Estimation of constrained parameters with guaranteed MSE improvement

A. Benavoli, L. Chisci *Member, IEEE*, A. Farina, *Fellow, IEEE*

Abstract

We address the problem of estimating an unknown parameter vector $\mathbf{x}$ in a linear model $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v}$ subject to the a-priori information that the true parameter vector $\mathbf{x}$ belongs to a known convex polytope $\mathbb{X}$. The proposed estimator has the parametrized structure of the *Maximum A-posteriori Probability (MAP)* estimator with prior Gaussian distribution, whose mean and covariance parameters are suitably designed via a *Linear Matrix Inequality* approach so as to guarantee, for any $\mathbf{x} \in \mathbb{X}$, an improvement of the *mean squared error (MSE)* matrix over the *least-squares (LS)* estimator. It is shown that this approach outperforms existing "superefficient" estimators for constrained parameters based on different parametrized structures and/or shapes of the parameter membership region $\mathbb{X}$.

I. Introduction

The problem of estimating a vector of unknown parameters $\mathbf{x}$ in a linear model $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v}$ from noisy observations $\mathbf{y}$, has a wide range of applications in science and engineering. It is well known that, if the measurement noise $\mathbf{v}$ is Gaussian, the *Least-Squares* (LS) (weighted by the inverse covariance of $\mathbf{v}$) or, equivalently, the *Maximum-Likelihood* (ML) estimator, is "efficient" in the sense that its covariance attains the *Cramer-Rao Lower Bound* (CRLB) for unbiased estimators. On the other hand, the LS estimator does not always provide a small *Mean-Squared Error* (MSE), especially for short data samples. For this reason, several so called "*superefficient*" or "*dominating*" estimators [1]-[16] have been developed. Superefficient estimators provide a MSE lower than the CRLB for unbiased estimators exploiting the relation "*MSE=variance + squared bias*" and trading-off bias for variance. Biased estimators have been introduced in order to provide regularized estimates for ill-conditioned estimation problems. Among

Authors' addresses: *A. Benavoli* and *L. Chisci*, DSI, Università di Firenze, Firenze, Italy, e-mail: {benavoli,chisci}@dsi.unifi.it; *A. Farina*, Analysis of Integrated Systems Group, SELEX - Sistemi Integrati S.p.A., Roma, Italy, e-mail: afarina@selex-si.com.

these are the Tikhonov regularization [13], the *ridge* estimator [14], the *shrunken* estimator [15] and the *Covariance Shaping LS* (CSLS) estimator [16].

A lot of work has also been devoted to estimators that dominate the LS estimator [1]-[5], [7]-[12] according to some pre-specified measure of the estimation performance (typically the MSE matrix or its trace), i.e. that provide a better performance than that of the LS estimator for any possible value of the unknown true parameter to be estimated. For linear regression models with Gaussian noise and adopting the MSE criterion, the concepts of superefficiency and LS-domination are clearly equivalent. The common idea on which superefficient or LS-dominating estimators rely is to modify the LS estimate by introducing appropriate design parameters. It is worth pointing out that such estimators achieve superefficiency (LS-domination) provided that the parameters are suitably chosen based on the knowledge of the true parameter vector $\mathbf{x}$ which, in turn, is the quantity to estimate. In many circumstances, some a-priori information on the membership set $\mathbb{X}$ of the unknown parameter vector $\mathbf{x}$ is available, e.g. lower and upper bounds on the components. The knowledge of $\mathbb{X}$ can, therefore, be exploited in order to select the design parameters of the estimator so as to achieve superefficiency. In this respect, Eldar et.al. [8], [9] presented a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ that dominates the LS estimator $\hat{\mathbf{x}}_{LS}$, in the sense that $E\left[\|\mathbf{x}-\hat{\mathbf{x}}\|^2\right] \leq E\left[\|\mathbf{x}-\hat{\mathbf{x}}_{LS}\|^2\right]$, for an ellipsoidal ($\mathbf{x}'\mathbf{T}\mathbf{x} \leq L^2$) membership set $\mathbb{X}$. The estimator's design in [8], [9] is based on a *minimax* (worst-case) approach which guarantees the LS-domination at the expense of some conservativeness on the average performance. In many engineering applications, prior knowledge on the parameters to be estimated is naturally expressed in terms of linear inequality constraints which imply a polytopic membership set $\mathbb{X}$. A recent contribution [12] generalizes the minimax MSE estimator to an arbitrary constraint set $\mathbb{X}$, showing that the convexity of the optimization problem and the LS-domination property (for the MSE's trace) are preserved irrespectively of $\mathbb{X}$.

In the present paper, an affine estimator and a polytopic membership set are considered. In addition, the stronger LS-domination $E\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})'\right] \leq E\left[(\mathbf{x}-\hat{\mathbf{x}}_{LS})(\mathbf{x}-\hat{\mathbf{x}}_{LS})'\right]$ is imposed and the conservative worst-case design approach is avoided. In particular, the proposed estimator is the *Maximum A-posteriori Probability* (MAP) estimator obtained by assuming that $\mathbf{x}$ is a random vector with prior Gaussian distribution, i.e. $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \overline{\boldsymbol{\Sigma}})$. If the unknown vector $\mathbf{x}$ were actually distributed in this way, this would yield the optimal *Minimum MSE* (MMSE) estimate. Since $\mathbf{x}$ is only known to belong to the polytope $\mathbb{X}$, the idea is to design $\bar{\mathbf{x}}$ and $\overline{\boldsymbol{\Sigma}}$, based on $\mathbb{X}$, so as to guarantee the above-mentioned LS-domination property for the matrix MSE. It is shown that this can be done by solving an appropriate *Linear Matrix Inequality* (LMI) problem [18]. It is worth pointing out that, unlike most of the existing literature on LS-dominating estimators, the present work imposes the stronger matrix domination instead of the usual trace domination.

The rest of the paper is organized as follows. Section II states the estimation problem of interest and, in particular, describes the "superefficiency" ("LS-domination") approach to the design of estimators for constrained parameters. Section III presents the LMI-design of a novel estimator for a polytopic parameter membership set. Section IV reviews the minimax MSE estimator with ellipsoidal constraints [9] and presents a modified version of the ridge estimator which takes into account polytopic constraints. In section V these estimators are compared via simulation results. Finally section VI ends the paper.

II. Constrained estimation and superefficiency

Consider the linear regression model

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v} \tag{II.1}$$

where: $\mathbf{x} \in \mathbb{R}^n$ is the variable to be estimated; $\mathbf{y} \in \mathbb{R}^N$ is the observed variable; $\mathbf{C} \in \mathbb{R}^{N \times n}$ is a known matrix; $\mathbf{v} \in \mathbb{R}^N$ is the measurement error assumed to be zero-mean and of covariance $\mathbf{\Sigma}_v$. In this work it will be assumed that the a-priori information on the parameters $\mathbf{x}$ is specified by a membership set $\mathbb{X}$ (hard constraints), i.e. $\mathbf{x} \in \mathbb{X}$. An estimator $\mathbf{f}(\mathbf{y})$ is said *superefficient* or, equivalently, *LS-dominating* in $\mathbb{X}$ if its MSE does not exceed the MSE of the LS estimator whichever be the true parameter vector in $\mathbb{X}$. Let $\mathbf{\Sigma}_{LS} \stackrel{\triangle}{=} \left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)^{-1}$ denote the covariance of the (unconstrained) LS estimator; then $\mathbf{f}(\mathbf{y})$ is *superefficient* in $\mathbb{X}$ if

$$E\left[\left(\mathbf{x} - \mathbf{f}(\mathbf{y})\right)\left(\mathbf{x} - \mathbf{f}(\mathbf{y})\right)'\right] \leq \mathbf{\Sigma}_{LS}, \qquad \forall \mathbf{x} \in \mathbb{X}$$

A sensible approach to the design of constrained estimators is, therefore, to impose a parametrized structure $\mathbf{f}(\mathbf{y};\mathbf{p})$ to the estimator and then find a suitable parameter $\mathbf{p}^*$ such that $\mathbf{f}(\mathbf{y};\mathbf{p}^*)$ is superefficient in $\mathbb{X}$. In this framework, different estimators can be obtained depending on the choices of the parametrized structure $\mathbf{f}(\mathbf{y};\mathbf{p})$ and of the geometrical shape (e.g. ellipsoidal, polytopic) of the constraint region $\mathbb{X}$.

