Analysis and Comparison of Biased Affine Estimators

Fernando Gama, *Student Member, IEEE*, Daniel Casaglia, and Bruno Cernuschi-Frías, *Senior Member, IEEE*

***Abstract*—Affine biased estimation is particularly useful when there is some a-priori knowledge on the parameters that can be exploited in adverse situations (when the number of samples is low, or the noise is high). Three different affine estimation strategies are discussed, namely the Deepest Minimum Criterion (DMC), the Min-Max (MM), and the Linear Matrix Inequality (LMI) strategies, and closed form expressions are obtained for all of them, for the case when the *a priori* knowledge is given in the form of ellipsoidal constraints on the parameter space, and when the covariance matrix of the unbiased estimator is constant. A relationship between affine estimation and Bayesian estimation of the mean of a multivariate Gaussian distribution with Gaussian prior is established and it is shown how affine estimation theory can help in the choice of the Gaussian prior distribution.**

***Index Terms*—Affine bias, Bayesian estimation, biased estimation, constrained estimators, least-squares methods, nonlinear optimization, parameter estimation, positive definite matrices.**

I. Introduction

IN the problem of parameter estimation, affine estimation has emerged as a technique that takes advantage of the bias-variance tradeoff in order to lower the mean squared error of estimation [1]–[4]. Essentially, the idea is to apply an affine transformation to an unbiased estimator, yielding a biased estimator with affine bias—an affine estimator—in such a way that the overall mean squared error of the estimator is reduced with respect to the unbiased estimator.

There are several techniques for choosing the optimal affine transformation to be applied. Mainly, the Min-Max (MM) technique [5], the Linear Matrix Inequality (LMI) method [6] and the Deepest Minimum Criterion (DMC) strategy [7]. Each technique looks to obtain the affine transformation following different optimality criteria. The MM strategy [5] is a conservative approach that tries to uniformly reduce the mean squared error while the LMI criterion [6] is a more strict strategy that lowers the mean squared error in every possible axis of the parameter space. The DMC [7] tries to lower the minimum value of the mean squared error of the biased estimator while ensuring that the affine estimator dominates over the unbiased estimator. Nonetheless, all three strategies yield admissible estimators. This means that no affine estimator is better than the other affine estimator over the whole region of interest; but all three affine estimators have a better mean squared error performance than the unbiased estimator they transform.

The key aspect for the affine estimators to be better than the unbiased estimator is the inclusion of some a-priori knowledge. This a-priori knowledge is given in the form of a subset of the parameter space to which the true parameter is known to belong to. Usually, the idea of a-priori knowledge is exclusive to Bayesian estimation, where the parameter is considered as a random variable and the a-priori information is given in the form of a prior distribution [8]–[11]. Although the posterior pdf is what it is usually sought in Bayesian estimation, in practice, if useful, only point estimators derived from it are generally used. It has been hinted that these estimators have a relationship with the affine estimator for the case of estimating the mean of a multivariate Gaussian distribution with Gaussian prior [6], [12].

The aim of this paper is manifold. First, an alternative proof of the optimal solution for the DMC problem is given, for the case of a-priori information given as an ellipsoid, and for unbiased estimators of constant covariance matrix (Section III). Second, a complete proof of the optimal solution of the MM strategy is given, also for the same case of ellipsoidal a-priori information and constant covariance matrix unbiased estimators (Section IV). Third, the LMI strategy is generalized to any parameter estimation problem rather than only linear regression problems. Fourth, the optimal solution for the LMI strategy is given, for the case of ellipsoidal a-priori information and constant covariance matrix estimators (Section V). Fifth, an explicit relationship between the Bayesian point estimator of the mean of a multivariate Gaussian distribution with Gaussian prior and affine estimators is obtained (Section VI). This relationship may serve as a bridge connecting both estimation philosophies for this case and will be of great help in the design of the a-priori information in either case. Finally, an illustrative simulated example is used to show the performance of the different affine estimators. This illustrative example is the estimation of a FIR lowpass filter (Section VII).

The general problem of affine biased estimation for the case of constant covariance matrix and ellipsoidal constraints is considered in Section II (for a more general treatment on the subject, see [7]). Also, conclusions can be found in Section VIII.

In the present paper, matrices will be uppercase boldface letters and vectors will be lowercase boldface letters. For any matrix $\mathbf{Z} \in \mathbb{R}^{M \times M}$, $\mathbf{Z}^T$ denotes its transpose and $\mathbf{Z} \geq 0$ denotes that $\mathbf{Z}$ is symmetric positive semidefinite (s.p.s.) and $\mathbf{Z} > 0$ stands for a symmetric positive definite (s.p.d.) matrix. Given two symmetric matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times M}$ the partial ordering defined on the symmetric positive semidefinite cone will be used. This is $\mathbf{X} \geq \mathbf{Y}$ is equivalent to $\mathbf{X} - \mathbf{Y}$ being s.p.s. and $\mathbf{X} > \mathbf{Y}$ is equivalent to $\mathbf{X} - \mathbf{Y}$ being s.p.d.

Manuscript received July 10, 2014; revised October 19, 2014; accepted December 09, 2014. Date of publication December 23, 2014; date of current version January 21, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Marano. This work was partially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, and the University of Buenos Aires, Argentina.

F. Gama and D. Casaglia are with the School of Engineering, University of Buenos Aires, Ciudad de Buenos Aires, Argentina (e-mail: fgama@seas.upenn.edu).

B. Cer nuschi-Frías is with the Instituto Argentino de Matemáticas, IAM, CONICET, and the School of Engineering, University of Buenos Aires, Ciudad de Buenos Aires, Argentina (e-mail: bcf@ieee.org).

Digital Object Identifier 10.1109/TSP.2014.2385663

II. Affine Biased Estimation

Let $\mathbf{h}_U(\mathfrak{X})$ be a given unbiased estimator of the true parameter $\boldsymbol{\theta}$ that generated the samples $\mathfrak{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, independent or not, drawn from some parametrized probability Distribution Function $F_{\mathfrak{X}}(\mathfrak{X}; \boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \mathbb{R}^M$ and $\mathbf{h}_U(\mathfrak{X}) \in \mathbb{R}^M$ so that $\mathbb{E}[\mathbf{h}_U(\mathfrak{X})] = \boldsymbol{\theta}$, see e.g., [8]. It is assumed that the covariance of the unbiased estimator exists for all values of n and $\boldsymbol{\theta}$. Consider the estimator

$$\mathbf{h}_B(\mathfrak{X}) = \mathbf{A}\,\mathbf{h}_U(\mathfrak{X}) + \mathbf{b} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times M}$ is a square $M \times M$ matrix, not necessarily symmetric, and $\mathbf{b} \in \mathbb{R}^M$.

Throughout this paper a given unbiased estimator $\mathbf{h}_U(\mathfrak{X})$ will be considered and, when compared, all affine estimators $\mathbf{h}_B(\mathfrak{X})$ of the form (1) will be transformations of the same unbiased estimator $\mathbf{h}_U(\mathfrak{X})$. Since minimizing squared errors is the problem here, if possible a MVUE should be the preferred unbiased estimator to choose.

The weighted mean squared error MSE of $\mathbf{h}_B(\mathfrak{X}) \in \mathbb{R}^M$ is defined as

$$\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) = \mathbb{E}[(\mathbf{h}_B(\mathfrak{X}) - \boldsymbol{\theta})^T \mathbf{W} (\mathbf{h}_B(\mathfrak{X}) - \boldsymbol{\theta})] \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{M\times M}$ is a *known* s.p.d. matrix that may give different weights to the components of the error vector, as considered in, for example ([10], p. 94, p. 135). Usually, $\mathbf{W}$ is taken to be the identity matrix and this will be assumed from now on as both the MM [5] and the LMI [6] estimators do not consider a weighted mean squared error. For the general case $\mathbf{W} \neq \mathbf{I}$ please refer to [7].

Let $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}) = \mathbb{E}[(\mathbf{h}_U(\mathfrak{X}) - \boldsymbol{\theta})(\mathbf{h}_U(\mathfrak{X}) - \boldsymbol{\theta})^T] \in \mathbb{R}^{M\times M}$ be the covariance matrix of the unbiased estimator $\mathbf{h}_U(\mathfrak{X})$. The matrix $\boldsymbol{\Sigma}_U(\boldsymbol{\theta})$ is, in general, a function of $\boldsymbol{\theta}$ and n. From (2), the mean squared error of the biased estimator is

$$\begin{aligned}\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) = \mathbf{tr}[\mathbf{A}\,\boldsymbol{\Sigma}_U(\boldsymbol{\theta})\,\mathbf{A}^T] \\ +[(\mathbf{I} - \mathbf{A})\boldsymbol{\theta} - \mathbf{b}]^T[(\mathbf{I} - \mathbf{A})\boldsymbol{\theta} - \mathbf{b}].\end{aligned} \tag{3}$$

From now on, it will be considered the practical case when $\det(\mathbf{I} - \mathbf{A}) \neq 0$, see [7]. This implies that there is room for some actual improvement of the biased affine estimator over the unbiased estimator. It is easy to observe that, if $\mathbf{A} = \mathbf{I}$ then, from (3), the MSE of the biased affine estimator will be higher than that of the unbiased estimator unless $\mathbf{b} = 0$, which yields $\mathbf{h}_U(\mathfrak{X}) = \mathbf{h}_B(\mathfrak{X})$ and there is no room for improvement. Generalizing this idea, if $\det(\mathbf{I} - \mathbf{A}) = 0$, then this same issue would arise in some subspace of the parameter space. In practice, should $\det(\mathbf{I} - \mathbf{A}) = 0$, the parameter space may be reformulated so as not to include the subspace where there is no room for improvement. So, from now on, it will be assumed that $\det(\mathbf{I} - \mathbf{A}) \neq 0$. This assumption implies that (3) can be rewritten as

$$\begin{aligned}\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) = \mathbf{tr}[\mathbf{A}\,\boldsymbol{\Sigma}_U(\boldsymbol{\theta})\,\mathbf{A}^T] \\ + (\boldsymbol{\theta} - (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b})^T(\mathbf{I} - \mathbf{A})^T \\ \cdot (\mathbf{I} - \mathbf{A})(\boldsymbol{\theta} - (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}).\end{aligned} \tag{4}$$

The interest is set on obtaining $\mathbf{A}$ and $\mathbf{b}$ that define a biased estimator $\mathbf{h}_B(\mathfrak{X}) = \mathbf{A}\,\mathbf{h}_U(\mathfrak{X}) + \mathbf{b}$ with less MSE than that of $\mathbf{h}_U(\mathfrak{X})$ inside a region of interest $\mathcal{V} \subseteq \mathbb{R}^M$ that will be called the *validation-region*, that is

$$\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) \leq \mathbf{MSE}[\mathbf{h}_U(\mathfrak{X})](\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \mathcal{V} \tag{5}$$

with strict inequality for at least one value of $\boldsymbol{\theta} \in \mathcal{V}$.

