i.e. a SINR gain is obtained in the presence of narrowband jammers and $\sigma_a^2 = 1$, when the smallest eigenvalue of the true covariance matrix is equal to 1.

Table 1.2: Maximum SINR gain (in dB) of the proposed estimator, both with the true condition number and its proposed predictor, with respect to the FML. The values refer to the simulations of Fig. 1.3.

case	1.3a	1.3b	1.3c	1.3d	1.3e	1.3f
true K_{max}	0	0	0.4	0.4	2	1.9
$K_{max}^{(*)}$	0	0	0.4	0.4	2.1	1.9

Table 1.3: Maximum SINR gain (in dB) of the proposed estimator, both with the true condition number and its proposed predictor, with respect to the FML. The values refer to the simulations of Fig. 1.4.

case	1.4a	1.4b	1.4c
true K_{max}	0	0.1	1.6
$K_{max}^{(*)}$	0	0.3	1.8

Table 1.4: Maximum SINR gain (in dB) of the proposed estimator, with the true condition number, with respect to the estimator with only a condition number constraint. The values refer to the simulations of Fig. 1.2.

case	1.2a	1.2b	1.2c	1.2d	1.2e	1.2f
true K_{max}	0.8	0.6	1	0.9	0.1	0.1

Table 1.5: Maximum SINR gain (in dB) of the proposed estimator, with the true condition number, with respect to the estimator with only a condition number constraint. The values refer to the simulations of Fig. 1.3.

case	1.3a	1.3b	1.3c	1.3d	1.3e	1.3f
true K_{max}	0.9	0.5	0.3	0.1	0	0

Table 1.6: Maximum SINR gain (in dB) of the proposed estimator, with the true condition number, with respect to the estimator with only a condition number constraint. The values refer to the simulations of Fig. 1.4.

case	1.4a	1.4b	1.4c
true K_{max}	0.6	0.7	0

output response in the jamming direction and that along the useful signal one

$$\mathrm{ORR} = \mathbb{E}\left[|\boldsymbol{w}^{\dagger}\boldsymbol{p}(\phi)|^2/|\boldsymbol{w}^{\dagger}\boldsymbol{p}|^2\right], \tag{1.25}$$

where $\boldsymbol{p}(\phi) = [1, \exp(j\phi), \exp(2j\phi), \ldots, \exp((N-1)j\phi)]^T$ is the steering vector in the direction ϕ and $\boldsymbol{w} = \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{p}$. Evidently, the smaller the available number of bit $n_b = \lceil \log_2 K_{max} \rceil$ [7, equation 4.102a p. 156], the worse the cancellation capabilities of the processor. This can be explained observing that the dynamic range of the eigenvalues in the estimated covariance matrix decreases as K_{max} decreases. As a consequence, the processor tends to treat all the directions in the same way or, equivalently, it has less degrees of freedom to set the depth of the null along the interference direction. Another implication of the eigenvalues dynamic range reduction (ruled by K_{max}) is a stabilization of the processor angular response. Otherwise stated, the statistical realizations of output angular pattern exhibit less and less fluctuations as K_{max} decreases. This is an important feature in practical applications because with a quite stable pattern, the disturbance is not very sensible to the modulation resulting from the spatial adaptivity; hence, it could be also cancelled with standard techniques like Moving Target Indicator (MTI) or extensions.

1.3.2 Doppler Processing in the Presence of Bimodal Clutter plus White Noise

The bimodal clutter model accounts for the presence of statistically independent ground and sea clutters in addition to the white noise. Assuming a Gaussian shaped PSD [80] for both the interfering sources, the (i,k)-th element of the overall normalized disturbance covariance matrix is given by

$$\begin{aligned}\boldsymbol{\Sigma}(i,k) = \mathrm{CNR}_S \rho_S^{(i-k)^2} \exp\left[-j2\pi(i-k)f_S\right] + \mathrm{CNR}_G \rho_G^{(i-k)^2} \\ + \sigma_a^2 \delta_{i,k},\end{aligned} \tag{1.26}$$

where CNR_S and CNR_G denote respectively the Clutter to Noise power Ratio for the sea and the ground clutter, ρ_S and ρ_G are respectively the one-lag correlation coefficients for the sea and the ground clutter, f_S is the normalized Doppler frequency of the sea clutter, and $\delta_{i,k}$ is the Kronecker delta function. The performance assessments for the case of Doppler processing refer to 3 different cases, where the actual power level of the white interference σ_a^2 assumes, respectively, the val-

ues 0, 5, 10 dB[9]. The considered temporal steering vector is given by $\boldsymbol{p} = [1, \exp(j2\pi f_d), \ldots, \exp(j2\pi(N-1)f_d)]^T$, with $f_d = 0.15$. Other simulations parameters are specified in the captions of Figs. 1.6 and 1.7.

In Fig. 1.6, the normalized average SINR is plotted versus the number of independent snapshots, for both the proposed algorithm, the one with only a condition number constraint, and the FML. The curves highlight that the constrained structured estimator can achieve, for the considered values of the parameters, an higher (or comparable) normalized average SINR than the FML algorithm (green dashed curve) and the technique of [48] (orange curve), both with the true condition number (blue curve with dots) and the proposed predictor (red curve with crosses). The curves, also, show that the SINR gain varies with the parameter σ_a^2, as shown in Tables 1.7-1.10. Indeed, the proposed estimator with the proposed predictor of K_{max} achieves a SINR gain of 2.24 dB with respect to the FML, in the scenario with $\sigma_a^2 = 10$ dB, $\text{CNR}_S = 10$ dB, and $\text{CNR}_G = 30$ dB. Moreover, the constrained structured estimator exhibits a SINR gain of 1.26 dB with respect to the algorithm of [48], in the scenario with $\sigma_a^2 = 0$ dB, $\text{CNR}_S = 10$ dB, and $\text{CNR}_G = 30$ dB.

Finally, in Fig. 1.7 the performance of the proposed estimator in a mixture of Gaussian ground clutter plus a compound-Gaussian sea clutter with a fully correlated Gamma texture among the training data (shape parameter 0.5 and a mean value 1) are analyzed. The result shows that the proposed estimator still outperforms the counterparts, namely it exhibits a certain robustness with respect to a departure from the nominal Gaussian model.

Table 1.7: Maximum SINR gain (in dB) of the proposed estimator, both with the true condition number and its proposed predictor, with respect to the FML. The values refer to the simulations of Fig. 1.6.

case	1.6a	1.6c	1.6e	1.6b	1.6d	1.6f
true K_{max}	0	0.6	2	0	0.4	2
$K_{max}^{(*)}$	0	0.8	2.2	0	0.2	1.5

[9]Two different cases have been considered. The former ($\sigma_a^2 = 0$ dB) is an ideal case, in which the power level of white interference (due to jammers and white interference) is perfectly known; while, in the latter more realistic case ($\sigma_a^2 \neq 0$ dB), the power level of the white term is not perfectly known.

Table 1.8: Maximum SINR gain (in dB) of the proposed estimator, both with the true condition number and its proposed predictor, with respect to the FML. The values refer to the simulations of Fig. 1.7.

case	1.7a	1.7c	1.7e	1.7b	1.7d	1.7f
true K_{max}	0	0.5	1.9	0	0.4	2
$K_{max}^{(*)}$	0	0.8	2.2	0	0.2	1.5

Table 1.9: Maximum SINR gain (in dB) of the proposed estimator, with the true condition number, with respect to the estimator with only a condition number constraint. The values refer to the simulations of Fig. 1.6.

case	1.6a	1.6c	1.6e	1.6b	1.6d	1.6f
true K_{max}	1.3	0.5	0	1.2	0.6	0

Table 1.10: Maximum SINR gain (in dB) of the proposed estimator, with the true condition number, with respect to the estimator with only a condition number constraint. The values refer to the simulations of Fig. 1.7.

case	1.7a	1.7c	1.7e	1.7b	1.7d	1.7f
true K_{max}	1.2	0.5	0	1.2	0.6	0

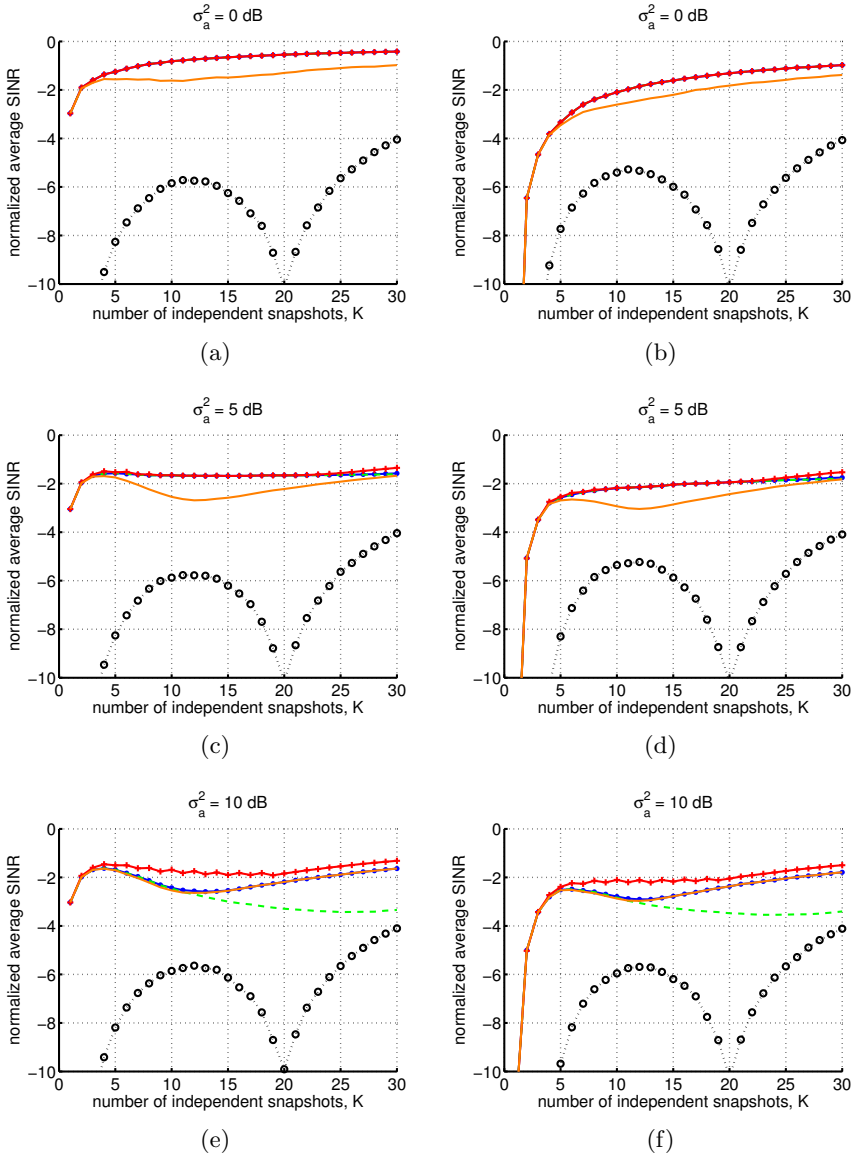


Figure 1.2: Spatial processing. SINR_{av} versus the number of independent snapshots (the blue curve with dots refers to the constrained structured estimator with the true condition number, the red curve with crosses to the constrained structured estimator with the proposed predictor of the condition number, the orange curve to the estimator with only a condition number constraint, the green dashed curve to the FML algorithm, whereas the black dotted line with circles to the PSMI). The analyzed environment includes 1 jammer with power $\sigma^2 = 30$ dB and phase angle $\phi = 20$ deg. Narrowband scenario ($B_f = 0$) (on the left). Wideband scenario ($B_f = 0.3$) (on the right). Three values of σ_a^2 are considered.

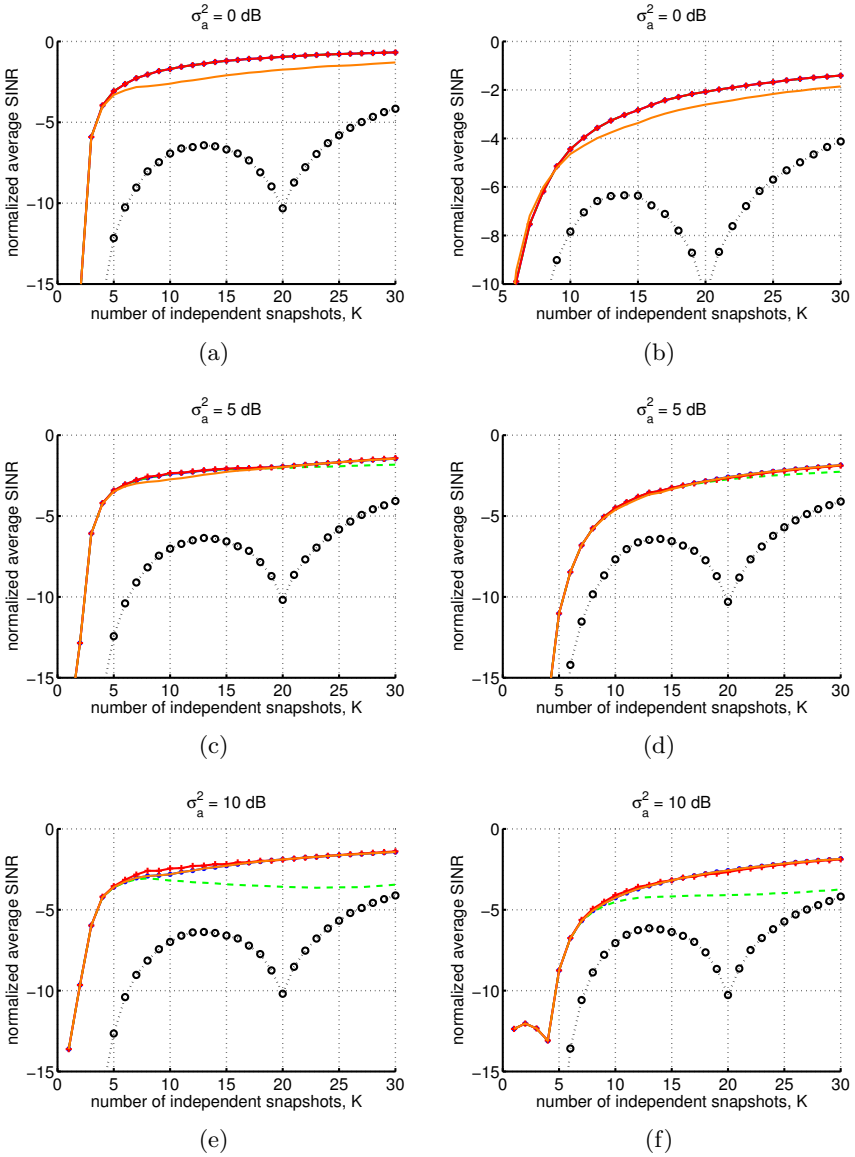


Figure 1.3: Spatial processing. SINR_{av} versus the number of independent snapshots (the blue curve with dots refers to the constrained structured estimator with the true condition number, the red curve with crosses to the constrained structured estimator with the proposed predictor of the condition number, the orange curve to the estimator with only a condition number constraint, the green dashed curve to the FML algorithm, whereas the black dotted line with circles to the PSMI). The analyzed environment includes 3 jammers with power $\sigma^2 = 30$ dB and phase angles $(\phi_1, \phi_2, \phi_3) = (20, 40, 60)$ deg. Narrowband scenario ($B_f = 0$) (on the left). Wideband scenario ($B_f = 0.3$) (on the right). Three values of σ_a^2 are considered.

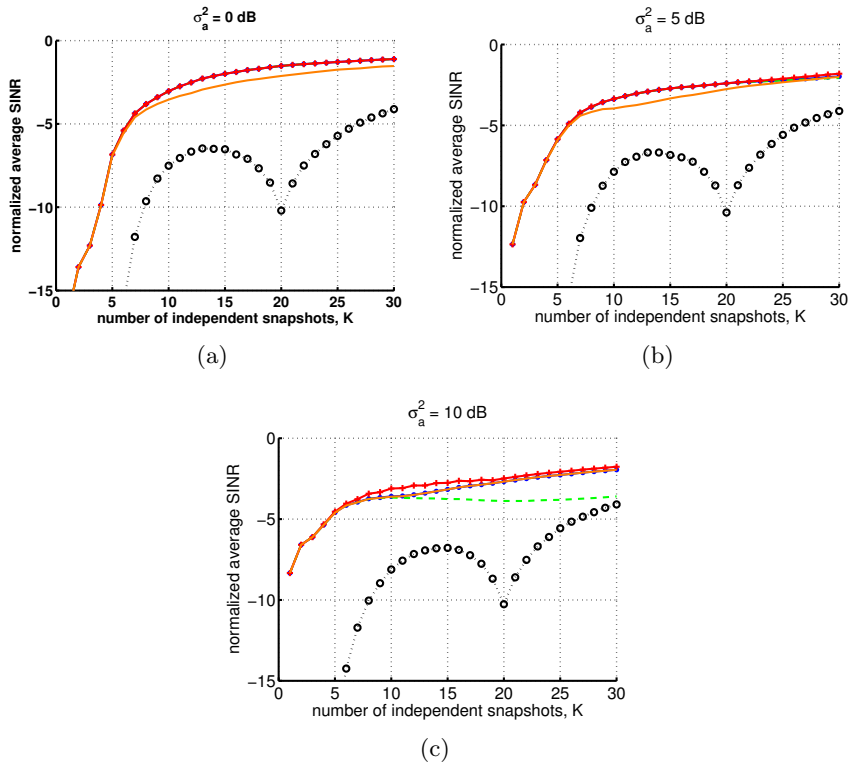


Figure 1.4: Spatial processing. SINR_{av} versus the number of independent snapshots (the blue curve with dots refers to the constrained structured estimator with the true condition number, the red curve with crosses to the constrained structured estimator with the proposed predictor of the condition number, the orange curve to the estimator with only a condition number constraint, the green dashed curve to the FML algorithm, whereas the black dotted line with circles to the PSMI). The analyzed environment includes 3 jammers with powers $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (10, 20, 30)$ dB, phase angles $(\phi_1, \phi_2, \phi_3) = (20, 40, 60)$ deg and fractional bandwidth $B_f = (0.2, 0, 0.3)$, respectively. Three values of σ_a^2 are considered.

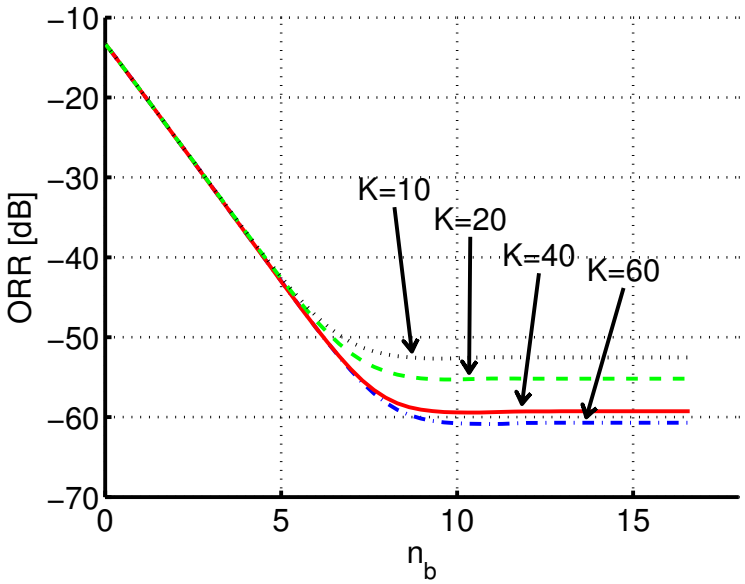


Figure 1.5: Spatial processing. Average ORR (expressed in dB) versus $\lceil \log_2(K_{max}) \rceil$, which is the minimum required wordlength. The analyzed environment includes 1 narrowband jammer with power $\sigma_1^2 = 30$ dB and phase angle $\phi = 25$ deg. The analysis has been conducted for different values of the sample support (i.e. $K = N/2$, $K = N$, $K = 2N$, and $K = 3N$).

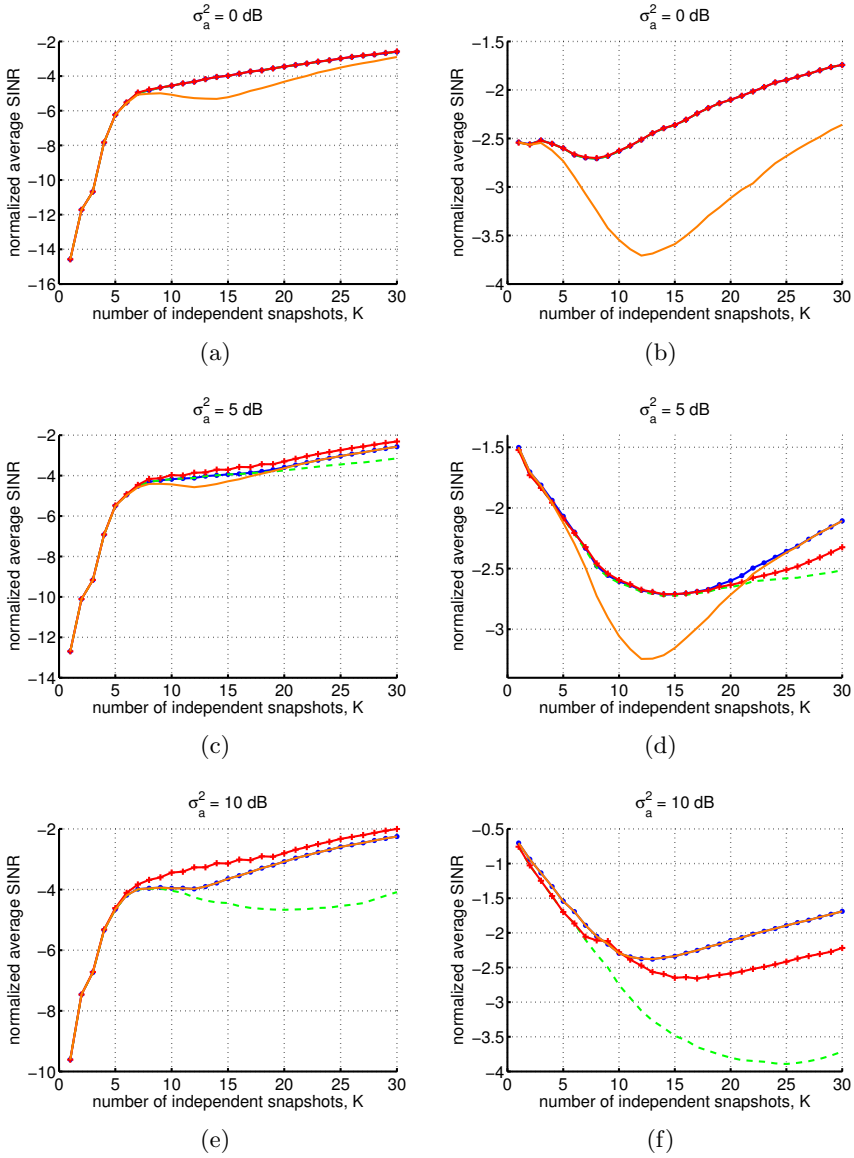


Figure 1.6: Doppler processing. SINR_{av} versus the number of independent snapshots (the blue curve with dots refers to the constrained structured estimator with the true condition number, the red curve with crosses to the constrained structured estimator with the proposed predictor of the condition number, the orange curve to the estimator with only a condition number constraint, whereas the green dashed curve to the FML algorithm). The analyzed environment parameters are $\rho_S = 0.6$ dB, $\rho_G = 0.99$, $f_S = 0.2$, and $\mathrm{CNR}_S = 10$ dB and $\mathrm{CNR}_G = 30$ dB for the curves on the left, and $\mathrm{CNR}_S = 5$ dB and $\mathrm{CNR}_G = 10$ dB for the curves on the right. Three values of σ_a^2 are considered.

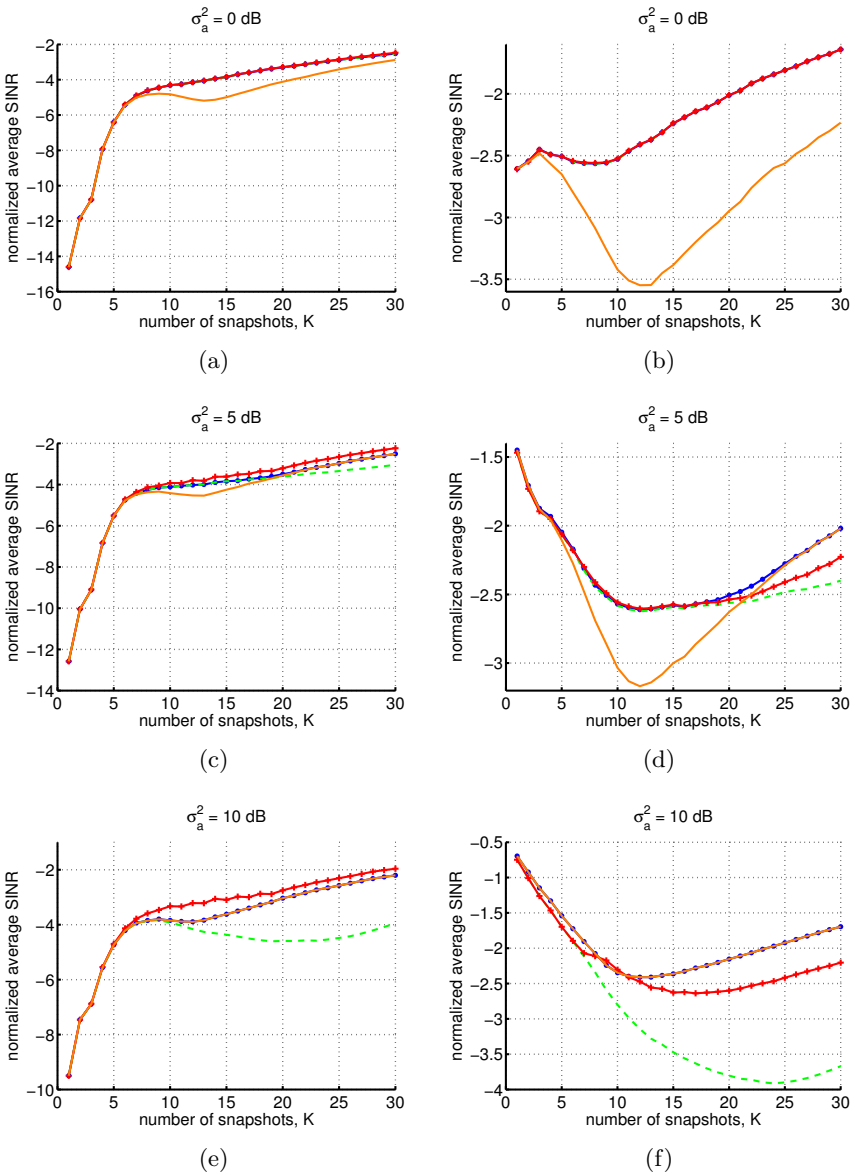


Figure 1.7: Doppler processing. SINR_{av} versus the number of snapshots (the blue curve with dots refers to the constrained structured estimator with the true condition number, the red curve with crosses to the constrained structured estimator with the proposed predictor of the condition number, the orange curve to the estimator with only a condition number constraint, whereas the green dashed curve to the FML algorithm). The analyzed environment parameters are $\rho_S = 0.6$ dB, $\rho_G = 0.99$, $f_S = 0.2$, and $\mathrm{CNR}_S = 10$ dB and $\mathrm{CNR}_G = 30$ dB for the curves on the left, and $\mathrm{CNR}_S = 5$ dB and $\mathrm{CNR}_G = 10$ dB for the curves on the right. Gamma texture shape parameter 0.5 and mean value 1. Three values of σ_a^2 are considered.