III. MAP estimator for polytopic constraints

In this section it is assumed that $\mathbb{X}$ is a convex polytope of vertices $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, i.e. $\mathbb{X} = Co\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ where $Co\{\cdots\}$ stands for *convex hull*. Further, for the estimator we adopt the following structure

$$\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y};\overline{\mathbf{x}},\overline{\mathbf{\Sigma}}) = \left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C} + \overline{\mathbf{\Sigma}}^{-1}\right)^{-1}\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{y} + \overline{\mathbf{\Sigma}}^{-1}\overline{\mathbf{x}}\right) \tag{III.1}$$

parametrized by $\mathbf{p} = \left(\overline{\mathbf{x}},\overline{\mathbf{\Sigma}}\right)$, where $\overline{\mathbf{x}} \in \mathbb{R}^n$ and $\overline{\mathbf{\Sigma}} \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. Notice that, in a Bayesian framework, (III.1) is nothing but the MAP (or equivalently MMSE) estimate

[17] obtained by assuming that $\mathbf{x}$ has a prior Gaussian distribution with mean $\overline{\mathbf{x}}$ and covariance $\overline{\mathbf{\Sigma}}$. By lengthy calculations reported in the appendix, the associated matrix-MSE, $\mathbf{\Sigma} \stackrel{\triangle}{=} E\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})'\right]$, turns out to be

$$\begin{aligned}\mathbf{\Sigma} = \mathbf{F}(\mathbf{x};\overline{\mathbf{x}},\overline{\mathbf{\Sigma}}) \;\; \stackrel{\triangle}{=} \;\; & \left(\mathbf{\Sigma}_{LS}^{-1}+\overline{\mathbf{\Sigma}}^{-1}\right)^{-1} \\ & \left[\overline{\mathbf{\Sigma}}^{-1}(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\overline{\mathbf{\Sigma}}^{-1}+\mathbf{\Sigma}_{LS}^{-1}\right] \\ & \left(\mathbf{\Sigma}_{LS}^{-1}+\overline{\mathbf{\Sigma}}^{-1}\right)^{-1}\end{aligned} \tag{III.2}$$

According to the above arguments, the objective is to design the parameters $\overline{\mathbf{x}}$ and $\overline{\mathbf{\Sigma}}$ of the prior distribution, based on the knowledge of the membership region $\mathbb{X}$, so as to impose a guaranteed improvement over the LS estimator, i.e. that $\mathbf{F}(\mathbf{x};\overline{\mathbf{x}},\overline{\mathbf{\Sigma}}) \le \mathbf{\Sigma}_{LS}$ for any $\mathbf{x} \in \mathbb{X}$. By calculations shown in the appendix, the above domination condition requires that $\overline{\mathbf{x}}$ and $\overline{\mathbf{\Sigma}}$ be chosen such that

$$\overline{\mathbf{\Sigma}} \ge \frac{1}{2}\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'-\mathbf{\Sigma}_{LS}\right], \quad \forall \mathbf{x} \in \mathbb{X} \tag{III.3}$$

Notice that (III.3) is a matrix inequality and represents a stronger domination condition than the one considered in the minimax approach [10]-[12], which only imposes that $tr\mathbf{\Sigma} \le tr\mathbf{\Sigma_{LS}}$ for all $\mathbf{x} \in \mathbb{X}$. Clearly, $\mathbf{\Sigma} \le \mathbf{\Sigma_{LS}}$ implies $tr\mathbf{\Sigma} \le tr\mathbf{\Sigma_{LS}}$ but the converse is not true, in general. This means that the *matrix* domination considered in this paper guarantees a MSE reduction for all the components of the parameter vector, while the *trace* domination, considered in [10]-[12], only guarantees an improvement for the sum of the MSEs of such components or equivalently for the RMSE $\sqrt{E\left[(\mathbf{x}-\hat{\mathbf{x}})'(\mathbf{x}-\hat{\mathbf{x}})\right]}$. The following lemma will be used in the subsequent developments.

Lemma 1: - Let $\mathbf{P} = \mathbf{P}' > 0$ and $\mathbf{x} \neq \mathbf{0}$, then

$$\mathbf{P} \ge \mathbf{x}\mathbf{x}' \quad \Longleftrightarrow \quad \mathbf{x}'\mathbf{P}^{-1}\mathbf{x} \le 1 \tag{III.4}$$

Proof - *Follows immediately from the below property of the Schur complement*

$$\textit{if} \;\; \mathbf{P} > \mathbf{0} \;\; \textit{and} \;\; \mathbf{C} > \mathbf{0} \;\; \textit{then}$$

$$\mathbf{P} \ge \mathbf{B}\mathbf{C}^{-1}\mathbf{B}' \iff \mathbf{C} \ge \mathbf{B}'\mathbf{P}^{-1}\mathbf{B}$$

by setting $\mathbf{B} = \mathbf{x}$ *and* $\mathbf{C} = 1$. ■

Let $\mathcal{E}(\mathbf{c},\mathbf{P}) \stackrel{\triangle}{=} \left\{\mathbf{x} : (\mathbf{x}-\mathbf{c})'\mathbf{P}^{-1}(\mathbf{x}-\mathbf{c}) \le 1\right\}$ denote the ellipsoid of center $\mathbf{c}$ and matrix $\mathbf{P}$. Then the lemma 1 states that the symmetric positive definite matrix $\mathbf{P}$ is greater than or equal to the dyad $\mathbf{x}\mathbf{x}'$ if and only if the vector $\mathbf{x}$ belongs to the ellipsoid $\mathcal{E}(\mathbf{0},\mathbf{P})$. Let $\mathbf{P} \stackrel{\triangle}{=} 2\overline{\mathbf{\Sigma}}+\mathbf{\Sigma}_{LS} \ge \mathbf{\Sigma}_{LS}$. In geometric terms, the condition (III.3) amounts, therefore, to imposing that the ellipsoid $\mathcal{E}(\overline{\mathbf{x}},\mathbf{P})$ contains the polytope $\mathbb{X}$ or, equivalently, its vertices $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m$. Summarizing the above derivation, the following result holds.

Theorem 1 - *The parametrized estimator (III.1) is superefficient in $\mathbb{X}$, i.e. satisfies (III.3), if the variables $\overline{\mathbf{x}}$ and $\overline{\boldsymbol{\Sigma}} = \overline{\boldsymbol{\Sigma}}'$ satisfy the following LMI:*

$$\begin{aligned} &\begin{bmatrix} 2\overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_{LS} & \mathbf{x}_i - \overline{\mathbf{x}} \\ \mathbf{x}_i' - \overline{\mathbf{x}}' & 1 \end{bmatrix} \geq 0 \qquad i = 1, 2, \ldots, m \\ &\overline{\boldsymbol{\Sigma}} \geq 0 \end{aligned} \tag{III.5}$$

Proof - *It must be proved that if $\overline{\boldsymbol{\Sigma}} \geq \frac{1}{2}\left[\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' - \boldsymbol{\Sigma}_{LS}\right]$ for every vertex $\tilde{\mathbf{x}}_i$ of $\tilde{\mathbb{X}}$, then $\overline{\boldsymbol{\Sigma}} \geq \frac{1}{2}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \boldsymbol{\Sigma}_{LS}\right]$ for every point $\tilde{\mathbf{x}} \in \tilde{\mathbb{X}}$, where $\tilde{\mathbf{x}} = \mathbf{x} - \overline{\mathbf{x}}$ and $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \overline{\mathbf{x}}$ ($i = 1, 2, \ldots, m$) have been obtained from a coordinate shift. The set $\tilde{\mathbb{X}}$ is defined by*

$$\begin{aligned} \tilde{\mathbb{X}} &= Co\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_m\} \\ &\triangleq \{\tilde{\mathbf{x}} : \ \tilde{\mathbf{x}} = \theta_1\ \tilde{\mathbf{x}}_1 + \theta_2\ \tilde{\mathbf{x}}_2 + \cdots + \theta_m\ \tilde{\mathbf{x}}_m | \theta_i \geq 0, \\ &\qquad \theta_1 + \theta_2 + \cdots + \theta_m = 1\} \end{aligned}$$

then

$$\begin{aligned} \frac{1}{2}\tilde{\mathbf{x}}\tilde{\mathbf{x}}' &= \frac{1}{2}\left(\sum_{i=1}^{m}\theta_i\ \tilde{\mathbf{x}}_i\right)\left(\sum_{i=1}^{m}\theta_i\ \tilde{\mathbf{x}}_i\right)' \\ &= \frac{1}{2}\left(\sum_{i=1}^{m}\theta_i^2\ \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' + \sum_{j=1}^{m}\sum_{k=1, k\neq j}^{m}\theta_j\theta_k\ \tilde{\mathbf{x}}_j\tilde{\mathbf{x}}_k'\right) \end{aligned} \tag{III.6}$$

Exploiting the relationship

$$(\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_k)(\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_k)' \geq 0 \ \Rightarrow \ \tilde{\mathbf{x}}_j\tilde{\mathbf{x}}_j' + \tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k' \geq \tilde{\mathbf{x}}_j\tilde{\mathbf{x}}_k' + \tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_j'$$

from (III.6) it follows that

$$\frac{1}{2}\tilde{\mathbf{x}}\tilde{\mathbf{x}}' \leq \frac{1}{2}\left(\sum_{i=1}^{m}\theta_i^2\ \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' + \sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\theta_j\theta_k\ (\tilde{\mathbf{x}}_j\tilde{\mathbf{x}}_j' + \tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k')\right) \tag{III.7}$$