In order to obtain a closed form solution, the a-priori information given in the validation region will take a special structure. In [6] it is considered to be a polytope (given by linear restrictions on the possible values of the parameters) and in [5] it is considered to be an ellipsoid. In this paper, the latter subset is considered. In any case, any other subset can be approximated by an ellipsoid, see [6]. Therefore, the a-priori known validation-region $\mathcal{V}$ that will be considered is the closed ellipsoidal-ball, see ([7], def. 1),

$$\mathcal{V} = \mathcal{E}(\mathbf{F}, \boldsymbol{\theta}_F) = \{\boldsymbol{\theta} \in \mathbb{R}^M : (\boldsymbol{\theta} - \boldsymbol{\theta}_F)^T \mathbf{F} (\boldsymbol{\theta} - \boldsymbol{\theta}_F) \leq 1\} \tag{6}$$

with $\mathbf{F} \in \mathbb{R}^{M\times M}$ s.p.d. and $\boldsymbol{\theta}_F \in \mathbb{R}^M$ both known. This region will be called the *validation-ellipsoid*. The validation regions are application-dependendent. However, in many cases, as suggested in [6], some upper and lower bounds on the values of the parameters to be estimated can be obtained. This bounded region can then be covered by a minimum-volume ellipsoid [13], see Section VII.

Also, in order to obtain a closed-form expression for the optimal solutions of all three strategies, it is further assumed that the covariance matrix $\boldsymbol{\Sigma}_U(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$, i.e., $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}) \equiv \boldsymbol{\Sigma}_U$. This case turns out to be quite useful in practice. In general, problems involving the estimation of the mean of Gaussian distributions [8], linear regression models [14], DC level in white noise [15], line fitting problems, range estimation [16], source localization [17], among many others make use of unbiased estimators of constant covariance matrix. The affine estimator developed in [6] only considers a constant covariance matrix unbiased estimator. Otherwise, if the covariance matrix is dependent on the parameter, then a reasonable constant bound to the covariance matrix may exist in the region of interest $\mathcal{V}$. With this assumption the MSE of the affine biased estimator (4) turns into

$$\begin{aligned}&\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) \\ &\quad = v_{\mathbf{A}}(\mathbf{A}) + (\boldsymbol{\theta} - \boldsymbol{\theta}_V)^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A})(\boldsymbol{\theta} - \boldsymbol{\theta}_V)\end{aligned} \tag{7}$$

where

$$v_{\mathbf{A}}(\mathbf{A}) = v_{\mathbf{A}} = \mathbf{tr}[\mathbf{A}\boldsymbol{\Sigma}_U\mathbf{A}^T]. \tag{8}$$

It is observed that (7) is the equation of an upwards paraboloid with axis in the direction of the MSE axis. The least MSE value corresponds to $\boldsymbol{\theta}_V = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ which is the $\boldsymbol{\theta}$-space coordinate of the vertex of the paraboloid, with MSE value $v_{\mathbf{A}}(\mathbf{A})$ given by (8). Paraboloid (7) will be called the *error-paraboloid*.

Considering that now $\mathbf{MSE}[\mathbf{h}_U(\mathfrak{X})](\boldsymbol{\theta}) = \mathbf{tr}[\boldsymbol{\Sigma}_U]$ and that (5) holds for actual improvement, then using (7) it yields

$$v_{\mathbf{A}} + (\boldsymbol{\theta} - \boldsymbol{\theta}_V)^T(\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A})(\boldsymbol{\theta} - \boldsymbol{\theta}_V) \leq \mathbf{tr}[\boldsymbol{\Sigma}_U] \tag{9}$$

for all $\boldsymbol{\theta} \in \mathcal{V}$ which is the condition that has to be satisfied by $\mathbf{A}$ and $\mathbf{b}$. Moreover, assuming that there exists a $\boldsymbol{\theta} \in \mathcal{V}$ for which the inequality in (5) is strict, then $\mathbf{tr}[\mathbf{\Sigma}_U] > v_{\mathbf{A}}$, and then (9) defines the ellipsoid

$$(\boldsymbol{\theta}-\boldsymbol{\theta}_V)^T\frac{(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})}{\mathbf{tr}[\mathbf{\Sigma}_U-\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]}(\boldsymbol{\theta}-\boldsymbol{\theta}_V)\le 1 \qquad (10)$$

for all $\boldsymbol{\theta}\in\mathcal{V}$. This ellipsoid will be called *error-ellipsoid* and it is a closed ellipsoidal-ball given by $\mathcal{E}(\mathbf{H}_{\mathbf{A}},\boldsymbol{\theta}_V)$, where

$$\mathbf{H}_{\mathbf{A}}=\frac{(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})}{\mathbf{tr}[\mathbf{\Sigma}_U-v_{\mathbf{A}}]}\quad\text{and}\quad \boldsymbol{\theta}_V=(\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}. \qquad (11)$$

Hence, the surface of the error-ellipsoid (10) is the intersection of the error-paraboloid (7) with the $\mathbf{tr}[\mathbf{\Sigma}_U]$-hyperplane.

It is important to recall that the only case of practical importance [7] is when the validation-region $\mathcal{V}$ contains a minimal error-spheroid $\mathcal{E}(\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U],\boldsymbol{\theta}_S)$ given by

$$\mathcal{E}(\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U],\boldsymbol{\theta}_S)=\left\{\boldsymbol{\theta}:\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}_S)^T(\boldsymbol{\theta}-\boldsymbol{\theta}_S)}{\mathbf{tr}[\mathbf{\Sigma}_U]}\le 1\right\} \qquad (12)$$

for some $\boldsymbol{\theta}_S\in\mathcal{V}$, see [18]. This is, $\mathcal{E}(\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U],\boldsymbol{\theta}_S)\subseteq\mathcal{V}$, which, in the case that the validation-region is the validation-ellipsoid (6), is equivalent to the following LMI ([7], obs. 3, (19))

$$\mathbf{F}\le\frac{\mathbf{I}}{\mathbf{tr}[\mathbf{\Sigma}_U]}. \qquad (13)$$

Besides its practical importance, (13) is an essential inequality for proving all the optimal solutions for any strategy. The existence of this upper bound is of paramount importance. Therefore, it will be assumed that $\mathcal{E}(\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U],\boldsymbol{\theta}_S)\subseteq\mathcal{V}$ throughout the rest of the paper.

The problem addressed here is if there is an optimal way to choose $\mathbf{A}$ and $\mathbf{b}$ such that (5) is met, for a given validation-region $\mathcal{V}$ that contains a minimal error-spheroid (12), with the restrictions $\det(\mathbf{I}-\mathbf{A})\neq 0$ and $\mathbf{tr}[\mathbf{\Sigma}_U]-v_{\mathbf{A}}>0$ (see [7]). The solution obtained depends on the optimality criterion selected. The most important difference between each criteria is the way the a-priori information is weighted with respect to the estimator based on the samples. It is of utmost importance to remark that the affine estimators that arise from all three strategies are all admissible, so no strategy is better over *all* the validation-region.

In what follows, the case considered will be that in which $\mathbf{\Sigma}_U(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$, that is $\mathbf{\Sigma}_U(\boldsymbol{\theta})=\mathbf{\Sigma}_U>0$. For a geometrical insight see ([7], Fig. 1) and for a geometrical comparison of the DMC and MM strategies applied to one-dimensional parameters, see ([19], Fig. 5).

III. THE DEEPEST MINIMUM OPTIMALITY CRITERION

The goal of the DMC strategy is to reduce the value $v_{\mathbf{A}}$ (8) of the MSE at the vertex of the error-paraboloid (7) as much as possible while condition (5) still holds. This implies that the validation-region has to be contained inside the error-ellipsoid (10), $\mathcal{V}\subseteq\mathcal{E}(\mathbf{H}_{\mathbf{A}},\boldsymbol{\theta}_V)$; otherwise, the affine estimator would not have less MSE than the unbiased estimator over the *whole* region $\mathcal{V}$. See ([7], Fig. 1) for a geometrical insight. The general problem can be formally stated as ([7], DMC-General Form)

DMC-Problem: Let $\mathbf{\Sigma}_U\in\mathbb{R}^{M\times M}$ be a s.p.d. matrix and let $\mathcal{V}$, where $\mathcal{V}\subseteq\mathbb{R}^M$, define a bounded validation-region such that it contains a minimal error-spheroid $\mathcal{E}(\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U],\boldsymbol{\theta}_S)$, see (12), for some spheroid center $\boldsymbol{\theta}_S$. Find $\mathbf{A}\in\mathbb{R}^{M\times M}$ and $\mathbf{b}\in\mathbb{R}^M$ such that the error-ellipsoid $\mathcal{E}(\mathbf{H}_{\mathbf{A}},\boldsymbol{\theta}_V)$, see (10), with $\mathbf{H}$ and $\boldsymbol{\theta}_V$ given by (11), contains the validation-region, $\mathcal{V}$, with $\det(\mathbf{I}-\mathbf{A})\neq 0$ and $\mathbf{tr}[\mathbf{\Sigma}_U-\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]>0$, and the value of the MSE at the vertex of the error-paraboloid (8), $v(\mathbf{A})=\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]$, is minimum.

For the particular case when the validation-region is given by the validation-ellipsoid (6), the following holds.

The optimal value of $\mathbf{b}$ is given by $\mathbf{b}_D^\star=(\mathbf{I}-\mathbf{A}_D^\star)\boldsymbol{\theta}_F$ ([7], Obs. 7), where $\mathbf{A}_D^\star\in\mathbb{R}^{M\times M}$ is the optimal value of $\mathbf{A}$ which can be obtained by solving the next problem ([7], E-DMC-General Form 2)

E-DMC-Problem: Let $\mathbf{\Sigma}_U,\mathbf{F}\in\mathbb{R}^{M\times M}$ be s.p.d. matrices, with $\mathbf{F}\le\mathbf{I}/\mathbf{tr}[\mathbf{\Sigma}_U]$. Find $\mathbf{A}\in\mathbb{R}^{M\times M}$ with $\det(\mathbf{I}-\mathbf{A})\neq 0,\mathbf{tr}[\mathbf{\Sigma}_U-\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]>0$ such that $v_{\mathbf{A}}=\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]$ is minimized, subject to

$$\frac{(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})}{\mathbf{tr}[\mathbf{\Sigma}_U-\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]}\le\mathbf{F}. \qquad (14)$$

This is a convex optimization problem of the form

$$\begin{array}{ll}\text{minimize} & \mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]\\ \text{subject to} & (\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})+\mathbf{F}\,\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T-\mathbf{\Sigma}_U]\le 0\end{array} \qquad (15)$$

The objective function $f_0(\mathbf{A})=\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T]=\mathbf{tr}[\mathbf{\Sigma}_U(\mathbf{A}^T\mathbf{A})]$ is convex because it is the composition of an increasing linear function and another convex function, ([13], ex. 3.46, ex. 3.48). The restriction function $f_1(\mathbf{A})=(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})+\mathbf{F}\,\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T-\mathbf{\Sigma}_U]$ is convex because it is the non-negative sum of two convex functions, ([13], ex. 3.49). This implies that problem (15) is convex [13].

The closed form solution to the DMC strategy is given in ([7], Main Theorem) and it was obtained through geometrical considerations. Here, a different proof will be given, using the Karush-Kuhn-Tucker (KKT) conditions of convex optimization problem (15). The proof given here is more straightforward, but it requires knowledge of the closed form expression. This closed form expression was obtained in a constructive way in [7].

