Hard-constrained vs. soft-constrained parameter estimation

A. Benavoli, L. Chisci *Member, IEEE*, A. Farina, *Fellow, IEEE*,
L. Ortenzi, G. Zappa

***Abstract*— The paper aims at contrasting two different ways of incorporating a-priori information in parameter estimation, i.e. hard-constrained and soft-constrained estimation. Hard-constrained estimation can be interpreted, in the Bayesian framework, as *Maximum A-posteriori Probability (MAP)* estimation with uniform prior distribution over the constraining set, and amounts to a constrained *Least-Squares (LS)* optimization. Novel analytical results on the statistics of the hard-constrained estimator are presented for a linear regression model subject to lower and upper bounds on a single parameter. This analysis allows to quantify the *Mean Squared Error (MSE)* reduction implied by constraints and to see how this depends on the size of the constraining set compared to the confidence regions of the unconstrained estimator. Contrastingly, soft-constrained estimation can be regarded as MAP estimation with Gaussian prior distribution and amounts to a less computationally demanding unconstrained LS optimization with a cost suitably modified by the mean and covariance of the Gaussian distribution. Results on the design of the prior covariance of the soft-constrained estimator for optimal MSE performance are also given. Finally, a practical case-study concerning a line fitting estimation problem is presented in order to validate the theoretical results derived in the paper as well as to compare the performance of the hard-constrained and soft-constrained approaches under different settings.**

***Index Terms*— Constrained estimation, Mean Squared Error, MAP estimator.**

I. INTRODUCTION

Consider the classical *Maximum Likelihood (ML)* estimation of a parameter vector, given a set of measurements. The estimate is provided by the minimization of the likelihood function; in this case, the *Cramer-Rao Lower Bound (CRLB)* can be computed by the inversion of the *Fisher Information Matrix (FIM)* and gives the minimum theoretically achievable variance of the estimate [2], [3]. In this paper, it will be investigated how to improve the estimation accuracy by exploiting a-priori information on the parameters to be estimated.

In [1] the CRLB is computed in presence of equality "constraints". The constrained estimation problem, in this case, can be reduced to an unconstrained problem by the use of Lagrange multipliers. Equality constraints [4], [5] express very precise a-priori information on the parameters to be estimated; unfortunately, in several practical applications, such precise information is not available. In most cases, "inequality" constraints are available [6] with a consequent complication of the estimation problem and of the computation of the CRLB.

Authors' addresses:
A. Benavoli, *L. Chisci* and *G. Zappa*, DSI, Università di Firenze, Firenze, Italy, e-mail: $\{benavoli,\ chisci\}$@dsi.unifi.it; *A. Farina* and *L. Ortenzi*, Engineering Division, SELEX - Sistemi Integrati S.p.A., Rome, Italy, e-mail: $\{afarina,\ lortenzi\}$@selex-si.com.

There are clearly many ways of expressing a-priori information. This paper will specifically address two possible approaches. A first approach, referred to as *"hard-constrained"* estimation, assumes inequality constraints (e.g. lower and upper bounds) on the parameters to be estimated and formulates, accordingly, the estimation problem as a constrained *Least Squares (LS)* optimization problem. The second approach, referred to as *"soft-constrained"* estimation, considers the parameters to be estimated as random variables with a Gaussian probability density function, whose statistical parameters (mean and covariance) are a-priori known. In this case the a-priori knowledge on the parameters simply modifies the LS cost functional and the resulting estimation problem amounts to a less computationally demanding unconstrained LS optimization. Hard-constrained and soft-constrained estimation can be interpreted, in a Bayesian framework, as *maximum a-posteriori probability (MAP)* estimation with a uniform and, respectively, Gaussian prior distribution. Another alternative could be constrained *Minimum Mean Squared Error (MMSE)* estimation which can be numerically approximated via *Monte Carlo methods* [7], [8]. This approach, which involves extensive computations for numerical integration, is computationally much more expensive and will, therefore, not be considered in this work.

A theoretical analysis on the statistics of the hard-constrained estimator for a linear regression model subject to hard bounds on a single parameter will be carried out. The obtained results allow to quantify the bias and MSE reduction implied by the hard bounds in terms of the width of the bounding interval and of the location of the true parameter in such an interval. To the best of the authors' knowledge there are no similar results in the literature giving an analytical quantification of the performance improvement achievable using a-priori information on parameters in terms of hard constraints. It will be outlined how an extension of such exact results to the case of multiple bounded parameters seems analytically intractable. It will also be shown how to use the single-bounded parameter analysis sequentially, i.e. parameter by parameter, in order to provide a, possibly conservative, estimate of the MSE reduction achievable under multiple bounded parameters. A further contribution concerns the design of the prior covariance of the soft-constrained estimator so as to make hard constraints and soft constraints information-theoretically comparable.