Since $\overline{\boldsymbol{\Sigma}} \geq \frac{1}{2}\left[\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' - \boldsymbol{\Sigma}_{LS}\right]$ one gets

$$\begin{aligned} \frac{1}{2}\tilde{\mathbf{x}}\tilde{\mathbf{x}}' &\leq \sum_{i=1}^{m}\theta_i^2\ \left(\overline{\boldsymbol{\Sigma}} + \frac{1}{2}\boldsymbol{\Sigma}_{LS}\right) \\ &+ \sum_{j=1}^{m-1}\sum_{k=j+1}^{m}2\theta_j\theta_k\ \left(\overline{\boldsymbol{\Sigma}} + \frac{1}{2}\boldsymbol{\Sigma}_{LS}\right) \\ &= \left(\sum_{i=1}^{m}\theta_i\right)^2\left(\overline{\boldsymbol{\Sigma}} + \frac{1}{2}\boldsymbol{\Sigma}_{LS}\right) \\ &= \overline{\boldsymbol{\Sigma}} + \frac{1}{2}\boldsymbol{\Sigma}_{LS} \end{aligned} \tag{III.8}$$

where the latter equality has been obtained from $\sum_{i=1}^{m} \theta_i = 1$. *This implies, for all* $\mathbf{x} \in \mathbb{X}$, *the superefficiency relationship* $\overline{\mathbf{\Sigma}} \geq \frac{1}{2} [\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \mathbf{\Sigma}_{LS}]$. ■

Therefore, given $\mathbb{X}$ and $\mathbf{\Sigma}_{LS}$, the parameters $(\overline{\mathbf{x}},\overline{\mathbf{\Sigma}})$ of the MAP estimator (III.1) can be designed as follows

$$\min_{\overline{\mathbf{x}},\, \overline{\mathbf{\Sigma}}} \; det\, \overline{\mathbf{\Sigma}} \quad \text{or} \quad tr\, \overline{\mathbf{\Sigma}} \qquad \text{subject to} \quad \text{(III.5)} \tag{III.9}$$

where det and tr denote the determinant and, respectively, the trace. The estimator (III.1) with parameters $(\overline{\mathbf{x}},\overline{\mathbf{\Sigma}})$ chosen as in (III.9) will be called *soft-constrained* MAP estimator as in [19] and denoted by MAP$_s$. The term *soft* means that the estimator can provide an estimate violating constraints, even if the estimator guarantees an improvement of performance (MSE matrix) for any true parameter vector satisfying constraints. Notice that, from a geometrical point of view, minimization of the determinant in (III.9) provides the minimum volume ellipsoid $\mathcal{E}(\overline{\mathbf{x}}, \overline{\mathbf{\Sigma}})$, while minimization of the trace provides an ellipsoid with minimum sum of the squares of the axes [18].

The choice of the design criterion adopted in (III.9) is justified by the following considerations.

- Both the determinant and the trace criterion in (III.9) give rise to standard convex LMI problems [18] which can be efficiently solved using reliable numerical tools [21].
- Both criteria in (III.9) amount to minimizing, in some sense, the size of the ellipsoid $\mathcal{E}(\overline{\mathbf{x}}, \overline{\mathbf{\Sigma}})$ which, in the Bayesian interpretation of the estimator (III.1), represents a confidence region under the prior Gaussian distribution $\mathcal{N}(\overline{\mathbf{x}}, \overline{\mathbf{\Sigma}})$. Hence, although minimizing the size of such a prior confidence region does not necessarily guarantee the minimum achievable MSE under the domination constraints (III.3), the choice (III.9) seems a reasonable compromise between estimation performance and numerical tractability of the optimization problem. As a matter of fact, the simulation results of section V will demonstrate the good performance improvement of the proposed estimator.
- A more sensible approach, from an estimation performance point of view, would be that of minimizing, subject to domination constraints (III.3), the MSE in (III.2), which clearly depends on the unknown true value $\mathbf{x} \in \mathbb{X}$, averaged over $\mathbb{X}$ assuming for instance that $\mathbf{x}$ is uniformly distributed in $\mathbb{X}$. This yields the following optimization problem:

  $$\min_{\overline{\mathbf{x}},\, \overline{\mathbf{\Sigma}}} \; E_{\mathbf{x}}[tr\; \mathbf{F}(\mathbf{x}; \overline{\mathbf{x}}, \overline{\mathbf{\Sigma}})] \qquad \text{subject to} \quad \text{(III.5)} \tag{III.10}$$

  Unfortunately, the dependence of the above criterion on the optimization variables $\overline{\mathbf{x}}$ and $\overline{\mathbf{\Sigma}}$ is much more complicated than in (III.9) and further investigation is needed to ascertain whether (III.10) is a tractable (convex or quasi-convex) problem.

Although the optimization problem (III.9) has two unknowns $\overline{\mathbf{x}}$ e $\overline{\boldsymbol{\Sigma}}$, intuition suggests that for symmetric constraints, $\overline{\mathbf{x}}$ should be chosen as the center of symmetry of $\mathbb{X}$. It is recalled that a convex set $\mathbb{X}$ has a center of symmetry $\mathbf{x}_c$ if for all $\mathbf{x} \in \mathbb{X}$, $2\mathbf{x}_c - \mathbf{x} \in \mathbb{X}$. Then, the following result holds.

Theorem 2 - *Let $\mathbb{X}$ have a center of symmetry $\mathbf{x}_c$ and let ($\overline{\mathbf{x}}$,$\overline{\boldsymbol{\Sigma}}$) be the optimal solution of (III.9), then $\overline{\mathbf{x}} = \mathbf{x}_c$.*

Proof - *From lemma 1 it follows that $(\mathbf{x} - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x} - \overline{\mathbf{x}}) \leq 1, \ \forall \ \mathbf{x} \in \mathbb{X}$, where $\mathbf{P} = 2\overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_{LS}$ and ($\overline{\mathbf{x}}$,$\overline{\boldsymbol{\Sigma}}$) is the optimal solution of (III.9). Let us consider*

$$Q \triangleq \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) \tag{III.11}$$

Adding and subtracting $\mathbf{x}_c$ to $\mathbf{x}_i - \overline{\mathbf{x}}$ one gets

$$Q = \sum_{i=1}^{m} (\mathbf{x}_i - \mathbf{x}_c + \mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c + \mathbf{x}_c - \overline{\mathbf{x}}) \tag{III.12}$$

The expression (III.12) can be rewritten in the following form

$$Q \quad = \quad \sum_{i=1}^{m} g(\mathbf{x}_i, \mathbf{x}_c, \overline{\mathbf{x}}, \mathbf{P}) \tag{III.13}$$

where

$$\begin{aligned} g(\mathbf{x}_i, \mathbf{x}_c, \overline{\mathbf{x}}, \mathbf{P}) \quad &\triangleq \quad \big[(\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c) \\ &+ \quad (\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}}) \\ &+ \quad (\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}}) \\ &+ \quad (\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c)\big] \leq 1 \end{aligned} \tag{III.14}$$

Notice that, being $\mathbf{x}_c$ the center of symmetry of $\mathbb{X}$, for every vertex $\mathbf{x}_i$ there exists a vertex $\mathbf{x}_j$ which is the symmetrical point of $\mathbf{x}_i$ with respect to $\mathbf{x}_c$, i.e. $\mathbf{x}_j = 2\mathbf{x}_c - \mathbf{x}_i$. Hence, for such a pair of symmetrical vertices $\mathbf{x}_i$ and $\mathbf{x}_j$, it follows that

$$\begin{aligned} \tfrac{g(\mathbf{x}_i, \mathbf{x}_c, \overline{\mathbf{x}}, \mathbf{P}) + g(\mathbf{x}_j, \mathbf{x}_c, \overline{\mathbf{x}}, \mathbf{P})}{2} \quad &= \quad (\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c) \\ &+ \quad (\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}}) \leq 1 \end{aligned} \tag{III.15}$$

Applying this fact to (III.13) it follows that

$$Q = \sum_{i=1}^{m} \big[(\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c) + (\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}})\big] \tag{III.16}$$

Therefore,

$$\begin{aligned} Q \quad &= \quad \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) \\ &= \quad \sum_{i=1}^{m} \big[(\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c) \\ &+ \quad (\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}})\big] \end{aligned} \tag{III.17}$$

If $\bar{\mathbf{x}} \neq \mathbf{x}_c$ *then* $(\mathbf{x}_c - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_c - \overline{\mathbf{x}}) > 0$*; hence it follows that*

$$\sum_{i=1}^{m} (\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}^{-1} (\mathbf{x}_i - \mathbf{x}_c) < \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}) \tag{III.18}$$

Thus, by continuity from (III.15) and (III.18), there exists a sufficiently small $\varepsilon > 0$ *such that* $\mathbf{x}_c$ *and* $\mathbf{P}_c \triangleq (1-\varepsilon)\mathbf{P}$ *satisfy* $(\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}_c^{-1} (\mathbf{x}_i - \mathbf{x}_c) \leq 1$ *for* $i = 1, 2, \ldots, m$ *and*

$$\sum_{i=1}^{m} (\mathbf{x}_i - \mathbf{x}_c)' \mathbf{P}_c^{-1} (\mathbf{x}_i - \mathbf{x}_c) < \sum_{i=1}^{m} (\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{P}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}). \tag{III.19}$$

Then, it follows that $(\mathbf{x}_c, \mathbf{\Sigma_c})$*, with* $2\mathbf{\Sigma_c} = \mathbf{P}_c - \mathbf{\Sigma}_{LS}$*, is an admissible solution of (III.5) and provides a trace* $tr(\mathbf{\Sigma}_c) = (1-\varepsilon)\ tr(\overline{\mathbf{\Sigma}}) < tr(\overline{\mathbf{\Sigma}})$ *smaller than any other solution* $(\bar{\mathbf{x}}, \bar{\mathbf{\Sigma}})$ *with* $\bar{\mathbf{x}} \neq \mathbf{x}_c$*; consequently the optimal solution* $(\overline{\mathbf{x}}, \overline{\mathbf{\Sigma}})$ *must satisfy* $\bar{\mathbf{x}} = \mathbf{x}_c$*. The same arguments hold for the case in which the determinant, instead of the trace, of* $\overline{\mathbf{\Sigma}}$ *is minimized.* ■

Figure 1 provides a pictorial interpretation in the 3-dimensional space of the constrained estimator design procedure. Notice that the ellipsoid $\mathcal{E}(\overline{\mathbf{x}}, \mathbf{P})$ circumscribes the polytope $\mathbb{X}$ (in this case a parallelepiped centered at the origin) i.e. the vertices of $\mathbb{X}$ belong to the boundary of $\mathcal{E}(\overline{\mathbf{x}}, \mathbf{P})$.