The KKT conditions for problem (15) are given by

KKT Conditions 1: ***DMC Problem.*** The KKT conditions for problem (15) are

$$\mathbf{L}_D^\star=\frac{(\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U}{1-\mathbf{tr}[(\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U\,\mathbf{F}]} \qquad (16)$$

$$\mathbf{tr}[\mathbf{L}_D^\star(\mathbf{I}-\mathbf{A}_D^\star)^T(\mathbf{I}-\mathbf{A}_D^\star)] =\mathbf{tr}[\mathbf{L}_D^\star\mathbf{F}]\mathbf{tr}[\mathbf{\Sigma}_U-\mathbf{A}_D^\star\mathbf{\Sigma}_U\mathbf{A}_D^{\star T}] \qquad (17)$$

$$\mathbf{L}_D^\star\ge 0 \qquad (18)$$

$$(\mathbf{I}-\mathbf{A}_D^\star)^T(\mathbf{I}-\mathbf{A}_D^\star)+\mathbf{F}\mathbf{tr}[\mathbf{A}_D^\star\mathbf{\Sigma}_U\mathbf{A}_D^{\star T}-\mathbf{\Sigma}_U]\le 0 \qquad (19)$$

where $\mathbf{A}_D^\star$ is the optimal value of the variable of the primal problem and $\mathbf{L}_D^\star\in\mathbb{R}^{M\times M}$ is the optimal value of the Lagrange multiplier $\mathbf{L}\in\mathbb{R}^{M\times M}$, $\mathbf{L}$ s.p.s., which is the variable of the dual problem.

Proof: The KKT conditions for problem (15) can be obtained as follows, ([13], p. 267). The Lagrangian is ([13], p. 264)

$$\mathcal{L}(\mathbf{A},\mathbf{L})=\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T] +\mathbf{tr}[\mathbf{L}((\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})+\mathbf{F}\,\mathbf{tr}[\mathbf{A}\mathbf{\Sigma}_U\mathbf{A}^T-\mathbf{\Sigma}_U])]$$

where $\mathbf{L} \in \mathbb{R}^{M\times M}$ is the Lagrange multiplier and, therefore, is s.p.s., ([13], p. 265). Condition (16) is obtained by setting the derivative of the Lagrangian $\mathcal{L}(\mathbf{A},\mathbf{L})$ with respect to $\mathbf{A}$ to zero for the optimal values $(\mathbf{A}_D^\star, \mathbf{L}_D^\star)$. Thus,

$$\frac{\partial}{\partial \mathbf{A}}\mathcal{L}(\mathbf{A},\mathbf{L}) = 2\mathbf{A}\mathbf{\Sigma}_U - 2\mathbf{L} + 2\mathbf{A}\mathbf{L} + \mathbf{tr}[\mathbf{L}\mathbf{F}]\, 2\mathbf{A}\mathbf{\Sigma}_U$$

Setting it equal to zero,

$$\frac{\mathbf{L}_D^\star}{1+\mathbf{tr}[\mathbf{L}_D^\star\mathbf{F}]} = (\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U$$

and, finally, by proposing $\mathbf{L}_D^\star = \alpha_D^\star(\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U, \alpha_D^\star \in \mathbb{R}$ condition (16) is obtained.

Condition (17) is the complementary slackness condition. Condition (18) is the feasibility of the dual problem and finally, condition (19) is the feasibility of the primal problem. □

The KKT conditions 1 are necessary and sufficient conditions the optimal solution must satisfy because the optimization problem (15) is convex and strong duality holds ([13], p. 267).

Theorem 1: ***Solution for the DMC Problem.*** The solution to problem (15) is

$$\begin{aligned}\mathbf{A}_D^\star &= \mathbf{I} - \frac{2\,\mathbf{tr}[(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{1/2}]}{1+\mathbf{tr}[\mathbf{\Sigma}_U\mathbf{F}]}(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{1/2}\mathbf{\Sigma}_U^{-1}\\ &= \mathbf{I} - \beta_D^\star\,(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{1/2}\mathbf{\Sigma}_U^{-1} \end{aligned} \tag{20}$$

Proof: Solution (20) is optimal as proved in ([7], Main Theorem), which is a constructive proof to get to the expression (20). As a double check, an alternative proof is given here by using KKT conditions.

Condition (16) yields the optimal value of the Lagrange multiplier

$$\begin{aligned}\mathbf{L}_D^\star &= \frac{\mathbf{\Sigma}_U(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{-1/2}\mathbf{\Sigma}_U}{\mathbf{tr}[(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{1/2}]} - \frac{2\mathbf{\Sigma}_U}{1+\mathbf{tr}[\mathbf{\Sigma}_U\mathbf{F}]}\\ &= \alpha_D^\star\left(\frac{1}{\beta_D^\star}\mathbf{\Sigma}_U(\mathbf{\Sigma}_U\mathbf{F}\mathbf{\Sigma}_U)^{-1/2}\mathbf{\Sigma}_U - \mathbf{\Sigma}_U\right)\end{aligned} \tag{21}$$

where $\alpha_D^\star = 2/(1+\mathbf{tr}[\mathbf{\Sigma}_U\mathbf{F}]) > 0$.

Condition (17) is verified by replacing the DMC solution (20) and the optimal value of $\mathbf{L}_D^\star$ (21) which yields

$$\begin{aligned}&\mathbf{tr}[\mathbf{L}_D^\star(\mathbf{I}-\mathbf{A}_D^\star)^T(\mathbf{I}-\mathbf{A}_D^\star)]\\ &\quad= \beta_D^{\star^2}\mathbf{tr}[\mathbf{L}_D^\star\mathbf{F}]\\ &\quad= \mathbf{tr}[\mathbf{L}_D^\star\mathbf{F}]\mathbf{tr}\left[\mathbf{\Sigma}_U - \mathbf{A}_D^\star\mathbf{\Sigma}_U\mathbf{A}_D^{\star^T}\right]\end{aligned}$$

For condition (18), note that

$$\mathbf{L}_D^\star = \alpha_D^\star(\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U \tag{22}$$

and that $\alpha_D^\star > 0$ and $(\mathbf{I}-\mathbf{A}_D^\star)^{-1}\mathbf{A}_D^\star\mathbf{\Sigma}_U$ is s.p.d. ([7], Theorem 2), which is obtained from the condition (13) imposed on $\mathbf{F}$.

Finally, condition (19) is immediate by replacing the DMC solution (20).

Hence, it has been checked that the DMC solution (20) satisfies the KKT conditions for the convex optimization problem (15) and therefore the optimal affine transformation according to the DMC criterion is given by $\mathbf{A}_D^\star$ in (20) and by $\mathbf{b}_D^\star = (\mathbf{I}-\mathbf{A}_D^\star)\boldsymbol{\theta}_F$. □

It should be observed that the DMC affine estimator is admissible, as it is not dominated by the Min-Max estimator which is also admissible ([5], p. 3827). This implies that the Min-Max estimator is not better than the DMC estimator over the whole region $\mathcal{V}$, or vice versa. Also, if the unbiased estimator that is transformed is efficient in the Cramér-Rao sense, then the affine biased estimator will achieve the bound on the MSE for biased estimators of deterministic parameters, see ([5], Theorem 1).

IV. The Min-Max Optimality Criterion

The Min-Max (MM) strategy [5] can be viewed as a two step process. First, for each $\mathbf{A}$ and $\mathbf{b}$ determine $\boldsymbol{\theta}_{\mathrm{M}}(\mathbf{A},\mathbf{b})$ which is the value of $\boldsymbol{\theta} \in \mathcal{V}$ that maximizes the MSE (3). Second, choose $\mathbf{A}_M^\star$ and $\mathbf{b}_M^\star$ for which the MSE at $\boldsymbol{\theta}_{\mathrm{M}}(\mathbf{A},\mathbf{b})$ is minimum.

Define $\mathbb{S}(\mathcal{V})$ as the surface of the validation-region. That is, for the case in which the validation-region is the validation-ellipsoid given by (6) it is

$$\mathbb{S}(\mathcal{V}) = \{\boldsymbol{\theta} \in \mathbb{R}^M : (\boldsymbol{\theta}-\boldsymbol{\theta}_F)^T\mathbf{F}(\boldsymbol{\theta}-\boldsymbol{\theta}_F) = 1\}. \tag{23}$$

In the $\boldsymbol{\theta}$-MSE space, (23) defines a cylinder, called the *validation-cylinder*, with axis parallel to the MSE axis and for which the perpendicular section is the validation-ellipsoid (6). Hence, the first step corresponds to the intersection of the validation-cylinder with the error-paraboloid for a given pair $(\mathbf{A},\mathbf{b})$, looking for the value of $\boldsymbol{\theta}$ inside the validation-ellipsoid with highest MSE. That maximum will occur on the surface $\mathbb{S}(\mathcal{V})$ of the validation-ellipsoid. With these definitions the first step of the Min-Max strategy corresponds to:

For each $\mathbf{A} \in \mathbb{R}^{M\times M}$ and $\mathbf{b} \in \mathbb{R}^M$, i.e., $\mathbf{h}_B(\mathfrak{X}) = \mathbf{A}\mathbf{h}_U(\mathfrak{X}) + \mathbf{b}$, determine $\boldsymbol{\theta}_{\mathrm{M}}(\mathbf{A},\mathbf{b}) \in \mathcal{V}$ that maximizes $\mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta})$, given by (3). That is, define f as:

$$\begin{aligned}f(\mathbf{A},\mathbf{b}) &= \max_{\boldsymbol{\theta}\in\mathcal{V}} \mathbf{MSE}[\mathbf{A}\mathbf{h}_U(\mathfrak{X})+\mathbf{b}](\boldsymbol{\theta})\\ &= \max_{\boldsymbol{\theta}\in\mathbb{S}(\mathcal{V})} \mathbf{MSE}[\mathbf{A}\mathbf{h}_U(\mathfrak{X})+\mathbf{b}](\boldsymbol{\theta}).\end{aligned} \tag{24}$$

The second step corresponds to:

$$(\mathbf{A}_M^\star, \mathbf{b}_M^\star) = \underset{\mathbf{A}\in\mathbb{R}^{M\times M},\mathbf{b}\in\mathbb{R}^M}{\arg\min} f(\mathbf{A},\mathbf{b}). \tag{25}$$

Then, the Min-Max strategy seeks an affine estimator that, in a sense, has uniformly deepest MSE on the surface $\mathbb{S}(\mathcal{V})$ of the validation-ellipsoid.

In [5], convex programming techniques [13] are used to find the solution of the Min-Max problem. In this direction, in [5], the Karush-Kuhn-Tucker (KKT) conditions that an optimal solution must satisfy are given for a validation-ellipsoid. In what follows, the closed form solution for this case is obtained, and then, as a double check, the KKT conditions in ([5], (58)) are verified for this solution.

A. Closed Form Solution for the Min-Max Strategy

To obtain the closed form solution the following argument is made. For a given $\mathbf{A}$ and $\mathbf{b}$, consider the hyper-surface $\mathcal{C}$ (curve for $\boldsymbol{\theta} \in \mathbb{R}^2$) obtained from the intersection of the error-paraboloid (7) and the validation-cylinder.