Chapter 2

Geometric Approaches to Covariance Estimation for Secondary Data Selection

In this chapter, two classes of covariance matrix estimators which does not depend on the probability distribution function of the sample support are proposed and analyzed. Precisely, any estimator of these two classes is associated to a suitable distance in the considered space and is defined, respectively, as the median matrix [23] or the geometric barycenter [81] of a set of covariance matrices, obtained from the available secondary data set. As to the considered distances, the focus is on Euclidean, Log-Euclidean, Root-Euclidean, and Power-Euclidean distances [82]. Furthermore, the basic covariance matrix estimates (used in the matrix median and geometric barycenter calculations) are computed exploiting some a-priori information about the covariance matrix structure [83]. Based on the new devised estimators, training data selectors, whose aim is to discard secondary data containing outliers, are proposed. The selection is based on the Generalized Inner Product (GIP) [8] exploiting the median matrices or the geometric barycenters in place of the classic sample covariance matrix.

At the analysis level, the performance of the new selection schemes are assessed, in terms of probability of correct outliers excision, comparing the systems exploiting the geometric barycenters with those exploiting the median matrices. The results show that data selectors exploiting geometric medians can outperform those based on geometric barycen-

ters but the former require a computational complexity higher than the latter.

The chapter is organized as follows. In Section 2.1, the system model is described and the two families of estimators are presented; the former are obtained as the solution of a convex optimization problem (i.e. the median matrices) and the latter are obtained in a closed form. In Section 2.2, training data selectors are presented, whereas in Section 2.3 performance analysis are provided.

2.1 Problem Formulation and Covariance Matrix Estimators

This section formalizes the problem of estimating the positive definite covariance matrix $\boldsymbol{\Sigma}$, of K secondary data $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K$, modeled as N-dimensional circularly symmetric zero-mean vectors, with an arbitrary joint statistical distribution, and sharing the same covariance matrix

$$\mathbb{E}\left[\boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}\right] = \boldsymbol{\Sigma}, \quad i = 1, \ldots, K, \tag{2.1}$$

assumed positive definite.

When the statistical characterization of the secondary data is not known, classic approaches, such as the ML or the Minimum Mean Square Error (MMSE) estimations, cannot be applied, and different families of covariance matrix estimators must be introduced.

The framework proposed in this chapter relies on the use of suitable types of distances in the positive definite matrix space, namely on the cone of the positive definite matrices as illustrated in Fig. 2.1. Otherwise

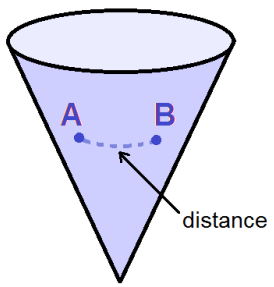


Figure 2.1: Cone of positive definite matrices.

stated, two classes of estimators are defined, based respectively on the geometric barycenter and the generalized median matrix of a set of basic covariance matrix estimates, obtained from the available secondary data set. Specifically, denoting by $\boldsymbol{S}_i$, $i = 1, \ldots, K$, the set of basic covariance matrix estimates and given a distance

$$d(\cdot,\cdot) : \boldsymbol{A} \succ \mathbf{0}, \boldsymbol{B} \succ \mathbf{0} \rightarrow [0, +\infty),$$

the corresponding geometric barycenter-based estimator is defined as

$$\widehat{\boldsymbol{\Sigma}} = \arg \min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \left\{ \sum_{i=1}^{K} w_i d^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}, \tag{2.2}$$

whereas the median-based estimator is

$$\widehat{\boldsymbol{\Sigma}}_M = \arg \min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \left\{ \sum_{i=1}^{K} w_i d(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}. \tag{2.3}$$

Notice that, the coefficients w_i (with $w_i > 0$ and $\sum_{i=1}^{K} w_i = 1$) allow to weight the secondary data, in order to account for their reliability, for instance their degree of homogeneity or temporal acquisition. The weights, w_i, can be also chosen on the basis of the similarity between the secondary data terrain and that of the cell under test (exploiting, for instance, the so-called National Land Cover Data, NLCD [76]). Of course, without any source of a-priori knowledge, it is reasonable to set equal weights and $\widehat{\boldsymbol{\Sigma}}_M$ ends up coincident with the median matrix induced by the metric $d(\cdot,\cdot)$ of the set of matrices $\boldsymbol{S}_i$, $i = 1, \ldots, K$. Moreover, when $w_i = 1/K$, $i = 1, \ldots, K$, and the matrices $\boldsymbol{S}_i$ are assumed i.i.d. random positive definite matrix, then

$$\widehat{\boldsymbol{\Sigma}} = \arg \min_{\boldsymbol{\Sigma}} \frac{1}{K} \left\{ \sum_{i=1}^{K} d^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\},$$

is an empirical (sample) Fréchet mean [84], [85], where a Fréchet mean of an $N \times N$ random covariance matrix $\boldsymbol{S}$ is given by [84], [85]

$$\widehat{\boldsymbol{\Omega}} = \arg \min_{\boldsymbol{\Omega}} \mathbb{E}\left[d^2(\boldsymbol{S}, \boldsymbol{\Omega}) \right]. \tag{2.4}$$

The idea of using the generalized median matrix in (2.3) stems from the well known robustness of the conventional median value with respect

to the presence of outliers in the data. To better explain this concept, it can be recalled that the median value of a real random variable X is defined as

$$\bar{x} = \arg\min_{m} \mathbb{E}\left[|X - m|\right], \quad \text{with } m \in \mathbb{R}. \tag{2.5}$$

It is well known that, for a continuous random variable, $\bar{x}$ is such that

$$\Pr\left(\{X \leq \bar{x}\}\right) = \frac{1}{2} \tag{2.6}$$

namely, the conventional median corresponds to the 50% percentile of the distribution of x. Let us now observe that for an empirical distribution defined by a set of real observations x_i, $i = 1, \ldots, N_0$, the empirical median value is given by

$$\bar{x}_e = \arg\min_{m} \frac{1}{N_0} \sum_{i=1}^{N_0} |x_i - m|, \quad \text{with } m \in \mathbb{R}. \tag{2.7}$$

Thus, denoting by $x_{(i)}$, $i = 1, \ldots, N_0$ the ordered observations in increasing order, it is easy to show that $\bar{x}_e = x_{\left(\lceil \frac{N}{2} \rceil\right)}$. Consequently, replacing an observation $x_i > \bar{x}$ with another (possibly much higher) $y > \bar{x}$, the new empirical median value does not change. In other words, there is a robustness of the median with respect to the presence of outliers. Therefore, leveraging on the above consideration and exploiting the concept of median matrix, which is a generalization of classic median definition, it is possible to devise robust covariance matrix estimators.

As to the set of basic covariance matrix estimates, it is assumed that each $\boldsymbol{S}_i$ is a function of the single secondary datum $\boldsymbol{r}_i$, namely $\boldsymbol{S}_i = \boldsymbol{S}_i(\boldsymbol{r}_i)$, $i = 1, \ldots, K$. In particular, a possible choice could be the rank-one sample matrix related to the vector $\boldsymbol{r}_i$, namely $\boldsymbol{S}_i = \boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}$. However, this is not acceptable since the aforementioned definition of distance requires that the considered matrices are positive definite; hence a-priori information has to be exploited to come up with positive definiteness. Toward this goal, the knowledge of a lower bound to the spectral density of the white noise is assumed (set, without loss of generality, to 0 dB), and the covariance matrix which minimizes the Euclidean distance (also called Frobenius distance) from $\boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}$ is associated to each vector of the sample support, under the constraint $\boldsymbol{S}_i \succeq \boldsymbol{I}$. The following theorem, provides the explicit expression of the aforementioned minimizer.

Theorem 2.1.1. *Let $\boldsymbol{r}_i$ be the i-th N-dimensional secondary datum, the matrix $\boldsymbol{S}_i$ which minimizes the Frobenius distance from $\boldsymbol{r}_i\boldsymbol{r}_i^\dagger$ under the constraint $\boldsymbol{S}_i \succeq \boldsymbol{I}$, i.e. the optimal solution to the optimization problem*

$$\mathcal{P}\begin{cases} \min\limits_{\boldsymbol{S}_i} & ||\boldsymbol{r}_i\boldsymbol{r}_i^\dagger - \boldsymbol{S}_i||^2 \\ s.t. & \boldsymbol{S}_i \succeq \boldsymbol{I} \end{cases}, \tag{2.8}$$

is given by

$$\boldsymbol{S}_i = \boldsymbol{U}_i\boldsymbol{\Lambda}_i\boldsymbol{U}_i^\dagger, \tag{2.9}$$

where

$$\boldsymbol{\Lambda}_i = \mathbf{diag}\left([\lambda_i, 1, \ldots, 1]\right) \qquad with \qquad \lambda_i = \max\left(1, ||\boldsymbol{r}_i||^2\right), \tag{2.10}$$

and $\boldsymbol{U}_i$ is a unitary matrix of the eigenvectors of $\boldsymbol{r}_i\boldsymbol{r}_i^\dagger$ with the first eigenvector corresponding to the eigenvalue $||\boldsymbol{r}_i||^2$.

Proof. See Appendix G.

Notice that, the eigenvalue decomposition is not necessary to compute (2.8) as

$$\boldsymbol{S}_i = \boldsymbol{r}_i\boldsymbol{r}_i^\dagger \max\left(\left[1 - \frac{1}{||\boldsymbol{r}_i||^2}\right], 0\right) + \boldsymbol{I}. \tag{2.11}$$

As shown in [83], the matrix $\boldsymbol{S}_i$ in (2.9) also minimizes the distance induced by the spectral norm from the matrix $\boldsymbol{r}_i\boldsymbol{r}_i^\dagger$ under the constraint $\boldsymbol{S}_i \succeq \boldsymbol{I}$. As to the considered distances, the focus is on the Euclidean, Log-Euclidean, Root-Euclidean, and Power-Euclidean distances. Thus, four covariance matrix estimators for each class are obtained, which are presented in the next subsection.

2.1.1 Covariance Matrix Estimators based on Geometric Barycenters and Median Matrices

In this subsection, the considered distances are described, and the corresponding estimators are derived. Specifically, let $\boldsymbol{A} \succ \boldsymbol{0}$ and $\boldsymbol{B} \succ \boldsymbol{0}$, the considered four distances and the corresponding estimators are formally defined as:

- Euclidean distance,

$$d_E(\boldsymbol{A}, \boldsymbol{B}) = \sqrt{\mathrm{tr}\ \{(\boldsymbol{A} - \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})^\dagger\}},$$

the corresponding barycenter estimator is

$$\widehat{\boldsymbol{\Sigma}}_E = \arg\min_{\boldsymbol{\Sigma}} \left\{ \sum_{i=1}^{K} w_i d_E^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\},$$

whereas the corresponding median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,E} = \arg\min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \left\{ \sum_{i=1}^{K} w_i d_E(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}.$$

- Log-Euclidean distance,

$$d_L(\boldsymbol{A}, \boldsymbol{B}) = \sqrt{\mathrm{tr}\ \{(\log \boldsymbol{A} - \log \boldsymbol{B})(\log \boldsymbol{A} - \log \boldsymbol{B})^\dagger\}},$$

the corresponding barycenter estimator is

$$\widehat{\boldsymbol{\Sigma}}_L = \arg\min_{\boldsymbol{\Sigma}} \left\{ \sum_{i=1}^{K} w_i d_L^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\},$$

whereas the corresponding median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,L} = \arg\min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \left\{ \sum_{i=1}^{K} w_i d_L(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}.$$

- Root-Euclidean distance,

$$d_H(\boldsymbol{A}, \boldsymbol{B}) = \sqrt{\mathrm{tr}\left\{ \left(\boldsymbol{A}^{1/2} - \boldsymbol{B}^{1/2} \right) \left(\boldsymbol{A}^{1/2} - \boldsymbol{B}^{1/2} \right)^\dagger \right\}},$$

and the corresponding barycenter estimator is

$$\widehat{\boldsymbol{\Sigma}}_H = \arg\min_{\boldsymbol{\Sigma}} \left\{ \sum_{i=1}^{K} w_i d_H^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\},$$

whereas the corresponding median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,H} = \arg\min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \left\{ \sum_{i=1}^{K} w_i d_H(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}.$$

- Power-Euclidean distance,

$$d_A(\boldsymbol{A},\boldsymbol{B}) = \sqrt{\text{tr }\{(\boldsymbol{A}^\alpha - \boldsymbol{B}^\alpha)(\boldsymbol{A}^\alpha - \boldsymbol{B}^\alpha)^\dagger\}},$$

and the corresponding barycenter estimator is

$$\widehat{\boldsymbol{\Sigma}}_A = \arg\min_{\boldsymbol{\Sigma}} \left\{ \sum_{i=1}^K w_i d_A^2(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\},$$

whereas the corresponding median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,A} = \arg\min_{\boldsymbol{\Sigma}\succ\mathbf{0}} \left\{ \sum_{i=1}^K w_i d_A(\boldsymbol{S}_i, \boldsymbol{\Sigma}) \right\}.$$

Let us now focus on the structure of the barycenter-based estimators; the following theorem allows to obtain the closed form expression of $\widehat{\boldsymbol{\Sigma}}_E$, $\widehat{\boldsymbol{\Sigma}}_L$, $\widehat{\boldsymbol{\Sigma}}_H$, and $\widehat{\boldsymbol{\Sigma}}_A$.

Theorem 2.1.2. *Given the set of matrices $\boldsymbol{A}_i \in \mathbb{C}^{N\times N}$, $i = 1,\ldots,K$, and the set of weights w_i, $i = 1,\ldots,K$, with $w_i > 0$ and $\sum_{i=1}^K w_i = 1$, the solution to the optimization problem*

$$\widehat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\{ \sum_{i=1}^K w_i ||\boldsymbol{A}_i - \boldsymbol{A}||^2 \right\}, \tag{2.12}$$

is unique, and can be written as

$$\widehat{\boldsymbol{A}} = \sum_{i=1}^K w_i \boldsymbol{A}_i. \tag{2.13}$$

Proof. See Appendix H.

Let us now specialize Theorem 2.1.2 to $\boldsymbol{A}_i = \boldsymbol{S}_i$, where $\boldsymbol{S}_i \succ \mathbf{0}$ is the i-th covariance matrix estimate, and $\boldsymbol{A} = \boldsymbol{\Sigma}$. Thus, the proposed Euclidean estimator is given by

$$\widehat{\boldsymbol{\Sigma}}_E = \sum_{i=1}^K w_i \boldsymbol{S}_i, \tag{2.14}$$

namely, it is the weighted mean of the covariance matrix estimates, $\boldsymbol{S}_i$. For real numbers and $w_i = 1/K$, $i = 1, \ldots, K$, this estimator becomes the arithmetic mean.

Similarly, defining $\boldsymbol{A}_i = \log \boldsymbol{S}_i$ and $\boldsymbol{A} = \log \boldsymbol{\Sigma}$, the Log-Euclidean estimator can be expressed as

$$\widehat{\boldsymbol{\Sigma}}_L = \exp \left\{ \sum_{i=1}^{K} w_i \log \boldsymbol{S}_i \right\}, \tag{2.15}$$

Notice that, for positive real numbers and $w_i = 1/K$, $i = 1, \ldots, K$, $\widehat{\boldsymbol{\Sigma}}_L$ represents a generalization of the geometric mean[1] of the matrices $\boldsymbol{S}_i$.

The Root-Euclidean estimator is obtained by substituting $\boldsymbol{A}_i = \boldsymbol{S}_i^{1/2}$ and $\boldsymbol{A} = \boldsymbol{\Delta}_H = \boldsymbol{\Sigma}^{1/2}$:

$$\widehat{\boldsymbol{\Sigma}}_H = \widehat{\boldsymbol{\Delta}}_H \widehat{\boldsymbol{\Delta}}_H^{\dagger} \qquad \text{where} \qquad \widehat{\boldsymbol{\Delta}}_H = \sum_{i=1}^{K} w_i \boldsymbol{S}_i^{1/2}. \tag{2.16}$$

Finally, the Power-Euclidean estimator follows, taking $\boldsymbol{A}_i = \boldsymbol{S}_i^{\alpha}$ and $\boldsymbol{A} = \boldsymbol{\Delta}_A = \boldsymbol{\Sigma}^{\alpha}$, i.e.

$$\widehat{\boldsymbol{\Sigma}}_A = (\widehat{\boldsymbol{\Delta}}_A)^{1/\alpha} \qquad \text{where} \qquad \widehat{\boldsymbol{\Delta}}_A = \sum_{i=1}^{K} w_i \boldsymbol{S}_i^{\alpha}. \tag{2.17}$$

The coefficient α usually lies in the set $(0, 1]$, and it is obvious that for $\alpha = 1/2$, the Power-Euclidean becomes the Root-Euclidean estimator, and for $\alpha = 1$ it becomes the Euclidean one.

It can be explicitly observed that the estimators (2.16) and (2.17)) extend to the positive definite matrix space the mean power definition of a set of positive real numbers. Indeed, for positive numbers x_i and $w_i = 1/K$, $i = 1, \ldots, K$, they reduce to $\left(\frac{1}{K} \sum_{i=1}^{K} x_i^{\alpha} \right)^{1/\alpha}$, $\alpha \in (0, 1]$.

Let us now focus on the computation of the median-based estimators; the following theorems allow to obtain $\widehat{\boldsymbol{\Sigma}}_{M,E}$, $\widehat{\boldsymbol{\Sigma}}_{M,L}$, $\widehat{\boldsymbol{\Sigma}}_{M,H}$, and $\widehat{\boldsymbol{\Sigma}}_{M,A}$ from the solution of a convex optimization problem.

[1] Given K positive numbers $(x_1, x_2, \ldots, x_K)$, the geometric mean, $\bar{x} = \sqrt[K]{x_1 \cdot x_2 \cdot \ldots \cdot x_K}$, minimizes the sum of the squared hyperbolic distances to the given positive numbers x_i:

$$\bar{x} = \arg\min_{x>0} \left\{ \sum_{i=1}^{K} |\log x - \log x_i|^2 \right\}.$$

Theorem 2.1.3. *Given the set of matrices $\boldsymbol{A}_i \in \mathbb{H}^N$, $i = 1, \ldots, K$, and the set of weights w_i, $i = 1, \ldots, K$, with $w_i > 0$ and $\sum_{i=1}^{K} w_i = 1$, the matrix*

$$\widehat{\boldsymbol{A}} = \arg \min_{\boldsymbol{A} \in \mathbb{H}^N} \left\{ \sum_{i=1}^{K} w_i ||\boldsymbol{A}_i - \boldsymbol{A}|| \right\}, \tag{2.18}$$

can be computed as the optimal solution to the convex optimization Semidefinite Programming (SDP) problem

$$\mathcal{P}_1 \left\{ \begin{array}{ll} \min\limits_{t_1,\ldots,t_K,\boldsymbol{A}} & \sum\limits_{i=1}^{K} w_i t_i \\ s.t. & \begin{pmatrix} t_i & (vec(\boldsymbol{A}) - vec(\boldsymbol{A}_i))^\dagger \\ (vec(\boldsymbol{A}) - vec(\boldsymbol{A}_i)) & t_i \boldsymbol{I} \end{pmatrix} \succeq \boldsymbol{0} \\ & \boldsymbol{A} \in \mathbb{H}^N \end{array} \right. , \quad i = 1, \ldots, K \tag{2.19}$$

Proof. See Appendix I.

Notice that, the computational complexity connected with the solution of the optimization problem $\mathcal{P}_1$ is $O(N^{3.5} \log(1/\eta))$, where η is a prescribed accuracy (see [74]).

The following proposition provides an interesting property of the median matrix in (2.18).

Proposition 2.1.4. *If $\boldsymbol{A}_i \succeq 0$, $i = 1, \ldots, K$, then $\widehat{\boldsymbol{A}}$ in (2.18) is positive semidefinite. Moreover, if $\boldsymbol{A}_i \succ 0$, $i = 1, \ldots, K$, then $\widehat{\boldsymbol{A}}$ in (2.18) is positive definite. Additionally, if the matrices $\boldsymbol{A}_i \in \mathbb{H}^N$ are not aligned, i.e. there are no $\boldsymbol{Q}_0 \in \mathbb{H}^N$ and $\boldsymbol{Q}_1 \in \mathbb{H}^N$ such that*

$$\boldsymbol{A}_i = t_i \boldsymbol{Q}_1 + \boldsymbol{Q}_0, \quad t_i \in \mathbb{R}, i = 1, \ldots, K, \tag{2.20}$$

the median matrix is unique.

Proof. See Appendix J.

Let us now specialize Theorem 2.1.3 to $\boldsymbol{A}_i = \boldsymbol{S}_i$, where $\boldsymbol{S}_i \succeq \boldsymbol{I}$ is the i-th covariance matrix estimate, and $\boldsymbol{A} = \boldsymbol{\Sigma}$. Thus, the proposed Euclidean median estimator $\widehat{\boldsymbol{\Sigma}}_{M,E}$ is obtained by the optimal solution to (2.18).

Similarly, defining $\boldsymbol{A}_i = \log \boldsymbol{S}_i \succeq \boldsymbol{0}$ and $\boldsymbol{A} = \log \boldsymbol{\Sigma}$, and denoting by $\widehat{\boldsymbol{A}}$ the corresponding optimal solution to (2.18), the Log-Euclidean median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,L} = \exp\left\{\widehat{\boldsymbol{A}}\right\}, \tag{2.21}$$

Now, defining $\boldsymbol{A}_i = \boldsymbol{S}_i^{1/2} \succeq \boldsymbol{I}$ and denoting by $\widehat{\boldsymbol{A}}$ the optimal solution to (2.18), the Root-Euclidean median estimator is

$$\widehat{\boldsymbol{\Sigma}}_{M,H} = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^{\dagger}. \tag{2.22}$$

Moreover, the Power-Euclidean median estimator follows taking $\boldsymbol{A}_i = \boldsymbol{S}_i^{\alpha} \succeq \boldsymbol{I}$ and defining

$$\widehat{\boldsymbol{\Sigma}}_{M,A} = (\widehat{\boldsymbol{A}})^{1/\alpha} \tag{2.23}$$

where $\widehat{\boldsymbol{A}}$ is the corresponding optimal solution to (2.18). The coefficient α lies in the set $(0, 1]$, and it is evident that for $\alpha = 1/2$, the Power-Euclidean degenerates into the Root-Euclidean median estimator whereas, for $\alpha = 1$, the Power-Euclidean becomes the Euclidean median estimator. Notice that, $(\cdot)^{\alpha}$ acts as a compressor on the eigenvalues and it is an infinite of order α. Moreover, the logarithm is an infinite of order lower than α, for each value of $\alpha > 0$. Hence, as $\alpha \to 0$, the Power-Euclidean selector tends to behave as the Log-Euclidean one.

Processing interpretation. First of all, the matrices $\boldsymbol{S}_i$, $i = 1, \ldots, K$, are transformed through the specific function (identity, log, power, root) defining the corresponding barycenter or median (Fig. 2.2). Then, the weighted average or the median of the transformed $\boldsymbol{S}_i$, $i = 1, \ldots, K$, is computed, solving the SDP problem (2.19) in the latter case. Finally, the inverse transformation of the barycenter or median specific function is applied to the average matrix or to the median matrix, respectively, computed at the previous step.

As a consequence, the matrices $\boldsymbol{S}_i$, $i = 1, \ldots, K$, are first processed into an ad-hoc domain and then, through the inverse transformation, the covariance estimate is re-mapped into the original positive definite matrix space. The benefits obtainable resorting to this processing chain could be explained observing that the log, power, and square-root transformations act as compressors of the eigenvalues and hence they try to de-emphasize the effect of outliers.

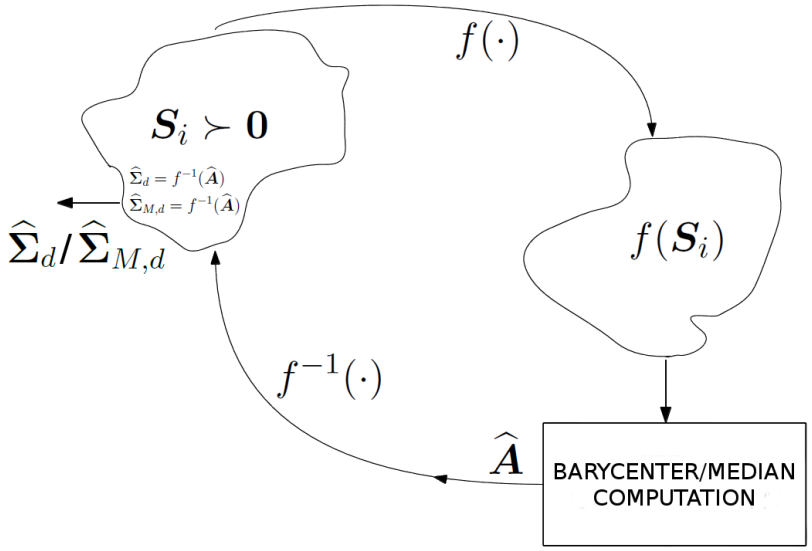


Figure 2.2: A pictorial illustration about the generalized geometric barycenter/median computation procedure.

2.2 Secondary Data Selection Design

In the present section, a secondary data selection scheme exploiting the proposed barycenter and median-based covariance matrix estimators is introduced. The aim of the data selector is to choose the most homogeneous secondary data, discarding vectors containing possible outliers. This screening is of paramount importance, since adaptive receivers achieve satisfactory performances when the available secondary data share the same spectral properties. In Fig. 2.3, the block-scheme of the proposed class of training data selectors is shown.