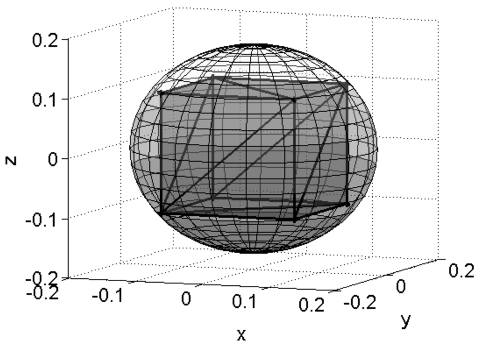

Fig. 1. The ellipsoid covers the set $\mathbb{X}$

IV. Ridge and minimax constrained estimators

In this section we consider the general form of the ridge estimator proposed by Hoerl and Kennard [14] (suitably modified so as to take into account the a-priori information) and the minimax MSE constrained estimator proposed by Eldar, Ben-Tal and Nemirovski [9]. The goal is to compare their performance with that of the MAP$_s$ estimator optimized via the LMI procedure described in the previous section.

A. Ridge estimator

In order to derive the general form of the ridge estimator, let us consider the eigenvector-eigenvalue decomposition of $\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-\mathbf{1}}\mathbf{C}$, i.e.

$$\boldsymbol{\Sigma}_{\mathbf{LS}}^{-\mathbf{1}} = \mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-\mathbf{1}}\mathbf{C} = \mathbf{V}'\boldsymbol{\Lambda}\mathbf{V} \tag{IV.1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix whose elements $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of $\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-\mathbf{1}}\mathbf{C}$ and $\mathbf{V}$ is the orthogonal eigenvector matrix. Further, defining the parameter transformation $\mathbf{z} = \mathbf{V}\mathbf{x}$ and the measurement transformation $\mathbf{s} = \boldsymbol{\Sigma}_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{y}$, one gets the transformed regression model

$$\mathbf{s} = \mathbf{H}\mathbf{z} + \mathbf{w} \tag{IV.2}$$

where $\mathbf{H} = \boldsymbol{\Sigma}_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{C}\mathbf{V}'$ and $\mathbf{w} = \boldsymbol{\Sigma}_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{v}$ has unit covariance. Hence, the ridge estimator is defined as follows

$$\begin{aligned} \hat{\mathbf{x}}_r = \mathbf{V}'\hat{\mathbf{z}}_r &= \mathbf{V}'\left(\mathbf{H}'\mathbf{H} + \mathbf{K}^{-1}\right)^{-1}\mathbf{H}'\mathbf{s} \\ &= \mathbf{V}'\left(\boldsymbol{\Lambda} + \mathbf{K}^{-1}\right)^{-1}\mathbf{H}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{y} \end{aligned} \tag{IV.3}$$

where $\mathbf{K}$ is a diagonal matrix with positive elements $k_1, k_2, \cdots, k_n$. The optimal matrix $\mathbf{K}$ can be selected by minimizing the trace of the MSE:

$$\begin{aligned} & \min_{\mathbf{K}} \;\; tr\ E\left[(\hat{\mathbf{x}}_r - \mathbf{x})'(\hat{\mathbf{x}}_r - \mathbf{x})\right] \\ = & \min_{\mathbf{K}} \;\; tr\ E\left[\left(\mathbf{V}'\left(\mathbf{H}'\mathbf{H} + \mathbf{K}^{-1}\right)^{-1}\mathbf{H}'\mathbf{s} - \mathbf{x}\right)\left(\cdots\right)'\right] \\ = & \min_{\mathbf{K}} \;\; tr\ \mathbf{V}'\left(\boldsymbol{\Lambda} + \mathbf{K}^{-1}\right)^{-1}\left(\boldsymbol{\Lambda} + \mathbf{K}^{-1}\mathbf{z}\mathbf{z}'\mathbf{K}^{-1}\right) \\ & \qquad \left(\boldsymbol{\Lambda} + \mathbf{K}^{-1}\right)^{-1}\mathbf{V} \end{aligned} \tag{IV.4}$$

Exploiting the diagonality of $\mathbf{K}$ and $\boldsymbol{\Lambda}$, the orthogonality of $\mathbf{V}$ and the relationship $tr(\mathbf{V}'\mathbf{M}\mathbf{V}) = tr(\mathbf{M}\mathbf{V}\mathbf{V}') = tr(\mathbf{M})$, (IV.4) can be simplified to

$$\min_{k_i} \;\; \sum_{i=1}^{n} \frac{\lambda_i + \frac{1}{k_i^2}\langle\mathbf{z}\rangle_i^{\,2}}{\left(\lambda_i + \frac{1}{k_i}\right)^2} \tag{IV.5}$$

where $\langle\mathbf{z}\rangle_i$ denotes the ith entry of vector $\mathbf{z}$. Differentiating (IV.5) w.r.t. k_i, the minimizing values are obtained as

$$k_i = \langle\mathbf{z}\rangle_i^2, \quad i = 1, \ldots, n \tag{IV.6}$$

Hence, the optimal $\mathbf{K}$ depends on $\mathbf{z}$ and, thus, on the true parameter vector $\mathbf{x}$ which is unknown. Hoerl and Kennard in [14] suggest the use of an iterative procedure to estimate k_i; conversely, here we use the a-priori information on $\mathbf{x}$. It is known a-priori that $\mathbf{z}$ belongs to the convex polytope $\mathbf{Z} = Co\left\{\mathbf{z}_1, \ldots, \mathbf{z}_m\right\}$

obtained by applying the orthogonal transformation $\mathbf{V}$ to the vertices of the original polytope $\mathbb{X}$. Then the ridge parameters k_i can be selected as

$$k_i \;=\; \max_{1 \le j \le m} \langle \mathbf{z}_j \rangle_i^2 \,, \qquad i = 1, \ldots, n \tag{IV.7}$$

i.e. k_i is taken as the square of the maximum ith entry among all transformed vertices $\mathbf{z}_j$.

Notice that (IV.3) can also be rewritten as

$$\hat{\mathbf{x}}_r \;=\; \left(\mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{V}}^{-1}\mathbf{C} + \overline{\boldsymbol{\Sigma}}^{-1} \right)^{-1} \mathbf{C}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\mathbf{y} \tag{IV.8}$$

where $\overline{\boldsymbol{\Sigma}} \stackrel{\triangle}{=} \mathbf{V}'\mathbf{K}\mathbf{V}$, from which it can be seen that the ridge estimator (IV.3) is a special case of the MAP$_s$ estimator (III.1) with $\overline{\boldsymbol{\Sigma}} = \mathbf{V}'\mathbf{K}\mathbf{V}$ and $\overline{\mathbf{x}} = \mathbf{0}$. Actually, the matrix $\overline{\boldsymbol{\Sigma}}$ of the generalized ridge estimator (IV.8) is not restricted to have the same eigenvectors of $\boldsymbol{\Sigma}_{LS}$ as in (IV.3). Hence, for symmetrical constraint sets, the design procedure for the MAP$_s$ estimator of section III yields $\overline{\mathbf{x}} = \mathbf{0}$ and provides, in turn, a good approach for the design of the ridge estimator (IV.8). Conversely, for unsymmetrical constraint sets, the design of the MAP$_s$ estimator provides $\overline{\mathbf{x}} \neq \mathbf{0}$ so that in this case the affine MAP$_s$ estimator (III.1) has clearly more degrees of freedom than the linear ridge estimator (IV.8).