Hence, the first step looks for the $\boldsymbol{\theta}$ with the highest MSE value on that surface (curve). Then, the optimal $\mathcal{C}^\star$ corresponding to $\mathbf{A}_M^\star$ and $\mathbf{b}_M^\star$, with $\mathbf{h}_M(\mathfrak{X}) =$

$\mathbf{A}_M^\star \mathbf{h}_U(\mathfrak{X}) + \mathbf{b}_M^\star$, obtained in the second step should be such that $\mathbf{MSE}[\mathbf{h}_M(\mathfrak{X})](\boldsymbol{\theta}) = \text{constant}, \quad \forall\, \boldsymbol{\theta} \in \mathbb{S}(\mathcal{V})$. If not a better solution $\mathbf{A}', \mathbf{b}'$ would be obtained by slightly shifting, rotating, and/or changing the relative axises of $\mathbf{A}_M^\star$, i.e., changing the error paraboloid, so that $f(\mathbf{A}', \mathbf{b}')$ results in a lower value than $f(\mathbf{A}_M^\star, \mathbf{b}_M^\star)$. Hence, taking into account ([7], obs. 2), the optimal error-paraboloid should be centered at the center of the validation-ellipsoid, and then $\mathbf{b}_M^\star = (\mathbf{I} - \mathbf{A}_M^\star)\, \boldsymbol{\theta}_F$. Since $\mathcal{C}^\star$ is parallel to the $\boldsymbol{\theta}$-hyperplane, i.e., it is equidistant from the validation-ellipsoid, then $\mathcal{C}^\star$ itself is the perpendicular section of the validation-cylinder, and then also it is the perpendicular section of the error-paraboloid. Recall that, by definition, the perpendicular section of the validation-cylinder is the validation-ellipsoid. Then, the Min-Max optimal error-paraboloid will intersect the $\boldsymbol{\Sigma}_U$-hyperplane, in an error-ellipsoid that will have the same relative axises, i.e., ratios of eigenvalues, than $\mathcal{C}^\star$. Since all the perpendicular sections of the error-paraboloid have the same relative axises, i.e., the same ellipsoid shape, then they should have the validation-ellipsoid shape. Hence the shape of the optimal Min-Max error-paraboloid may be obtained from the result of ([7], Theorem 4). Finally, it only remains to determine the size β of the Min-Max optimal error-ellipsoid.

Call $\mathbf{A}_\beta = \mathbf{I} - \mathbf{B}_\beta$, and $\mathbf{b}_\beta = (\mathbf{I} - \mathbf{A}_\beta)\, \boldsymbol{\theta}_F = \mathbf{B}_\beta\, \boldsymbol{\theta}_F$, where $\mathbf{B}_\beta = \beta(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2} \boldsymbol{\Sigma}_U^{-1}$.

Then, the optimal Min-Max value $\beta_M^\star$ may be obtained as:

$$\beta_M^\star = \arg\min_{\beta \in \mathbb{R}} \mathbf{MSE}[\mathbf{A}_\beta \mathbf{h}_U(\mathfrak{X}) + \mathbf{b}_\beta](\boldsymbol{\theta}')$$

for any $\boldsymbol{\theta}' \in \mathbb{S}(\mathcal{V})$.

Since $(\mathbf{I} - \mathbf{A}_\beta)^T (\mathbf{I} - \mathbf{A}_\beta) = \mathbf{B}_\beta^T \mathbf{B}_\beta = \beta^2 \mathbf{F}$, and $\mathbf{tr}[\mathbf{A}_\beta \boldsymbol{\Sigma}_U \mathbf{A}_\beta^T] = \mathbf{tr}[\boldsymbol{\Sigma}_U - 2\mathbf{B}_\beta \boldsymbol{\Sigma}_U + \mathbf{B}_\beta \boldsymbol{\Sigma}_U \mathbf{B}_\beta^T] = \beta^2 \mathbf{tr}[\mathbf{F}\boldsymbol{\Sigma}_U] - 2\beta \mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}] + \mathbf{tr}[\boldsymbol{\Sigma}_U]$, then from (3), with $\mathbf{b}_\beta = \mathbf{B}_\beta\, \boldsymbol{\theta}_F$, it results $\mathbf{MSE}[\mathbf{A}_\beta\, \mathbf{h}_U(\mathfrak{X}) + \mathbf{b}_\beta](\boldsymbol{\theta}') = \beta^2 (\boldsymbol{\theta} - \boldsymbol{\theta}_F)^T \mathbf{F} (\boldsymbol{\theta} - \boldsymbol{\theta}_F) + \beta^2 \mathbf{tr}[\mathbf{F}\boldsymbol{\Sigma}_U] - 2\beta \mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}] + \mathbf{tr}[\boldsymbol{\Sigma}_U]$. Since for any $\boldsymbol{\theta}' \in \mathbb{S}(\mathcal{V})$, it is $(\boldsymbol{\theta}' - \boldsymbol{\theta}_F)^T \mathbf{F} (\boldsymbol{\theta}' - \boldsymbol{\theta}_F) = 1$, then

$$\beta_M^\star = \arg\min_{\beta \in \mathbb{R}} \{\beta^2 (1 + \mathbf{tr}[\mathbf{F}\boldsymbol{\Sigma}_U]) - 2\beta \mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}] + \mathbf{tr}[\boldsymbol{\Sigma}_U]\}$$

so that

$$\beta_M^\star = \frac{\mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}]}{1 + \mathbf{tr}[\mathbf{F}\boldsymbol{\Sigma}_U]}. \tag{26}$$

Hence, the closed form solution for the Min-Max estimator, under an ellipsoidal constraint (6) with covariance matrix, $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_U$ independent of $\boldsymbol{\theta}$, is given by

$$\mathbf{B}_M^\star = \mathbf{I} - \mathbf{A}_M^\star = \frac{\mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}]}{1 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]} (\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2} \boldsymbol{\Sigma}_U^{-1} \tag{27}$$

and $\mathbf{b}_M^\star = \mathbf{B}_M^\star\, \boldsymbol{\theta}_F$.

Note that the scalar coefficient in (27) is one half the scalar coefficient in $\mathbf{B}_D^\star = \mathbf{I} - \mathbf{A}_D^\star$, with $\mathbf{A}_D^\star$ given by (20). Thus, $\mathbf{B}_M^\star = \frac{1}{2} \mathbf{B}_D^\star$.

B. KKT Conditions for the Min-Max Problem

As a double check, the KKT conditions in ([5], (58)) are verified for the closed form solution given by (27).

The Min-Max problem (24)–(25) is equivalent to convex optimization problem ([5], Lemma 4) when the validation region $\mathcal{V}$ is given by (6). It is also strictly feasible ([5], proposition 3) and, therefore, strong duality holds. Thus, KKT conditions are necessary and sufficient conditions for a solution to be optimal.

In ([5], (58)) the KKT conditions were first obtained for problem ([5], Lemma 4) and then they were solved for $\mathbf{A}_M^\star$ and $\mathbf{b}_M^\star$. Conditions ([5], (58)) are given for a general quadratic covariance matrix unbiased estimator, but here will be used for constant covariance matrix unbiased estimators (obtained by replacing $\mathbf{B}_j = 0, j = 1, \ldots, \ell$; $\mathbf{C}_i = 0, \mathbf{z}_i = 0, i = 1, \ldots, k$; and $\mathbf{A} = \boldsymbol{\Sigma}_U$ in ([5], (22), (24) and (33))).

It has already been proved in the previous section that solution (27), together with $\mathbf{b}_M^\star = \mathbf{B}_M^\star\, \boldsymbol{\theta}_F$ is optimal. As an alternative proof, it can be easily shown that these values satisfy the KKT conditions ([5], (58)) for a constant covariance matrix unbiased estimator.

In particular, the value of the Lagrange Multiplier related variables $\boldsymbol{\Pi}$ and $\mathbf{w}$ are

$$\boldsymbol{\Pi}^\star = \frac{1 + \mathbf{tr}(\boldsymbol{\Sigma}_U \mathbf{F})}{\mathbf{tr}\left((\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{\frac{1}{2}}\right)} \boldsymbol{\Sigma}_U (\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{-\frac{1}{2}} \boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_U + \boldsymbol{\theta}_F {\boldsymbol{\theta}_F}^T \tag{28}$$

$$\mathbf{w}^\star = \boldsymbol{\theta}_F. \tag{29}$$

C. Comparison Between Min-Max and DMC Affine Biased Estimators

For the Min-Max solution, from (27) and (8) it results:

$$v_{\mathbf{A}}(\mathbf{A}_M^\star) = v_{\mathbf{B}}(\mathbf{B}_M^\star) = v_M^\star = \mathbf{tr}[\boldsymbol{\Sigma}_U] - \left(\frac{\mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}]}{1 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]}\right)^2 (2 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]) \tag{30}$$

with error-ellipsoid given by

$$\left\{\boldsymbol{\theta} \in \mathbb{R}^M : (\boldsymbol{\theta} - \boldsymbol{\theta}_V)^T \frac{\mathbf{F}}{2 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]} (\boldsymbol{\theta} - \boldsymbol{\theta}_V) \leq 1\right\} \tag{31}$$

while for the DMC-solution, from (20) and (8) it results

$$v_{\mathbf{A}}(\mathbf{A}_D^\star) = v_{\mathbf{B}}(\mathbf{B}_D^\star) = v_D^\star = \mathbf{tr}[\boldsymbol{\Sigma}_U] - \left(\frac{2\mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}]}{1 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]}\right)^2 \tag{32}$$

with error-ellipsoid coinciding with the validation-ellipsoid (6) ([7], Theorem 4)

$$\{\boldsymbol{\theta} \in \mathbb{R}^M : (\boldsymbol{\theta} - \boldsymbol{\theta}_F)^T \mathbf{F} (\boldsymbol{\theta} - \boldsymbol{\theta}_F) \leq 1\}.$$

Since $0 < \mathbf{F} \leq \mathbf{I}/\mathbf{tr}[\boldsymbol{\Sigma}_U]$ it is $0 < \boldsymbol{\Sigma}_U^{1/2}\, \mathbf{F} \boldsymbol{\Sigma}_U^{1/2} \leq \boldsymbol{\Sigma}_U/\mathbf{tr}[\boldsymbol{\Sigma}_U]$, so that $0 < \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}] = \mathbf{tr}[\boldsymbol{\Sigma}_U^{1/2}\, \mathbf{F} \boldsymbol{\Sigma}_U^{1/2}] \leq 1$. Hence $v_D^\star < v_M^\star$.