The selector screens among the K training data excising the K_0 vectors with the highest GIP [8], [86], [87] computed using one of the new devised covariance matrix estimators (described in Subsection 2.1.1) in place of the classic sample covariance matrix. Precisely, denoting by $\Omega = \{1, 2, \ldots, K\}$ the set of the secondary data indices and by $\Omega_0 = \{i_1, i_2, \ldots, i_{K-K_0}\} \subseteq \Omega$ the subset, of cardinality $K - K_0$, of the

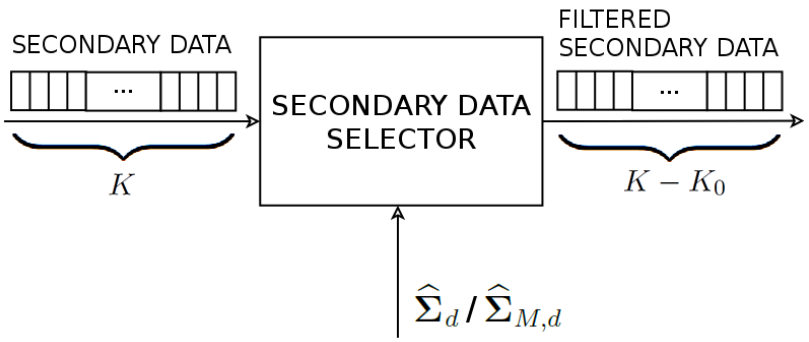


Figure 2.3: Pictorial representation of the training data selector scheme.

selected training data indices, the proposed screening procedure based on the estimator $\widehat{\boldsymbol{\Sigma}}_d$, where d accounts for the considered distance, can be summarized as follows:

- compute the covariance matrix estimate $\widehat{\boldsymbol{\Sigma}}_d$ exploiting the secondary data $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_K$;
- $\forall\, i \in \Omega$ construct the quadratic form (GIP)

$$\beta_i^d = \boldsymbol{r}_i^\dagger \widehat{\boldsymbol{\Sigma}}_d^{-1} \boldsymbol{r}_i \quad i = 1, \ldots, K, \tag{2.24}$$

  and sort β_i^ds in decreasing order;
- The set Ω_0 of the selected $K - K_0$ indices, is obtained censoring from Ω the indices corresponding to the K_0 highest values β_i^d.

Notice that β_i^d measures the energy of the vector $\widehat{\boldsymbol{\Sigma}}_d^{-\frac{1}{2}} \boldsymbol{r}_i$. Thus, the screening procedure excises the K_0 data having the highest energy in the "quasi-whitened-space", where the whitening operation is tied up to the specific covariance matrix estimate. Notice also that, the procedure here illustrated is the same for the medians, substituting $\widehat{\boldsymbol{\Sigma}}_{M,d}$ in place of $\widehat{\boldsymbol{\Sigma}}_d$.

Whenever the secondary data have been selected, the sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{SM} = \frac{1}{K - K_0} \sum_{j \in \Omega_0} \boldsymbol{r}_j \boldsymbol{r}_j^\dagger \tag{2.25}$$

can be computed from the filtered data, and it can be used in adaptive detectors like AMF, [6], or Kelly, [5].

As to the selection of K_0, heuristic rules, based on the available a-priori information, can be considered. Some guidelines are now presented. Exploiting digital terrain maps, such as the National Land Cover Data [76], the illuminated environment can be classified in terms of its homogeneity. Precisely, through the NLCD interrogation, the environment illuminated by the radar can be classified as urban, open sea, desert, and consequently it can be associated to each classified environment a suitable number of secondary data to be discarded, i.e. K_0. For instance, if the analyzed environment is classified as high intensity residential urban area, an high value of K_0 can be considered, due to outliers corresponding to cars, or street lamps. On the contrary, if the environment is classified as open sea, K_0 can be set to a lower value. Furthermore, a cognitive approach can be followed. Precisely, resorting to previous acquired observations (data scans), a prediction of the number of discrete outliers, present in the scene, can be obtained.

In the following section, the capabilities of the data selectors based on barycenters and on median matrices to excise training data containing outliers are analyzed. For comparison purpose, also the selector which exploits the classical sample covariance matrix is considered

$$\widehat{\boldsymbol{\Sigma}}_S = \frac{1}{K}\sum_{i=1}^{K} \boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}, \tag{2.26}$$

to compute the GIP. The covariance matrix estimates are obtained considering equal weights $w_i = \frac{1}{K}$, $i = 1, \ldots, K$, namely no a-priori information is assumed. Moreover, the parameter α of the Power-Euclidean distance based estimators (2.17) and (2.23) is set to 0.8.

2.3 Analysis of the Selection Properties

To evaluate the performance of the proposed training data selectors, the probability of correct selection of the secondary data, $P_{selection}$ (defined as the probability that all the selected data do not contain any outliers), is considered as figure of merit.

A Doppler processing is considered and the covariance matrix of the disturbance is modeled as the sum of two contributions, i.e. a colored

matrix accounting for the clutter and an identity matrix accounting for the white thermal noise

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \boldsymbol{I}. \tag{2.27}$$

Assuming a Gaussian shaped clutter PSD [80], the covariance matrix term due to the clutter is modeled as

$$\boldsymbol{\Sigma}_0(i,k) = \sigma_c^2 \rho^{|i-k|} e^{j2\pi f_{d_c}(i-k)}, \qquad i,k = 1,\dots,N, \tag{2.28}$$

where ρ is the one-lag correlation coefficient, σ_c^2 is the Clutter to Noise power Ratio (CNR) and f_{d_c} is the clutter normalized Doppler frequency. The considered simulation setting is $\rho = 0.95$, $\sigma_c^2 = 20$ dB, and $f_{d_c} = 0.05$. With reference to the secondary data to be processed, the availability of $K = 24$ training vectors is assumed. In four vectors of the training set some outliers are injected, whose temporal steering signatures are

$$\boldsymbol{p}_i = \alpha_i [1, e^{j2\pi f_{d_{o,i}}}, \dots, e^{j2\pi(N-1) f_{d_{o,i}}}]^T, \quad i = 1,\dots,4, \tag{2.29}$$

where $f_{d_{o,i}}$ is the normalized Doppler frequency of the i-th outlier, and $|\alpha_i|^2$ accounts for the outlier power. In the following, the average power of the outliers is denoted by

$$|\alpha|_{av}^2 = \frac{1}{4}\sum_{i=1}^{4} |\alpha_i|^2. \tag{2.30}$$

Due to the lack of a closed form expression for $P_{selection}$, the performance analysis is conducted resorting to Monte Carlo simulation, evaluating the number of times that a correct selection occurs over the total number of trials MC, set to 500.

In Fig. 2.4, for each considered data selector, $P_{selection}$ is plotted versus the average power of the outliers $|\alpha|_{av}^2$, for $N = 8$, $K = 24$, $f_{d_{o,1}} = 0.10$, $f_{d_{o,2}} = 0.20$, $f_{d_{o,3}} = 0.18$, $f_{d_{o,4}} = 0.22$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$, i.e. equal power outliers. As to the number of training data to be excised, four situations are considered, namely $K_0 = 4, 5, 6, 7$, whose results are respectively shown in the subplots (a), (b), (c) and (d) of the figure. The curves highlight that the Log-Euclidean barycenter-based estimator outperforms the other barycenters and the Log-Euclidean median-based estimator outperforms the other medians; it can be justified observing that the log function is the best compressor among the considered estimators, namely it should achieve

the best effect of outliers de-emphasization. Moreover, increasing K_0, higher values of $P_{selection}$ are achieved, and all the estimators tend to approach the same performance. Obviously, K_0 trades-off the homogeneity of the selected data with the actual amount of secondary data to be used for the subsequent adaptation process. Moreover, the figure shows that, given a specific distance, the selector based on the median matrix generally achieve a higher probability of correct outliers excision than that ensured by the geometric barycenter. This behavior highlights the expected robustness of the median with respect to the presence of outliers. For instance, comparing the Euclidean median and the Euclidean barycenter, in correspondence of $|\alpha|^2_{av} = 25$ dB and $K_0 = 4$, the former ensures a $P_{selection}$ higher than 0.9, whereas the latter a $P_{selection}$ almost equal to 0.8. Moreover, the conducted analysis shows that the Log-Euclidean median-based and the Log-Euclidean barycenter-based estimators achieve almost the same selection performance. This result can be justified by the outliers de-emphasization property of the log transformation. Nevertheless, it is worth pointing out that the better performance of the selectors based on the geometric medians is paid with an increased computational complexity. This is due to the fact that, while matrix barycenters can be evaluated through simple analytic expressions, the computation of median matrices involves the solution of an SDP problem. As to the classic sample covariance matrix based screener, it experiences a performance degradation with respect to both geometric barycenters-based and geometric medians-based selectors; this behaviour is not surprising and was actually expected being the sample covariance matrix the ML estimator under the assumption of homogeneity for the training data.

In Fig. 2.5, $P_{selection}$ is plotted versus $|\alpha|^2_{av}$ assuming non equal power outliers. Precisely, the same simulation setup of Fig. 2.4 is considered, with the only difference that $|\alpha_2|^2 = \frac{2}{3}|\alpha_1|^2$, $|\alpha_3|^2 = |\alpha_1|^2$, and $|\alpha_4|^2 = \frac{3}{2}|\alpha_1|^2$. Interestingly, the Log-Euclidean median estimator still achieves the best performance (which is almost the same as that of the corresponding barycenter-based estimator), whereas the Root-Euclidean median estimator still achieves an effective rejection of the outliers. Moreover, Fig. 2.5 highlights a degradation in the selection performance of the other estimators. The present analysis also confirms the fact that, for each considered distance, the estimator based on the geometric median reaches a higher $P_{selection}$ than the one based on ge-

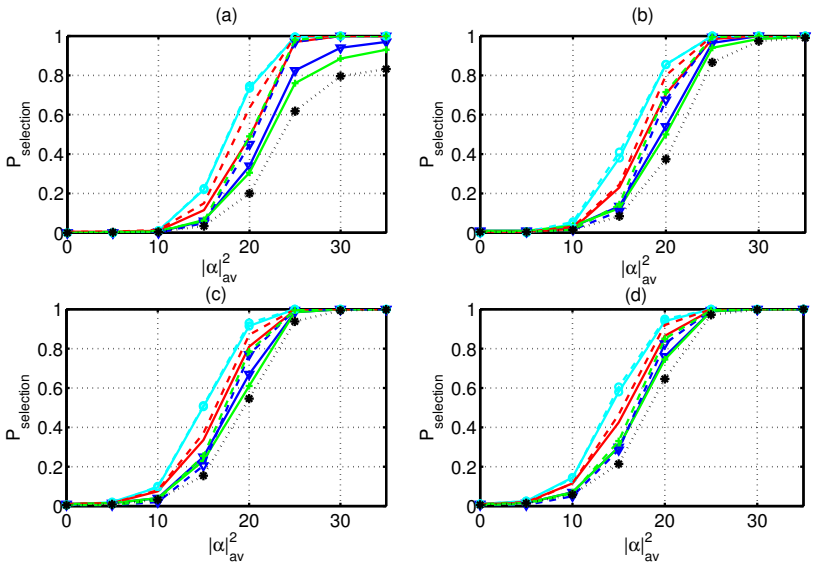


Figure 2.4: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$, whereas $K_0 = (4, 5, 6, 7)$ respectively for subplots (a), (b), (c) and (d). The dashed curves refer to the geometric medians, i.e. $\widehat{\mathbf{\Sigma}}_{M,E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. Euclidean $\widehat{\mathbf{\Sigma}}_E$ ($\bigtriangledown$-marked curve), Log-Euclidean $\widehat{\mathbf{\Sigma}}_L$ ($\circ$-marked curve), Root-Euclidean $\widehat{\mathbf{\Sigma}}_H$ (no marked curve), Power-Euclidean $\widehat{\mathbf{\Sigma}}_A$ ($+$-marked curve). Finally, the dotted curve refers to $\widehat{\mathbf{\Sigma}}_S$ ($*$-marked curve).

ometric barycenter. Also in this case the Log-Euclidean median-based estimator and the Log-Euclidean barycenter estimator present almost the same selection performance.

Finally, in Fig. 2.6, under the same simulation setup of Fig. 2.5, the outliers configuration $|\alpha_1|^2 = |\alpha_2|^2$ and $|\alpha_3|^2 = |\alpha_4|^2 = |\alpha_1|^2/10$ is analyzed. The best performances are still ensured by the Log-Euclidean estimators (again, the Log-Euclidean median-based and Log-Euclidean barycenter-based estimators present comparable performances). All the other estimators exhibit a severe selection performance degradation. Also in this figure it is clear that, the performances achieved by the estimators based on geometric medians are generally better than those achieved by the corresponding estimators based on geometric barycenters. For instance, for $|\alpha|^2_{av} = 25$ dB and $K_0 = 4$, the Root-Euclidean

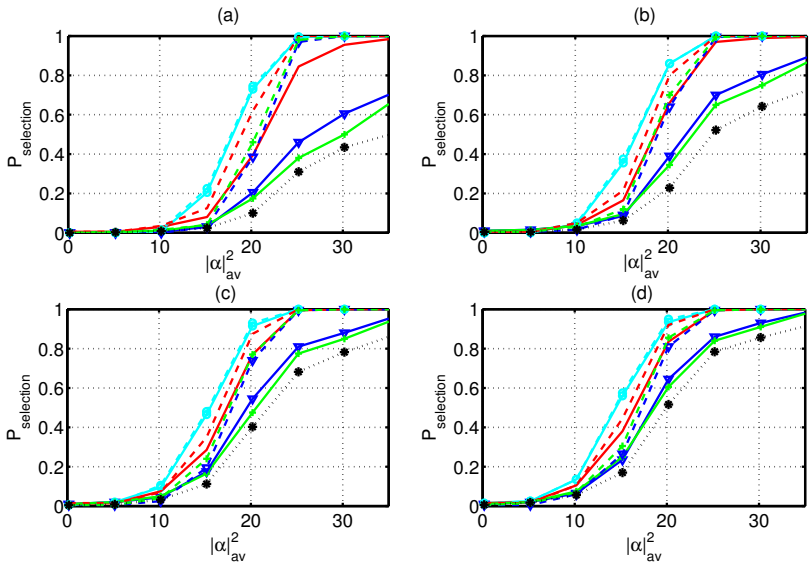


Figure 2.5: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $|\alpha_2|^2 = \frac{2}{3}|\alpha_1|^2$, $|\alpha_3|^2 = |\alpha_1|^2$, and $|\alpha_4|^2 = \frac{3}{2}|\alpha_1|^2$, whereas $K_0 = (4, 5, 6, 7)$ respectively for subplots (a), (b), (c) and (d). The dashed curves refer to the geometric medians, i.e. $\widehat{\boldsymbol{\Sigma}}_{M,E}$ ($\triangledown$-marked curve), $\widehat{\boldsymbol{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\boldsymbol{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\boldsymbol{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. $\widehat{\boldsymbol{\Sigma}}_{E}$ ($\triangledown$-marked curve), $\widehat{\boldsymbol{\Sigma}}_{L}$ ($\circ$-marked curve), $\widehat{\boldsymbol{\Sigma}}_{H}$ (no marked curve), $\widehat{\boldsymbol{\Sigma}}_{A}$ ($+$-marked curve). Finally, the dotted curve refers to $\widehat{\boldsymbol{\Sigma}}_{S}$ ($*$-marked curve).

median has a $P_{selection}$ of about 0.7, whereas the Root-Euclidean barycenter presents a $P_{selection}$ of about 0.2.

To further analyze the selection properties of the barycenter and median-based estimators, the focus is on another outlier scenario [87], [88]. Specifically, given K complex Gaussian zero-mean secondary data, it is assumed that $K - 4$ vectors share the covariance matrix $\boldsymbol{\Sigma}$, given in (2.27), whereas 4 are characterized by a perturbed covariance matrix, defined as

$$\boldsymbol{\Sigma}_{o,i} = \boldsymbol{\Sigma} + \boldsymbol{p}_i \boldsymbol{p}_i^{\dagger} \tag{2.31}$$

where $\boldsymbol{p}_i$ is the steering vector of the i-th outlier, defined in (2.29).

In Fig. 2.7, for each considered data selector, $P_{selection}$ is plotted versus the average power of the outliers $|\alpha|^2_{av}$, for $N = 8$, $K = 24$, $f_{d_{o,i}} = 0.15$, for $i = 1, \ldots, 4$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$. As to

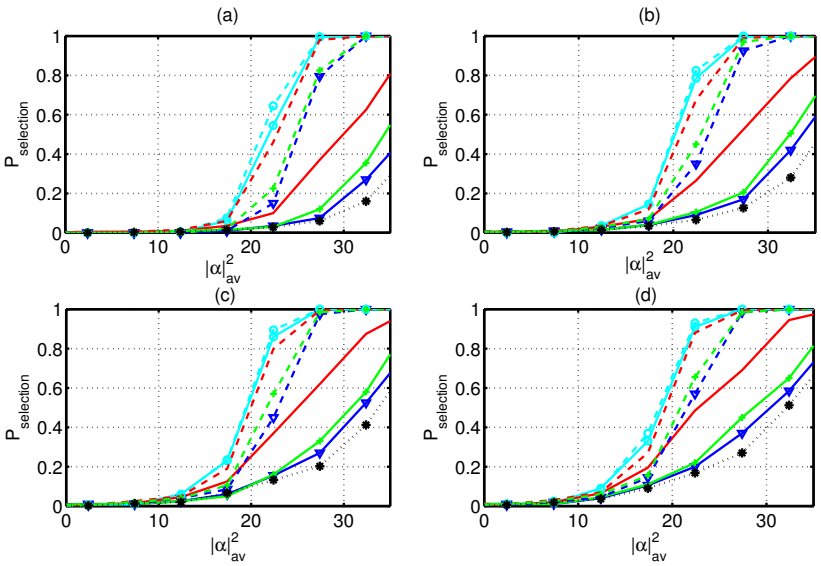


Figure 2.6: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $|\alpha_1|^2 = |\alpha_2|^2$ and $|\alpha_3|^2 = |\alpha_4|^2 = |\alpha_1|^2/10$, whereas $K_0 = (4, 5, 6, 7)$ respectively for subplots (a), (b), (c) and (d). The dashed curves refer to the geometric medians, i.e. $\widehat{\mathbf{\Sigma}}_{M,E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. $\widehat{\mathbf{\Sigma}}_{E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{A}$ ($+$-marked curve). Finally, the dotted curve refers to $\widehat{\mathbf{\Sigma}}_{S}$ ($*$-marked curve).

the number of training data to be excised, it is considered $K_0 = 6$. The curves confirm that the Log-Euclidean median-based estimator outperforms the others. Additionally, given a specific distance, the selectors exploiting the median matrices generally achieve an higher probability of correct outliers excision than those exploiting geometric barycenters. The Log-Euclidean median-based selector exhibits a performance gain of the order of 1 dB with respect to the Log-Euclidean barycenter screener, which has a robust behaviour with respect to presence of outliers. Finally, the selector based on the classical sample covariance matrix is characterized by the worst selection capabilities.

In Fig. 2.8, for each considered data selector, $P_{selection}$ is plotted versus the average power of the outliers $|\alpha|^2_{av}$, for $N = 8$, $K = 24$, $f_{d_{o,1}} = f_{d_{o,1}} = 0.15$, $f_{d_{o,3}} = 0.1$ and $f_{d_{o,4}} = 0.2$, and $|\alpha_1|^2 = |\alpha_2|^2 =$

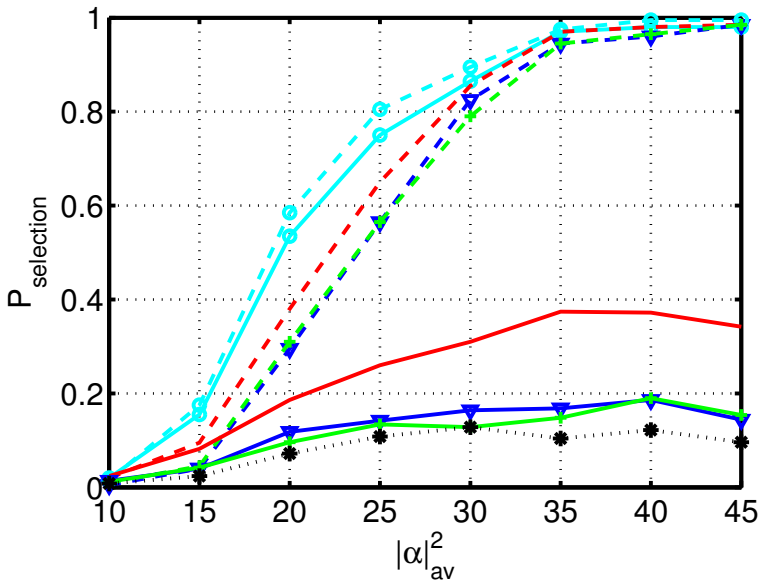


Figure 2.7: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $f_{d_{o,i}} = 0.15$, for $i = 1, \ldots, 4$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$, whereas $K_0 = 6$. The dashed curves refer to the geometric medians, i.e. $\widehat{\mathbf{\Sigma}}_{M,E}$ ($\triangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. $\widehat{\mathbf{\Sigma}}_{E}$ ($\triangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{A}$ ($+$-marked curve). Finally, the dotted curve refers to $\widehat{\mathbf{\Sigma}}_{S}$ ($*$-marked curve).

$|\alpha_3|^2 = |\alpha_4|^2$. As to the number of training data to be excised, it is still considered $K_0 = 6$. The curves attest the trend observed in the previous analyzed cases, even if the barycenters show an increment in their selection performance with respect to the scenario considered in Fig. 2.7.

In Fig. 2.9, for each considered data selector, $P_{selection}$ is plotted versus the average power of the outliers $|\alpha|^2_{av}$, for $N = 8$, $K = 24$, $K_0 = 6$, $f_{d_{o,1}} = 0.1$, $f_{d_{o,2}} = 0.2$, $f_{d_{o,3}} = 0.3$ and $f_{d_{o,4}} = 0.4$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$. The plots highlight that, even if the curves tend to approach the same performance, the Log-Euclidean median-based estimator still outperforms the counterparts.

Summarizing, the obtained results show that the median-based estimators generally outperform the barycenter-based estimators, at the

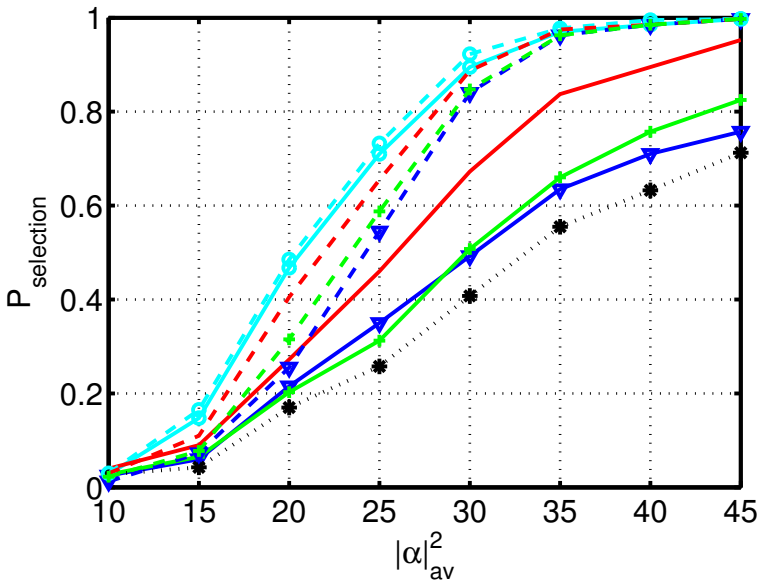


Figure 2.8: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $f_{d_{o,1}} = 0.1$, $f_{d_{o,2}} = 0.2$, $f_{d_{o,3}} = 0.3$ and $f_{d_{o,4}} = 0.4$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$, whereas $K_0 = 6$. The dashed curves refer to the geometric medians, i.e. $\widehat{\mathbf{\Sigma}}_{M,E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. $\widehat{\mathbf{\Sigma}}_{E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{A}$ ($+$-marked curve). Finally, the dotted curve refers $\widehat{\mathbf{\Sigma}}_{S}$ ($*$-marked curve).

price of a higher computational burden.
However, the results also suggest, from a practical point of view, the use of the Log-Euclidean barycenter-based estimator, since it achieves almost the same selection capabilities of the Log-Euclidean median-based estimator, while strongly reducing the computational effort.

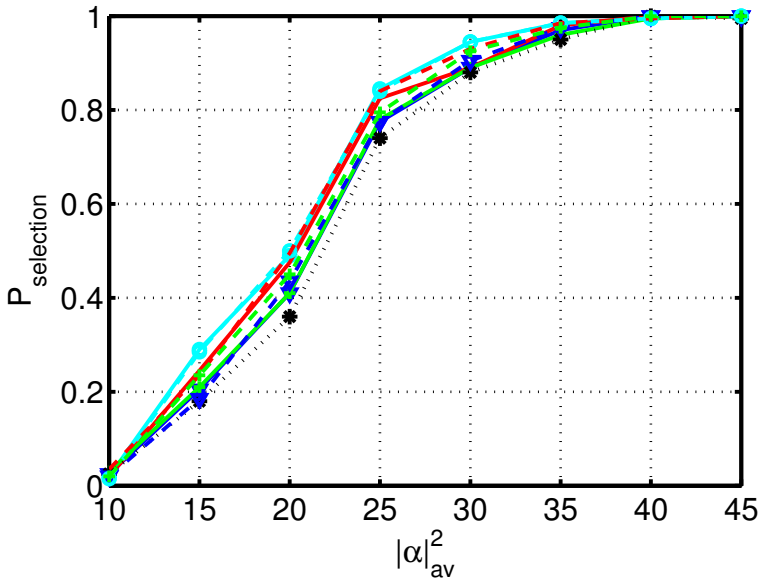


Figure 2.9: $P_{selection}$ versus $|\alpha|^2_{av}$. The curves refer to 4 outliers with $f_{d_{o,1}} = 0.1$, $f_{d_{o,2}} = 0.2$, $f_{d_{o,3}} = 0.3$ and $f_{d_{o,4}} = 0.4$, and $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = |\alpha_4|^2$, whereas $K_0 = 6$. The dashed curves refer to the geometric medians, i.e. $\widehat{\mathbf{\Sigma}}_{M,E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{M,H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{M,A}$ ($+$-marked curve). The continuous curves refer to the geometric barycenters, i.e. $\widehat{\mathbf{\Sigma}}_{E}$ ($\bigtriangledown$-marked curve), $\widehat{\mathbf{\Sigma}}_{L}$ ($\circ$-marked curve), $\widehat{\mathbf{\Sigma}}_{H}$ (no marked curve), $\widehat{\mathbf{\Sigma}}_{A}$ ($+$-marked curve). Finally, the dotted curve refers to $\widehat{\mathbf{\Sigma}}_{S}$ ($*$-marked curve).