B. Minimax constrained estimator

The minimax MSE constrained estimator [9] is the linear estimator $\hat{\mathbf{x}}_{mm} = \mathbf{G}\mathbf{y}$ that minimizes the worst-case trace of the MSE among all possible values of $\mathbf{x}$ satisfying $\mathbf{x}'\mathbf{T}\mathbf{x} \le L^2$

$$\begin{aligned} & \min_{\mathbf{G}} \max_{\mathbf{x}'\mathbf{T}\mathbf{x} \le L^2} \; tr \; E\left[(\hat{\mathbf{x}}_{mm} - \mathbf{x})(\hat{\mathbf{x}}_{mm} - \mathbf{x})' \right] \\ = \; & \min_{\mathbf{G}} \max_{\mathbf{x}'\mathbf{T}\mathbf{x} \le L^2} \left\{ tr(\mathbf{G}\boldsymbol{\Sigma}_v\mathbf{G}') + \mathbf{x}'(\mathbf{I} - \mathbf{G}\mathbf{C})'(\mathbf{I} - \mathbf{G}\mathbf{C})\mathbf{x} \right\} \end{aligned} \tag{IV.9}$$

where $\mathbf{T}$ is a symmetric positive definite matrix and L^2 is a positive scalar. In the special case $\mathbf{T} = \mathbf{I}$, the solution of the minimax problem [9] is given by

$$\hat{\mathbf{x}}_{mm} = \frac{L^2}{L^2 + \gamma_0} \; \left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} \right)^{-1} \mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{y} \tag{IV.10}$$

where $\gamma_0 = tr \left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} \right)^{-1}$ is the variance of the LS estimator. The estimator (IV.10) is a special case of the *shrunken* estimator proposed by Mayer and Willke [15]

$$\hat{\mathbf{x}}_{sh} = \beta \; \left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} \right)^{-1} \mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{y} \tag{IV.11}$$

with an optimal choice of the shrinkage factor β. The minimax estimator (IV.10) minimizes the worst-case MSE w.r.t. all vectors $\mathbf{x}$ such that $\mathbf{x}'\mathbf{x} \le L^2$ and also guarantees [11] that the resulting trace of the MSE does not exceed the MSE's trace of the LS estimator, γ_0, for any $\mathbf{x}$ satisfying $\mathbf{x}'\mathbf{x} \le L^2$.

C. Comparison with LS

Notice that both the above estimators have been designed minimizing the trace of the MSE, without guaranteeing matrix-MSE improvement over the LS estimator. The ridge estimator yields a matrix-MSE smaller than the matrix-MSE of the LS estimator, i.e. $MSE_r(\mathbf{x}) \leq \mathbf{\Sigma}_{LS}$ for all $\mathbf{x} \in \mathbb{X}$, if the following condition is satisfied:

$$\begin{aligned} MSE_r(\mathbf{x}) &= \mathbf{V}'\left(\mathbf{\Lambda}+\mathbf{K}^{-1}\right)^{-1}\left(\mathbf{\Lambda}+\mathbf{K}^{-1}\mathbf{z}\mathbf{z}'\mathbf{K}^{-1}\right) \\ & \quad \left(\mathbf{\Lambda}+\mathbf{K}^{-1}\right)^{-1}\mathbf{V} \\ &\leq \mathbf{\Sigma}_{LS} = \mathbf{V}'\mathbf{\Lambda}^{-1}\mathbf{V}. \end{aligned} \tag{IV.12}$$

By easy calculations the above inequality can be simplified to

$$\mathbf{K} \geq \frac{1}{2}\left(\mathbf{z}\mathbf{z}' - \mathbf{\Lambda}^{-1}\right) \tag{IV.13}$$

The condition (IV.13) must be satisfied by $\mathbf{K}$ in order to guarantee superefficiency of the ridge estimator (IV.3). It can be noticed that $\mathbf{K}$ in (IV.7), although does not allow to satisfy the above matrix inequality, guarantees that the diagonal elements of $MSE_r(\mathbf{x})$ are not smaller than the diagonal elements of $\mathbf{\Sigma}_{LS}$. In fact $\mathbf{K}$ is a diagonal matrix whose elements are always not smaller than the diagonal elements of $\mathbf{z}\mathbf{z}'$, see equation (IV.7), and therefore (IV.13) holds for the diagonal elements. This means that the ridge estimator, defined by (IV.3) and (IV.7), always gives a MSE improvement over the LS estimator for all components of the parameter vector.

On the other hand, the matrix-MSE of the shrunken estimator (IV.11) is not bigger than the matrix-MSE of the LS estimator provided that the following matrix inequality is satisfied:

$$\mathbf{x}\mathbf{x}'(\beta-1)^2 + \beta^2\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)^{-1} \leq \left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)^{-1} \tag{IV.14}$$

or, equivalently,

$$\mathbf{x}\mathbf{x}'(\beta-1)^2 + (\beta^2-1)\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)^{-1} \leq 0 \tag{IV.15}$$

The above expression can be rewritten in this form:

$$\frac{(1-\beta^2)}{(1-\beta)^2}\;\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)^{-1} - \mathbf{x}\mathbf{x}' \geq 0 \tag{IV.16}$$

From lemma 1 the inequality (IV.16) holds if and only if

$$\frac{(1-\beta^2)}{(1-\beta)^2}\;\frac{1}{\mathbf{x}'\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)\mathbf{x}} - 1 \geq 0 \tag{IV.17}$$

This requires that β be chosen such that

$$\frac{\mathbf{x}'\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)\mathbf{x} - 1}{\mathbf{x}'\left(\mathbf{C}'\mathbf{\Sigma}_v^{-1}\mathbf{C}\right)\mathbf{x} + 1} \leq \beta \leq 1 \tag{IV.18}$$

Notice that for $\beta = 1$ the shrunken estimator coincides with the LS estimator and for

$$\beta = \frac{\mathbf{x}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \mathbf{x} - 1}{\mathbf{x}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \mathbf{x} + 1}$$

gives a biased estimate with the same MSE of LS. Hence, since (IV.17) is a quadratic inequality in β, choosing β as the center of the interval (IV.18), i.e.

$$\beta = \frac{1}{2} \left(1 + \frac{\mathbf{x}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \mathbf{x} - 1}{\mathbf{x}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \mathbf{x} + 1} \right), \tag{IV.19}$$

the shrunken estimator gives the maximum matrix-MSE reduction with respect to LS. The lower bound in (IV.18) depends on $\mathbf{x}$ that is unknown, but one can maximize such a bound using the a-priori information $\mathbf{x}'\mathbf{x} \le L^2$, considering therefore the worst-case value of $\mathbf{x}$. Denoting by $\check{\mathbf{x}}$ this worst-case value, the shrinkage factor $\beta = L^2/(L^2 + \gamma_0)$, provided by the minimax constrained estimator, guarantees improvement over the LS estimator, in the matrix MSE sense, if and only if

$$\frac{\check{\mathbf{x}}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \check{\mathbf{x}} - 1}{\check{\mathbf{x}}' \left(\mathbf{C}' \mathbf{\Sigma}_v^{-1} \mathbf{C} \right) \check{\mathbf{x}} + 1} \le \frac{L^2}{L^2 + \gamma_0} \tag{IV.20}$$

Notice that since $(\rho - 1)(\rho + 1)^{-1}$ is an increasing function of ρ for $\rho \in [0, \infty)$, maximization of the LHS of (IV.20) is equivalent to maximization of $\rho(\mathbf{x}) = \mathbf{x}'\mathbf{\Sigma}_{LS}^{-1}\mathbf{x}$. By trivial geometric arguments, the maximum of $\rho(\mathbf{x}) = \mathbf{x}'\mathbf{\Sigma}_{LS}^{-1}\mathbf{x}$ subject to $\mathbf{x}'\mathbf{x} \le L^2$ must lie on the boundary of the hypersphere $\mathbf{x}'\mathbf{x} \le L^2$ (i.e. must satisfy $\mathbf{x}'\mathbf{x} = L^2$) and can be easily found. Hence, the condition (IV.20) can be checked in order to assess whenever the minimax estimator (IV.10) is LS-dominating in the matrix sense. Thus, if (IV.20) does not hold, the minimax estimator ensures trace-domination but not matrix-domination. Examination of (IV.20) reveals that this inequality is satisfied for L^2 sufficiently small w.r.t. $\mathbf{\Sigma}_{LS}$, i.e. whenever the constrained region is sufficiently small w.r.t. the confidence regions of the LS estimator. This lack of guarantee of superefficiency can be overcome by choosing the shrinkage factor β of the shrunken estimator (IV.11) as in (IV.19) with $\mathbf{x} = \check{\mathbf{x}}$.

In the next section the novel estimator presented in section 3 will be compared with the ridge, the minimax and the shrunken estimators described in this section via simulation examples.

V. Numerical examples

The goal is to fit position measurements of a target obtained at different time instants via a linear regression of n-th order. The measurement $y(t)$ at sample time t is modeled by

$$y(t) \ = \ A_0 \ + \ A_1 \ T \ t \ + \ \ldots \ + \ A_n \ (T \ t)^n \ + \ v(t) \tag{V.1}$$

where $t = 0, 1, \ldots, N-1$; $N = 5$ is the number of measurements; $T = 1$ is the sampling interval; A_i for $i = 0, 1, \ldots, n$ are the parameters to be estimated; $v(t)$ is the measurement noise which is assumed to be normally distributed, white, with zero mean and standard deviation $\sigma = 0.05$. The N available data can be collected in a vector $\mathbf{y}$, thus giving

$$\mathbf{y} \; = \; \mathbf{C}\mathbf{x} + \mathbf{v} \tag{V.2}$$

with

$$\mathbf{y} = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix}, \; \mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & T & \dots & T^n \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (N-1)T & \dots & (N-1)^n T^n \end{bmatrix},$$
$$\mathbf{x} = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_n \end{bmatrix}, \; \mathbf{v} = \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix}. \tag{V.3}$$

where $\mathbf{x}$ is the parameter vector to be estimated and the measurement noise has covariance matrix $E\left[\mathbf{v}\mathbf{v}'\right] \; = \; \sigma^2 \mathbf{I}_N$, $\mathbf{I}_N$ denoting the $N \times N$ identity matrix.