The error-paraboloid for the MM estimator is given by

$$v_M^\star + \left(\frac{\mathbf{tr}[(\boldsymbol{\Sigma}_U \mathbf{F} \boldsymbol{\Sigma}_U)^{1/2}]}{1 + \mathbf{tr}[\boldsymbol{\Sigma}_U \mathbf{F}]}\right)^2 (\boldsymbol{\theta} - \boldsymbol{\theta}_V)^T \mathbf{F} (\boldsymbol{\theta} - \boldsymbol{\theta}_V) \tag{33}$$

while for the DMC estimator is given by

$$v_D^\star + \left(\frac{2\,\mathbf{tr}[(\boldsymbol{\Sigma}_U\mathbf{F}\boldsymbol{\Sigma}_U)^{1/2}]}{1+\mathbf{tr}[\boldsymbol{\Sigma}_U\mathbf{F}]}\right)^2 (\boldsymbol{\theta}-\boldsymbol{\theta}_V)^T\mathbf{F}(\boldsymbol{\theta}-\boldsymbol{\theta}_V). \quad (34)$$

The intersection of the two paraboloids defines the *intersection-ellipsoid* given by

$$\frac{3}{2-\mathbf{tr}[\boldsymbol{\Sigma}_U\mathbf{F}]}(\boldsymbol{\theta}-\boldsymbol{\theta}_V)^T\mathbf{F}(\boldsymbol{\theta}-\boldsymbol{\theta}_V) = 1. \quad (35)$$

This intersection is well defined since $0 < \mathbf{tr}[\boldsymbol{\Sigma}_U\mathbf{F}] \leq 1$, as previously discussed.

Since

$$\frac{1}{2+\mathbf{tr}[\boldsymbol{\Sigma}_U\mathbf{F}]}\,\mathbf{F} < \mathbf{F} < \frac{3}{2-\mathbf{tr}[\boldsymbol{\Sigma}_U\mathbf{F}]}\mathbf{F} \quad (36)$$

then the intersection-ellipsoid is inside the validation ellipsoid, which in turn is inside the Min-Max error-ellipsoid.

Hence four regions are defined:

- Region I, outside the Min-Max error ellipsoid. The Min-Max estimator is better that the DMC estimator, but both are worse than the unbiased estimator;
- Region II, inside the Min-Max estimator and outside the validation-ellipsoid. Both the Min-Max estimator and the unbiased estimator are better than the DMC estimator, and the Min-Max estimator is better than the unbiased estimator.

The important situation corresponds to the following cases since the true parameter is assumed to be inside the validation-ellipsoid.

- Region III, inside the validation-ellipsoid and outside the intersection-ellipsoid. Both the DMC and the Min-Max are better than the unbiased estimator, and the Min-Max estimator is better than the DMC estimator.
- Region IV, inside the intersection-ellipsoid. Both the DMC and the Min-Max are better than the unbiased estimator, and the DMC estimator is better than the Min-Max estimator.

Special care should be taken with this classification, in terms of the estimators, since it is valid asymptotically when the number of samples in $\mathfrak{X}$ is large. When the number of samples is finite and small, which is the case of special interest for affine biased estimators, the biased estimate may fall outside the validation region to which the true parameter belongs. This can happen with any strategy DMC, MM or any other. If that is the case some criterion should be set to choose some value of the estimator on the surface of the validation-region. Such a situation requires careful analysis using large deviation theory.

V. THE LMI CRITERION

In [6] the problem of affine estimation is considered for linear regression models. In this section, those ideas are extended to a general deterministic parameter estimation problem, i.e., a sample $\mathfrak{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, independent or not, drawn from some parametrized probability Distribution Function $F_{\mathfrak{X}}(\mathfrak{X};\boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \mathcal{V} \subseteq \mathbb{R}^M$. The affine estimator proposed ([6], (III.1)) has the structure

$$\mathbf{h}_B(\mathfrak{X}) = \mathbf{A}_L\mathbf{h}_U(\mathfrak{X}) + \mathbf{b}_L \quad (37)$$

with $\mathbf{h}_U(\mathfrak{X})$ an unbiased estimator with constant covariance matrix[1] $\boldsymbol{\Sigma}_U$ and with

$$\begin{aligned}\mathbf{A}_L &= \mathbf{A}_L(\bar{\boldsymbol{\Sigma}}) \\ &= \left(\boldsymbol{\Sigma}_U^{-1} + \bar{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\boldsymbol{\Sigma}_U^{-1} \quad (38)\\ \mathbf{B}_L &= \mathbf{B}_L(\bar{\boldsymbol{\Sigma}}) = \mathbf{I} - \mathbf{A}_L = \left(\boldsymbol{\Sigma}_U^{-1} + \bar{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\bar{\boldsymbol{\Sigma}}^{-1} \quad (39)\\ \mathbf{b}_L &= \mathbf{b}_L(\bar{\boldsymbol{\Sigma}},\bar{\boldsymbol{\theta}}) = \mathbf{B}_L(\bar{\boldsymbol{\Sigma}})\bar{\boldsymbol{\theta}}. \quad (40)\end{aligned}$$

where $\bar{\boldsymbol{\Sigma}} \in \mathbb{R}^{M\times M}$ and $\bar{\boldsymbol{\theta}} \in \mathbb{R}^M$ are the design variables. As it will be discussed in Section VI, estimator (37) is first proposed with the structure of the Bayesian estimator of the mean of a multivariate Gaussian random variable with Gaussian prior, hinting at the idea of an existing relationship between affine estimators and Bayesian estimation. Also, it is observed that in the definition of $\mathbf{B}_L$ (39), it was already assumed in [6] that $\mathbf{B}_L\boldsymbol{\Sigma}_U$ has to be s.p.d., the same as in ([7], Theorem 1).

The aim of the LMI strategy is to achieve mean square error matrix domination ([6], p. 1265) over the validation-region $\mathcal{V}$, and not only MSE domination (5). This implies

$$\boldsymbol{\Sigma}_B(\boldsymbol{\theta}) \leq \boldsymbol{\Sigma}_U \quad \forall \boldsymbol{\theta} \in \mathcal{V} \quad (41)$$

with strict inequality for at least one value of $\boldsymbol{\theta} \in \mathcal{V}$. In (41), $\boldsymbol{\Sigma}_B(\boldsymbol{\theta}) \in \mathbb{R}^{M\times M}$ s.p.d. is defined as

$$\begin{aligned}\boldsymbol{\Sigma}_B(\boldsymbol{\theta}) &= \mathbb{E}[(\mathbf{h}_B(\mathfrak{X})-\boldsymbol{\theta})(\mathbf{h}_B(\mathfrak{X})-\boldsymbol{\theta})^T] \quad (42)\\ \boldsymbol{\Sigma}_B(\boldsymbol{\theta}) &= (\mathbf{I}-\mathbf{B})\boldsymbol{\Sigma}_U(\mathbf{I}-\mathbf{B})^T \\ &\quad + \mathbf{B}(\boldsymbol{\theta}-\mathbf{B}^{-1}\mathbf{b})(\boldsymbol{\theta}-\mathbf{B}^{-1}\mathbf{b})^T\mathbf{B}^T \quad (43)\end{aligned}$$

where $\mathbf{B} = \mathbf{I}-\mathbf{A}$ and where the practical restriction that $\det(\mathbf{I}-\mathbf{A}) = \det(\mathbf{B}) \neq 0$ discussed earlier is used. It is observed that $\mathbf{tr}[\boldsymbol{\Sigma}_B(\boldsymbol{\theta})] = \mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta})$ and that $\boldsymbol{\Sigma}_B(\boldsymbol{\theta}) \leq \boldsymbol{\Sigma}_U \Rightarrow \mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) \leq \mathbf{MSE}[\mathbf{h}_U(\mathfrak{X})](\boldsymbol{\theta})$.

Using (43) in (41) together with the Schur-complement based lemma ([6], Lemma 1), then the restriction (41) is equivalent to

$$\begin{aligned}(\boldsymbol{\theta}-\mathbf{B}^{-1}\mathbf{b})^T\mathbf{B}^T\cdot[\boldsymbol{\Sigma}_U - (\mathbf{I}-\mathbf{B})\boldsymbol{\Sigma}_U(\mathbf{I}-\mathbf{B})^T]^{-1} \\ \cdot\mathbf{B}(\boldsymbol{\theta}-\mathbf{B}^{-1}\mathbf{b}) \leq 1 \quad (44)\end{aligned}$$

and this has to be true for every $\boldsymbol{\theta} \in \mathcal{V}$.

For the validation-ellipsoid (6) $\bar{\boldsymbol{\theta}}^\star = \mathbf{B}_L^{\star^{-1}}\mathbf{b}_L^\star = \boldsymbol{\theta}_F$ ([6], Theorem 2). It is observed that if (44) has to be valid for every $\boldsymbol{\theta} \in \mathcal{V}$, then the validation-ellipsoid should be contained by the ellipsoid given by (44). Then, for the validation-ellipsoid, restriction (41) is equivalent to

$$\mathbf{B}_L^{\star^T}[\boldsymbol{\Sigma}_U - (\mathbf{I}-\mathbf{B}_L^\star)^T\boldsymbol{\Sigma}_U(\mathbf{I}-\mathbf{B}_L^\star)]^{-1}\mathbf{B}_L^\star \leq \mathbf{F}. \quad (45)$$

for the optimal $\mathbf{B}_L^\star$.

The optimal value of $\mathbf{B}_L$ can be found by solving the problem ([6], (III.9))

$$\begin{aligned}\underset{\mathbf{B}_L}{\text{minimize}} \quad & \mathbf{tr}\left[\left(\mathbf{B}_L^{-1}-\mathbf{I}\right)\boldsymbol{\Sigma}_U\right] \\ \text{subject to} \quad & \mathbf{B}_L^T[\boldsymbol{\Sigma}_U - (\mathbf{I}-\mathbf{B}_L)^T\boldsymbol{\Sigma}_U(\mathbf{I}-\mathbf{B}_L)]^{-1}\mathbf{B}_L \leq \mathbf{F} \\ & \mathbf{B}_L^{-1}\boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_U \geq 0 \quad (46)\end{aligned}$$

[1] In the linear regression problem, there exists a MVUE estimator with constant covariance matrix [15]. Here the theory is extended to *any* unbiased estimator with constant covariance matrix.

which is valid for the general estimation problem given in Section II, not only for a linear regression model used in [6].

From now on, a practical restriction analogous to (13), see [7], is imposed on the validation-ellipsoid

$$\mathbf{F} \leq \boldsymbol{\Sigma}_U^{-1} \tag{47}$$

which guarantees that the measurements introduce new information, besides that given by the a-priori information given by the validation-ellipsoid and it is worthwhile to perform the estimation based on those measurements.

Note that because $\mathbf{B}_L^{-1}\boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_U$ has to be s.p.s. (second restriction in problem (46)), and the case of $\mathbf{B}_L = \mathbf{I}$ is not of interest here (it is the Hodges-Le Cam estimator, see [7]), then the second restriction implies that the solution $\mathbf{B}_L^\star$ has to be such that $\mathbf{B}_L^{-1}\boldsymbol{\Sigma}_U$ is s.p.d. as in ([7], Theorem 1).