Chapter 3

Radar Distributed Targets Detection in Homogeneous Interference with Covariance Matrix Unitary Constraints

In this chapter, the problem of detecting an extended target (or distributed target), embedded in Gaussian noise with unknown but structured covariance matrix, is addressed. It is supposed that the data are collected by N channels (temporal, spatial, or spatial-temporal) and that the possible target is sought within K range cells. The target echo from each range bin is modeled as a deterministic signal times a deterministic but unknown scaling factor which accounts for the target response and may possibly vary from cell to cell. Moreover, it isn't supposed the availability of any secondary data set, free of signal components; whereas, some a-priori knowledge about the operating environment are exploited to enforce the covariance matrix to belong to a suitable uncertainty set. Specifically, it is considered that a properly transformed inverse disturbance covariance matrix belongs to a nonempty and bounded set described through unitary invariant continuous functions, namely continuous functions depending only on the eigenvalues of their matrix argument. Hence, the constrained ML estimates of the unknown parameters

(the target range responses and the covariance matrix) are derived and the GLRT detector for the hypothesis test under consideration is designed. The devised constrained ML estimators involve the eigenvalue decomposition of both the transformed sample covariance matrix as well as its modified version computed from the data projected in the null space of the transformed useful signal. Additionally, two optimization problems have to be solved. Many covariance matrix uncertainty sets of practical interest (compliant with the proposed model) lead to convex optimization problems, that can be solved in the worst case with a polynomial complexity in the number of channels. This is, for instance, the case of the structural constraint in [47], the condition number upper bound constraint in [48] and Chapter 1, and similarity constraints like that in [11]. Moreover, there are also situations where the optimization problems to be solved are non-convex but their optimal solutions can be still efficiently obtained; an example of such a case arises for the constraint set introduced in [89].

At the analysis stage, the performance of the devised class of GLRT-based detectors is assessed, for some interference scenarios arising in angle and Doppler processing applications. Specifically, for each analyzed situation, the Detection Probability (P_D) of the GLRT decision rule, devised exploiting a specific a-priori knowledge which suitably constrains the interference plus noise covariance matrix, is evaluated. The results highlight that the proper use of the a-priori information can lead to a detection performance quite close to the optimum receiver, which supposes the perfect knowledge of the interference plus noise covariance matrix. Additionally, as expected, they also confirm the intuition that the better the covariance matrix uncertainty set characterization, the better the P_D.

The chapter is organized as follows. In Section 3.1, the problem is formulated and both the target and disturbance model are introduced; furthermore, the constrained ML estimates of the unknown parameters are devised and the GLRT is designed. In Section 3.2, some practical examples of covariance matrix uncertainty sets compliant with the proposed model are given and ad-hoc solution algorithms to compute efficiently the corresponding ML estimates are presented. In Section 3.3, the performances of the proposed class of GLRT-based detectors are assessed for some covariance matrix uncertainty sets.

3.1 Problem Formulation & System Design

In this section, the problem of detecting the presence of a target distributed across K range cells, $\boldsymbol{r}_i$, $i = 1, \ldots, K$ is dealt, considering a radar that collects data from N channels. This problem can be formulated as the following binary hypothesis test

$$\begin{cases} H_0 : \ \boldsymbol{r}_i = \boldsymbol{n}_i & i = 1, \ldots, K \\ H_1 : \ \boldsymbol{r}_i = \boldsymbol{n}_i + \alpha_i \boldsymbol{p} & i = 1, \ldots, K \end{cases}, \tag{3.1}$$

where $\boldsymbol{p} \in \mathbb{C}^N$ denotes the known unitary norm steering vector and α_i's, $i = 1, \ldots, K$, are (unknown) deterministic complex parameters accounting for both target reflectivity and channel propagation effects. As to the disturbance random vectors, it is assumed that $\boldsymbol{n}_i$'s, $i = 1, \ldots, K$, are independent, complex, zero-mean, circular Gaussian vectors sharing the same unknown covariance matrix $\boldsymbol{\Sigma}$, namely

$$\mathbb{E}[\boldsymbol{r}_i \boldsymbol{r}_i^{\dagger}] = \boldsymbol{\Sigma} \succ \boldsymbol{0}, \quad i = 1, \ldots, K.$$

Resorting to a matrix representation, the hypothesis test (3.1) can be equivalently written in a more compact form as

$$\begin{cases} H_0 : \ \boldsymbol{R} = \boldsymbol{N} \\ H_1 : \ \boldsymbol{R} = \boldsymbol{N} + \boldsymbol{p}\boldsymbol{\alpha}^{\dagger} \end{cases}, \tag{3.2}$$

where $\boldsymbol{R} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_K] \in \mathbb{C}^{N,K}$, $\boldsymbol{N} = [\boldsymbol{n}_1, \ldots, \boldsymbol{n}_K] \in \mathbb{C}^{N,K}$, and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^{\dagger} \in \mathbb{C}^K$.

According to the Neyman-Pearson criterion, the optimum solution to the above hypothesis testing problem is the Likelihood Ratio Test (LRT). However, for the case under consideration, this procedure does not lead to a Uniformly Most Powerful (UMP) test and the resulting detector requires the knowledge of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$, which reasonably are assumed to be unknown.

A possible way to cope with the aforementioned a-priori uncertainty is to resort to adaptive design procedures where the unknown parameters appearing in nominal decision statistics are replaced by their ML estimates. The most well known example of such a design framework is the GLRT which is tantamount to substituting the unknown parameters, appearing in the LRT, with their ML estimates under each hypothesis [90].

Subsequent developments require specifying the complex multivariate pdf of the observable matrix $\boldsymbol{R}$ under both the hypotheses. Previous assumptions imply that

$$f_{\boldsymbol{R}}(\boldsymbol{R}|\boldsymbol{X}, H_0) = \frac{[\det(\boldsymbol{X})]^K}{\pi^{NK}} \exp\left(-\mathrm{tr}\left\{\boldsymbol{X}\boldsymbol{R}\boldsymbol{R}^\dagger\right\}\right), \tag{3.3}$$

and

$$\begin{aligned} f_{\boldsymbol{R}}(\boldsymbol{R}|\boldsymbol{X}, H_1) = & \frac{[\det(\boldsymbol{X})]^K}{\pi^{NK}} \cdot \\ & \exp\left(-\mathrm{tr}\left\{\boldsymbol{X}\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)^\dagger\right\}\right), \end{aligned} \tag{3.4}$$

where $\boldsymbol{X} = \boldsymbol{\Sigma}^{-1} \succ \boldsymbol{0}$ denotes the inverse covariance matrix of the interference plus noise. Notice that, since it is not assumed the availability of secondary data free of useful signal, the likelihood function is unbounded under the hypothesis H_1, [41], namely the ML estimates of the unknown parameters do not exist. In order to overcome this drawback, in the following, a constrained ML estimation is considered; precisely, it is assumed that $\boldsymbol{X}$ belongs to the uncertainty set Ω defined as

$$\Omega = \{\boldsymbol{X} \succ \boldsymbol{0} : \ f_i(\boldsymbol{A}\boldsymbol{X}\boldsymbol{A}) \leq 0, \quad i = 1, \ldots, M\}, \tag{3.5}$$

where $\boldsymbol{A} \succ \boldsymbol{0}$ and each function $f_i(\boldsymbol{B})$, $i = 1, \ldots M$, applied on a positive definite matrix $\boldsymbol{B}$, is a unitary invariant continuous function, i.e. it is a continuous function depending only on the eigenvalues of $\boldsymbol{B}$. This is equivalent to exploiting some a-priori knowledge about the operating environment in order to enforce the covariance matrix to belong to a suitable uncertainty set. In Section 3.2, some practical examples of uncertainty sets compliant with the general model (3.5) are described.

From the above considerations, it is clear that the key ingredient to devise the GLRT decision rule, is the capability to compute the constrained ML estimates of the unknown parameters under both the hypotheses. This is the goal of the next subsection.

3.1.1 ML Estimates of the Unknown Parameters Under H_0 and H_1

The constrained ML estimates of the unknown parameters, under the hypotheses H_0 and H_1, are, respectively, optimal solutions to the

optimization problems $\mathcal{P}_{H_0}$ and $\mathcal{P}_{H_1}$, defined as

$$\mathcal{P}_{H_k} \begin{cases} \min\limits_{\boldsymbol{\alpha},\boldsymbol{X}} & \operatorname{tr}\left\{\boldsymbol{X}\frac{1}{K}\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)^\dagger\right\} - \log\det\left(\boldsymbol{X}\right) \\ \text{s.t.} & f_i(\boldsymbol{A}\boldsymbol{X}\boldsymbol{A}) \leq 0, \quad i=1,\ldots,M \\ & \boldsymbol{X} \succ \boldsymbol{0} \\ & \boldsymbol{\alpha} \in \Theta_{H_k} \end{cases}, \quad k=0,1, \tag{3.6}$$

where $\Theta_{H_0} = \{\boldsymbol{0}\}$ and $\Theta_{H_1} = \mathbb{C}^K$.

Let us focus on $\mathcal{P}_{H_0}$, denote by $\boldsymbol{S}_0 = \frac{1}{K}\boldsymbol{R}\boldsymbol{R}^\dagger$ the sample covariance matrix of the received data, and introduce the eigenvalue decomposition of $\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}$

$$\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1} = \boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{d}_0\right)\boldsymbol{V}_0^\dagger, \tag{3.7}$$

where $\boldsymbol{V}_0$ is a unitary matrix containing the eigenvectors of $\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}$, and $\boldsymbol{d}_0 = \boldsymbol{\lambda}(\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}) \in \mathbb{R}^N$, with $d_{0N} \geq 0$. The following proposition gives an optimal solution to $\mathcal{P}_{H_0}$:

Proposition 3.1.1. *Let the constraint set Ω in (3.5) be nonempty and bounded. An optimal solution $\hat{\boldsymbol{X}}_{H_0}$ to $\mathcal{P}_{H_0}$ is given by*

$$\hat{\boldsymbol{X}}_{H_0} = \boldsymbol{A}^{-1}\boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{\lambda}_0^\star\right)\boldsymbol{V}_0^\dagger\boldsymbol{A}^{-1}, \tag{3.8}$$

where $\boldsymbol{\lambda}_0^\star$ is an optimal solution to the optimization problem

$$\mathcal{P}(\boldsymbol{d}_0) \begin{cases} \min\limits_{\boldsymbol{\lambda}} & \boldsymbol{d}_0^\dagger\boldsymbol{\lambda} - \sum\limits_{i=1}^{N}\log\lambda_i \\ s.t. & \overline{f_i}(\boldsymbol{\lambda}) \leq 0, \quad i=1,\ldots,M \\ & \lambda_i > 0, \quad i=1,\ldots,N \end{cases}, \tag{3.9}$$

with $\overline{f_i}(\boldsymbol{\lambda}) = f_i(\mathbf{diag}\left(\boldsymbol{\lambda}\right))$. Additionally, if the functions $\overline{f_i}(\boldsymbol{\lambda})$, $i = 1,\ldots,M$ are convex, the optimal solution to $\mathcal{P}(\boldsymbol{d}_0)$ is unique.

Proof. See Appendix K. □

Notice that the ML estimate of $\boldsymbol{X}$ under H_0, given by (3.8), defines a generalized shrinkage covariance matrix estimator; indeed, when the matrix $\boldsymbol{A}$ is proportional to the identity matrix then $\hat{\boldsymbol{X}}_{H_0}$ is exactly a

shrinkage estimator [48], otherwise it is obtained properly transforming a shrinkage estimator of a pre-processed sample covariance matrix.

Let us now consider the optimization problem $\mathcal{P}_{H_1}$; let us denote by

$$\boldsymbol{p}_1 = \frac{\boldsymbol{A}^{-1}\boldsymbol{p}}{\|\boldsymbol{A}^{-1}\boldsymbol{p}\|}, \tag{3.10}$$

the transformed useful signal,

$$\boldsymbol{S}_1 = \frac{1}{K}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\right)\boldsymbol{A}^{-1}\boldsymbol{R}\boldsymbol{R}^{\dagger}\boldsymbol{A}^{-1}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\right), \tag{3.11}$$

the sample covariance matrix of the transformed observations projected in the null space of $\boldsymbol{p}_1$, and introduce the eigenvalue decomposition of $\boldsymbol{S}_1$

$$\boldsymbol{S}_1 = \boldsymbol{V}_1\mathbf{diag}\left(\boldsymbol{d}_1\right)\boldsymbol{V}_1^{\dagger}$$

where $\boldsymbol{V}_1$ is a unitary matrix containing the eigenvectors of $\boldsymbol{S}_1$, and $\boldsymbol{d}_1 = \boldsymbol{\lambda}(\boldsymbol{S}_1) \in \mathbb{R}^N$, with $d_{1N} = 0$. The following proposition gives an optimal solution to $\mathcal{P}_{H_1}$:

Proposition 3.1.2. *Let the constraint set Ω in (3.5) be nonempty and bounded. An optimal solution $\left(\hat{\boldsymbol{X}}_{H_1}, \hat{\boldsymbol{\alpha}}_{H_1}\right)$ to $\mathcal{P}_{H_1}$ is given by*

$$\hat{\boldsymbol{\alpha}}_{H_1} = \frac{\boldsymbol{R}^{\dagger}\boldsymbol{A}^{-1}\boldsymbol{p}_1}{\|\boldsymbol{A}^{-1}\boldsymbol{p}\|} = \frac{\boldsymbol{R}^{\dagger}\hat{\boldsymbol{X}}_{H_1}\boldsymbol{p}}{\boldsymbol{p}^{\dagger}\hat{\boldsymbol{X}}_{H_1}\boldsymbol{p}}, \tag{3.12}$$

$$\hat{\boldsymbol{X}}_{H_1} = \boldsymbol{A}^{-1}\boldsymbol{V}_1\mathbf{diag}\left(\boldsymbol{\lambda}_1^{\star}\right)\boldsymbol{V}_1^{\dagger}\boldsymbol{A}^{-1}, \tag{3.13}$$

where $\boldsymbol{\lambda}_1^{\star}$ is an optimal solution to the optimization problem

$$\mathcal{P}(\boldsymbol{d}_1)\begin{cases} \min\limits_{\boldsymbol{\lambda}} & \boldsymbol{d}_1^{\dagger}\boldsymbol{\lambda} - \sum\limits_{i=1}^{N}\log\lambda_i \\ s.t. & \overline{f_i}(\boldsymbol{\lambda}) \leq 0, \quad i = 1,\ldots,M \\ & \lambda_i > 0, \quad i = 1,\ldots,N \end{cases}, \tag{3.14}$$

with $\overline{f_i}(\boldsymbol{\lambda}) = f_i(\mathbf{diag}\left(\boldsymbol{\lambda}\right))$. Additionally, if the functions $\overline{f_i}(\boldsymbol{\lambda})$, $i = 1,\ldots,M$, are convex, the optimal solution to $\mathcal{P}(\boldsymbol{d}_1)$ is unique.

Proof. See Appendix L. □

Let us observe that the role of the transformed sample covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}$ in the ML estimate of $\boldsymbol{X}$ under H_0 (given in (3.8)) is now played by $\left(\boldsymbol{I}-\boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\right)\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\left(\boldsymbol{I}-\boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\right)$. This means that $\hat{\boldsymbol{X}}_{H_1}$ gives up exploiting the information provided by the observed data along the direction of the useful signal $\boldsymbol{p}_1$, since this direction could be contaminated by the useful target presence.

As to the computational complexity of the devised estimators, it involves the eigenvalue decomposition of both the transformed sample covariance matrix $\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}$ as well as its modified version $\boldsymbol{S}_1$ based on the data projected in the null space of the transformed useful signal. Additionally, optimal solutions to the optimization problems $\mathcal{P}(\boldsymbol{d}_0)$ and $\mathcal{P}(\boldsymbol{d}_1)$ have to be evaluated. If the functions $\overline{f_i}(\boldsymbol{\lambda})$, $i=1,\ldots,M$, are convex, both $\mathcal{P}(\boldsymbol{d}_0)$ and $\mathcal{P}(\boldsymbol{d}_1)$ admit a unique solution which can be computed in the worst case with a polynomial complexity in the number of channels N. As it will be shown in Section 3.2, many covariance matrix uncertainty sets of practical interest can be recast in terms of convex functions $\overline{f_i}(\boldsymbol{\lambda})$, $i=1,\ldots,M$, for which ad-hoc solution algorithms can be devised. Finally, there are also situations where the inverse covariance matrix uncertainty set is non-convex, and the associated functions $\overline{f_i}(\boldsymbol{\lambda})$, $i=1,\ldots,M$ are non-convex. However, often computationally efficient algorithms to solve the associated eigenvalue optimization problems still exist (see for instance Subsection 3.2). In Fig. 3.1 a block diagram of the whole ML estimation process is shown. Therein, the block "regularization process" accounts for the constrained ML estimation of the covariance eigenvalues, namely the solution to problems $\mathcal{P}(\boldsymbol{d}_0)$ and $\mathcal{P}(\boldsymbol{d}_1)$.

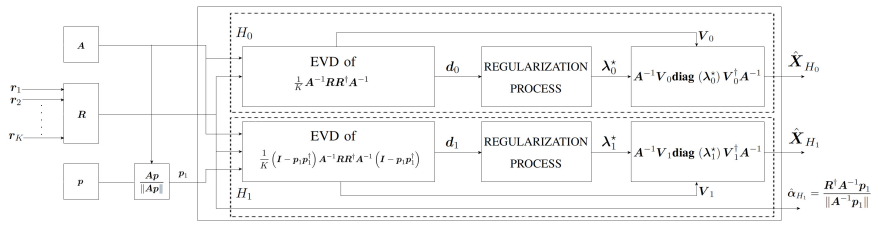


Figure 3.1: Block scheme of the ML estimation process.

Based on the derived ML estimates (Proposition 3.1.1 and Proposi-

tion 3.1.2) in the following subsection the GLRT detector is designed, for problem (3.2).

3.1.2 GLRT Based Detector

The GLRT detector for the hypothesis test (3.2) replaces the unknown parameters in the LRT with their ML estimates, namely it is the following decision rule

$$\frac{\max\limits_{\boldsymbol{\alpha},\, \boldsymbol{X} \in \Omega} [\det(\boldsymbol{X})]^K \exp\left(-\mathrm{tr}\left\{\boldsymbol{X}\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)\left(\boldsymbol{R}-\boldsymbol{p}\boldsymbol{\alpha}^\dagger\right)^\dagger\right\}\right)}{\max\limits_{\boldsymbol{X} \in \Omega} [\det(\boldsymbol{X})]^K \exp\left(-\mathrm{tr}\left\{\boldsymbol{X}\boldsymbol{R}\boldsymbol{R}^\dagger\right\}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{3.15}$$

where η is the detection threshold. Based on Proposition 3.1.1 and Proposition 3.1.2, $\hat{\boldsymbol{X}}_{H_0}$ shares the same eigenvectors as $\boldsymbol{R}\boldsymbol{R}^\dagger$ and $\hat{\boldsymbol{X}}_{H_1}$ shares the same eigenvectors as $(\boldsymbol{R}-\boldsymbol{p}\hat{\boldsymbol{\alpha}}_{H_1}^\dagger)(\boldsymbol{R}-\boldsymbol{p}\hat{\boldsymbol{\alpha}}_{H_1}^\dagger)^\dagger$; as a consequence, the GLRT in (3.15) can be recast as

$$\sum_{i=1}^{N}\left\{\log\left(\frac{\lambda_{1i}^\star}{\lambda_{0i}^\star}\right)+d_{0i}\lambda_{0i}^\star-d_{1i}\lambda_{1i}^\star\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{3.16}$$

where $\boldsymbol{\lambda}_0^\star$ and $\boldsymbol{\lambda}_1^\star$ are respectively optimal solutions to $\mathcal{P}(\boldsymbol{d}_0)$ and $\mathcal{P}(\boldsymbol{d}_1)$, while η is a suitable modification of the original threshold in (3.15).

3.2 Uncertainty Sets Defined Through Unitary Invariant Functions

In this section, some practical examples of covariance matrix uncertainty sets compliant with the general model (3.5) are given. Additionally, ad-hoc solution algorithms, with low computational complexity, for the optimization problem $\mathcal{P}(\boldsymbol{d})$, with $d_i \geq 0$, $i = 1, \ldots, N$, are presented.

Structured Covariance Matrix with a Lower Bound on the White Disturbance Power Level

The constraint set

$$\begin{cases} \boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I} + \boldsymbol{R}_e \\ \boldsymbol{R}_e \succeq \boldsymbol{0} \\ \sigma_n^2 \geq \sigma^2 \end{cases}, \tag{3.17}$$

where $\boldsymbol{R}_e$ accounts for colored interference and clutter, σ_n^2 is the power of the white disturbance term, and $\sigma^2 > 0$ is a lower bound on the white disturbance power, is studied in [47], [83]. Moreover, the corresponding constrained GLRT detector is derived in [43].

It is not difficult to show that the set (3.17), can be equivalently recast as

$$\begin{cases} \lambda_i(\boldsymbol{X}) \leq \frac{1}{\sigma^2}, \quad i = 1, \ldots, N \\ \boldsymbol{X} \succ \boldsymbol{0} \end{cases}. \tag{3.18}$$

Since each function $\lambda_i(\boldsymbol{X})$ is unitary invariant and continuous [75, Appendix D], it is possible to resort to Proposition 3.1.1 and Proposition 3.1.2 to obtain the required constrained ML estimates. Furthermore, the optimal solution to problem $\mathcal{P}(\boldsymbol{d})$, associated with the constraint set (3.18), is given by [47]

$$\lambda_i^\star = \min\left(\frac{1}{d_i}, \frac{1}{\sigma^2}\right), \quad i = 1, \ldots, N. \tag{3.19}$$

Structured Covariance Matrix with a Condition Number Constraint

The constraint set

$$\begin{cases} \boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I} + \boldsymbol{R}_e \\ \boldsymbol{R}_e \succeq \boldsymbol{0} \\ \sigma_n^2 \geq \sigma^2 \\ \frac{\lambda_1(\boldsymbol{\Sigma})}{\lambda_N(\boldsymbol{\Sigma})} \leq K_{max} \end{cases}, \tag{3.20}$$

where $\boldsymbol{R}_e$, σ_n^2, and σ^2 are defined as in (3.17), whereas K_{max} is an upper bound to the condition number, is considered in Chapter 1. Therein, it is shown that the constraint set (3.20) can be equivalently expressed as

$$\begin{cases} \lambda_i(\boldsymbol{X}) \leq \frac{1}{\sigma^2}, \quad i = 1, \ldots, N \\ \lambda_1(\boldsymbol{X}) - K_{max}\lambda_N(\boldsymbol{X}) \leq 0 \\ \boldsymbol{X} \succ \boldsymbol{0} \end{cases}. \tag{3.21}$$

Since each function $\lambda_i(\boldsymbol{X})$ is unitary invariant and continuous, the proposed theoretical framework can be applied to obtain the required constrained ML estimates. Additionally, the optimal solution to the corresponding problem $\mathcal{P}(\boldsymbol{d})$ is given by

$$\lambda_i^\star = \min\left(\min\left(K_{max}u^\star, \frac{1}{\sigma^2}\right), \max\left(u^\star, \frac{1}{d_i}\right)\right), \quad i = 1, \ldots, N,$$

where $u^\star$ is the optimal solution to the optimization problem,

$$\begin{cases} \min\limits_{u} & \sum\limits_{i=1}^{N} G_i(u) \\ \text{s.t.} & 0 < u \leq 1 \end{cases}, \tag{3.22}$$

with $G_i(u) = d_i\lambda_i^\star(u) - \log\lambda_i^\star(u)$, $i = 1, \ldots, N$. As shown in Chapter 1, $u^\star$ can be obtained with a linear complexity. Notice that, the covariance matrix uncertainty set accounting only for a condition number upper bound [48] can be obtained from (3.20) setting $\sigma^2 = 0$.

Structured Covariance Matrix with a Disturbance Power Constraint

The constraint set

$$\begin{cases} \boldsymbol{\Sigma} = \sigma_n^2\boldsymbol{I} + \boldsymbol{R}_e \\ \boldsymbol{R}_e \succeq \boldsymbol{0} \\ \sigma_n^2 \geq \sigma^2 \\ \text{tr}\{\boldsymbol{\Sigma}\} \leq P \end{cases}, \tag{3.23}$$

is a suitable model to account both for a structured covariance matrix ($\boldsymbol{R}_e$, σ_n^2, and σ^2 are defined as in (3.17)) and an upper bound on the whole disturbance power[1]. It could be important to highlight that a prediction of the parameter P can be obtained in a cognitive way jointly exploiting clutter maps, digital terrain maps (for instance the NLCD [76]), and ESM system.

It is not difficult to show that the constraint set (3.23) is equivalent to

$$\begin{cases} \sum\limits_{i=1}^{N} \dfrac{1}{\lambda_i(\boldsymbol{X})} \leq P \\ \lambda_i(\boldsymbol{X}) \leq \frac{1}{\sigma^2}, \quad i = 1, \ldots, N \\ \boldsymbol{X} \succ \boldsymbol{0} \end{cases}. \tag{3.24}$$

[1]Notice that, the set (3.23) is nonempty if and only if $\sigma^2 N \leq P$.

Since each function $\lambda_i(\boldsymbol{X})$ is unitary invariant and continuous, Proposition 3.1.1 and Proposition 3.1.2 can be exploited to compute the required constrained ML estimates. Besides, observing that the constraint set (3.24) can be recast as

$$\begin{cases} \sum_{i=1}^{N} a_i \leq P \\ s.t. \begin{pmatrix} a_i & 1 \\ 1 & \lambda_i(\boldsymbol{X}) \end{pmatrix} \succeq \mathbf{0}, \quad i = 1, \ldots, N \\ a_i > 0, \quad i = 1, \ldots, N \\ \lambda_i(\boldsymbol{X}) \leq \frac{1}{\sigma^2}, \quad i = 1, \ldots, N \\ \boldsymbol{X} \succeq \mathbf{0} \end{cases}, \tag{3.25}$$

the corresponding problem $\mathcal{P}(\boldsymbol{d})$ falls in the class of MAXDET convex optimization problems [91], which can be efficiently solved in polynomial time using interior point methods.