Further, it has been assumed that a-priori information $\mathbf{x} \in \mathbb{X}$ is available on the parameters to be estimated. In the numerical examples, presented in the next sub-sections, two types of membership sets $\mathbb{X}$ have been hypothesized:

- **Sphere**: $\mathbb{X}_s = \left\{\mathbf{x} : \; \mathbf{x}'\mathbf{x} \leq L^2\right\}$
- **Polytope**: $\mathbb{X}_p = \{\mathbf{x} : \; \mathbf{x} \in Co\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}\}$

where L is the sphere's radius and $\mathbf{x}_i$, for $i = 1, \dots, m$, are the vertices of the polytope. Whenever the true membership set, available from a-priori information, is a sphere ($\mathbb{X} = \mathbb{X}_s$), a polytope $\widehat{\mathbb{X}} = \mathbb{X}_p$ outer-approximating the sphere is built in order to design the MAP$_s$ estimator. Conversely, if the true membership set is a polytope ($\mathbb{X} = \mathbb{X}_p$), a sphere $\widehat{\mathbb{X}} = \mathbb{X}_s$ outer-approximating the polytope is built in order to design the minimax estimator. Hence, using the true membership set, $\mathbb{X} = \mathbb{X}_s$ or $\mathbb{X} = \mathbb{X}_p$, and its outer-approximation, $\widetilde{\mathbb{X}} = \mathbb{X}_p$ or $\widetilde{\mathbb{X}} = \mathbb{X}_s$ respectively, the following estimators have been designed.

$$
\begin{array}{rccl}
\textbf{LS:} & \hat{\mathbf{x}}_{LS} & = & \left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{y} \\
\textbf{MAP}_s\textbf{:} & \hat{\mathbf{x}}_{MAP_s} & = & \left(\dfrac{\mathbf{C}'\mathbf{C}}{\sigma^2} + \overline{\mathbf{\Sigma}}^{-1}\right)^{-1}\left(\dfrac{\mathbf{C}'\mathbf{y}}{\sigma^2} + \overline{\mathbf{\Sigma}}^{-1}\overline{\mathbf{x}}\right) \\
\textbf{Minimax:} & \hat{\mathbf{x}}_{mm} & = & \dfrac{L^2}{L^2 + \sigma^2\ tr\ (\mathbf{C}'\mathbf{C})^{-1}}\left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{y} \\
\textbf{Shrunken:} & \hat{\mathbf{x}}_{sh} & = & \beta\left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\mathbf{y}\ \text{ with} \\
 & & & \beta = \frac{1}{2}\left(1 + \dfrac{\check{\mathbf{x}}'\mathbf{C}'\mathbf{C}\check{\mathbf{x}} - \sigma^2}{\check{\mathbf{x}}'\mathbf{C}'\mathbf{C}\check{\mathbf{x}} + \sigma^2}\right) \\
\textbf{Ridge:} & \hat{\mathbf{x}}_{r} & = & \mathbf{V}'\left(\mathbf{\Lambda} + \mathbf{K}^{-1}\right)^{-1}\dfrac{\mathbf{V}\mathbf{C}'\mathbf{y}}{\sigma^2} \\
\textbf{MAP}_h\textbf{:} & \hat{\mathbf{x}}_{\mathrm{MAP}_h} & = & \arg\min\limits_{\mathbf{x}\in\mathbb{X}}\left(\mathbf{y} - \mathbf{C}\mathbf{x}\right)'\left(\mathbf{y} - \mathbf{C}\mathbf{x}\right).
\end{array}
\tag{V.4}
$$

The design parameters $\overline{\mathbf{x}}$, $\overline{\mathbf{\Sigma}}$, L, $\check{\mathbf{x}}$, $\mathbf{V}$, $\mathbf{\Lambda}$ and $\mathbf{K}$ can be obtained using the available a-priori information, as specified in sections III and IV. In particular, remind that $\overline{\mathbf{x}}$ and $\overline{\mathbf{\Sigma}}$ are based on the polytope $\mathbb{X}_p$, L is based on the sphere $\mathbb{X}_s$ while $\check{\mathbf{x}}$, $\mathbf{V}$, $\mathbf{\Lambda}$, $\mathbf{K}$ are based on the true membership set $\mathbb{X}$. Hence, depending on the shape of $\mathbb{X}$ (either spherical or polytopic), either the MAP$_s$ or the minimax estimator is affected by some conservativeness in the representation of constraints. The specific values of all the design parameters used in the examples will be reported in the next subsections. Notice that an alternative approach to constrained estimation is to explicitly impose the constraint $\mathbf{x} \in \mathbb{X}$ in the MAP estimation. This gives rise to the *hard-constrained* MAP estimator [19], denoted as MAP$_h$, which has been implemented and compared with the LS, MAP$_s$, minimax, shrunken and ridge estimators.

The MSE of the compared estimators has been evaluated via Monte Carlo simulations. More specifically, 300 vectors of parameters $\mathbf{x}$ have been randomly generated, uniformly over $\mathbb{X}$, and for each vector 1000 independent trials have been run by varying the measurement noise realization.

A. Three-dimensional case with box/spherical constraints

In this example $\mathbf{x} = [A_0, A_1, A_2]'$. First it is assumed that $\mathbf{x}$ belongs to the three-dimensional box

$$
\mathbb{X} = \mathbb{X}_p = \{\mathbf{x} : A_0 \in [-0.1, 0.1],\ A_1 \in [-0.1, 0.1],\ A_2 \in [-0.1, 0.1]\} \tag{V.5}
$$

The polytopic membership set $\mathbb{X}$ has been approximated with the minimum volume sphere $\widehat{\mathbb{X}} = \mathbb{X}_s$ circumscribing the polytope, that is a sphere with center $\mathbf{0}$ and radius $L = \sqrt{0.03}$. In this case, the design parameters of the estimators (V.4) are

MAP$_s$: $\bar{\mathbf{x}} = \mathbf{0}$ (by symmetry of $\mathbb{X}$) and

$$\overline{\boldsymbol{\Sigma}} = 10^{-3} \begin{bmatrix} 13.90 & 0.96 & 0.18 \\ 0.96 & 13.45 & 0.35 \\ 0.18 & 0.35 & 14.91 \end{bmatrix}$$

Ridge: $\boldsymbol{\Lambda} = 10^5 \; diag(0.0021, 0.0139, 1.54)$,

$\mathbf{K} = diag(0.0241, 0.0252, 0.0173)$ and

$$\mathbf{V}' = \begin{bmatrix} -0.6081 & -0.7895 & 0.0828 \\ 0.7759 & -0.5691 & 0.2724 \\ -0.1679 & 0.2299 & 0.9586 \end{bmatrix}$$

Shrunken: $\check{\mathbf{x}} = [0.1, \; 0.1, \; 0.1]'$

Minimax: $L^2 = 0.03$

The simulation results have been reported in table I (box constraints). More precisely, table I and the subsequent tables report the MSE of each component of the parameter vector and the MSE trace (both averaged w.r.t. the true parameter vector and the measurement noise) of the LS estimator as well as the $\%$ MSE reductions, w.r.t. the LS estimator, of the various constrained estimators (MAP$_s$, ridge, shrunken, minimax, MAP$_h$). Further, the best-case ($\mathbf{x} = 0$) and the worst-case (average over the vertices $\mathbf{x}_i$ of the 3-D box) are also reported; these results are labelled as *best* and, respectively, *worst* in the table while the label *random* refers to averages w.r.t randomly generated $\mathbf{x}$. From table I (box constraints) it can be noted that, in the random case, the MAP$_h$ estimator provides the best results (it uses in a direct way the constraints), but the MSE of the MAP$_s$ estimator, optimized according to the LMI procedure of section III, is quite similar and is obtained with a lower computational burden. In fact, the MAP$_h$ approach requires the solution of a constrained optimization problem, precisely a quadratic programming problem. Conversely, in the MAP$_s$ approach, the solution of the LMI problem does not depend on the measurements but only on $\mathbb{X}$ and, therefore, can be calculated off-line. Notice that the minimax estimator can give for some scalar parameter (cf. 13.14% increase of MSE for $\hat{A}_2$) an estimation performance even worse than LS. In this case this can be due to the approximation of the membership set $\mathbb{X}$ with the circumscribed sphere, but in general it can also depend on:

- the violation of the condition (IV.20), i.e. the minimax estimator guarantees trace-domination but not matrix-domination;
- the use of a minimax approach, which chooses the design parameters based on the worst-case vector $\mathbf{x} \in \mathbb{X}$ and, therefore, is very likely to be worse on average but slightly better in some cases of low

probability.

In fact, from table I (box constraints) it can be seen that in the best-case the minimax estimator gives for all parameters a better performance than LS, although also in this case the MSE of the MAP$_s$ estimator is lower. Finally notice that, in the worst-case, the MAP$_h$ provides a 75% MSE reduction w.r.t. LS. In fact, whenever $\mathbf{x}$ is on the boundary of the constrained region, this estimator yields the best performance [19].