By proposing the invertible change of variables, $\mathbf{X}_L = \boldsymbol{\Sigma}_U^{-1}\mathbf{B}_L$, and noting that looking for matrix $\mathbf{B}_L$ that minimizes $\mathbf{tr}[\mathbf{B}_L^{-1}\boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_U]$ is the same as the one that minimizes $\mathbf{tr}[\mathbf{B}_L^{-1}\boldsymbol{\Sigma}_U]$, then the problem can be reformulated as

LMI-Problem: Let $\boldsymbol{\Sigma}_U, \mathbf{F} \in \mathbb{R}^{M\times M}$ be s.p.d. matrices such that (47) holds. Find $\mathbf{X}_L \in \mathbb{R}^{M\times M}$ s.p.d. that solves the convex optimization problem

$$\begin{aligned} \underset{\mathbf{X}_L}{\text{minimize}} \quad & \mathbf{tr}\left[\mathbf{X}_L^{-1}\right] \\ \text{subject to} \quad & \mathbf{X}_L - 2(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1} \leq 0. \end{aligned} \tag{48}$$

LMI-Problem (48) is convex (SDP) because $\mathbf{tr}[\mathbf{X}_L^{-1}]$ is a convex ([13], p. 116) decreasing function ([13], p. 109), provided that $\mathbf{X}$ is s.p.d. which is guaranteed. Also, the restriction of problem (48) is an affine function of the variable $\mathbf{X}_L$.

The solution to the LMI-Problem (48) can be readily found by geometric arguments. Since $\mathbf{tr}[\mathbf{X}_L^{-1}]$ is a decreasing function, the minimum will be attained at the maximum possible value of $\mathbf{X}_L$ which is

$$\mathbf{X}_L^\star - 2(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1} = 0 \tag{49}$$

because the restriction is an increasing (affine) function of $\mathbf{X}_L$.

A. KKT Conditions for the LMI Criterion Problem

Solution (49) can also be proved optimal because it satisfies the KKT conditions which are necessary and sufficient because the problem is convex and strong duality holds.

KKT Conditions 2: ***LMI Problem.*** The KKT conditions for problem (48) are

$$\mathbf{L}^\star = \mathbf{X}_L^{\star^{-2}} \tag{50}$$
$$\mathbf{tr}\left[\mathbf{X}_L^{\star^{-1}} - 2\mathbf{X}_L^{\star^{-2}}(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1}\right] = 0 \tag{51}$$
$$\mathbf{X}_L^{\star^{-2}} \geq 0 \tag{52}$$
$$\mathbf{X}_L^\star - 2(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1} \leq 0 \tag{53}$$

where $\mathbf{X}_L^\star$ is the optimal $\mathbf{X}_L$ and $\mathbf{L}^\star$ is the optimal Lagrange multiplier $\mathbf{L} \in \mathbb{R}^{M\times M}$.

Proof: Condition (50) is obtained by differentiating the Lagrangian $\mathcal{L}(\mathbf{X}_L, \mathbf{L})$ with respect to $\mathbf{X}_L$ ([13], ex. 3.37) and setting its derivative to zero for optimal $(\mathbf{X}_L^\star, \mathbf{L}^\star)$.

Condition (51) is the complementary slackness condition obtained by using (50) in $\mathbf{L}^\star$.

Conditions (52) and (53) are the feasibility condition of dual problem and primal problem respectively. □

Theorem 2: ***Solution for the LMI Problem.*** The optimal solution for problem (48) is given by (49) and, therefore, the optimal values according to the LMI strategy are

$$\mathbf{B}_L^\star = 2\boldsymbol{\Sigma}_U(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1} \tag{54}$$
$$\mathbf{b}_L^\star = \mathbf{B}_L^\star \boldsymbol{\theta}_F. \tag{55}$$

Proof: The KKT conditions 2 are necessary and sufficient because the problem is convex and strong duality holds, so the optimal solution has to satisfy those conditions. By replacing $\mathbf{X}_L^\star = 2(\mathbf{F}^{-1} + \boldsymbol{\Sigma}_U)^{-1}$ given by (49) in conditions (50)–(53) the result is straightforward.

Given that $\mathbf{X}_L = \boldsymbol{\Sigma}_U^{-1}\mathbf{B}_L$, then by using (49) $\mathbf{B}_L^\star$ is obtained to be (54). Using ([6], Theorem 2) $\bar{\boldsymbol{\theta}}^\star = \mathbf{B}_L^{\star^{-1}}\mathbf{b}_L^\star = \boldsymbol{\theta}_F$ because validation-ellipsoid $\mathcal{V}$ has a center of symmetry and thus, (55) is obtained. □

B. Comparison of the LMI Estimator With the DMC Estimator and the Min-Max Estimator

Geometrically, the most striking difference between the LMI estimator and both the DMC and MM estimators is that, while the latter two have an error-paraboloid (and therefore, an error-ellipsoid) with the same orientation as the validation-ellipsoid, the LMI estimator does not.

The error-paraboloid for the DMC strategy is given by (34) while for the MM strategy, it is (33). It is observed that for these two strategies, the matrix of the error-paraboloid are scaled versions of $\mathbf{F}$.

However, for the LMI estimator, the error paraboloid is obtained by replacing (54) and (55) in (7), recalling that $\mathbf{B} = \mathbf{I} - \mathbf{A}$,

$$\begin{aligned} \mathbf{MSE}[\mathbf{h}_B(\mathfrak{X})](\boldsymbol{\theta}) = {} & v_L^\star \\ & + 4(\boldsymbol{\theta} - \boldsymbol{\theta}_F)^T\mathbf{F}\left(\mathbf{F} + \boldsymbol{\Sigma}_U^{-1}\right)^{-2}\mathbf{F}(\boldsymbol{\theta} - \boldsymbol{\theta}_F) \end{aligned} \tag{56}$$

with $v_L^\star = v_{\mathbf{B}}(\mathbf{B}_L^\star)$, obtained by replacing (54) in (8).

It is clear then, that in order to ensure matrix domination, the error-paraboloid has to take into account the actual matrix $\boldsymbol{\Sigma}_U$ for the problem and change its axis accordingly.

It is also worth noting that both the LMI and the DMC strategies minimize the minimum MSE as much as possible, see Courant-Fischer theorem, ([20], Theorem 4.2.11) and ([21], Theorem 7.7). This implies that, if the DMC problem (15) is solved for a different validation region $\tilde{\mathcal{V}} = \{\boldsymbol{\theta} \in \mathbb{R}^M : (\boldsymbol{\theta} - \boldsymbol{\theta}_F)^T\tilde{\mathbf{F}}(\boldsymbol{\theta} - \boldsymbol{\theta}_F) \leq 1\}$ given by

$$\tilde{\mathbf{F}} = \frac{\mathbf{F}(\mathbf{F} + \boldsymbol{\Sigma}_U^{-1})^{-2}\mathbf{F}}{\mathbf{tr}\left[\boldsymbol{\Sigma}_U\mathbf{F}\left(\mathbf{F} + \boldsymbol{\Sigma}_U^{-1}\right)^{-1}\left(\mathbf{I} - \left(\mathbf{F} + \boldsymbol{\Sigma}_U^{-1}\right)^{-1}\mathbf{F}\right)\right]} \tag{57}$$

then the DMC solution (20) obtained for $\tilde{\mathcal{V}}$ is the same as the LMI solution (54). In other words, the LMI problem is the same as the DMC problem but considering a different validation region $\tilde{\mathcal{V}}$ such that it takes into account the covariance matrix $\boldsymbol{\Sigma}_U$ and achieves matrix domination instead of only MSE domination.

VI. Relationship With Bayes' Estimator

In the Bayesian setting, the inclusion of a-priori knowledge about the unknown parameters is achieved by considering them as *random variables* drawn from a completely known probability distribution, [8], [9]. In the case in which the samples have a normal distribution with a known covariance and the mean is considered as a random variable with a completely known normal distribution, the posterior distribution of the parameter given the samples is normal and its mean is an affine transformation of the sample mean, [8]. This mean corresponds to the MAP estimator as well as the conditional expectation estimator. See [11] for an in depth discussion of this model. Alternatively the Bayesian approach may be considered as the problem of estimating a random variable given a realization of a second random variable, [10], for which the conditional expectation is the optimal estimator that minimizes the mean square error.

There is an interesting relationship between the affine estimators and the Bayes' estimator for the mean of a multivariate Gaussian random variable with Gaussian prior, [8]. In [6] a deterministic parameter context is considered, but an estimator with the structure of this bayesian estimator is used, hinting at the idea that there is a relationship between affine estimators in deterministic parameter setting and Bayesian estimation of the mean of a multivariate Gaussian distribution with Gaussian prior. Also, it is shown in [12] that the Min-Max estimation criterion can be formulated in terms of a Bayesian approach and it is equivalent to choosing the least favorable prior.

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be an M-dimensional real random variable with Gaussian distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^M$ unknown and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$, s.p.d., known. Consider $\boldsymbol{\mu}$ as a random parameter with Gaussian prior distribution, $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\mu}_0 \in \mathbb{R}^M$ known and $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{M\times M}$, s.p.d., known. Let $\mathfrak{X}$ be a sample of size n of the random variable $\mathbf{X}$, then the posterior distribution of $\boldsymbol{\mu}$ given the sample is [8],

$$\boldsymbol{\mu} \,|\, \mathfrak{X} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_n, \boldsymbol{\Sigma}_n) \tag{58}$$

with

$$\hat{\boldsymbol{\mu}}_n = \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \mathbf{m}_n + \frac{\boldsymbol{\Sigma}}{n}\left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \boldsymbol{\mu}_0 \tag{59}$$

and

$$\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \frac{\boldsymbol{\Sigma}}{n} \tag{60}$$

where

$$\mathbf{m}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i. \tag{61}$$

For this case, both the conditional expectation estimator, $\mathbb{E}[\boldsymbol{\mu} \,|\, \mathfrak{X}]$ and the MAP estimator coincide and are equal to $\hat{\boldsymbol{\mu}}_n$ given by (59).

As such, (59) may be considered as another estimate of $\boldsymbol{\theta}$, abstracting from how it was obtained. It is clear that this estimator is strongly consistent, and then weakly consistent, asymptotically unbiased, and asymptotically efficient in the Cramér-Rao sense.

For finite n it is biased and the question of minimum variance must be analyzed, and will be discussed below.

The rationale of Bayes estimation indicates that for small n the a-priori knowledge should help with respect to the lack of enough samples, and that for large n, the Bayesian estimator should become the MSE estimator [8] which is indeed the case for (59).

It is readily observed that the estimator $\hat{\boldsymbol{\mu}}_n$ (59) is an affine transformation of the unbiased estimator $\mathbf{m}_n$ (61). Equation (59) shows that the form of the Bayesian estimator for i.i.d. Gaussian samples with Gaussian prior is of the form

$$\hat{\boldsymbol{\mu}}_n = \mathbf{A}_n \mathbf{m}_n + \mathbf{b}_n \tag{62}$$

with $\mathbf{A}_n = \boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n})^{-1}$ and $\mathbf{b}_n = \frac{\boldsymbol{\Sigma}}{n}(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n})^{-1}\boldsymbol{\mu}_0$, where $\mathbf{h}_U(\mathfrak{X}) = \mathbf{m}_n$ is the unbiased estimator of $\boldsymbol{\mu} = \boldsymbol{\theta}$, with covariance matrix $\boldsymbol{\Sigma}_U = \boldsymbol{\Sigma}/n$ independent of $\boldsymbol{\theta}$.