Structured Covariance Matrix with a Rank Constraint

The constraint set

$$\begin{cases} \boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I} + \boldsymbol{R}_e \\ \boldsymbol{R}_e \succeq \mathbf{0} \\ \text{Rank}\,(\boldsymbol{R}_e) \leq r \end{cases}, \tag{3.26}$$

where $\boldsymbol{R}_e$ and σ_n^2 are defined as in (3.17), whereas r is the maximum rank allowed to the matrix $\boldsymbol{R}_e$, is considered in [89]. The set in (3.26) can be recast as

$$\begin{cases} \lambda_i(\boldsymbol{X}) = \frac{1}{\sigma_n^2}, \quad i = 1, \ldots, N - r \\ \lambda_i(\boldsymbol{X}) \leq \frac{1}{\sigma_n^2}, \quad i = N - r + 1, \ldots, N \\ \boldsymbol{X} \succ \mathbf{0} \end{cases}. \tag{3.27}$$

Since the functions $\lambda_i(\boldsymbol{X})$ are unitary invariant and continuous, the proposed framework can be exploited to evaluate the required constrained ML estimates. In this case $\mathcal{P}(\boldsymbol{d})$ is a non-convex optimization problem. However, it can be efficiently solved rearranging the elements of the vector $\boldsymbol{d}$ in increasing order, obtaining

$$\lambda_i^\star = \begin{cases} \frac{1}{\sigma_n^2} & i = 1, \ldots, N - r \\ \min\left(\frac{1}{d_i}, \frac{1}{\sigma_n^2}\right) & i = N - r + 1, \ldots, N \end{cases}. \tag{3.28}$$

Similarity Constraint with a Knowledge-Based Prior Covariance Estimate

To account for some a-priori knowledge on the radar interference scenario, a suitable constraint set is given by

$$\begin{cases} \|\boldsymbol{\Sigma}_0^{\frac{1}{2}}\boldsymbol{X}\boldsymbol{\Sigma}_0^{\frac{1}{2}} - \boldsymbol{I}\|_2 \leq \epsilon \\ \boldsymbol{X} \succ \boldsymbol{0} \end{cases}, \tag{3.29}$$

where $\boldsymbol{\Sigma}_0$ is a knowledge-based prior covariance estimate, assumed positive definite, and $\epsilon > 0$ is the parameter ruling the degree of similarity. Otherwise stated, ϵ allows to control the confidence level (reliability) associated to the prior covariance matrix estimate $\boldsymbol{\Sigma}_0$. As to the prior model $\boldsymbol{\Sigma}_0$, it can be derived from some physical scattering models of the terrain and/or of the environment [11]. Additionally, recently proposed knowledge-based covariance models, [92], [68], could be exploited. Now, introducing the matrix $\boldsymbol{A} = \boldsymbol{\Sigma}_0^{\frac{1}{2}}$, the uncertainty set (3.29) can be recast, in terms of the transformed inverse covariance matrix $\boldsymbol{X}_1 = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}$, as

$$\begin{cases} \|\boldsymbol{X}_1 - \boldsymbol{I}\|_2 \leq \epsilon \\ \boldsymbol{X}_1 \succ \boldsymbol{0} \end{cases}. \tag{3.30}$$

It is not difficult to show that (3.31) can be equivalently expressed as

$$\begin{cases} \lambda_1(\boldsymbol{X}_1) \leq 1 + \epsilon \\ \lambda_N(\boldsymbol{X}_1) \geq 1 - \epsilon \\ \boldsymbol{X}_1 \succ \boldsymbol{0} \end{cases}. \tag{3.31}$$

Hence, exploiting the unitary invariance and continuity of the functions $\lambda_1(\boldsymbol{X}_1)$ and $\lambda_N(\boldsymbol{X}_1)$, the results of Proposition 3.1.1 and Proposition 3.1.2 can be utilized to compute the required constrained ML estimates. Finally, the optimal solution to problem $\mathcal{P}(\boldsymbol{d})$, associated with the constraint set (3.31), is given by

$$\lambda_i^\star = \min\left(1+\epsilon, \max\left(1-\epsilon, \frac{1}{d_i}\right)\right), \quad i = 1, \ldots, N.$$

3.3 Performance Analysis

In this section, the performance analysis of the devised GLRT is dealt, focusing on some covariance matrix uncertainty sets described

in Subsection 3.1.2. Specifically, for a specific covariance model, the performance of the synthesized GLRT detector is assessed in terms of P_D for a given Probability of False Alarm (P_{FA}) level. To this end, the detection threshold is set resorting to Monte Carlo simulations based on $100/P_{FA}$ independent trials. Moreover, to limit the computational effort, the nominal P_{FA} is fixed to 10^{-4}.

3.3.1 Spatial Processing in the Presence of Jamming and White Interference

A radar system equipped with a uniform linear array antenna of $N = 20$ elements is considered, with a spacing between them equal to one half the radar operating wavelength, λ_0, whose main beam points in the boresight direction. The overall jammer plus thermal noise covariance matrix is $\boldsymbol{\Sigma} = \boldsymbol{R}_J + \sigma_a^2 \boldsymbol{I}$, where σ_a^2 denotes the actual power level of the white disturbance term, whereas $\boldsymbol{R}_J$ is the covariance matrix associated to J jammers, [41], [43], [47], [7],

$$\boldsymbol{R}_J(n,m) = \frac{1}{N}\sum_{i=1}^{J} \sigma_i^2 \mathrm{sinc}\left[\frac{1}{2}B_f(n-m)\phi_i\right] e^{j(n-m)\phi_i}, \tag{3.32}$$

with $B_f = B/f_0$ the fractional bandwidth, B the instantaneous bandwidth of the desired signal (coinciding with the jammer's bandwidth), σ_i^2 the power of the i-th jammer, and ϕ_i the jammer phase angle with respect to the antenna phase center (namely, $\phi = 2\pi d(\sin\theta)/\lambda_0$, and θ the angle off-boresight of the jammer).

The SINR is defined as

$$\mathrm{SINR} = \sum_{i=1}^{K} |\alpha_i|^2 \boldsymbol{p}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{p}, \tag{3.33}$$

where $\boldsymbol{p} = (1/\sqrt{N})[1, 1, \ldots, 1]^T$ is the target steering vector. The actual target[2] is composed of $N_s = 2$ equal power dominant scatterers, i.e. $|\alpha_1| = |\alpha_2|$ and $\alpha_i = 0$, $i = 3, \ldots, K$.

[2]It is worth noting that the proposed detector does not suffer for the so called "collapsing loss", i.e. its detection performance depends on $\boldsymbol{\alpha}$ only through $\|\boldsymbol{\alpha}\|$. In fact, the decision statistic (3.16) depends on received data $\boldsymbol{r}_i$, $i = 1, \ldots, K$, only via the sample correlation matrix $\boldsymbol{R}\boldsymbol{R}^\dagger$, whose pdf functionally depends on $\boldsymbol{\alpha}$ only via $\|\boldsymbol{\alpha}\|$. Specifically, under H_1, $\boldsymbol{R}\boldsymbol{R}^\dagger$ is distributed as a Noncentral Wishart random matrix with non central parameter $\boldsymbol{p}\boldsymbol{\alpha}^\dagger$ and matrix parameter $\boldsymbol{\Sigma}$ [93, equation 67].

With reference to the interfering scenario, the presence of 3 narrowband jammers (i.e. $B_f = 0$ for each jammer), with powers $\sigma_i^2 = 30$ dB, $i = 1, \ldots, 3$, and phase angles $\phi_1 = 20°$, $\phi_2 = 40°$, and $\phi_3 = 60°$, is assumed. Moreover, the knowledge of a lower bound on the white disturbance term is considered, namely it is set $\sigma_0^2 = 0$ dB (see Table 3.1).

Table 3.1: Angular Processing. Simulation parameters.

N_s	J	σ_i^2 [dB]	ϕ_i	σ_0^2 [dB]	N
2	3	30	[20, 40, 60]	0	20

As to the covariance matrix uncertainty model, the constraint sets described in Subsections 3.2 and 3.2 are considered. For comparison purposes, in these simulations, the MGLRT detector[3] has been also considered [41], defined as

$$\frac{\prod_{i=1}^{N}\lambda_i\left[\frac{1}{K}\boldsymbol{R}\boldsymbol{R}^{\dagger}\right]}{\prod_{i=1}^{N-1}\lambda_i\left[\frac{1}{K}\left(\boldsymbol{I}-\frac{\boldsymbol{p}\boldsymbol{p}^{\dagger}}{||\boldsymbol{p}||^2}\right)\boldsymbol{R}\boldsymbol{R}^{\dagger}\left(\boldsymbol{I}-\frac{\boldsymbol{p}\boldsymbol{p}^{\dagger}}{||\boldsymbol{p}||^2}\right)\right]} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_M, \tag{3.34}$$

where η_M is the detection threshold. Notice that, the MGLRT is not defined when the number of data K is less than the number of channels N. Finally, as benchmark, the optimal receiver is considered, which supposes the perfect covariance matrix knowledge, i.e.

$$\sum_{i=1}^{K}\frac{|\boldsymbol{r}_i^{\dagger}\boldsymbol{\Sigma}^{-1}\boldsymbol{p}|^2}{\boldsymbol{p}^{\dagger}\boldsymbol{\Sigma}^{-1}\boldsymbol{p}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0. \tag{3.35}$$

where η_0 denotes the detection threshold.

Structured Covariance Matrix with a Condition Number Constraint

In Fig. 3.2, the P_D is plotted versus SINR for the GLRT detector compliant with the constraint set of Subsection 3.2. The subplots (a)

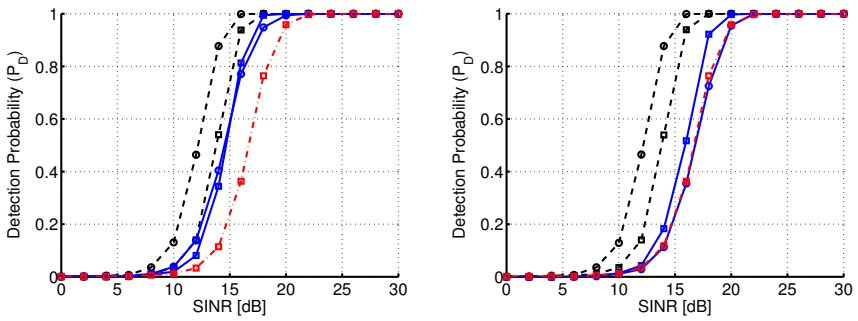
[3]The MGLRT is derived resorting to the *method of sieves*, namely restricting the parameter space $(\boldsymbol{\Sigma}, \boldsymbol{\alpha})$ to a subspace such that the ML estimate of the parameters exists and is unique; see [41] for more details.

and (b) are obtained setting K_{max} equal to the condition number of the disturbance covariance matrix $\mathbf{\Sigma}$. Conversely, the subplots (c) and (d) are related to a different value of the parameter K_{max}, i.e. it is equal to twice the condition number of $\mathbf{\Sigma}$. To further investigate the behavior of the considered GLRT-based detector, two different situations are presented: the former refers to $\sigma_a^2 = 0$ dB, namely there is no mismatch between the actual white noise power level and its known lower bound (subplots (a) and (c)); the latter considers $\sigma_a^2 = 10$ dB (subplots (b) and (d)). By doing so, it is possible to study the effect of a possible mismatch between the nominal value σ_0^2 and the actual value σ_a^2 on the performance of the synthesized receiver. Finally, both $K = 10$ and $K = 30$ are considered.

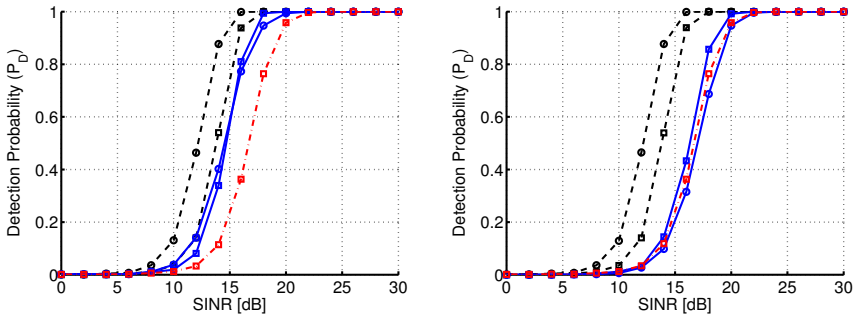
The curves highlight that the new GLRT outperforms the MGLRT, when they exploit the same number of data, i.e. $K = 30$. This behavior can be justified observing that the MGLRT does not suppose any a-priori structure for the interference covariance. Additionally, increasing K the optimal detector experiences a performance degradation; this is not surprising since the integrated amount of noise is an increasing function of K whereas the useful energy does not change. As to the GLRT detector, it is not possible to predict the impact of K on its performance. In fact, K has a double effect: from one side it increases the amount of integrated noise; on the other side it allows for a better covariance matrix estimation. For the analyzed scenarios, it can be observed that the GLRT synthesized for $K = 30$ outperforms the GLRT devised for $K = 10$, when the SINR is high enough. Thus, it could be argued that the estimation accuracy improvement, due to the larger number of available cells, dominates the deleterious effect due to the increased amount of integrated noise. Finally, studying P_D versus the K_{max} value, it can be observed that the better the covariance uncertainty characterization, the better the detection performance.

Structured Covariance Matrix with a Rank Constraint

In Fig. 3.3, P_D is plotted versus SINR for the GLRT detector compliant with the constraint set introduced in Subsection 3.2, in correspondence of the operating environment described in Table 3.1. Subplot (a) is obtained setting the parameter r equal to the rank of the covariance matrix $\boldsymbol{R}_J$ in equation (3.32). Subplot (b), refers to a parameter r equal to twice the rank of the covariance matrix $\boldsymbol{R}_J$. Again, both $K = 10$ and



(a) $\sigma_a^2 = 0$ dB. (b) $\sigma_a^2 = 10$ dB.

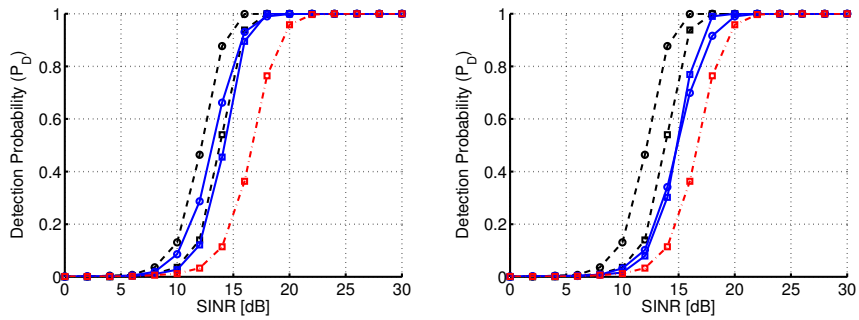


(c) $\sigma_a^2 = 0$ dB. (d) $\sigma_a^2 = 10$ dB.

Figure 3.2: P_D versus SINR for the optimum receiver (dashed curves), MGLRT (dot-dashed curves), and the proposed GLRT based on a structured covariance matrix with a condition number constraint (solid curves). □-marked curves refer to $K = 30$, whereas ∘-marked curves refer to $K = 10$. Subplots (a) and (c) refer to a matched scenario, i.e. $\sigma_a^2 = 0$ dB, whereas subplots (b) and (d) refer to a mismatched scenario, i.e. $\sigma_a^2 = 10$ dB. Finally, subplots (a) and (b) assume $K_{max} = \lambda_{max}(\mathbf{\Sigma})/\lambda_{min}(\mathbf{\Sigma})$, whereas subplots (c) and (d) assume $K_{max} = 2\lambda_{max}(\mathbf{\Sigma})/\lambda_{min}(\mathbf{\Sigma})$.

$K = 30$ are considered.

As in Fig. 3.2, the curves highlight that the proposed GLRT detector outperforms the MGLRT. This is a confirmation that a correct use of the a-priori information can produce performance improvements with respect to the unstructured case (MGLRT). Again, the available number of cells trades off the covariance estimation accuracy with the overall integrated noise. For the analyzed scenario, the GLRT devised for $K = 10$ achieves a better detection performance than the one devised for



(a) $r = 3$ (b) $r = 6$

Figure 3.3: P_D versus SINR for the optimum receiver (dashed curves), MGLRT (dot-dashed curves), and the proposed GLRT based on a structured covariance matrix with a rank constraint (solid curves). □-marked curves refer to $K = 30$, whereas ○-marked curves refer to $K = 10$. Finally, subplot (a) assumes $r = 3$, whereas subplot (b) assumes $r = 6$.

$K = 30$. However, this improvement decreases as the rank constraint parameter changes from $r = 3$ to $r = 6$.

Finally, comparing the curves of subplots (a) and (b), it can be seen how a more accurate covariance model can lead to better performance levels.

3.3.2 Doppler Processing in the Presence of Bimodal Clutter Plus White Noise

In this section, a monostatic radar system equipped with an antenna transmitting a coherent burst of N pulses is considered. The focus is on a bimodal clutter model which accounts for the presence of statistically independent ground and sea clutters as well as white noise. Moreover, a Gaussian shaped Power Spectral Density (PSD) [80] for both the interfering sources is assumed; consequently, the (i,k)-th element of the overall disturbance covariance matrix is modeled as, [80], [87],

$$\mathbf{\Sigma}(i,k) = \mathrm{CNR}_S \rho_S^{(i-k)^2} \exp\left[-j2\pi(i-k)f_S\right] + \mathrm{CNR}_G \rho_G^{(i-k)^2} + \delta_{i,k}\,, \tag{3.36}$$

where CNR_S and CNR_G denote respectively the Clutter to Noise power Ratio for the sea and the ground clutter, whereas ρ_S and ρ_G are respec-

tively the one-lag correlation coefficient for the sea and the ground clutter. Finally, f_S is the normalized Doppler frequency of the sea clutter, and $\delta_{i,k}$ is the Kronecker delta function. In Table 3.2, all the parameters considered in the simulation setup are summarized.

Table 3.2: Doppler processing. Simulation parameters.

	$\mathbf{\Sigma}_0$	$\mathbf{\Sigma}$ (case 1)	$\mathbf{\Sigma}$ (case 2)
CNR_S [dB]	10	10	10
CNR_G [dB]	30	30	30
ρ_S	0.5	$0.5 + \mathcal{U}[-0.05, 0.05]$	$0.5 + \mathcal{U}[-0.1, 0.1]$
ρ_G	0.99	0.99	0.99
f_S	0.2	$0.2 + \mathcal{U}[-0.025, 0.025]$	$0.2 + \mathcal{U}[-0.05, 0.05]$
N	20	20	20

As to the knowledge-based covariance matrix $\mathbf{\Sigma}_0$, it is assumed that it is described by model (3.36) with the parameter values reported in Table 3.2. The actual covariance matrix $\mathbf{\Sigma}$ still complies with model (3.36) but the values of some parameters are changed randomly, from a simulation run to another, as specified in the last two lines[4] of Table 3.2. By doing so, mismatches among the a-priori covariance model $\mathbf{\Sigma}_0$ and the actual covariance matrix $\mathbf{\Sigma}$ we accounted for. Case 1 refers to a more reliable a-priori model than case 2.

The considered target is composed of $N_s = 2$ equal power dominant scatterers, i.e. $|\alpha_1| = |\alpha_2|$ and $\alpha_i = 0$, $i = 3, \ldots, K$, sharing a temporal steering vector given by

$$\boldsymbol{p} = \frac{1}{\sqrt{N}} \left[1, \exp(j2\pi f_d), \ldots, \exp(j2\pi(N-1)f_d)\right]^T,$$

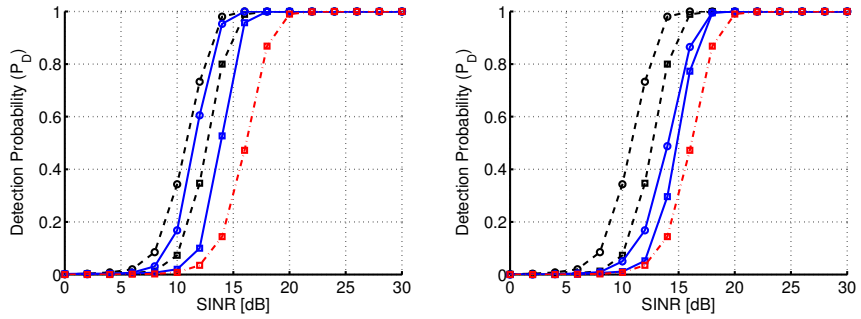
where f_d is the normalized Doppler frequency, set equal to 0.15 in the simulations. Furthermore, the following definition for the SINR is adopted

$$\text{SINR} = \sum_{i=1}^{K} |\alpha_i|^2 \boldsymbol{p}^\dagger \mathbf{\Sigma}_0^{-1} \boldsymbol{p}. \tag{3.37}$$

In Fig. 3.4, P_D is plotted versus SINR for the GLRT detector compliant with the similarity constraint of Subsection 3.2. Subplots (a) and

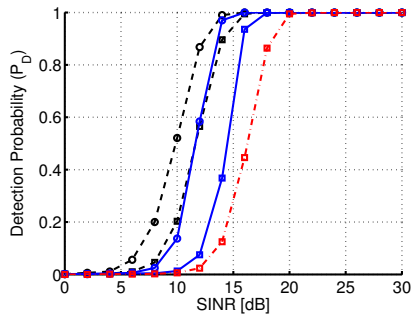
[4] $x \sim \mathcal{U}(a, b)$ denotes a uniformly distributed random variable within the interval $[a, b]$.

(b) refer to case 1 described in Table 3.2, whereas subplots (c) and (d) refer to case 2 of Table 3.2. The subplots on the left account for $\epsilon = 0.1$, while the subplots on right assume $\epsilon = 1$.

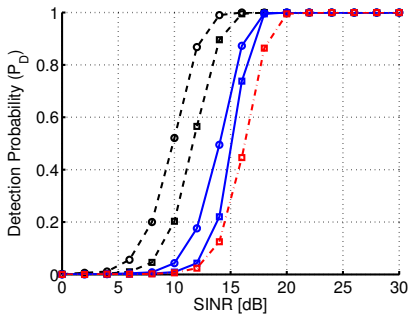


(a) $\epsilon = 0.1$.

(b) $\epsilon = 1$.



(c) $\epsilon = 0.1$.



(d) $\epsilon = 1$.

Figure 3.4: P_D versus SINR for the optimum receiver (dashed curves), MGLRT (dot-dashed curves), and the proposed GLRT based on a similarity constraint (solid curves). □-marked curves refer to $K = 30$, whereas ∘-marked curves refer to $K = 10$. Subplots (a) and (c) assume $\epsilon = 0.1$, whereas subplots (b) and (d) assume $\epsilon = 1$. Finally, subplots (a) and (b) refer to the case 1 of Table 3.2, whereas subplots (c) and (d) refer to the case 2 of Table 3.2.

As in Fig. 3.2 and Fig. 3.3, the curves show that the proposed GLRT outperforms the MGLRT. Otherwise stated, the results confirm the intuition that a proper exploitation of the a-priori information can lead to performance improvements. An interesting remark on the GLRT performance concerns the degradation experienced increasing the number of integrated cells. It can be claimed that, owing to the quite reli-

able a-priori model $\mathbf{\Sigma}_0$, even a small number of observations allows for an accurate covariance estimation; thereby, the noise integration effect dominates the performance behavior as K increases. Inspection of the curves also highlights that the detectors synthesized for $\epsilon = 0.1$ outperform those devised for $\epsilon = 1$, which is a further confirmation of the a-priori covariance model $\mathbf{\Sigma}_0$ reliability. Finally, comparing the curves of subplots (a) and (c), it can be seen how a more accurate covariance model can reduce the performance gap with respect to the one of the optimum detector. In fact, subplot (a) refers to a scenario with a much smaller clutter covariance unpredictability than in subplot (c).

Conclusions

In this thesis, the problem of covariance matrix estimation for radar signal processing applications has been addressed. In particular, both a statistical and a geometric approach have been followed, a robust adaptive radar receiver and a secondary data selection scheme have been respectively proposed. Moreover, the problem of detecting extended targets has been considered, enforcing structural constraints on the estimator of the disturbance covariance matrix.

Precisely, in Chapter 1, the focus has been on ML covariance matrix estimation. At the design stage, the covariance matrix $\mathbf{\Sigma}$ has been modeled as the sum of an unknown positive semi-definite matrix (describing the colored covariance of the jamming signals and clutter) and a (partially known) matrix proportional to the identity (accounting for the white interference). Moreover, an upper bound on the condition number of the estimator has been also enforced. The problem has been formulated in terms of a convex MAXDET optimization problem, solvable efficiently through interior point methods. Remarkably, an algorithm to find a closed form solution to the aforementioned problem has been developed. It requires the computation of the sample covariance matrix eigenvalue decomposition and the solution of a scalar convex optimization problem, whose complexity is linear with respect to the number of sample eigenvalues greater than one. As a consequence, the overall complexity of the proposed estimator is dominated by the computational effort connected with the sample covariance matrix eigenvalue decomposition. At the analysis stage, the performance of the proposed estimator has been evaluated in terms of normalized average SINR behavior. The results have highlighted the capabilities of the new technique to provide, under some circumstances, an higher normalized average SINR than the FML [47], which only accounts for the information about the structure of the covariance matrix. The new method can also outperform the al-

gorithm of [48] which does not impose any special covariance structure, enforcing only a condition number constraint.

In Chapter 2, two new classes of covariance matrix estimators, independent of the statistical characterization of the secondary data and exploiting the properties of the positive definite matrix space, have been devised. Precisely, each estimator has been designed so as to be the geometric barycenter or the geometric median, with respect to a suitable distance in the considered space, of a set of covariance matrix estimates. Exploiting the devised estimators, a training data selection scheme has been considered, which resorts to the GIP computed from the proposed covariance matrix estimators, in order to select the most homogeneous secondary data. At the analysis stage, the performance of each devised barycenter or median-based estimator has been assessed, evaluating the selection capabilities of the corresponding data selector. The obtained results have highlighted that the Log-Euclidean median-based estimator outperforms all the others, in terms of $P_{selection}$. Furthermore, considering a specific distance, the corresponding median-based estimator generally achieves a higher $P_{selection}$ than the corresponding barycenter-based one. As a consequence, the analysis has shown the presence of a trade-off between the computational burden and the selection capability. Indeed, the median-based estimators are characterized by an high computational burden, due the lack of a closed form expression, whereas the barycenter-based estimators can be evaluated with simple analytic expressions. Finally, $P_{selection}$ analysis has highlighted that the Log-Euclidean median-based and the Log-Euclidean barycenter-based estimators share almost the same performance. Consequently, accounting for both the computational effort and selection properties, a reasonable choice, from a practical point of view, is to use the Log-Euclidean barycenter-based estimator for data selection applications.