The remaining part of the paper is organized as follows. Section 2 formulates the parameter estimation problem for a linear regression model and reviews classical results of unconstrained estimation. Section 3 deals with the hard-constrained approach and provides novel results on the statistics of the hard-constrained estimator. Section 4 deals with the soft-constrained approach and provides results on the choice of the design parameters (prior covariance) of the soft-constrained estimator. Section 5 examines a practical case-study concerning a line fitting estimation problem in order to validate the presented theoretical results as well as to compare the performance improvement achievable with the hard-constrained and soft-constrained approaches. Finally, section 6 summarizes the main results of the paper drawing some concluding remarks.

II. UNCONSTRAINED ESTIMATION

Notation

Throughout the paper the following notation will be adopted.

- Lower-case, boldface lower-case and boldface upper-case letters denote scalars, vectors and, respectively, matrices;
- Prime denotes transpose and tr denotes trace.
- $\|\mathbf{v}\|^2_{\mathbf{M}} \triangleq \mathbf{v}'\mathbf{M}\mathbf{v}$.
- Given a vector $\mathbf{v}$: v_i denotes its ith entry, $\mathbf{v}_{i:j} = [v_i, \ldots, v_j]'$ if $i \leq j$ or $\mathbf{v}_{i:j} = \mathbf{0}$ if $i > j$. Similarly, given a matrix $\mathbf{M}$: $\mathbf{m}_i$ denotes its ith column and $\mathbf{M}_{i:j} = [\mathbf{m}_i, \ldots, \mathbf{m}_j]$ if $i \leq j$ or $\mathbf{M}_{i:j} = \mathbf{0}$ when $i > j$.
- Given the estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, $\tilde{\mathbf{x}} \triangleq \hat{\mathbf{x}} - \mathbf{x}$ is the associated estimation error.
- $\text{erf}(s) \triangleq \frac{2}{\sqrt{\pi}} \int_0^s e^{-r^2} dr$ is the error function.
- $\mathbf{v} \sim \mathcal{N}(\mathbf{m}_\mathbf{v}, \mathbf{\Sigma}_\mathbf{v})$ means that $\mathbf{v}$ is normally distributed with mean $\mathbf{m}_\mathbf{v}$ and covariance $\mathbf{\Sigma}_\mathbf{v}$; $\mathbf{v} \sim \mathcal{U}(\mathbf{V})$ means that $\mathbf{v}$ is uniformly distributed over the set $\mathbf{V}$.
- Given the set $\mathbf{X}$, $\mathcal{I}_\mathbf{X}$ is the *indicator function* of $\mathbf{X}$ defined as $\mathcal{I}_\mathbf{X}(x) = 1$ for $x \in \mathbf{X}$ and $\mathcal{I}_\mathbf{X}(x) = 0$ for $x \notin \mathbf{X}$; $\delta(\cdot)$ is the *delta of Dirac*.

Problem formulation

Consider the linear regression model

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1}$$

where: $\mathbf{x} \in \mathbb{R}^n$ is the variable to be estimated; $\mathbf{z} \in \mathbb{R}^N$ is the observed variable; $\mathbf{H} \in \mathbb{R}^{N\times n}$ is a known matrix; $\mathbf{w} \in \mathbb{R}^N$ is the measurement noise satisfying $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{w})$. It is well known that in the *unconstrained* case (i.e. the case in which no a-priori information on the parameters is assumed) the optimal estimate of $\mathbf{x}$ - in the weighted LS or, equivalently, ML, MAP or MMSE sense - is given by

$$\hat{\mathbf{x}} \;=\; \left(\mathbf{H}'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{z}. \tag{2}$$

It is also well known that $\hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{x}, \mathbf{\Sigma})$, i.e. $\hat{\mathbf{x}}$ is normally distributed with mean $\mathbf{x}$ (unbiased estimate) and covariance matrix [2], [3]

$$\mathbf{\Sigma} \;\triangleq\; E\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\right] = \left(\mathbf{H}'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1} \tag{3}$$

For the sake of comparison with the *constrained* case that will be addressed in the next two sections, let us consider the situation in which one scalar parameter (without loss of generality the first entry of $\mathbf{x}$) is known. To this end, let x_1 denote the first entry of $\mathbf{x}$ and introduce the partitioning