As it is, this estimator is a biased estimator and, its mean squared error considering that a specific realization $\boldsymbol{\mu}$ generated the n samples is, using (4),

$$\begin{aligned}
&\mathbf{MSE}[\hat{\boldsymbol{\mu}}_n \,|\, \boldsymbol{\mu}] \\
&= \mathbf{tr}\left[\boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \frac{\boldsymbol{\Sigma}}{n} \left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \boldsymbol{\Sigma}_0\right] \\
&\quad + (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \left(\frac{\boldsymbol{\Sigma}}{n}\left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1}\right)^T \\
&\quad \cdot \left(\frac{\boldsymbol{\Sigma}}{n}\left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1}\right)(\boldsymbol{\mu} - \boldsymbol{\mu}_0).
\end{aligned} \tag{63}$$

Since $\mathbf{B}_n \equiv \mathbf{I} - \mathbf{A}_n = \frac{\boldsymbol{\Sigma}}{n}(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n})^{-1}$ is such that $\mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n} = \frac{\boldsymbol{\Sigma}}{n}(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n})^{-1}\frac{\boldsymbol{\Sigma}}{n}$ is s.p.d. and since

$$\begin{aligned}
\boldsymbol{\Sigma}_0 > 0 &\Leftrightarrow 0 < \frac{\boldsymbol{\Sigma}}{n} < \boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n} \\
&\Leftrightarrow 0 < \left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1} < \left(\frac{\boldsymbol{\Sigma}}{n}\right)^{-1} \\
&\Leftrightarrow 0 < \frac{\boldsymbol{\Sigma}}{n}\left(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n}\right)^{-1}\frac{\boldsymbol{\Sigma}}{n} < \frac{\boldsymbol{\Sigma}}{n} \\
&\Leftrightarrow 0 < \mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n} < \frac{\boldsymbol{\Sigma}}{n}
\end{aligned} \tag{64}$$

then, using theorem ([7], Theorem 2), $\mathbf{tr}[\boldsymbol{\Sigma}/n] > \mathbf{tr}[2\,\mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n} - \mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n}\mathbf{B}_n^T] > 0$ so that the corresponding error-ellipsoid (10) is given by the s.p.d. matrix

$$\mathbf{F}_n = \frac{\mathbf{B}_n^T\mathbf{B}_n}{\mathbf{tr}\left[2\,\mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n} - \mathbf{B}_n\frac{\boldsymbol{\Sigma}}{n}\mathbf{B}_n^T\right]} \tag{65}$$

with center $\boldsymbol{\theta}_F = \boldsymbol{\mu}_0$ (as observed from (63); see also ([6], Theorem 2)).

Conversely, if a validation-ellipsoid $\mathbf{F}_n \leq \mathbf{I}/\mathbf{tr}(\boldsymbol{\Sigma}_U)$ is given for some fixed n, and a $\boldsymbol{\Sigma}_0$ in the Bayesian setting is to be found, such that the corresponding error-ellipsoid for the estimator given by (59) coincides with the validation-ellipsoid $\mathbf{F}_n$, then, equivalently $\mathbf{B}_n = \frac{\boldsymbol{\Sigma}}{n}(\boldsymbol{\Sigma}_0 + \frac{\boldsymbol{\Sigma}}{n})^{-1}$ must be found. Hence, ([7], Theorem 4) is used to obtain $\mathbf{B}_n = \beta_n(\frac{\boldsymbol{\Sigma}}{n}\mathbf{F}_n\frac{\boldsymbol{\Sigma}}{n})^{1/2}(\frac{\boldsymbol{\Sigma}}{n})^{-1}$ where

$$\beta_n = \frac{2\mathbf{tr}\left[\frac{\boldsymbol{\Sigma}}{n}\mathbf{F}_n\frac{\boldsymbol{\Sigma}}{n}\right]}{1 + \mathbf{tr}\left[\frac{\boldsymbol{\Sigma}}{n}\mathbf{F}_n\right]} \tag{66}$$

so that

$$\frac{\mathbf{\Sigma}}{n}\left(\mathbf{\Sigma}_0+\frac{\mathbf{\Sigma}}{n}\right)^{-1}\frac{\mathbf{\Sigma}}{n}=\mathbf{B}_n\frac{\mathbf{\Sigma}}{n}$$
$$=\beta_n\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{1/2}\left(\frac{\mathbf{\Sigma}}{n}\right)^{-1}\frac{\mathbf{\Sigma}}{n}$$
$$=\beta_n\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{1/2}.$$

Hence

$$\left(\mathbf{\Sigma}_0+\frac{\mathbf{\Sigma}}{n}\right)^{-1}=\beta_n\left(\frac{\mathbf{\Sigma}}{n}\right)^{-1}\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{1/2}\left(\frac{\mathbf{\Sigma}}{n}\right)^{-1}$$

and then

$$\mathbf{\Sigma}_0=\frac{1}{\beta_n}\frac{\mathbf{\Sigma}}{n}\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{-1/2}\frac{\mathbf{\Sigma}}{n}-\frac{\mathbf{\Sigma}}{n}. \tag{67}$$

But, by ([7], Theorem 4)

$$\mathbf{B}_n\frac{\mathbf{\Sigma}}{n}=\beta_n\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{1/2}<\frac{\mathbf{\Sigma}}{n}$$

then, using the fact that for arbitrary s.p.d. square matrices $\mathbf{Z},\mathbf{Y}\in\mathbb{R}^{M\times M}$, then $0<\mathbf{Z}\leq\mathbf{Y}$ if and only if $0<\mathbf{Y}^{-1}\leq\mathbf{Z}^{-1}$ ([20], Corollary 7.7.4),

$$\left(\frac{\mathbf{\Sigma}}{n}\right)^{-1}<\frac{1}{\beta_n}\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{-1/2}$$
$$\Rightarrow\quad\frac{\mathbf{\Sigma}}{n}=\frac{\mathbf{\Sigma}}{n}\left(\frac{\mathbf{\Sigma}}{n}\right)^{-1}\left(\frac{\mathbf{\Sigma}}{n}\right)$$
$$<\frac{1}{\beta_n}\frac{\mathbf{\Sigma}}{n}\left(\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}\right)^{-1/2}\frac{\mathbf{\Sigma}}{n} \tag{68}$$

so that using (68) in (67), it results $\mathbf{\Sigma}_0>0$, i.e., $\mathbf{\Sigma}_0$ is s.p.d.

Hence, (67) permits to find the covariance of the prior Gaussian density from the geometrical knowledge of the validation-region. Note that when $\mathbf{\Sigma}$ does not commute with $\mathbf{F}_n$, (67) gives a non-trivial solution for $\mathbf{\Sigma}_0$. Then, if geometrical a-priori knowledge about $\mathbf{F}_n$, equivalently the validation-region, is known, the above analysis shows how (67) may become a useful tool to help to choose the parameters of the Gaussian prior distribution for the estimation of the mean of a Guassian random variable with known covariance matrix. In other words, if the a-priori knowledge is given as an ellipsoid in the $\boldsymbol{\theta}$-space and Bayesian estimation methods are to be used, then (67) can be used to convert from known matrix $\mathbf{F}$ to prior covariance matrix $\mathbf{\Sigma}_0$. Conversely, if the a-priori information is given in the form of a Gaussian prior distribution but a deterministic setting is to be used, then (65) can be used to obtain the matrix $\mathbf{F}$ of a validation-ellipsoid from the known prior covariance matrix $\mathbf{\Sigma}_0$. This shows an explicit link between the Gaussian prior distribution given in a Bayesian setting for the estimation of the mean of a Gaussian random variable and the validation-ellipsoid for a corresponding DMC estimator in a deterministic setting. Nonetheless, it is necessary to remark that in a Bayesian context, the main interest is set in the whole posterior probability density function of the parameter and not only in estimator (59).

This can be readily generalized for any $\beta_n>0$ (66). That is, for any $\beta_n>0$ a suitable $\mathbf{\Sigma}_0$ can be obtained through (67). Each β_n will give a different way in which the a-priori information is considered. For instance, the MM estimator can also be converted to the Bayesian estimator by taking $\beta_n=(\mathbf{tr}[\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n\frac{\mathbf{\Sigma}}{n}])/(1+\mathbf{tr}[\frac{\mathbf{\Sigma}}{n}\mathbf{F}_n])$.

The final remark is that (65)–(67) is not the only way to link Bayesian estimation of the mean of a Gaussian random variable with known covariance matrix and Gaussian prior with deterministic estimation. The affine estimator structure proposed in [6] is, indeed, the structure of a Bayesian estimator with prior covariance matrix $\bar{\mathbf{\Sigma}}$. Then, by using the LMI solution (54) together with (39), a different link between geometrical a-priori knowledge and the Gaussian prior distribution can be established.

All in all, in this section, several ways to establish connections between a-priori knowledge given in the form of geometrical knowledge of the validation-region and given in the form of a prior Gaussian density have been explicitly obtained.

VII. Illustrative Example

As an illustrative example, the problem of estimating a discrete-time lowpass FIR filter is considered. The idea is to exploit the special structure of discrete-time LTI lowpass FIR filters designed by windowing methods in order to improve the estimation over the unbiased estimator considered. This is achieved through the use of a-priori information in affine estimators.

A lowpass causal and stable FIR filter $\{h(k)\}_{k=0}^{r-1}$, designed by the windowing method, is obtained as [22]

$$h(k)=\frac{\sin\left(\omega_c(k-\frac{r-1}{2})\right)}{\pi(k-\frac{r-1}{2})}w(k),\ k=0,1,\ldots,r-1 \tag{69}$$

where $\omega_c\in(0,\pi)$ is the cutoff frequency, $\{w(k)\}_{k=0}^{r-1}$ is the chosen window, r is the length of the window (and consequently, of the FIR filter) and where a delay of $(r-1)/2$ is introduced in order to make the filter causal. This delay introduces a linear phase. Linear phase filters are very useful because they do not distort the desired filtered signal, they only delay it [22]. The window sequence $\{w(k)\}_{k=0}^{r-1}$ is usually chosen such that $0\leq w(k)\leq 1, k=0,1,\ldots,r-1$ and also such that it is symmetric around $(r-1)/2$. This implies that the filter $\{h(k)\}$ is a generalized linear phase FIR filter, symmetric, and will be a type I filter if r is odd or a type II filter if r is even [22]. In this example, it will be considered that r is odd.

The aim of this illustrative example is to estimate the sequence $\{h(k)\}$ using a known input sequence and measuring the output of the system. Let $\mathfrak{X}=\{x(k)\}_{k=0}^{n-1}$ be a set of measurements, with $x(k)=\sum_{t=0}^{r-1}h(t)u(k-t)+v(k), k=0,1,\ldots,n-1$, where $\{u(k)\}$ is the known input sequence and where $\{v(k)\}$ is the measurement noise assumed to be Gaussian and white, with zero mean and σ_v^2 variance.