In Chapter 3, the problem of detecting targets extended across a number of adjacent range cells, embedded in Gaussian interference with an unknown but structured covariance matrix, has been considered. The possible target echo from each range bin has been modeled as a deterministic signal, with an unknown scaling factor accounting for the target response. At the design stage, some a-priori knowledge about the operating environment has been exploited to suitable constraint the covariance matrix. Specifically, it has been assumed that a properly transformed inverse disturbance covariance matrix belongs to a set described via

unitary invariant continuous functions. Hence, the constrained ML estimates of the unknown parameters have been derived and the GLRT for the hypothesis test under consideration has been designed. The proposed architecture involves the eigenvalue decomposition of both a properly transformed sample covariance matrix as well as its modified version based on the data projected in the null space of the transformed useful signal. Additionally, it requires the solution of two optimization problems, which often can be solved with a polynomial computational complexity with respect to the number of channels. At the analysis stage, the performance of the devised GLRT has been assessed, focusing on some covariance matrix uncertainty sets of practical relevance. The results have shown that the proper use of the a-priori information can lead to a detection performance quite close to the optimum receiver, which supposes the perfect knowledge of the interference plus noise covariance matrix. Finally, the developed analysis have confirmed the intuition that the better the covariance matrix uncertainty set characterization, the better the detection performance.

Concluding, if there is the availability of homogeneous data, even if limited, the approach proposed in Chapter 1 represents the best choice to consider. However, if there is some a-priori knowledge about the presence of a limited number of outliers in the data or it is not possible to find a statistical characterization of the training data, the techniques proposed in Chapter 2 have to be considered. Finally, if there is a great amount of outliers in the secondary data or the echo due to the target can no longer be considered as contained only in the cell under test, the best solution to be applied is the one studied in Chapter 3.

Possible future research tracks might concern the analysis of all the algorithms proposed in this thesis on real radar data. Moreover, as to the covariance matrix estimator proposed in Chapter 1, the possibility to exploit other condition number predictors could be taken into account. It might also be of interest to consider a minimum mean square error shrinkage design approach, as in [69], accounting for both a structural constraint and condition number constraint. Last but not least, it might be challenging to design techniques jointly estimating both the structured covariance matrix and the array manifold. As to the geometric barycenters and median matrices given in Chapter 2, it is possible to consider the use of suitable distances which explicitly try to account for other environment heterogeneities (for instance clutter power varia-

tions). However, the design of selection schemes jointly exploiting statistical and geometrical properties represents definitely an issue of special interest. Finally, as to the detection of extended targets, future research tracks might concern the inspection of other structural covariance constraints compliant with the considered model proposed in Chapter 3.

Appendix A

Proof of Proposition 1.2.1

Proof. Let $\left(\bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{R}}, \bar{\sigma}_n^2\right)$ be a feasible point for problem $\mathcal{P}$.
Since $\left(\bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{R}} + (\bar{\sigma}_n^2 - \sigma^2)\boldsymbol{I}, \sigma^2\right)$, is a feasible point for problem $\mathcal{P}$, which achieves the same objective value as $\left(\bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{R}}, \bar{\sigma}_n^2\right)$, it follows that $\mathcal{P}$ is equivalent to problem $\mathcal{P}'$

$$\mathcal{P}' \begin{cases} \underset{\boldsymbol{\Sigma}, \boldsymbol{R}}{\min} & \text{tr}\left(\boldsymbol{S}_1 \boldsymbol{\Sigma}^{-1}\right) - \log\det\left(\boldsymbol{\Sigma}^{-1}\right) \\ \text{s.t.} & \frac{\lambda_{max}(\boldsymbol{\Sigma})}{\lambda_{min}(\boldsymbol{\Sigma})} \leq K_{max} \\ & \sigma^2 \boldsymbol{I} + \boldsymbol{R} = \boldsymbol{\Sigma} \\ & \boldsymbol{R} \succeq \boldsymbol{0} \end{cases} . \tag{A.1}$$

Thus, without loss of generality, it is possible to focus on $\mathcal{P}'$. Since the constraint set

$$\begin{cases} \sigma^2 \boldsymbol{I} + \boldsymbol{R} = \boldsymbol{\Sigma} \\ \boldsymbol{R} \succeq \boldsymbol{0} \end{cases},$$

can be expressed as

$$\begin{cases} \boldsymbol{\Sigma} \succeq \sigma^2 \boldsymbol{I} \\ \boldsymbol{R} = \boldsymbol{\Sigma} - \sigma^2 \boldsymbol{I} \end{cases},$$

letting $\boldsymbol{X} = \left(\frac{\boldsymbol{\Sigma}}{\sigma^2}\right)^{-1}$, $\mathcal{P}'$ can be equivalently recast as

$$\mathcal{P}'' \begin{cases} \underset{\boldsymbol{X}}{\min} & \text{tr}\left(\boldsymbol{S}\boldsymbol{X}\right) - \log\det\left(\boldsymbol{X}\right) \\ \text{s.t.} & \frac{\lambda_{max}(\boldsymbol{X})}{\lambda_{min}(\boldsymbol{X})} \leq K_{max} \\ & \boldsymbol{0} \prec \boldsymbol{X} \preceq \boldsymbol{I} \end{cases}, \tag{A.2}$$

where $\boldsymbol{S} = \frac{\boldsymbol{S}_1}{\sigma^2}$. Now, the constraint set

$$\begin{cases} \frac{\lambda_{max}(\boldsymbol{X})}{\lambda_{min}(\boldsymbol{X})} \leq K_{max} \\ \boldsymbol{0} \prec \boldsymbol{X} \preceq \boldsymbol{I} \end{cases},$$

is equivalent to

$$\begin{cases} u\boldsymbol{I} \preceq \boldsymbol{X} \preceq uK_{max}\boldsymbol{I} \\ \boldsymbol{X} \preceq \boldsymbol{I} \\ 0 < u \leq 1 \end{cases},$$

and problem $\mathcal{P}''$ becomes

$$\mathcal{P}_1 \begin{cases} \underset{\boldsymbol{X},u}{\min} & \text{tr}\,(\boldsymbol{S}\boldsymbol{X}) - \log\det(\boldsymbol{X}) \\ \text{s.t.} & u\boldsymbol{I} \preceq \boldsymbol{X} \preceq uK_{max}\boldsymbol{I} \\ & \boldsymbol{X} \preceq \boldsymbol{I} \\ & 0 < u \leq 1 \end{cases}, \tag{A.3}$$

Since

$$\text{tr}\,(\boldsymbol{S}\boldsymbol{X}) - \log\det(\boldsymbol{X}) \geq -\sum_{i=1}^{N}\log(\lambda_i) \geq -\log(\lambda_{min}(\boldsymbol{X})),$$

where the first inequality stems from $\text{tr}\,(\boldsymbol{S}\boldsymbol{X}) \geq 0$ and the last inequality is due to $\lambda_i \leq 1$, $i = 1, \ldots, N$, (λ_i, $i = 1, \ldots, N$, are the eigenvalues of $\boldsymbol{X}$), when $\boldsymbol{X}$ tends to a rank deficient matrix the objective function tends to $+\infty$. This means that there exists $\epsilon > 0$ such that $\mathcal{P}_1$ is equivalent to

$$\mathcal{P}'_1 \begin{cases} \underset{\boldsymbol{X},u}{\min} & \text{tr}\,(\boldsymbol{S}\boldsymbol{X}) - \log\det(\boldsymbol{X}) \\ \text{s.t.} & u\boldsymbol{I} \preceq \boldsymbol{X} \preceq uK_{max}\boldsymbol{I} \\ & \epsilon\boldsymbol{I} \preceq \boldsymbol{X} \preceq \boldsymbol{I} \\ & \epsilon \leq u \leq 1 \end{cases}. \tag{A.4}$$

Notice that the objective function of problem $\mathcal{P}'_1$ is a continuous function, and the constraint set defines a compact set. Hence, from the Weierstrass Theorem, there exists a feasible point $(\boldsymbol{X}^\star, u^\star)$ for $\mathcal{P}'_1$ such that $v(\mathcal{P}'_1) = \text{tr}\left(\boldsymbol{S}\boldsymbol{X}^{\star-1}\right) - \log\det\left(\boldsymbol{X}^{\star-1}\right)$. Thus, resorting to the previous equivalence, both $\mathcal{P}_1$ and $\mathcal{P}$ are solvable and an optimal solution to $\mathcal{P}$ is given by

$$\left(\sigma^2\boldsymbol{X}^{\star-1}, \sigma^2\boldsymbol{X}^{\star-1} - \sigma^2\boldsymbol{I}, \sigma^2\right).$$

□

Appendix B

Proof of Lemma 1.2.2

Proof. $\forall \bar{u} \in]0,1]$, let us consider the sub-problem $\mathcal{P}_1(\bar{u})$ given in (1.5), obtained from $\mathcal{P}_1$ fixing $u = \bar{u}$. Before proceeding further, the following lemma, whose proof can be found in [75, Theorem 7.4.10, p. 433], are given:

Lemma B.0.1. *Let $\boldsymbol{A} \succeq \mathbf{0}$ and $\boldsymbol{B} \succeq \mathbf{0}$. Let $\lambda_1(\boldsymbol{A}) \geq \ldots \geq \lambda_N(\boldsymbol{A})$ and $\lambda_1(\boldsymbol{B}) \geq \ldots \geq \lambda_N(\boldsymbol{B})$ the eigenvalues of, respectively, $\boldsymbol{A}$ and $\boldsymbol{B}$ in decreasing order. Then, there exists a permutation π of the integers $1, 2, \ldots, N$ such that*

$$\text{tr } (\boldsymbol{AB}) = \sum_{i=1}^{N} \lambda_i(\boldsymbol{A}) \lambda_{\pi(i)}(\boldsymbol{B}). \tag{B.1}$$

Let us now consider[1] a feasible point $\bar{\boldsymbol{X}}$ for problem $\mathcal{P}_1(\bar{u})$ and define $\tilde{\boldsymbol{X}} = \boldsymbol{V}\,\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}})\right)\boldsymbol{V}^\dagger$, with $\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}})$ the vector of the permutated eigenvalues of $\bar{\boldsymbol{X}}$ satisfying the relation

$$\text{tr } \left(\boldsymbol{S}\bar{\boldsymbol{X}}\right) = \sum_{i=1}^{N} \lambda_i(\boldsymbol{S}) \lambda_{\pi(i)}(\bar{\boldsymbol{X}}).$$

Since the LMI's

$$\bar{u}\boldsymbol{I} \preceq \boldsymbol{X} \preceq \bar{u}K_{max}\boldsymbol{I} \quad \text{and} \quad \boldsymbol{X} \preceq \boldsymbol{I},$$

[1] Without loss of generality, the focus is on the set of matrices compliant with Lemma B.0.1 [75] since for an arbitrary pair of matrices $\boldsymbol{A}_1 \succeq \mathbf{0}$ and $\boldsymbol{A}_2 \succeq \mathbf{0}$, the lower bound $\text{tr } (\boldsymbol{A}_1\boldsymbol{A}_2) \geq \sum_{i=1}^{N} \lambda_i(\boldsymbol{A}_1)\lambda_{N-i}(\boldsymbol{A}_2)$ is achieved inside the considered set, through a unitary transformation on $\boldsymbol{A}_2$.

do not involve the eigenvectors of $\boldsymbol{X}$, but only the eigenvalues, $\tilde{\boldsymbol{X}}$ is also a feasible point for $\mathcal{P}_1(\bar{u})$. Moreover,

$$\begin{aligned}\text{tr}\left(\boldsymbol{S}\bar{\boldsymbol{X}}\right) - \log\det\left(\bar{\boldsymbol{X}}\right) &= \sum_{i=1}^{N}\lambda_i(\boldsymbol{S})\lambda_{\pi(i)}(\bar{\boldsymbol{X}}) - \log\det\left(\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}})\right)\right)\\ &= \text{tr}\left(\boldsymbol{V}\mathbf{diag}\left(\boldsymbol{d}\right)\boldsymbol{V}^\dagger\boldsymbol{V}\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}})\right)\boldsymbol{V}^\dagger\right)\\ &- \log\det\left(\boldsymbol{V}\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}})\right)\boldsymbol{V}^\dagger\right)\\ &= \text{tr}\left(\boldsymbol{S}\tilde{\boldsymbol{X}}\right) - \log\det\left(\tilde{\boldsymbol{X}}\right). \qquad \text{(B.2)}\end{aligned}$$

Consequently, for any feasible matrix $\bar{\boldsymbol{X}}$ of problem $\mathcal{P}_1(\bar{u})$ there exists a matrix $\tilde{\boldsymbol{X}} = \boldsymbol{V}\mathbf{diag}(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{X}}))\boldsymbol{V}^\dagger$, which is feasible and achieves the same objective value of $\bar{\boldsymbol{X}}$. Hence, the optimal matrix $\boldsymbol{X}^\star$ is given by $\boldsymbol{X}^\star = \boldsymbol{V}\mathbf{diag}\left(\boldsymbol{\lambda}^\star(\bar{u})\right)\boldsymbol{V}^\dagger$, with $\boldsymbol{\lambda}^\star = \boldsymbol{\lambda}^\star(\bar{u})$ the optimal solution to the following optimization problem:

$$\bar{\mathcal{P}}_1(\bar{u})\begin{cases}\min\limits_{\boldsymbol{\lambda}} & \sum_{i=1}^{N}\left(\lambda_i d_i - \log\lambda_i\right)\\ \text{s.t.} & \bar{u} \le \lambda_i \le \bar{u}K_{max}, \quad i = 1,\ldots,N\\ & \lambda_i \le 1\end{cases}. \qquad \text{(B.3)}$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N] \in \mathbb{R}^N$. For any fixed $\bar{u}$, the objective function and the constraint functions of $\bar{\mathcal{P}}_1(\bar{u})$ are separable functions of the variables λ_i, $i = 1,\ldots,N$. This implies that the optimum point to problem $\bar{\mathcal{P}}_1(\bar{u})$ is obtained solving the N scalar optimization problems:

$$\bar{\mathcal{P}}_1^i(\bar{u})\begin{cases}\min\limits_{\lambda_i} & \lambda_i d_i - \log\lambda_i\\ \text{s.t.} & \bar{u} \le \lambda_i \le \bar{u}K_{max}\\ & \lambda_i \le 1\end{cases}, \qquad \text{(B.4)}$$

where $i = 1,\ldots,N$. Now, to solve each $\bar{\mathcal{P}}_1^i(\bar{u})$, the following constrained optimization problem has to be studied

$$\tilde{\mathcal{P}}_1\begin{cases}\min\limits_{x} & xy - \log x\\ \text{s.t.} & 0 < a \le x \le b\end{cases}, \qquad \text{(B.5)}$$

where the variables x, y, a, b, for each $\bar{\mathcal{P}}_1^i(\bar{u})$, play the role of[2]

$$x = \lambda_i, \quad y = d_i, \quad a = \bar{u}, \quad b = \min\left(K_{max}\bar{u}, 1\right). \qquad \text{(B.6)}$$

[2]Let us observe that the pair of constraints $\bar{u} \le \lambda_i \le \bar{u}K_{max}$ and $\lambda_i \le 1$ is equivalent to $\bar{u} \le \lambda_i \le \min\left(\bar{u}K_{max}, 1\right)$.

Since the function $xy - \log x$, with $y > 0$, is monotonically decreasing for $0 < x \leq \frac{1}{y}$, and monotonically increasing for $x \geq \frac{1}{y}$, the minimizer is given by

$$x = \min\left(b, \max\left(a, \frac{1}{y}\right)\right). \tag{B.7}$$

Thus, using (B.7) and (B.6), the minimizer of $\bar{\mathcal{P}}_1^i(\bar{u})$ is given by[3]

$$\lambda_i^\star(\bar{u}) = \min\left(\min\left(K_{max}\bar{u}, 1\right), \max\left(\bar{u}, \frac{1}{d_i}\right)\right). \tag{B.8}$$

□

[3] If $d_i = 0$, then $\lambda_i^\star(\bar{u}) = \min\left(K_{max}\bar{u}, 1\right)$.

Appendix C

Proof of Theorem 1.2.3

Proof. Let $\boldsymbol{X}^\star(\bar{u})$ be the optimal solution of $\mathcal{P}_1(\bar{u})$, given in (1.5), for any fixed $u = \bar{u}$. From [94], $(\boldsymbol{X}^\star, u^\star) = (\boldsymbol{X}^\star(u^\star), u^\star)$, with $u^\star$ an optimal solution to the optimization problem

$$\mathcal{P}_2' \left\{ \begin{array}{ll} \min\limits_{u} & \mathrm{tr}\,(\boldsymbol{S}\boldsymbol{X}^\star(u)) - \log\det(\boldsymbol{X}^\star(u)) \\ \text{s.t.} & 0 < u \leq 1 \end{array} \right. , \tag{C.1}$$

is an optimal solution to $\mathcal{P}_1$. Using Lemma 1.2.2, $\mathcal{P}_2'$ can be reformulated as

$$\mathcal{P}_2 \left\{ \begin{array}{ll} \min\limits_{u} & \sum_{i=1}^N (\lambda_i^\star(u) d_i - \log \lambda_i^\star(u)) \\ \text{s.t.} & 0 < u \leq 1 \end{array} \right. , \tag{C.2}$$

where $\lambda_i^\star(u)$, $i = 1, \ldots, N$, are the entries of the vector function $\boldsymbol{\lambda}^\star(u)$ defined in (1.8). Thus, using

$$\lambda_i^\star(u) = \min\left(\min\left(K_{max}u, 1\right), \max\left(u, \frac{1}{d_i}\right)\right) \quad i = 1, \ldots, N,$$

$u^\star$ can be obtained as an optimal solution to

$$\mathcal{P}_2 \left\{ \begin{array}{ll} \min\limits_{u} & \sum_{i=1}^N G_i(u) \\ \text{s.t.} & 0 < u \leq 1 \end{array} \right. , \tag{C.3}$$

with, for any $i = 1, \ldots, N$, $G_i(u) = d_i\lambda_i^\star(u) - \log\lambda_i^\star(u)$, i.e.

$$G_i(u) \;=\; \left\{ \begin{array}{ll} -\log K_{max} - \log u + K_{max} d_i u & \text{if } \; u \leq \frac{1}{K_{max}} \\ d_i & \text{if } \; \frac{1}{K_{max}} \leq u \leq 1 \end{array} \right.$$

if $d_i \leq 1$, and

$$G_i(u) = \begin{cases} -\log K_{max} - \log u + K_{max} d_i u & \text{if } 0 < u \leq \frac{1}{K_{max} d_i} \\ \log d_i + 1 & \text{if } \frac{1}{K_{max} d_i} \leq u \leq \frac{1}{d_i} \\ -\log u + d_i u & \text{if } \frac{1}{d_i} \leq u \leq 1 \end{cases}$$

if $d_i > 1$. □

Appendix D

Proof of Theorem 1.2.4

Proof. Let us assume that $d_1 \leq 1$; thus, for all $i = 1, \ldots, N$, $d_i \leq 1$ and $G_i(u)$ is given by (1.10). Thus, for any $i = 1, \ldots, N$, $G_i(u)$ is a strictly decreasing function for $u \in]0, \frac{1}{K_{max}}]$ and a constant function for $\frac{1}{K_{max}} \leq u \leq 1$. Hence, a minimizer for $\mathcal{P}_2$ is given by $u^\star = \frac{1}{K_{max}}$.
Let us consider the case $1 < d_1 \leq K_{max}$. Let $I = \{i : d_i > 1\}$ be the set of indexes such that the corresponding eigenvalues are greater than 1. Since, for any $i \in I$, $G_i(u)$ is an increasing function in the interval $\frac{1}{K_{max} d_i} \leq u \leq 1$

$$\sum_{i \in I} G_i(u)$$

is an increasing function in $\frac{1}{K_{max}} \leq u \leq 1$ $\left(\frac{1}{K_{max} d_i} \leq \frac{1}{K_{max}} < 1\right)$. Moreover, since for $d_i \leq 1$, $G_i(u) = d_i$ in the interval $\frac{1}{K_{max}} \leq u \leq 1$, it follows that $G(u)$ is an increasing function for $\frac{1}{K_{max}} \leq u \leq 1$.
Note that, since for all $i \in I$, $G_i(u)$ is a decreasing function in the interval $0 < u \leq \frac{1}{d_i}$

$$\sum_{i \in I} G_i(u)$$

is a decreasing function in the interval $0 < u \leq \frac{1}{d_1}$ $\left(0 < \frac{1}{d_1} \leq \frac{1}{d_i} < 1\right)$. Furthermore, since $G_i(u)$ is a decreasing function if $d_i \leq 1$, it follows that $G(u)$ is a decreasing function in the interval $0 < u \leq \frac{1}{d_1}$. Thus, since $\frac{1}{K_{max}} \leq \frac{1}{d_1}$, it follows that there exists a minimizer for $\mathcal{P}_2$ which is given by $u^\star = \frac{1}{d_1}$.
Finally, in the case $d_1 > K_{max}$, using the previous considerations, it

results that $G(u)$ is a decreasing function in the interval $u \in \left]0, \frac{1}{d_1}\right]$ and an increasing function in the interval $\left[\frac{1}{K_{max}}, 1\right]$; then, if the minimum exists, it has to belong to $u \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$. Since $G(u)$ is a continuous function, and $\left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$ a compact set, Weierstrass Theorem ensures that $u^\star \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$. □

Appendix E

Proof of Lemma 1.2.5

Proof. Let us consider the case $d_i > 1$; evaluating the derivative of $G_i(u)$, the following result is obtained

$$\frac{dG_i(u)}{du} = \begin{cases} -\frac{1}{u} + K_{max} d_i & \text{if } 0 < u < \frac{1}{K_{max} d_i} \\ 0 & \text{if } \frac{1}{K_{max} d_i} < u < \frac{1}{d_i} \\ -\frac{1}{u} + d_i & \text{if } \frac{1}{d_i} < u < 1 \end{cases} . \quad \text{(E.1)}$$

Let us observe that in each subinterval $u \in \left]0, \frac{1}{K_{max} d_i}\right[$, $u \in \left]\frac{1}{K_{max} d_i}, \frac{1}{d_i}\right[$, and $u \in \left]\frac{1}{d_i}, 1\right[$, $\frac{dG_i(u)}{du}$ is a continuous function. Moreover, in correspondence of the points $u_1 = \frac{1}{K_{max} d_i}$ and $u_2 = \frac{1}{d_i}$ the right derivative is equal to the left derivative. Then, the whole derivative function is a continuous function in the interval $u \in]0, 1]$. Let us now consider the case $d_i \leq 1$; in this case, the derivative of $G_i(u)$ in the interval $u \in \left]0, \frac{1}{K_{max}}\right]$ is given by

$$\frac{dG_i(u)}{du} = -\frac{1}{u} + K_{max} d_i \quad \text{if } u \leq \frac{1}{K_{max}}. \quad \text{(E.2)}$$

Thus, $\frac{dG_i(u)}{du}$ is a continuous function on the interval $u \in \left]0, \frac{1}{K_{max}}\right]$. Consequently, $\frac{dG(u)}{du}$ is a continuous function on the interval $u \in \left]0, \frac{1}{K_{max}}\right]$. As to the convexity of $G(u)$, in the interval $u \in \left]0, \frac{1}{K_{max}}\right]$, let us recall that the sum of convex functions is a convex function [94]. Hence, it is sufficient to study the convexity of each term $G_i(u)$, $i = 1, \ldots, N$. If

$d_i \leq 1$, since $\frac{d^2 G_i(u)}{du^2} > 0$, $u \in]0, \frac{1}{K_{max}}]$, from [94], $G_i(u)$ is a convex function. Moreover, in the case $d_i > 1$, recalling that a derivable function is convex if and only if its derivative is an increasing function, [95, Theorem 6, p.144], since $\frac{dG_i(u)}{du}$ is an increasing function over the interval $]0, \frac{1}{K_{max}}]$, $G_i(u)$ is a convex function. □

Appendix F

Proof of Theorem 1.2.6

Proof. 1) From Lemma 1.2.5, $G(u)$ is a convex and derivable function on $\left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$. This means that $\frac{dG(u)}{du}$ $\left(u \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]\right)$, is an increasing function. Since by hypothesis $\left.\frac{dG(u)}{du}\right|_{u=\frac{1}{d_1}} = 0$, it follows that

$$\frac{dG(u)}{du} \geq 0,\ u \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right],$$

i.e. $G(u)$ is an increasing function. Thus, a minimum to problem $\mathcal{P}_2$ is achieved in $u^\star = \frac{1}{d_1}$. Notice that if $\left.\frac{dG(u)}{du}\right|_{u=\frac{1}{d_1}} > 0$, $u^\star \neq \frac{1}{d_1}$ and the optimal solution to problem $\mathcal{P}_2$ is unique. In fact, in the interval $u \in \left]0, \frac{1}{d_1}\right]$ $G(u)$ is strictly decreasing, in the interval $u \in \left]\frac{1}{K_{max}}, 1\right]$ $G(u)$ is strictly increasing. Finally, there does not exists any interval $I \subseteq \left]\frac{1}{d_1}, \frac{1}{K_{max}}\right]$ such that $\frac{d^2G(u)}{du^2} = 0\ \forall u \in I$.

2) Let us, now, assume that

$$\left.\frac{dG(u)}{du}\right|_{u=\frac{1}{d_1}} < 0 \ \text{ and } \ \left.\frac{dG(u)}{du}\right|_{u=\frac{1}{K_{max}}} \leq 0. \tag{F.1}$$

Again, since by Lemma 1.2.5, $G(u)$ is a convex and derivable function on $\left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$, $\frac{dG(u)}{du}$, $u \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$, is an increasing func-

tion. Consequently, from (F.1),

$$\frac{dG(u)}{du} \leq 0, \quad u \in \left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$$

and $G(u)$ is a decreasing function in $\left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$. Then, its optimal value on the interval $\left[\frac{1}{d_1}, \frac{1}{K_{max}}\right]$ is attained at $u^\star = \frac{1}{K_{max}}$.