For a better interpretation of the results concerning the minimax estimator, further simulations have been carried out assuming that $\mathbf{x}$ actually belongs to the sphere $\mathbb{X} = \mathbb{X}_s = \left\{\mathbf{x} : \mathbf{x}'\mathbf{x} \leq L^2\right\}$ with $L = 0.1$, while the polytope (V.5) is in this case only an outer-approximation $\widehat{\mathbb{X}}$ of $\mathbb{X}$. The design parameters, in this case, turn out to be:

Ridge: $\mathbf{K} = diag(0.01, 0.01, 0.01)$

Shrunken: $\check{\mathbf{x}} = [0.0083, 0.0272, 0.0959]'$

Minimax: $L^2 = 0.01$

Notice that, since the polytope and the regression problem are unchanged, the parameters of the MAP$_s$ estimator and the parameters $\mathbf{V}$ and $\mathbf{\Lambda}$ of the ridge estimator remain the same as before. Table I (spherical constraints) shows the obtained simulation results for the random-case, best-case ($\mathbf{x} = 0$) and worst-case ($\mathbf{x}$ uniformly generated over the sphere's boundary). For the random-case it can be seen from table I (spherical constraints) that, again, the minimax estimator provides an MSE greater than LS for the parameter A_2. In fact, even if the minimax estimator is designed w.r.t. the correct membership set $\mathbb{X} = \mathbb{X}_s$, the condition (IV.20) is violated and, therefore, matrix MSE domination is not guaranteed. However in this example the minimax estimator provides the lowest MSE's trace. Also notice that the MAP$_h$ estimator is not the best estimator; this is probably due to the practical difficulty for this estimation approach of solving a nonlinear optimization problem with quadratic cost and constraints.

B. Two-dimensional case with circle constraints

In this example $\mathbf{x} = [A_0, A_1]'$ is known to belong to the circle of radius $L = 0.1$, that is $\mathbb{X} = \mathbb{X}_s = \left\{\mathbf{x} : \mathbf{x}'\mathbf{x} \leq L^2\right\}$. To implement the MAP$_s$ estimator, this a-priori information is approximated, in a crude way, with a regular hexagon (see fig. 2) of center $\mathbf{0}$ and edge length $2L/\sqrt{3}$, i.e.

$$\begin{aligned} \widehat{\mathbb{X}} \;&=\; \mathbb{X}_p = \Big\{\mathbf{x} = [A_0, A_1] : |A_1| \leq 0.1, \;\; \Big|A_1 + \tfrac{\sqrt{3}}{10L} A_0\Big| \leq 0.2, \\ &\qquad \Big|A_1 - \tfrac{\sqrt{3}}{10L} A_0\Big| \leq 0.2\Big\}. \end{aligned}$$

The true membership set $\mathbb{X}$ and the approximated set $\widehat{\mathbb{X}}$ have been exploited in order to derive the design parameters of the various estimators in (V.4). Such parameters turn out to be:

TABLE I

SIMULATION RESULTS FOR THE SECOND-ORDER REGRESSION: BOX AND SPHERICAL CONSTRAINTS.

box constraints							
	MSE	LS	MAP$_s$	ridge	shrunken	minimax	MAP$_h$
random	$\hat{A}_0$	0.0022	-35.2%	-23.7%	-0.1%	-25.1%	-35.6%
	$\hat{A}_1$	0.0031	-39.4%	-26.6%	-0.1%	-25.8%	-39.8%
	$\hat{A}_2$	0.0002	-33.3%	-22.4%	-0.1%	$+13.1\%$	-33.9%
	MSE trace	0.0055	-37.5%	-25.8%	-0.1%	-24.0%	-37.9%
worst	$\hat{A}_0$	0.0022	-26.3%	-20.5%	-0.1%	-17.4%	-70.1%
	$\hat{A}_1$	0.0031	-28.9%	-23.3%	-0.1%	-21.4%	-80.1%
	$\hat{A}_2$	0.0002	-25.4%	-19.6%	-0.1%	$+50.8\%$	-77.2%
	MSE trace	0.0055	-28.2%	-22.0%	-0.1%	-15.5%	-75.1%
best	$\hat{A}_0$	0.0022	-39.2%	-25.0%	-0.1%	-28.5%	-9.5%
	$\hat{A}_1$	0.0031	-44.1%	-28.0%	-0.1%	-28.5%	-13.0%
	$\hat{A}_2$	0.0002	-36.9%	-23.5%	-0.1%	-28.5%	-11.3%
	MSE trace	0.0055	-42.1%	-27.0%	-0.1%	-28.5%	-11.7%
spherical constraints							
	MSE	LS	MAP$_s$	ridge	shrunken	minimax	MAP$_h$
random	$\hat{A}_0$	0.0022	-33.2%	-44.6%	-0.1%	-53.5%	-24.3%
	$\hat{A}_1$	0.0031	-38.1%	-42.0%	-0.1%	-55.3%	-28.0%
	$\hat{A}_2$	0.0002	-32.6%	-42.6%	-0.1%	$+29.9\%$	-24.1%
	MSE trace	0.0055	-35.9%	-43.4%	-0.1%	-51.3%	-26.4%
worst	$\hat{A}_0$	0.0022	-36.5%	-42.3%	-0.1%	-45.6%	-30.9%
	$\hat{A}_1$	0.0031	-41.1%	-46.8%	-0.1%	-51.7%	-42.5%
	$\hat{A}_2$	0.0002	-34.7%	-39.1%	-0.1%	$+78.0\%$	-39.6%
	MSE trace	0.0055	-33.9%	-44.7%	-0.1%	-35.5%	-37.5%
best	$\hat{A}_0$	0.0022	-39.2%	-45.5%	-0.1%	-58.3%	-9.5%
	$\hat{A}_1$	0.0031	-44.1%	-50.07%	-0.1%	-58.3%	-13.0%
	$\hat{A}_2$	0.0002	-36.9%	-23.5%	-0.1%	-58.3%	-11.3%
	MSE trace	0.0055	-42.1%	-47.1%	-0.1%	-58.3%	-11.7%

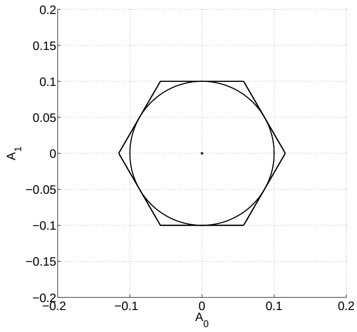

Fig. 2. A-priori information (circle) and its approximation (hexagon)

MAP$_s$**:** $\bar{\mathbf{x}} = \mathbf{0}$ (by symmetry of $\mathbb{X}$) and $\overline{\boldsymbol{\Sigma}} = 10^{-3} \begin{bmatrix} 5.9 & 0.3 \\ 0.3 & 6.5 \end{bmatrix}$

Ridge: $\boldsymbol{\Lambda} = 10^4 \; diag(0.0597, 1.3403)$,

$$\mathbf{K} = diag(0.01, 0.01), \; \mathbf{V}' = \begin{bmatrix} -0.9436 & 0.3310 \\ 0.3310 & 0.9436 \end{bmatrix}$$

Shrunken: $\check{\mathbf{x}} = [0.1047, \;\; 0.2984]'$

Minimax: $L^2 = 0.01$

The simulation results are reported in table II. From this table it can be seen that the MAP$_s$ estimator is the best, although it has been designed using conservative hexagonal constraints, i.e. $\widehat{\mathbb{X}}$, instead of the exact circle constraints, i.e. $\mathbb{X}$. Notice that, as in the previous example, the MAP$_h$ is not the best estimator (also in this case it must solve a nonlinear optimization problem with quadratic cost and constraints). The minimax estimator provides a MSE reduction for both parameters, but yields an higher MSE's trace than MAP$_s$.

TABLE II

SIMULATION RESULTS FOR THE FIRST-ORDER REGRESSION: CIRCLE CONSTRAINTS.

MSE	LS	MAP$_s$	ridge	shrunken	minimax	MAP$_h$
$\hat{A}_0$	0.0015	-32%	-24%	-0.1%	-24%	-19%
$\hat{A}_1$	0.0002	-24%	-18%	-0.1%	-12%	-14%
MSE trace	0.0017	-31%	-23%	-0.1%	-22%	-19%

C. *Two-dimensional case with irregular polytope constraints*

In this example $\mathbf{x} = [A_0, A_1]'$ is assumed to belong to the irregular polytope (see fig. 3):

$$\begin{aligned}\mathbb{X} = \mathbb{X}_p \quad &= \quad \{\mathbf{x} = [A_0, A_1]' : \ A_1 \geq 0.1, \ 0.1 \leq A_0 \leq 0.3, \\ &\qquad A_1 - 2A_0 \leq 0.1, \ A_1 + 3A_0 \leq 1.1\} \,.\end{aligned}$$

The design parameters for the MAP$_s$ estimator turn out to be:

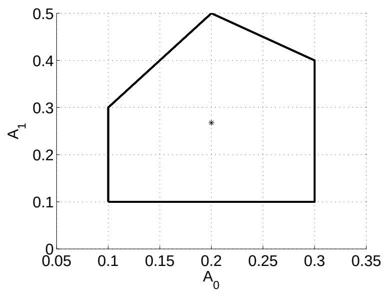

Fig. 3. A-priori information irregular polytope

$$\overline{\mathbf{\Sigma}} = 10^{-3} \begin{bmatrix} 5.9 & 0.3 \\ 0.3 & 6.5 \end{bmatrix}, \qquad \bar{\mathbf{x}} = \begin{bmatrix} 0.2 \\ 0.2676 \end{bmatrix}.$$

Since the polytope $\mathbb{X}$ is not symmetrical, the parameter $\bar{\mathbf{x}}$ has also been found by solving the LMI problem. The resulting performance of the MAP$_s$ and MAP$_h$ estimators are compared in table III.