The following linear regression model can be set up

$$\mathbf{x}=\mathbf{U}\,\boldsymbol{\theta}+\mathbf{v} \tag{70}$$

with $\mathbf{x}=[x(0)\ \ldots\ x(n-1)]^T\in\mathbb{R}^n,\boldsymbol{\theta}=[h(0)\ \ldots\ h(\frac{r-1}{2})]^T\in\mathbb{R}^M$ is the parameter vector to be estimated and it comprises only of $M=(r-1)/2+1$

filter coefficients due to the symmetry of type I filters, $\mathbf{v} = [v(0) \quad \dots \quad v(n-1)]^T \in \mathbb{R}^n$ and

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}^T(0) \\ \vdots \\ \mathbf{u}^T(n-1) \end{bmatrix} \tag{71}$$

with

$$\mathbf{u}(k) = \begin{bmatrix} u(k) + u(k-r+1) \\ u(k-1) + u(k-r+2) \\ \vdots \\ u\left(k - \frac{r-1}{2} + 1\right) + u\left(k - \frac{r-1}{2} - 1\right) \\ u\left(k - \frac{r-1}{2}\right) \end{bmatrix} \in \mathbb{R}^M \tag{72}$$

where the matrix $\mathbf{U} \in \mathbb{R}^{n \times M}$ has been adapted to exploit the filter symmetry.

There exists an unbiased estimator for $\boldsymbol{\theta}$ given by $\mathbf{h}_U(\mathfrak{X}) = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\,\mathbf{x} \in \mathbb{R}^M$ with constant covariance matrix $\boldsymbol{\Sigma}_U = \sigma_v^2(\mathbf{U}^T\mathbf{U})^{-1}$ that achieves the Cramér-Rao bound [15]. In this example, this will be the unbiased estimator considered and all three affine estimators will be transformations of this same unbiased estimator.

For including the a-priori information it is observed that the filter sequence $\{h(k)\}$ is upper and lower bounded by the sequences $\{h_H(k)\} = \{\frac{\omega_H}{\pi|k-(r-1)/2|}\}$ and $\{h_L(k)\} = \{-\frac{\omega_H}{\pi|k-(r-1)/2|}\}$ with $h_H((r-1)/2) = \omega_H/\pi$ and $h_L((r-1)/2) = \omega_L/\pi$; and where some bounds on the cutoff frequency have been assumed $\omega_L < \omega_c < \omega_H$; $\omega_L, \omega_H \in [0, \pi]$, both known. If not, then the natural bounds on the cutoff frequency can be used, $\omega_L = 0$ and $\omega_H = \pi$. These bounds hold as long as the window used for design meets the conditions mentioned before. Now, each coefficient of the filter is bounded to be in between two values $h_L(k) \leq h(k) \leq h_H(k)$. These linear constraints represent a polytope (an hyperrectangle) in the parameter space $\boldsymbol{\theta}$. This is the convex validation-region $\mathcal{V}$ needed for the affine estimators. For the closed form solutions of these estimators to be used, then the minimum volume ellipsoid that contains the hyperrectangle can be found, and this ellipsoid will be the validation ellipsoid. The center vector $\boldsymbol{\theta}_F$ will be the lowpass FIR filter with cutoff frequency $(\omega_L + \omega_H)/2$ and matrix $\mathbf{F}$ is obtained by solving a convex optimization problem ([13], p. 222).

For the simulations, it will be considered that $r = 17, M = 9, n = 15, \omega_L = \pi/6$ and $\omega_H = 5\pi/6$.

The first simulation is carried out as a function of the noise variance σ_v^2 which take values in the set $\{0.0251,\ 0.0158,\ 0.0100,\ 0.0063,\ 0.0040,\ 0.0025, 0.0016, 0.0010\}$. The true value of the filter considered is $\omega_c = 0,3\pi$ and the filter was designed with a Blackman-Harris window [22]. The input sequence $\{u(k)\}$ is a pseudo random noise (PRN) input sequence, such that $u(k) = 0, k < 0$. For each value of the variance, 1000 iterations are carried out in order to obtain an estimate of the mean squared errors. The idea of this simulation is to illustrate the performance of the estimators as a function of the variance. Results are shown in Fig. 1.

First of all, it is observed that all three affine estimators are better than the unbiased estimator, in terms of MSE, especially

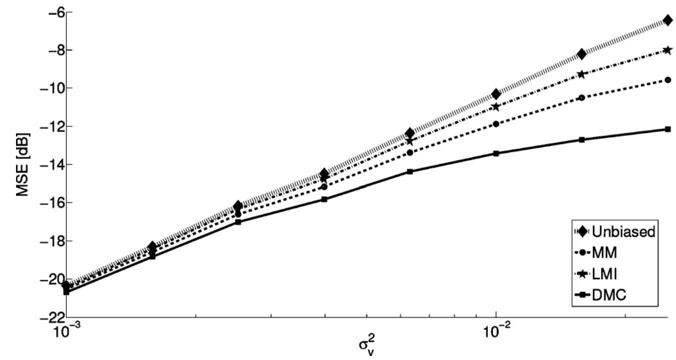


Fig. 1. Simulation results for $\mathbf{MSE}$ in the case when $\omega_c = 0,3\pi$ is the true cutoff frequency, as a function of σ_v^2.

TABLE I
MONTE CARLO SIMULATION. IMPROVEMENT FOR $\sigma_v^2 = 0.0158$, FOR AN IMPULSE INPUT SEQUENCE

Estimator	Best	Worst	Average	Median
MM	45%	29%	35%	35%
LMI	26%	11%	17%	16%
DMC	70%	42%	55%	55%

for the adverse case of high noise variance. As the variance lowers, all three affine estimators converge to the unbiased estimator. This happens because when the noise variance is high, the measurements are unreliable and the affine estimator weighs the a-priori information more important than the measurement-based unbiased estimator. For the highest noise variance, i.e., $\sigma_v^2 = 0.0251$, the DMC estimator improvement over the unbiased estimator is 73%, the MM improvement is 52%, and the LMI improvement is 30%. This improvement is measured as $[\mathbf{MSE}(\mathbf{h}_U(\mathfrak{X})) - \mathbf{MSE}(\mathbf{h}_B(\mathfrak{X}))]/\mathbf{MSE}(\mathbf{h}_U(\mathfrak{X})) \cdot 100\%$.

For the second simulation, a Monte Carlo analysis is carried out. For a fixed noise variance of $\sigma_v^2 = 0.0158, 300$ cutoff frequencies are selected at random from the interval $[\omega_L; \omega_H]$ and 1000 iterations are run for each of the 300 cutoff frequencies selected, with unit-impulse input sequence. The improvement for the worst performance realization, the best performance realization, the average and the median performance for each estimator can be found in Table I. It can be observed that all affine estimators perform better than the unbiased estimator.

VIII. CONCLUSION

This paper has addressed several aspects of affine estimation under ellipsoidal constraints, and how this is related to Bayesian estimation and the choice of a prior distribution.

Three different strategies, namely the DMC, the MM and the LMI criteria, have been discussed, analyzed and compared. Closed form expressions for all three strategies have been obtained for the case when the validation-region is a known validation-ellipsoid and when the covariance matrix of the unbiased estimator is independent of the parameter to be estimated. The explicit expression of the intersection-ellipsoid that determines which estimator between the DMC and the Min-Max is better within each subregion of the validation-ellipsoid has been obtained. Also, it has been observed that the LMI criterion rotates the error-paraboloid according to the covariance matrix of the unbiased estimator in order to ensure matrix domination. Finally, it was shown that both the DMC and the LMI seek to minimize the minimum MSE but for different validation regions. The relationship between these validation regions was explicitly obtained.

The relationship between the affine estimators and the Bayesian estimation of the mean of a multivariate Gaussian distribution with Gaussian prior has been discussed. It was shown that the theory of affine estimation can be of great help when choosing a suitable Gaussian prior and how a-priori information can be related between different estimation contexts. This may serve as a practical bridge connecting both estimation philosophies for this case.

Finally, an illustrative simulated example estimating a FIR lowpass filter has been used to show the characteristics of the three affine estimation strategies.

REFERENCES

[1] S. Uhlich and B. Yang, "Efficient recursive estimators for a linear, time-varying Gaussian model with general constraints," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4910–4915, Sep. 2010.

[2] B. Demissie and S. Kreuzer, "Performance predictions for parameter estimators that minimize cost-functions using Wirtinger calculus with application to cm blind equalization," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3685–3698, Aug. 2011.

[3] C. Hametner and S. Jakubek, "Nonlinear identification with local model networks using GTLS techniques and equality constraints," *IEEE Trans. Neural Netw.*, vol. 22, no. 9, pp. 1406–1418, Sep. 2011.

[4] E. Stathakis, J. Jalden, L. K. Rasmussen, and M. Skoglund, "Uniformly improving maximum-likelihood SNR estimation of known signals in Gaussian channels," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 156–167, Jan. 2014.

[5] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3824–3836, Aug. 2008.

[6] A. Benavoli, L. Chisci, and A. Farina, "Estimation of constrained parameters with guaranteed MSE improvement," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1264–1274, Apr. 2007.

[7] B. Cernuschi-Frías, F. Gama, and D. Casaglia, "Deepest minimum criterion for biased affine estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2437–2449, May 2014.

[8] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, 1st ed. New York: Wiley, 1973.

[9] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, 1st ed. Reading, MA, USA: Addison-Wesley, 1973.

[10] B. O. Anderson and J. B. Moore, *Optimal Filtering*, 1st ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[11] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, 1st ed. Piscataway, NJ, USA: Wiley/ IEEE, 2007.

[12] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Statist. Plan. Inference*, vol. 13, pp. 297–318, 1986.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[14] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[16] W. C. Knight, R. G. Pridham, and S. M. Kay, "Digital signal processing for sonar," *Proc. IEEE*, vol. 69, no. 11, pp. 1451–1506, Nov. 1981.

[17] A. Belouchrani, M. G. Amin, N. Thirion-Moreau, and Y. Zhang, "Back to results source separation and localization using time-frequency distributions: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 97–107, Nov. 2013.

[18] L. Le Cam, "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates," *Univ. Calif. Publ. Statist.*, vol. 1, no. 11, pp. 277–330, 1953.

[19] F. Gama, D. Casaglia, and B. Cernuschi-Frías, "Alternative admissible biased estimators," *IEEE Latin Amer. Trans.*, vol. 11, no. 1, pp. 40–47, Feb. 2013.

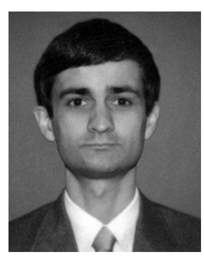
[20] R. A. Horn and C. R. Johson, *Matrix Analysis*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[21] F. Zhang, *Matrix Theory: Basic Results and Techniques*, 1st ed. New York, NY, USA: Springer-Verlag, 1999.

[22] A. V. Oppenheim, R. W. Schafer, and R. J. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

Fernando Gama (S'14) received the electronic engineering degree from the School of Engineering, University of Buenos Aires, Argentina.

He is currently pursuing the Ph.D. degree in the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, in the area of networking theory. He has been awarded a Fulbright Scholarship for international students.



Daniel Casaglia received the Ingeniero Electromecánico Orientación Electrónica degree from the University of Buenos Aires.

He is a researcher at Biological Signal Processing Group, Instituto de Ingeniería Biomédica, FIUBA.

Bruno Cernuschi-Frías (S'77–M'78–SM'01) received the Ph.D. degree in electrical engineering from Brown University, Providence, RI.

He is currently a full tenured Professor with the Department of Electronics, School of Engineering, University of Buenos Aires, Argentina, and Principal Researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).