3) Let us, finally, consider the case

$$\frac{dG(u)}{du}\Big|_{u=\frac{1}{d_1}} < 0 \quad \text{and} \quad \frac{dG(u)}{du}\Big|_{u=\frac{1}{K_{max}}} > 0.$$

Thus the optimal point has to be an interior point of $[\frac{1}{d_1}, \frac{1}{K_{max}}]$. Since from lemma 1.2.5, $G(u)$ is a convex and differentiable function, $u^\star$ is the optimal point if and only if, [94],

$$\frac{dG(u)}{du} = 0\Big|_{u=u^\star}. \tag{F.2}$$

Let us now characterize the necessary and sufficient condition (F.2). Let $u^\star$ the optimal point. Assume that $v_i \neq v_j$ and $v_i \neq K_{max}v_j$ for any $i \neq j$ with $1 \leq i, j \leq \bar{N}+1$, which hold with probability 1. Let $\alpha \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ be the largest index such that $\frac{1}{v_\alpha} < u^\star$ ($\alpha \geq 1$ since $u^\star > \frac{1}{v_1}$ and $\alpha \leq \bar{N}$ since $u^\star < \frac{1}{K_{max}} < \frac{1}{v_{\bar{N}+1}}$), and let $\beta \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ the smallest index such that $\frac{1}{v_\beta K_{max}} > u^\star$ ($\beta \leq \bar{N}+1$ since $u^\star < \frac{1}{K_{max}} = \frac{1}{K_{max}v_{\bar{N}+1}}$ and $\beta \geq 2$ since $\frac{1}{K_{max}v_1} < \frac{1}{v_1} < u^\star$). Furthermore, $\alpha < \beta$ since assuming $\beta \leq \alpha$ then $\frac{1}{K_{max}v_\beta} \leq \frac{1}{K_{max}v_\alpha} < \frac{1}{v_\alpha} < u^\star$, which can not comply with $\frac{1}{K_{max}v_\beta} > u^\star$.
Thus, there exists a neighborhood $B_{u^\star}$ of $u^\star$, contained in the interval $\left]\frac{1}{v_\alpha}, \frac{1}{K_{max}v_\beta}\right[$, such that $G(u)$ can be expressed as

$$\begin{aligned} G(u) &= \sum_{i=1}^{\alpha} (-\log u + d_i u) + \sum_{i=\alpha+1}^{\beta-1} (\log d_i + 1) \\ &+ \sum_{i=\beta}^{N} (-\log K_{max} - \log u + K_{max} d_i u), \end{aligned} \tag{F.3}$$

$\forall u \in B_{u^\star}$. Performing the derivative of (F.3) with respect to u and imposing the optimality condition (F.2), the minimizer $u^\star$ is then given by

$$u^\star = \frac{N + \alpha - \beta + 1}{\sum_{i=1}^{\alpha} d_i + \sum_{i=\beta}^{N} K_{max} d_i}. \tag{F.4}$$

Conversely, let $\bar{u}$ be a point such that

$$\bar{u} = \frac{N + \bar{\alpha} - \bar{\beta} + 1}{\sum_{i=1}^{\bar{\alpha}} d_i + \sum_{i=\bar{\beta}}^{N} K_{max} d_i}, \tag{F.5}$$

with $\bar{\alpha} \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ the largest index such that $\frac{1}{v_{\bar{\alpha}}} < \bar{u}$, and $\bar{\beta} \in \{1, 2, \ldots, \bar{N}, \bar{N}+1\}$ the smallest index such that $\frac{1}{v_{\bar{\beta}} K_{max}} > \bar{u}$.
This means that there exists a neighborhood $B_{\bar{u}}$ of $\bar{u}$, contained in the interval $\left]\frac{1}{v_{\bar{\alpha}}}, \frac{1}{K_{max} v_{\bar{\beta}}}\right[$ such that $\forall u \in B_{\bar{u}}$, $G(u)$ is given by

$$\begin{aligned} G(u) &= \sum_{i=1}^{\bar{\alpha}} (-\log u + d_i u) + \sum_{i=\bar{\alpha}+1}^{\bar{\beta}-1} (\log d_i + 1) \\ &+ \sum_{i=\bar{\beta}}^{N} (-\log K_{max} - \log u + K_{max} d_i u)\,. \end{aligned} \tag{F.6}$$

Now, performing the derivative of (F.6), condition (F.5) means that

$$\left.\frac{dG(u)}{du}\right|_{u=\bar{u}} = 0,$$

i.e. $\bar{u} = u^\star$.

□

Appendix G

Proof of Theorem 2.1.1

Proof. Since the squared Euclidian norm $||\cdot||^2$ defines a strictly convex function in the set of complex matrices $\boldsymbol{B} \in \mathbb{C}^{N\times N}$, there exists a unique optimal solution to problem $\mathcal{P}$. Furthermore, problem $\mathcal{P}$ is equivalent to

$$\mathcal{P}' \left\{ \begin{array}{ll} \min\limits_{\boldsymbol{S}_k} & -2\text{tr}\left\{\boldsymbol{S}_k \boldsymbol{r}_k \boldsymbol{r}_k^\dagger\right\} + \text{tr}\left\{\boldsymbol{S}_k^2\right\} \\ \text{s.t.} & \boldsymbol{S}_k \succeq \mathbf{I} \end{array} \right. . \tag{G.1}$$

Hence, the following lemma, whose proof can be found in [75, Theorem 7.4.10, p. 433], is given:

Lemma G.0.2. *Let $\boldsymbol{A} \succeq \mathbf{0}$ and $\boldsymbol{B} \succeq \mathbf{0}$. Let $\lambda_1(\boldsymbol{A}) \geq \ldots \geq \lambda_N(\boldsymbol{A})$ and $\lambda_1(\boldsymbol{B}) \geq \ldots \geq \lambda_N(\boldsymbol{B})$ be the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, in decreasing order. There exists a permutation π of the integers $1, 2, \ldots, N$ such that*

$$\text{tr}\,(\boldsymbol{AB}) = \sum_{i=1}^{N} \lambda_i(\boldsymbol{A})\lambda_{\pi(i)}(\boldsymbol{B}). \tag{G.2}$$

Since $\text{tr}\left\{\boldsymbol{S}_k^2\right\}$ does not depend on the eigenvectors of $\boldsymbol{S}_k$, exploiting Lemma G.0.2, the optimal solution to problem $\mathcal{P}'$ is $\boldsymbol{U}_k \mathbf{diag}(\boldsymbol{\lambda}) \boldsymbol{U}_k^\dagger$, where $\boldsymbol{U}_k$ is a unitary matrix of the eigenvectors of $\boldsymbol{r}_k \boldsymbol{r}_k^\dagger$ with the first eigenvector corresponding to the eigenvalue $||\boldsymbol{r}_k||^2$, and the vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_K]$ is the optimal solution to problem

$$\mathcal{P}'' \left\{ \begin{array}{ll} \min\limits_{\boldsymbol{\lambda}} & -2||\boldsymbol{r}_k||^2\lambda_1 + \sum_{i=1}^{K} \lambda_i^2 \\ \text{s.t.} & \lambda_i \geq 1, \quad i = 1, \ldots K \end{array} \right. . \tag{G.3}$$

Consequently,

$$\lambda_i = 1, \quad i = 2, \ldots, K,$$

and λ_1 is the optimal solution to

$$\mathcal{P}_1'' \begin{cases} \min\limits_{\lambda_1} & -2||\boldsymbol{r}_k||^2\lambda_1 + \lambda_1^2 \\ \text{s.t.} & \lambda_1 \geq 1 \end{cases} . \tag{G.4}$$

Hence, it follows that

$$\lambda_1 = \max\left(1, ||\boldsymbol{r}_k||^2\right), \tag{G.5}$$

and

$$\boldsymbol{\lambda} = \left[\max\left(1, ||\boldsymbol{r}_k||^2\right), 1, \ldots, 1\right] \tag{G.6}$$

□

Appendix H

Proof of Theorem 2.1.2

Proof. Since $\sum_{i=1}^{K} w_i ||\boldsymbol{A}_i - \boldsymbol{A}||^2 \to +\infty$ when $||\boldsymbol{A}|| \to +\infty$ and the squared Euclidian norm $||\cdot||^2$ defines a strictly convex function in the set of complex matrices $\boldsymbol{B} \in \mathbb{C}^{N\times N}$, the matrix

$$\widehat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\{ \sum_{i=1}^{K} w_i ||\boldsymbol{A}_i - \boldsymbol{A}||^2 \right\} \tag{H.1}$$

is unique. Furthermore, it is possible to observe that

$$\begin{aligned}
\widehat{\boldsymbol{A}} &= \arg\min_{\boldsymbol{A}} \left\{ \sum_{i=1}^{K} w_i \text{tr}\left((\boldsymbol{A}_i - \boldsymbol{A})(\boldsymbol{A}_i - \boldsymbol{A})^\dagger \right) \right\} = \\
&= \arg\min_{\boldsymbol{A}} \sum_{i=1}^{K} w_i \left\{ \text{tr}\left(\boldsymbol{A}_i \boldsymbol{A}_i^\dagger \right) + \text{tr}\left(\boldsymbol{A}\boldsymbol{A}^\dagger \right) - 2\text{Re}\left\{ \text{tr}\left(\boldsymbol{A}_i \boldsymbol{A}^\dagger \right) \right\} \right\} = \\
&= \arg\min_{\boldsymbol{A}} \left\{ \sum_{i=1}^{K} w_i \text{tr}\left(\boldsymbol{A}_i \boldsymbol{A}_i^\dagger \right) + \sum_{i=1}^{K} w_i \text{tr}\left(\boldsymbol{A}\boldsymbol{A}^\dagger \right) \right. \\
&\left. -2\text{Re}\left\{ \text{tr}\left(\sum_{i=1}^{K} w_i \boldsymbol{A}_i \boldsymbol{A}^\dagger \right) \right\} \right\} = \\
&= \arg\min_{\boldsymbol{A}} \left\{ \text{tr}\left(\boldsymbol{A}\boldsymbol{A}^\dagger \right) - 2\text{Re}\left\{ \text{tr}\left(\overline{\boldsymbol{A}} \boldsymbol{A}^\dagger \right) \right\} \right\},
\end{aligned} \tag{H.2}$$

where $\overline{\boldsymbol{A}} = \sum_{i=1}^{K} w_i \boldsymbol{A}_i$. Notice that

$$\text{Re}\left\{ \text{tr}\left(\overline{\boldsymbol{A}} \boldsymbol{A}^\dagger \right) \right\} \leq \left| \text{tr}\left(\overline{\boldsymbol{A}} \boldsymbol{A}^\dagger \right) \right| \leq ||\boldsymbol{A}||||\overline{\boldsymbol{A}}||, \tag{H.3}$$

where the second inequality stems from the Schwartz inequality. Since both the inequalities in (H.3) become equalities if and only if $\boldsymbol{A} = \alpha\overline{\boldsymbol{A}}$ with $\alpha > 0$, $\widehat{\boldsymbol{A}} = \widehat{\alpha}\overline{\boldsymbol{A}}$ with $\widehat{\alpha}$ given by

$$\widehat{\alpha} = \arg\min_{\alpha>0} \left\{ \alpha^2 \mathrm{tr}\left(\overline{\boldsymbol{A}}\ \overline{\boldsymbol{A}}^{\dagger}\right) - 2\alpha \mathrm{tr}\left(\overline{\boldsymbol{A}}\ \overline{\boldsymbol{A}}^{\dagger}\right) \right\} = 1. \tag{H.4}$$

Thus, $\widehat{\boldsymbol{A}} = \overline{\boldsymbol{A}} = \sum_{i=1}^{K} w_i \boldsymbol{A}_i$. □

Appendix I

Proof of Theorem 2.1.3

Proof. Problem (2.18) is equivalent to

$$\begin{cases} \min\limits_{t_1,\dots,t_K,\boldsymbol{A}} & \sum_{i=1}^K w_i t_i \\ \text{s.t.} & \sqrt{\text{tr}\left\{(\boldsymbol{A}_i-\boldsymbol{A})(\boldsymbol{A}_i-\boldsymbol{A})^\dagger\right\}} \leq t_i \quad i=1,\dots,K \\ & \boldsymbol{A}\in\mathbb{H}^N \end{cases} . \tag{I.1}$$

Let us now observe that $\forall i \in \{1,\dots,K\}$

$$\sqrt{\text{tr}\left\{(\boldsymbol{A}_i-\boldsymbol{A})(\boldsymbol{A}_i-\boldsymbol{A})^\dagger\right\}} \leq t_i \tag{I.2}$$

is equivalent to

$$\text{tr}\left\{(\boldsymbol{A}_i-\boldsymbol{A})(\boldsymbol{A}_i-\boldsymbol{A})^\dagger\right\} \leq t_i^2 \tag{I.3}$$

which can be recast in

$$\begin{pmatrix} t_i & (\text{vec}(\boldsymbol{A})-\text{vec}(\boldsymbol{A}_i))^\dagger \\ (\text{vec}(\boldsymbol{A})-\text{vec}(\boldsymbol{A}_i)) & t_i\boldsymbol{I} \end{pmatrix} \succeq \boldsymbol{0}. \tag{I.4}$$

As a consequence, problem (I.1) is equivalent to the convex optimization SDP problem

$$\begin{cases} \min\limits_{t_1,\dots,t_K,\boldsymbol{A}} & \sum_{i=1}^K w_i t_i \\ \text{s.t.} & \begin{pmatrix} t_i & (\text{vec}(\boldsymbol{A})-\text{vec}(\boldsymbol{A}_i))^\dagger \\ (\text{vec}(\boldsymbol{A})-\text{vec}(\boldsymbol{A}_i)) & t_i\boldsymbol{I} \end{pmatrix} \succeq \boldsymbol{0} \\ & \boldsymbol{A}\in\mathbb{H}^N \end{cases} . \quad i=1,\dots,K \tag{I.5}$$

□

Appendix J

Proof of Proposition 2.1.4

Proof. 1. **Positive semidefinite nature of the median**

- Let us assume that $\boldsymbol{A}_i \succeq \boldsymbol{0}$, $i = 1, \ldots, K$, and let us suppose that $\boldsymbol{A} = \boldsymbol{U}\mathbf{diag}(\boldsymbol{\lambda})\boldsymbol{U}^\dagger$ (where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N]$ is the vector containing its eigenvalues and $\boldsymbol{U}$ is the unitary matrix of the corresponding eigenvectors) has some negative eigenvalues, then

$$\begin{aligned} \|\boldsymbol{A}_i - \boldsymbol{A}\| &= \|\overline{\boldsymbol{A}}_i - \mathbf{diag}(\boldsymbol{\lambda})\| \\ &= \sqrt{\mathrm{tr}\,(\overline{\boldsymbol{A}}_i^2) + \sum_{i=1}^{N} \lambda_i^2 - 2\sum_{l=1}^{N} \overline{\boldsymbol{A}}_i(l,l)\lambda_l} \\ &\geq \sqrt{\mathrm{tr}\,(\overline{\boldsymbol{A}}_i^2) + \sum_{i=1}^{N} \lambda_i^2 - 2\sum_{l=1}^{N} \overline{\boldsymbol{A}}_i(l,l)|\lambda_l|} \\ &= \|\overline{\boldsymbol{A}}_i - |\mathbf{diag}(\boldsymbol{\lambda})|\| = \|\boldsymbol{A}_i - |\boldsymbol{A}|\| \end{aligned} \tag{J.1}$$

  where $\overline{\boldsymbol{A}}_i = \boldsymbol{U}^\dagger \boldsymbol{A}_i \boldsymbol{U}$ and it has been exploited the fact that $\overline{\boldsymbol{A}}_i(l,l) \geq 0$, $\forall i \in \{1, \ldots, K\}$ and $\forall l \in \{1, \ldots, N\}$.

- Let now assume that $\boldsymbol{A}_i \succ \boldsymbol{0}$, $i = 1, \ldots, K$, and suppose that $\boldsymbol{A}$ has some zero eigenvalues λ_l, $l = N - H, \ldots, N$; hence,

there is

$$\begin{aligned}\|\boldsymbol{A}_i - \boldsymbol{A}\| \geq \sqrt{\operatorname{tr}(\overline{\boldsymbol{A}}_i^2) + \sum_{i=1}^{N} \lambda_i^2 + H\left[\min_{i,l}(\overline{\boldsymbol{A}}_i(l,l))^2\right] - \cdots} \\ \overline{\cdots - 2\sum_{l=1}^{N-H} \overline{\boldsymbol{A}}_i(l,l)\lambda_l - \cdots} \\ \overline{\cdots - 2\sum_{l-H+1}^{N} \overline{\boldsymbol{A}}_i(l,l)\left[\min_{i,l}(\overline{\boldsymbol{A}}_i(l,l))\right]} \\ = \|\boldsymbol{A}_i - \boldsymbol{A}_E\| \quad i = 1, \ldots, K\end{aligned} \tag{J.2}$$

where $\boldsymbol{A}_E$ is obtained from $\boldsymbol{A}$ setting its zero eigenvalues to the value $\min_{i,l}(\overline{\boldsymbol{A}}_i(l,l)) > 0$.

2. **Uniqueness of the solution** As shown in [96], if the distribution of a random vector is not concentrated on a line then the median is unique. As a consequence, in this case, it is sufficient to assume that the matrices $\boldsymbol{A}_i \in \mathbb{H}^N$, $i = 1, \ldots, K$, are not aligned (as specified in (2.20)) to ensure the existence and uniqueness of the median matrix.

□

Appendix K

Proof of Proposition 3.1.1

Proof. Performing the change of variable $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}$, problem $\mathcal{P}_{H_0}$ becomes equivalent to

$$\mathcal{P}_0' \begin{cases} \min\limits_{\boldsymbol{Y}} & \text{tr}\left\{\boldsymbol{S}_0\boldsymbol{A}^{-1}\boldsymbol{Y}\boldsymbol{A}^{-1}\right\} - \log\det\left(\boldsymbol{Y}\right) \\ \text{s.t.} & f_i(\boldsymbol{Y}) \leq 0, \quad i = 1, \ldots, M \\ & \boldsymbol{Y} \succ \boldsymbol{0} \end{cases} . \tag{K.1}$$

Indeed, given an optimal solution $\boldsymbol{Y}^\star$ to $\mathcal{P}_0'$, then $\boldsymbol{A}^{-1}\boldsymbol{Y}^\star\boldsymbol{A}^{-1}$ is an optimal solution to $\mathcal{P}_{H_0}$, and viceversa, given an optimal solution $\boldsymbol{X}^\star$ to $\mathcal{P}_{H_0}$, then $\boldsymbol{A}\boldsymbol{X}^\star\boldsymbol{A}$ is an optimal solution to $\mathcal{P}_0'$. Let us show that the problem $\mathcal{P}_0'$ is solvable[1], namely the ML estimate is well defined. Since Ω is a bounded set, the quantity

$$C = \sup_{\boldsymbol{Y}\in\Omega} \text{tr}\left\{\boldsymbol{Y}\right\},$$

is a bounded positive real number. Furthermore, there is

$$\text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\boldsymbol{Y}\right\} - \log\det\left(\boldsymbol{Y}\right) \geq -\log(\lambda_N(\boldsymbol{Y})) - (N-1)\log(C), \tag{K.2}$$

where the inequality stems from $\text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\boldsymbol{Y}\right\} \geq 0$ and $\lambda_i(\boldsymbol{Y}) \leq C$, $i = 1, \ldots, N$. As a result, when $\boldsymbol{Y}$ tends to a rank deficient matrix the objective function tends to $+\infty$. This means that there exists $\epsilon > 0$

[1] By "solvable", it is meant that the problem is feasible and bounded, and the optimal value is attained, see [74, p. 13].

such that $\mathcal{P}_0'$ is equivalent to

$$\mathcal{P}_0'' \left\{ \begin{array}{ll} \min\limits_{\boldsymbol{Y}} & \text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\boldsymbol{Y}\right\} - \log\det\left(\boldsymbol{Y}\right) \\ \text{s.t.} & f_i(\boldsymbol{Y}) \leq 0, \quad i = 1, \ldots, M \\ & \boldsymbol{Y} \succeq \epsilon\mathbf{I} \end{array} \right. . \tag{K.3}$$

Notice that the objective function of problem $\mathcal{P}_0''$ is a continuous function and the constraint set defines a compact set. Indeed, the constraint set is bounded (it is a subset of Ω) and closed (the intersection of closed sets). Hence, Weierstrass Theorem ensures the existence of a feasible point $\boldsymbol{Y}^\star$ for $\mathcal{P}_0'$ such that $\text{v}(\mathcal{P}_0') = \text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\boldsymbol{Y}^\star\right\} - \log\det\left(\boldsymbol{Y}^\star\right)$.

Before proceeding further with the study of problem $\mathcal{P}_0'$, the following lemma, whose proof can be found in [75, Theorem 7.4.10, p. 433], is given:

Lemma K.0.3. *Let $\boldsymbol{B}_1 \succeq \mathbf{0}$ and $\boldsymbol{B}_2 \succeq \mathbf{0}$. There exists a permutation π of the integers $1, 2, \ldots, N$ such that*

$$\text{tr}\left\{\boldsymbol{B}_1\boldsymbol{B}_2\right\} = \sum_{i=1}^{N} \lambda_i(\boldsymbol{B}_1)\lambda_{\pi(i)}(\boldsymbol{B}_2). \tag{K.4}$$

Let us now consider a feasible point $\bar{\boldsymbol{Y}} = \bar{\boldsymbol{V}}\mathbf{diag}\left(\boldsymbol{\lambda}(\bar{\boldsymbol{Y}})\right)\bar{\boldsymbol{V}}^\dagger$ for problem $\mathcal{P}_0'$, where $\bar{\boldsymbol{V}}$ is a unitary matrix containing the eigenvectors of $\bar{\boldsymbol{Y}}$, and define $\tilde{\boldsymbol{Y}} = \boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}})\right)\boldsymbol{V}_0^\dagger$, with $\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}})$ the vector of the permutated eigenvalues of $\bar{\boldsymbol{Y}}$ satisfying the relation

$$\text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\bar{\boldsymbol{Y}}\right\} = \sum_{i=1}^{N} \lambda_i(\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1})\lambda_{\pi(i)}(\bar{\boldsymbol{Y}}).$$

Since by assumption $f_i(\boldsymbol{Y})$, $i = 1, \ldots, M$, are unitary invariant functions,

$$f_i(\tilde{\boldsymbol{Y}}) = f_i(\bar{\boldsymbol{V}}\boldsymbol{V}_0^\dagger\tilde{\boldsymbol{Y}}\boldsymbol{V}_0\bar{\boldsymbol{V}}^\dagger) = f_i(\bar{\boldsymbol{Y}}) \leq 0, \quad i = 1, \ldots, M,$$

namely $\tilde{\boldsymbol{Y}}$ is a feasible point for $\mathcal{P}_0'$. Moreover,

$$\begin{aligned}
&\text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\bar{\boldsymbol{Y}}\right\} - \log\det\left(\bar{\boldsymbol{Y}}\right) = \\
&= \sum_{i=1}^{N} \lambda_i(\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1})\lambda_{\pi(i)}(\bar{\boldsymbol{Y}}) - \log\det\left(\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}})\right)\right) \\
&= \text{tr}\left\{\boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{d}_0\right)\boldsymbol{V}_0^\dagger\boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}})\right)\boldsymbol{V}_0^\dagger\right\} \\
&- \log\det\left(\boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}})\right)\boldsymbol{V}_0^\dagger\right) \\
&= \text{tr}\left\{\boldsymbol{A}^{-1}\boldsymbol{S}_0\boldsymbol{A}^{-1}\tilde{\boldsymbol{Y}}\right\} - \log\det\left(\tilde{\boldsymbol{Y}}\right).
\end{aligned} \tag{K.5}$$

Consequently, for any feasible matrix $\bar{\boldsymbol{Y}}$ to problem $\mathcal{P}_0'$, there exists a matrix $\tilde{\boldsymbol{Y}} = \boldsymbol{V}_0\mathbf{diag}(\boldsymbol{\lambda}_\pi(\bar{\boldsymbol{Y}}))\boldsymbol{V}_0^\dagger$, which is feasible and achieves the same objective value as $\bar{\boldsymbol{Y}}$. Hence, an optimal solution $\boldsymbol{Y}^\star$ to problem $\mathcal{P}_0'$ exhibits the structure $\boldsymbol{Y}^\star = \boldsymbol{V}_0\mathbf{diag}\left(\boldsymbol{\lambda}^\star\right)\boldsymbol{V}_0^\dagger$, with $\boldsymbol{\lambda}^\star$ an optimal solution to the optimization problem $\mathcal{P}(\boldsymbol{d}_0)$. Since the objective function of problem $\mathcal{P}(\boldsymbol{d}_0)$ is strictly convex, the assumption that the functions $\overline{f}_i(\boldsymbol{\lambda})$ $i = 1, \ldots, M$, are convex ensures the uniqueness of the solution. □

Appendix L

Proof of Proposition 3.1.2

Performing the change of variables $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{A}$, $\boldsymbol{\beta} = \boldsymbol{\alpha}\|\boldsymbol{A}^{-1}\boldsymbol{p}\|$, problem $\mathcal{P}_{H_1}$ becomes equivalent to

$$\mathcal{P}_1' \left\{ \begin{array}{ll} \min\limits_{\boldsymbol{\beta},\boldsymbol{Y}} & \operatorname{tr}\left\{\frac{1}{K}\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}^\dagger\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}^\dagger\right)^\dagger \boldsymbol{Y}\right\} - \log\det\left(\boldsymbol{Y}\right) \\ \text{s.t.} & f_i(\boldsymbol{Y}) \leq 0, \quad i = 1,\ldots,M \\ & \boldsymbol{Y} \succ \boldsymbol{0} \end{array} \right. , \tag{L.1}$$

where $\boldsymbol{R}_1 = \boldsymbol{A}^{-1}\boldsymbol{R}$. In fact, given an optimal solution $(\boldsymbol{Y}^\star, \boldsymbol{\beta}^\star)$ to $\mathcal{P}_0'$, then $\left(\boldsymbol{A}^{-1}\boldsymbol{Y}^\star\boldsymbol{A}^{-1}, \frac{\boldsymbol{\beta}^\star}{\|\boldsymbol{A}^{-1}\boldsymbol{p}\|}\right)$ is an optimal solution to $\mathcal{P}_{H_0}$, and vicevversa, given an optimal solution $(\boldsymbol{X}^\star, \boldsymbol{\alpha}^\star)$ to $\mathcal{P}_{H_0}$, then $(\boldsymbol{A}\boldsymbol{X}^\star\boldsymbol{A}, \boldsymbol{\alpha}^\star\|\boldsymbol{A}^{-1}\boldsymbol{p}\|)$ is an optimal solution to $\mathcal{P}_0'$. Notice that, using the same arguments as in the proof of Proposition 3.1.1, for any $\bar{\boldsymbol{\beta}}$, it can be shown that $\mathcal{P}_1'$ admits an optimal solution

$$\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}} = \boldsymbol{U}_{\bar{\boldsymbol{\beta}}}\mathbf{diag}\left(\boldsymbol{\lambda}_{\bar{\boldsymbol{\beta}}}\right)\boldsymbol{U}_{\bar{\boldsymbol{\beta}}}^\dagger, \tag{L.2}$$

where $\boldsymbol{U}_{\bar{\boldsymbol{\beta}}}$ is a unitary matrix containing the eigenvectors of

$$\boldsymbol{S}_{\bar{\boldsymbol{\beta}}} = \frac{1}{K}\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\bar{\boldsymbol{\beta}}^\dagger\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\bar{\boldsymbol{\beta}}^\dagger\right)^\dagger = \boldsymbol{U}_{\bar{\boldsymbol{\beta}}}\mathbf{diag}\left(\boldsymbol{d}_{\bar{\boldsymbol{\beta}}}\right)\boldsymbol{U}_{\bar{\boldsymbol{\beta}}}^\dagger \tag{L.3}$$

corresponding to the eigenvalues $\boldsymbol{d}_{\bar{\boldsymbol{\beta}}} = \boldsymbol{\lambda}(\boldsymbol{S}_{\bar{\boldsymbol{\beta}}})$. Furthermore, a ML estimate of the eigenvalues $\boldsymbol{\lambda}_{\bar{\boldsymbol{\beta}}}$ is an optimal solution to the optimization

problem

$$\mathcal{P}_1'' \begin{cases} \min\limits_{\boldsymbol{\lambda}} & \boldsymbol{d}_{\bar{\boldsymbol{\beta}}}^{\dagger}\boldsymbol{\lambda} - \sum\limits_{i=1}^{N} \log \lambda_i \\ \text{s.t.} & \overline{f_i}(\boldsymbol{\lambda}) \leq 0, \quad i = 1, \ldots, M \\ & \lambda_i > 0, \quad i = 1, \ldots, N \end{cases} . \tag{L.4}$$

As for Proposition 3.1.1, since the objective function of problem $\mathcal{P}_1''$ is strictly convex, the assumption that the functions $\overline{f_i}(\boldsymbol{\lambda})$ $i = 1, \ldots, M$, are convex guarantees the uniqueness of the solution.