TABLE III

SIMULATION RESULTS FOR THE FIRST-ORDER REGRESSION: NON SYMMETRICAL POLYTOPIC CONSTRAINTS.

MSE	LS	MAP$_s$	MAP$_h$
$\hat{A}_0$	0.0015	-22.8%	-23.7%
$\hat{A}_1$	0.0002	-15.9%	-16.2%
MSE trace	0.0017	-22.1%	-22.8%

VI. CONCLUSIONS

The paper has addressed parameter estimation in linear regression models under polytopic constraints on the parameter vector to be estimated, following a superefficiency approach. This approach consists of fixing a parametrized structure of the estimator and then designing the free parameters of the estimator so as to guarantee, for any admissible value of the unknown parameter vector to be estimated, an improvement of the matrix-MSE w.r.t. the standard LS estimator. The superefficiency, also called LS-domination, approach provides significant computational savings, which are relevant for applications with stringent real time requirements, with respect to alternative approaches for estimation of constrained parameters based on either constrained optimization or Monte Carlo methods [20]. The proposed technique adopts a MAP estimator with prior Gaussian distribution on the unknown parameter and selects the mean and covariance of such a distribution via solution of an LMI optimization problem, so as to guarantee a lower matrix-MSE than the LS estimator. The resulting estimator has been compared with existing estimators for constrained parameters and a significant performance improvement has been found, especially when the size of the constrained region is small compared to the confidence regions of the LS estimate (i.e. for short data samples and/or for high noise variance).

With respect to the vast amount of work on superefficient or LS-dominating estimators, this paper is characterized by the following distinguishing aspects.

1) It allows to design, based on LMIs, estimators that dominate the LS estimator in a stronger sense, i.e. that give a lower MSE matrix while most existing design methods guarantee only a lower trace of the MSE matrix.
2) It allows to exploit in a simple way linear inequality constraints, ubiquitous in engineering applications, without conservativeness.
3) It avoids the worst-case design criterion adopted in the minimax approach.
4) It exploits the available degrees of freedom by minimizing a simple optimality criterion.

Future work will be devoted to investigate optimality criteria more related to the estimation performance (e.g. the MSE trace) and to ascertain the solvability of the resulting optimization problem.

VII. Appendix - Derivation of (III.2) and (III.3)

The MSE for the estimator (III.1) is given by

$$\begin{aligned}\boldsymbol{\Sigma} &\triangleq E\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})'\right] \\ &= E\left[\left(\mathbf{x} - \left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} + \overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{y} + \overline{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{x}}\right)\right)\left(\dots\right)'\right]\end{aligned} \tag{VII.1}$$

Replacing $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v}$ and exploiting the identity $\mathbf{x} = \left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} + \overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} + \overline{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{x}$, (VII.1) yields:

$$\boldsymbol{\Sigma} = E\left[\left(\left(\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} + \overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\left(\overline{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}-\overline{\mathbf{x}}) - \mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{v}\right)\right)\left(\dots\right)'\right] \tag{VII.2}$$

Hence, taking into account that $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_v)$ and $\mathbf{C}'\boldsymbol{\Sigma}_v^{-1}\mathbf{C} = \boldsymbol{\Sigma}_{LS}^{-1}$, the MSE's expression (III.2) is obtained from (VII.2) by straightforward calculations. To derive the domination condition (III.3) it must be imposed that $\boldsymbol{\Sigma} \leq \boldsymbol{\Sigma}_{LS}$, that is

$$\begin{aligned} \boldsymbol{\Sigma}_{LS} \quad \geq \quad & \left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1} \\ & \left[\overline{\boldsymbol{\Sigma}}^{-1}(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})'\overline{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_{LS}^{-1}\right] \\ & \left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1} \end{aligned} \tag{VII.3}$$

Hence, left and right multiplying (VII.3) by $\left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right)$, it follows that

$$\overline{\boldsymbol{\Sigma}}^{-1}(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})'\overline{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_{LS}^{-1} \leq \left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right)\boldsymbol{\Sigma}_{LS}\left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right) \tag{VII.4}$$

Next, left and right multiplying (VII.4) by $\overline{\boldsymbol{\Sigma}}$, yields

$$\begin{aligned} (\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})' + \overline{\boldsymbol{\Sigma}}\ \boldsymbol{\Sigma}_{LS}^{-1}\ \overline{\boldsymbol{\Sigma}} & \leq \overline{\boldsymbol{\Sigma}}\left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right) \\ & \qquad \boldsymbol{\Sigma}_{LS}\ \left(\boldsymbol{\Sigma}_{LS}^{-1} + \overline{\boldsymbol{\Sigma}}^{-1}\right)\overline{\boldsymbol{\Sigma}} \\ (\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})' + \overline{\boldsymbol{\Sigma}}\ \boldsymbol{\Sigma}_{LS}^{-1}\ \overline{\boldsymbol{\Sigma}} & \leq \overline{\boldsymbol{\Sigma}}\ \boldsymbol{\Sigma}_{LS}^{-1}\ \overline{\boldsymbol{\Sigma}} + 2\overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_{LS} \\ (\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})' & \leq 2\overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_{LS} \end{aligned}$$

from which (III.3) directly follows. ■

REFERENCES

[1] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 197-206, University of California Press, Berkeley, 1956.

[2] W. James, C Stein, "Estimation of quadratic loss", *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 361-379, University of California Press, Berkeley, 1961.

[3] W.E. Strawderman, "Proper Bayes minimax estimators of a multivariate normal mean", *Ann. Math. Statist.*, vol. 42, pp. 385-388, 1971.

[4] K. Alam, "A family of admissible minimax estimators of the mean of a multivariate normal distribution", *Ann. Statist.*, vol. 1, pp. 517-525, 1973.

[5] J.O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss", *Ann. Statist.*, vol. 4, n. 1, pp. 223-226, 1976.

[6] P. Stoica, B. Ottersten, "The Evil of Superefficiency", *Signal Processing*, vol. 55, pp. 133-136, 1996.

[7] E.L. Lehmann, G. Casella, *Theory of Point Estimation*, Springer, New York, 2nd edition, 1999.

[8] Y. C. Eldar, A. Ben-Tal, A. Nemirovski, "Linear Minimax Regret Estimation of deterministic parameters with Bounded Data Uncertainties", *IEEE Trans. on Signal Processing*, vol. 52, pp. 2177-2188, August 2004

[9] Y. C. Eldar, A. Ben-Tal, A. Nemirovski, "Robust Mean-Squared Error Estimation in the Presence of Model Uncertainties", *IEEE Trans. on Signal Processing*, vol. 53, pp. 168-181, January 2005

[10] Z. Ben-Haim, Y.C. Eldar, "Blind minimax estimators: improving on least-squares estimation", *IEEE Workshop on Statistical Signal Processing (SSP'05)*, Bordeaux, France, 2005.

[11] Z. Ben-Haim, Y.C. Eldar, "Maximum set estimators with bounded estimation error", *IEEE Trans. on Signal Processing*, vol. 53, pp. 3172-3182, 2005.

[12] Y.C. Eldar, “Comparing between estimation approaches. Admissible and dominating linear estimators”, to appear in *IEEE Trans. on Signal Processing*; available at the URL http://www.ee.technion.ac.il/Sites/People/YoninaEldar/.

[13] A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, 1977

[14] A.E. Hoerl, R.W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, *Technometrics*, 12, 55-67, 1970.

[15] L.S. Mayer, T.A. Willke, “On Biased Estimation in Linear Models”, *Technometrics*, 15, 497-508, 1973.

[16] Y. C. Eldar, A. V. Oppenheim, “Covariance Shaping Least-Squares Estimation”, *IEEE Trans. on Signal Processing*, vol. 51, pp. 686-697, 2003

[17] S.M. Kay, *Fundamentals of Statistical Signal Processing: estimation theory*, Prentice Hall, 1993.

[18] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, 1994.

[19] A. Benavoli, L. Chisci, A. Farina, L. Ortenzi, G. Zappa, “Hard-constrained vs. soft-constrained parameter estimation”, to appear in *IEEE Trans. on Aerospace and Electronic Systems*; also available at the URL http://www.dsi.unifi.it/users/chisci/.

[20] B. Ristic, S. Arulampalam, N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.

[21] P. Gahinet, A. Nemirovshi, A. Laub, M. Chilali, “MATLAB LMI Control Toolbox”, Natick, MA: MathWorks, May 1995