It is now proved that the objective function value achieved by $\left(\boldsymbol{Y}_{\boldsymbol{\beta}_1}, \boldsymbol{\beta}_1\right)$, with $\boldsymbol{\beta}_1 = \boldsymbol{R}_1^{\dagger}\boldsymbol{p}_1$ and $\boldsymbol{Y}_{\boldsymbol{\beta}_1}$ the corresponding ML inverse covariance matrix estimate, is lower than or equal to the objective function value achieved by $\left(\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}}, \bar{\boldsymbol{\beta}}\right)$, where $\bar{\boldsymbol{\beta}}$ is an arbitrary complex vector and $\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}}$ is the ML inverse covariance matrix estimate sharing the structure (L.2)[1]. To this end, let us consider the following feasible solution to problem $\mathcal{P}_1'$

$$\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}, \boldsymbol{\beta}_1\right)$$

where

$$\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1} = \boldsymbol{U}_{\boldsymbol{\beta}_1} \mathbf{diag}\left(\boldsymbol{\lambda}_{\bar{\boldsymbol{\beta}}}\right) \boldsymbol{U}_{\boldsymbol{\beta}_1}^{\dagger},$$

with $\boldsymbol{U}_{\boldsymbol{\beta}_1}$ a unitary matrix containing along its columns the eigenvectors of

$$\boldsymbol{S}_{\boldsymbol{\beta}_1} = \frac{1}{K}\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}_1^{\dagger}\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}_1^{\dagger}\right)^{\dagger} = \boldsymbol{U}_{\boldsymbol{\beta}_1} \mathbf{diag}\left(\boldsymbol{d}_{\boldsymbol{\beta}_1}\right) \boldsymbol{U}_{\boldsymbol{\beta}_1}^{\dagger},$$

corresponding to the eigenvalues $\boldsymbol{d}_{\boldsymbol{\beta}_1} = \boldsymbol{\lambda}(\boldsymbol{S}_{\boldsymbol{\beta}_1})$. Now, observing that

$$\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}_1^{\dagger} = \boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\boldsymbol{R}_1 = \left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^{\dagger}\right)\boldsymbol{R}_1,$$

[1]Notice that, for a fixed $\bar{\boldsymbol{\beta}}$, there could exist some optimal inverse covariance matrices, in the ML sense; however all these solutions share the same optimal likelihood value.

, it follows that

$$\begin{aligned}\boldsymbol{S}_{\boldsymbol{\beta}_1} &= \frac{1}{K}\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}_1^\dagger\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\boldsymbol{\beta}_1^\dagger\right)^\dagger \\ &= \frac{1}{K}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\boldsymbol{R}_1\boldsymbol{R}_1^\dagger\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)^\dagger \\ &= \frac{1}{K}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\bar{\boldsymbol{\beta}}\right)\left(\boldsymbol{R}_1 - \boldsymbol{p}_1\bar{\boldsymbol{\beta}}\right)^\dagger\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)^\dagger \\ &= \left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)^\dagger,\end{aligned} \tag{L.5}$$

with $\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}$ given in (L.3). Since

$$\boldsymbol{\lambda}\left(\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)^\dagger\right) = \boldsymbol{\lambda}\left(\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}^{1/2}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}^{1/2}\right) \tag{L.6}$$

and

$$\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}^{1/2}\left(\boldsymbol{I} - \boldsymbol{p}_1\boldsymbol{p}_1^\dagger\right)\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}^{1/2} \preceq \boldsymbol{S}_{\bar{\boldsymbol{\beta}}}, \tag{L.7}$$

from [74, Theorem A.7.4], it follows that

$$\lambda_i(\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}) \geq \lambda_i(\boldsymbol{S}_{\boldsymbol{\beta}_1}), \quad i = 1, \ldots, N. \tag{L.8}$$

As a consequence, it follows that

$$\begin{aligned}\operatorname{tr}\left\{\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}\boldsymbol{S}_{\boldsymbol{\beta}_1}\right\} - \log\det\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}\right) &= \sum_{i=1}^{N}\lambda_{\bar{\boldsymbol{\beta}}_i}\lambda_i(\boldsymbol{S}_{\boldsymbol{\beta}_1}) - \log\det\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}\right) \\ &\leq \sum_{i=1}^{N}\lambda_{\bar{\boldsymbol{\beta}}_i}\lambda_i(\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}) - \log\det\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}\right) \\ &= \operatorname{tr}\left\{\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}}\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}\right\} - \log\det\left(\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}}\right),\end{aligned} \tag{L.9}$$

having exploited the inequalities $\lambda_i(\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}) \geq \lambda_i(\boldsymbol{S}_{\boldsymbol{\beta}_1})$, $i = 1, \ldots, N$, and having used (L.2) and (L.3) (the eigenvectors of $\boldsymbol{Y}_{\bar{\boldsymbol{\beta}}}$ are those of $\boldsymbol{S}_{\bar{\boldsymbol{\beta}}}$). Thus, for any $\bar{\boldsymbol{\beta}}$, $\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}, \boldsymbol{\beta}_1\right)$ achieves an objective function lower than or equal to the one achieved by $\left(\bar{\boldsymbol{Y}}_{\bar{\boldsymbol{\beta}}}, \bar{\boldsymbol{\beta}}\right)$. Finally, since the pair $\left(\boldsymbol{Y}_{\boldsymbol{\beta}_1}, \boldsymbol{\beta}_1\right)$

is not worse than the pair $\left(\bar{\boldsymbol{Y}}_{\boldsymbol{\beta}_1}, \boldsymbol{\beta}_1\right)$, in the ML sense, it can be written

$$\hat{\boldsymbol{\alpha}}_{H_1} = \frac{\boldsymbol{R}^{\dagger}\boldsymbol{A}^{-1}\boldsymbol{p}_1}{\|\boldsymbol{A}^{-1}\boldsymbol{p}\|} \tag{L.10}$$

and

$$\hat{\boldsymbol{X}}_{H_1} = \boldsymbol{A}^{-1}\boldsymbol{V}_1 \mathbf{diag}\left(\boldsymbol{\lambda}_1^{\star}\right)\boldsymbol{V}_1^{\dagger}\boldsymbol{A}^{-1}, \tag{L.11}$$

where $\boldsymbol{\lambda}_1^{\star}$ is an optimal solution to the optimization problem $\mathcal{P}(\boldsymbol{d}_1)$.

Bibliography

[1] M. C. Wicks, M. Rangaswamy, R. Adve, and T. D. Hale. Space-Time Adaptive Processing: a Knowledge-Based Perspective for Airborne Radar. *IEEE Signal Processing Magazine*, 23(1):51–65, 2006.

[2] J. Ward. "Space-Time Adaptive Processing for Airborne Radar". Technical Report 1015, December 1994.

[3] E. D'Addio, A. Farina, and F. A. Studer. "Performance Comparison of Optimum and Conventional MTI and Doppler Processors". *IEEE Trans. on Aerospace and Electronic Systems*, 20(6):707–715, November 1984.

[4] I. S. Reed, J. D. Mallett, and L. E. Brennan. "Rapid Convergence Rate in Adaptive Arrays". *IEEE Trans. on Aerospace and Electronic Systems*, 10(6):853–863, November 1974.

[5] E. J. Kelly. "An Adaptive Detection Algorithm". *IEEE Trans. on Aerospace and Electronic Systems*, 22(1):115–127, March 1986.

[6] F. C. Robey, D. R. Fuhrmann, R. Nitzberg, and E. J. Kelly. "A CFAR Adaptive Matched Filter Detector". *IEEE Trans. on Aerospace and Electronic Systems*, 28(1):208–216, January 1992.

[7] A. Farina. *"Antenna-Based Signal Processing Techniques for Radar System"*. Artech House, 1992.

[8] P. Chen, W. L. Melvin, and M. C. Wicks. "Screening Among Multivariate Normal Data". *Journal of Multivariate Analysis*, 69:10–29, 1999.

[9] A. De Maio, Y. Huang, D. P. Palomar, S. Zhang, and A. Farina. "Fractional QCQP With Applications in ML Steering Direction Estimation for Radar Detection". *IEEE Trans. on Signal Processing*, 59(1):172–185, January 2011.

[10] A. Wiesel. "Unified Framework to Regularized Covariance Estimation in Scaled Gaussian Models". *IEEE Trans. on Signal Processing*, 60(1):29–38, January 2012.

[11] A. De Maio, S. De Nicola, L. Landi, and A. Farina. "Knowledge-Aided Covariance Matrix Estimation: a MAXDET Approach". *IET Radar Sonar & Navigation*, 3(4):341–356, August 2009.

[12] B. C. Armstrong, H. D. Griffiths, C. J. Baker, and R. G. White. "Performance of Adaptive Optimal Doppler Processors in Heterogeneous Clutter". *IEE Proc.-Radar Sonar Navig.*, 142(4):179–190, April 1995.

[13] B. Himed and W. L. Melvin. "Analyzing Space-Time Adaptive Processors Using Measured Data". In *Asilomar Conference on Signals, Systems & Computers*, pages 930–935, Pacific Grove (CA), USA, November 1997.

[14] F. Gini (Editor). "Knowledge-Based Systems for Adaptive Radar: Detection, Tracking, and Classification". *IEEE Signal Processing Magazine*, 23(1):14–76, January 2006.

[15] J. R. Guerci and W. L. Melvin (Editors). "Special Section on Knowledge-Aided Sensor Signal and Data Processing". *IEEE Trans. on Aerospace and Electronic Systems*, 42(3):983–1120, July 2006.

[16] C. T. Capraro, G. T. Capraro, A. De Maio, A. Farina, and M. Wicks. "Demonstration of Knowledge-Aided Space-Time Adaptive Processing Using Measured Airborne Data". *IEE Proc.-Radar Sonar Navig.*, 153(6):487–494, December 2006.

[17] K. Gerlach and M. L. Picciolo. "Airborne/Spacebased Radar STAP Using a Structured Covariance Matrix". *IEEE Trans. on Aerospace and Electronic Systems*, 39(1):269–281, January 2003.

[18] A. De Maio, A. Farina, and M. Wicks. "KB-GLRT: Exploiting Knowledge of the Clutter Ridge in Airborne Radar". *IEE Proc.-Radar Sonar Navig.*, 152(6):412–428, December 2005.

[19] P. Stoica, J. Li, X. Zhu, and J. R. Guerci. "On Using A-Priori Knowledge in Space-Time Adaptive Processing". *IEEE Trans. on Signal Processing*, 56(6):2598–2602, June 2008.

[20] S. M. Kay. *"Fundamentals of Statistical Signal Processing, Vol. I, Estimation Theory"*. Prentice Hall PTR, 1993.

[21] F. Barbaresco. "Robust Statistical Processing in Fréchet Metric Space: OS-HDR-CFAR and OS-STAP Processing in Siegel Homogeneous Bounded Domains". In *International Radar Symposium*, Leipzig, Germany, September 2011.

[22] M. Arnaudon, L. Yang, and F. Barbaresco. "Stochastic Algorithms for Computing p-Means of Probability Measures, Geometry of Radar Toeplitz Covariance Matrices and Applications to HR Doppler Processing". In *International Radar Symposium*, Leipzig, Germany, September 2011.

[23] F. Barbaresco. *Matrix Information Geometry*, chapter "Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median", pages 199–255. Springer Link, 2013.

[24] B. I. L. Dryden, A. Koloydenko, and D. Zhou. "Non-Euclidean Statistics for Covariance Matrices, with Applications to Diffusion Tensor Imaging". *The Annals of Applied Statistics*, 3(3):1102–1123, 2009.

[25] C. G. Backman. "Some Recent Developments in RCS Measurements Techniques". *Proceedings of the IEEE*, 53(8):962–972, August 1965.

[26] Q. Li, E. J. Rothwell, K. M. Chen, and D. P. Nyquist. "Scattering Center Analysis of Radar Targets Using Fitting Scheme and Genetic Algorithm". *IEEE Trans. on Antennas and Propagation*, 44(2):198–207, February 1996.

[27] T. T. Moon and P. J. Bawden. "High Resolution RCS Measurements of Boats". *IEE Proceedings-F*, 138(3):218–222, June 1991.

[28] A. Farina, F. Scannapieco, and F. Vinelli. "Target Detection and Classification with Very High Range Resolution Radar". In *International Conference on Radar*, pages 20–25, Versailles, France, April 1989.

[29] A. Farina, F. Scannapieco, and F. Vinelli. *Direct and Inverse Methods in Radar Polarimetry, Part I*, chapter Target Detection and Classification with Polarimetric High Resolution Range Radar, pages 1021–1041. Kluwer Academic Publishers. Boston.

[30] Y. I. Abramovich, V. M. Koshevoy, and V. P. Lavrinenko. "Detection of Spatially-Distributed Targets". *Izvestiia vysshikh uchebnykh zavedenii Radioelektronika*, 19(9):96–98, 1976. English translation by Y. I. Abramovich. Typesetting by G. San Antonio.

[31] P. K. Hughes II. "A High Resolution Radar Detection Strategy". *IEEE Trans. on Aerospace and Electronic Systems*, 19(5):663–667, September 1983.

[32] E. J. Kelly and K. Forsythe. Adaptive Detection and Parameter Estimation for Multidimensional Signal Models. Technical Report 848, Lincoln Laboratories, April 1989.

[33] K. A. Burgess and B. D. Van Veen. "Subspace-Based Adaptive Generalized Likelihood Ratio Detection". *IEEE Trans. on Signal Processing*, 44(4):912–927, April 1996.

[34] S. Bose and A. O. Steinhardt. "Adaptive Array Detection of Uncertain Rank One Waveforms". *IEEE Trans. on Signal Processing*, 44(11):2801–2808, November 1996.

[35] E. Conte, A. De Maio, and G. Ricci. "GLRT-based Adaptive Detection Algorithms for Range-Spread Targets". *IEEE Trans. on Signal Processing*, 49(7):1336–1348, July 2001.

[36] F. Bandiera, A. De Maio, A. S. Greco, and G. Ricci. "Adaptive Radar Detection of Distributed Targets in Homogeneous and Partially Homogeneous Noise Plus Subspace Interference". *IEEE Trans. on Signal Processing*, 57(9):1223–1237, April 2007.

[37] F. Bandiera, O. Besson, D. Orlando, G. Ricci, and L. L. Scharf. "GLRT-Based Direction Detectors in Homogeneous Noise and Subspace Interference". *IEEE Trans. on Signal Processing*, 55(6):2386–2394, June 2007.

[38] C. Hao, J. Yang, X. Ma, C. Hou, and D. Orlando. "Adaptive Detection of Distributed Targets with Orthogonal Rejection". *IET Radar Sonar & Navigation*, 6(6):483–493, 2012.

[39] F. Bandiera and G. Ricci. "Adaptive Detection and Interference Rejection of Multiple Point-Like Radar Targets". *IEEE Trans. on Signal Processing*, 54(12):4510–4518, December 2006.

[40] F. Bandiera, D. Orlando, and G. Ricci. "CFAR Detection of Extended and Multiple Point-Like Targets without Assignment of Secondary Data". *IEEE Signal Processing Letters*, 13(4):240–243, April 2006.

[41] K. Gerlach and M. J. Steiner. "Adaptive Detection of Range Distributed Targets". *IEEE Signal Processing Letters*, 47(7):1844–1851, July 1999.

[42] A. De Maio, A. Farina, and K. Gerlach. "Adaptive Detection of Range Spread Targets with Orthogonal Rejection". *IEEE Trans. on Aerospace and Electronic Systems*, 43(2):738–752, April 2007.

[43] K. Gerlach and M. J. Steiner. "Fast Converging Adaptive Detection of Doppler-Shifted, Range-Distributed Targets". *IEEE Trans. on Signal Processing*, 48(9):2686–2690, September 2000.

[44] P. L. Shui, S. W. Xu, and H. W. Liu. "Range-Spread Target Detection using Consecutive HRRPs". *IEEE Trans. on Aerospace and Electronic Systems*, 47(1):647–665, January 2011.

[45] P. Wang, H. Li, T. R. Kavala, and B. Himed. "Generalised Parametric Rao Test for Multi-Channel Adaptive Detection of Range-Spread Targets". *IET Radar Sonar & Navigation*, 6(5):404–412, 2012.

[46] P. Wang, H. Li, and B. Himed. "A New Parametric GLRT for Multichannel Adaptive Signal Detection". *IEEE Trans. on Signal Processing*, 58(1):317–325, January 2010.

[47] M. Steiner and K. Gerlach. "Fast Converging Adaptive Processors for a Structured Covariance Matrix". *IEEE Trans. on Aerospace and Electronic Systems*, 36(4):1115–1126, October 2000.

[48] J. H. Won, J. Lim, S. J. Kim, and B. Rajaratnam. "Maximum Likelihood Covariance Estimation with a Condition Number Constraint". Technical Report No. 2009-10, Department of Statistics, Stanford University, August 2009. Downloadable at http://statistics.stanford.edu/∼ckirby/techreports/GEN/2009/2009-10.pdf.

[49] L. Vandenberghe, S. Boyd, and S. P. Wu. "Determinant Maximization with Linear Matrix Inequality Constraints". *SIAM Journal on Matrix Analysis and Applications*, 19:499–533, April 1998.

[50] E. Conte, A. De Maio, and G. Ricci. "Recursive Estimation of the Covariance Matrix of a Compound-Gaussian Process and its Application to Adaptive CFAR Detection". *IEEE Trans. on Signal Processing*, 50(8):1908–1915, August 2002.

[51] A. De Maio. "Rao Test for Adaptive Detection in Gaussian Interference with Unknown Covariance Matrix". *IEEE Trans. on Signal Processing*, 55(7):3577–3584, July 2007.

[52] C. D. Richmond. "Performance of a Class of Adaptive Detection Algorithms in Nonhomogeneous Environments". *IEEE Trans. on Signal Processing*, 48(5):1248–1262, May 2000.

[53] W. L. Melvin. "Space-Time Adaptive Radar Performance in Heterogeneous Clutter". *IEEE Trans. on Aerospace and Electronic Systems*, 36(2):621–633, April 2000.

[54] R. Nitzberg. "Application of Maximum Likelihood Estimation of Persymmetric Covariance Matrices to Adaptive Processing". *IEEE Trans. on Aerospace and Electronic Systems*, 16(1):124–127, January 1980.

[55] D. R. Fuhrmann. "Application of Toeplitz Covariance Estimation to Adaptive Beamforming and Detection". *IEEE Trans. on Signal Processing*, 39(10):2194–2198, October 1991.

[56] H. Li, P. Stoica, and J. Li. "Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices". *IEEE Trans. on Signal Processing*, 40(5):1314–1323, May 1999.

[57] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spence. "Positive-Definite Toeplitz Completion in DOA Estimation for Nonuniform Linear Antenna Arrays Part I: Fully Augmentable Arrays". *IEEE Trans. on Signal Processing*, 46(9):2458–2471, September 1998.

[58] E. Conte, M. Lops, and G. Ricci. "Adaptive Detection Schemes in Compound-Gaussian Clutter". *IEEE Trans. on Aerospace and Electronic Systems*, 34(4):1058–1069, October 1998.

[59] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels. "Parametric Adaptive Matched Filter for Airborne Radar Applications". *IEEE Trans. on Aerospace and Electronic Systems*, 36(2):677–692, April 2000.

[60] P. Wang, H. Li, and B. Himed. "A Simplified Parametric GLRT for STAP Detection". In *IEEE International Radar Conference*, pages 1–5, Pasadena, CA, May 2009.

[61] A. L. Kraay and A. B. Baggeroer. "A Physically Constrained Maximum-likelihood Method for Snapshot-Deficient Adaptive Array Processing". *IEEE Trans. on Signal Processing*, 55(8):4048–4063, August 2007.

[62] A. De Maio and A. Farina. "Adaptive Radar Detection: A Bayesian Approach". In *International Radar Symposium*, pages 85–88, Krakow, Poland, May 2006.

[63] A. De Maio, A. Farina, and G. Foglia. "Adaptive Radar Detection: A Bayesian Approach". In *IEEE International Radar Conference*, pages 624–629, Boston (MA), USA, April 2007.

[64] O. Besson, J.-Y. Tourneret, and S. Bidon. "Knowledge-Aided Bayesian Detection in Heterogeneous Environments". *IEEE Signal Processing Letters*, 14(5):355–358, May 2007.

[65] J. R. Guerci and E. Baranoski. "Knowledge-Aided Adaptive Radar at DARPA". *IEEE Signal Processing Magazine*, 23(1):41–50, January 2006.

[66] P. Wang, H. Li, and B. Himed. "A Bayesian Parametric Test for Multichannel Adaptive Signal Detection in Nonhomogeneous Environment". *IEEE Signal Processing Letters*, 17(4):351–354, April 2010.

[67] P. R. Gurram and N. A. Goodman. "Spectral-Domain Covariance Estimation with A Priori Knowledge". *IEEE Trans. on Aerospace and Electronic Systems*, 42(3):1110–1120, July 2006.

[68] W. L. Melvin and G. A. Showman. "An Approach to Knowledge-Aided Covariance Estimation". *IEEE Trans. on Aerospace and Electronic Systems*, 42(3):1021–1042, July 2006.

[69] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero. "Shrinkage Algorithms for MMSE Covariance Estimation". *IEEE Trans. on Signal Processing*, 58(10):5016–5029, October 2010.

[70] O. Ledoit and M. Wolf. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection". *Journal on Empirical Finance*, 10(5):603–621, December 2003.

[71] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo. "Robust Adaptive Beamforming Using Worst-Case Performance Optimization: A Solution to the Signal Mismatch Problem". *IEEE Trans. on Signal Processing*, 51(2):313–324, February 2003.

[72] D. G. Manolakis, V. K. Ingle, and S. M. Kogon. *"Statistical and Adaptive Signal Processing"*. Artech House, 2005.

[73] G. H. Golub and C. F. Van Loan. *"Matrix Computations"*. John Hopkins Studies in Mathematical Sciences, 3rd Edition, 1996.

[74] A. Nemirovski. *"Lectures on Modern Convex Optimization"*. http://www.isye.gatech.edu/faculty-staff/profile.php?entry=an63.

[75] R. A. Horn and C. R. Johnson. *"Matrix Analysis"*. Cambridge University Press, 1985.

[76] *"National Land Cover Data (NLCD)"*. available at http://landcover.usgs.gov/prodescription.asp.

[77] M. Skolnik. *"Radar Handbook"*. Mc Graw Hill, 2008.

[78] R. A. Monzingo and T. W. Miller. *"Introduction to Adaptive Arrays"*. John Wiley Interscience Publication, 1980.

[79] A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson. "An Estimate for the Condition Number of a Matrix". *SIAM Journal on Numerical Analysis*, 16(2):368–375, April 1979.

[80] A. Farina, F. Gini, M. V. Greco, and P. H. Y. Lee. "Improvement Factor for Real Sea-Clutter Doppler Frequency Spectra". *IEE Proc.-Radar Sonar Navig.*, 143(5):341–344, October 1996.

[81] S. I. Ohta. "Barycenters in Alexandrov Spaces of Curvature Bounded Below". Technical report, December 2011. Downloadable at http://www.math.kyoto-u.ac.jp/∼sohta/papers/bary.pdf.

[82] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. "Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices". *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.

[83] A. Aubry, A. De Maio, and V. Carotenuto. "Optimality Claims for the FML Covariance Estimator with respect to two Matrix Norms". *IEEE Trans. on Aerospace and Electronic Systems*, 49(3):2055–2057, 2013.

[84] W. S. Kendall and H. Le. "Limit Theorems for Empirical Fréchet Means of Independent and Non-Identically Distributed Manifold-Valued Random Variables". *Brazilian Journal of Probability and Statistics*, 25(3):323–352, 2011.

[85] A. Bhattacharya and R. Bhattacharya. "Nonparametric Statistics on Manifolds with Applications to Shape Spaces". *Institute of Mathematical Statistics*, 3:282–301, 2008.

[86] W. L. Melvin and M. C. Wicks. "Improving Practical Space-Time Adaptive Radar". In *IEEE International Radar Conference*, pages 48–53, Syracuse, New York, May 1997.

[87] E. Conte, A. De Maio, A. Farina, and G. Foglia. "Design and Analysis of a Knowledge-Aided Radar Detector for Doppler Processing". *IEEE Trans. on Aerospace and Electronic Systems*, 42(3):1058–1079, July 2006.

[88] K. Gerlach. "Outlier Resistant Adaptive Matched Filtering". *IEEE Trans. on Aerospace and Electronic Systems*, 38(3):885–901, July 2002.

[89] V. Monga and M. Rangaswamy. "Rank Constrained ML Estimation of Structured Covariance Matrices with Applications in Radar Target Detection". In *IEEE International Radar Conference*, pages 475–480, Atlanta, GE, May 2012.

[90] H. L. Van Trees. *Detection, Estimation, and Modulation Theory.* John Wiley & Sons, New York, USA, 1968.

[91] L. Vanderberghe, S. Boyd, and S. P. Wu. "Determinant Maximization with Linear Matrix Inequality Constraints". *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.

[92] J. B. Billingsley. *Low-Angle Radar Land Clutter.* Scitech Publishing, 2002.

[93] A. T. James. "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples". *The Annals of Mathematical Statistics*, 35(2):475–501, 1964.

[94] S. Boyd and L. Vandenberghe. *"Convex Optimization"*. Cambridge University Press, 2004.

[95] J. R. Magnus and H. Neudecker. *"Matrix Differential Calculus with Applications in Statistics and Econometrics"*. John Wiley & Sons, 3rd Edition, 2007.

[96] P. Milasevic and G. R. Ducharme. "Uniqueness of the Spatial Median". *The Annals of Statistics*, 15(3):1332–1333, 1987.