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{H} = [\mathbf{h}_1,\ \mathbf{H}_2], \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \boldsymbol{\sigma}_{21}' \\ \boldsymbol{\sigma}_{21} & \mathbf{\Sigma}_2 \end{bmatrix}$$

The optimal estimate of $\mathbf{x}_2$ for known x_1 is trivially

$$\hat{\mathbf{x}}_{2|1} = \left(\mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}_2\right)^{-1}\mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\left(\mathbf{z} - \mathbf{h}_1 x_1\right) \tag{4}$$

and the corresponding MSE is

$$\mathbf{\Sigma}_{2|1} = E\left[\tilde{\mathbf{x}}_{2|1}\tilde{\mathbf{x}}_{2|1}'\right] \;=\; \left(\mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}_2\right)^{-1} \tag{5}$$

Obviously the knowledge of x_1 provides a MSE reduction on $\mathbf{x}_2$, i.e. $\mathbf{\Sigma}_{2|1} \leq \mathbf{\Sigma}_2$. More precisely, via matrix algebra, it can be easily checked that

$$\mathbf{\Sigma}_{2|1} = \mathbf{\Sigma}_2 \;-\; \mathbf{L}\mathbf{L}'\ \sigma_1^2 \tag{6}$$

where

$$\mathbf{L} \triangleq \left(\mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}_2\right)^{-1}\mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{h}_1 = \mathbf{\Sigma}_{2|1}\ \mathbf{H}_2'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{h}_1 \tag{7}$$

The relationship (6) expresses the fact that the knowledge of x_1 implies a MSE reduction on $\mathbf{x}_2$ which is proportional, via the matrix gain $\mathbf{L}\mathbf{L}'$, to the MSE σ_1^2 that would be obtained for x_1 in the absence of such knowledge. In practice, perfect knowledge of a parameter is never available and the MSE reduction in (6) represents a theoretical limitation. In the next two sections we shall investigate analytically the MSE reduction implied by a partial knowledge of the parameter x_1 in terms of either (upper and lower) hard bounds (*hard-constrained estimation*) or a Gaussian prior distribution (*soft-constrained estimation*).

III. HARD-CONSTRAINED ESTIMATION

In this section, it will be assumed that the a-priori information on the parameters $\mathbf{x}$ is specified by a membership set $\mathbf{X}$ (hard constraints), i.e. it is assumed that $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$. A possible choice is

$$\begin{aligned} \mathbf{X} \;=\; & \left\{\mathbf{x} = [x_1, x_2, \ldots, x_n]' :\right. \\ & \left. x_{i,min} \leq x_i \leq x_{i,max},\ i = 1, 2, \ldots, n\right\} \end{aligned} \tag{8}$$

There exist several approaches to encompass the a-priori information provided by the hard bounds in the inference process. In particular one may assume a *uniform prior distribution on* the parameter vector $\mathbf{x}$, i.e. a prior *probability density function (pdf)*

$$f_0(\mathbf{x}) \;=\; \begin{cases} \dfrac{1}{\mu(\mathbf{X})}, & \mathbf{x} \in \mathbf{X} \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

where $\mu(\mathbf{X})$ is the Lebesgue measure of $\mathbf{X}$, assumed finite. Consequently, the posterior pdf of $\mathbf{x}$ conditioned on the observation $\mathbf{z}$ becomes

$$\begin{aligned} f(\mathbf{x}|\mathbf{z}) \;&=\; \frac{f_\mathbf{w}(\mathbf{z} - \mathbf{H}\mathbf{x})\ f_0(\mathbf{x})}{\displaystyle\int_\mathbf{X} f_\mathbf{w}(\mathbf{z} - \mathbf{H}\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x}} \\ &=\; \begin{cases} \dfrac{1}{c}\ exp\left(-\|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{\Sigma}_\mathbf{w}^{-1}}\right), & \mathbf{x} \in \mathbf{X} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

(10)

where

$$c = \int_{\mathbf{X}} exp\left(-\left\|\mathbf{z}-\mathbf{H}\mathbf{x}\right\|^2_{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}\right) d\mathbf{x}$$

is a normalizing constant. In this case it is well known that the MMSE estimate of $\mathbf{x}$ is provided by the conditional mean

$$\hat{\mathbf{x}}_{MMSE} = E\left[\mathbf{x}|\mathbf{z}\right] = \frac{1}{c}\int_{\mathbf{X}} \mathbf{x}\; exp\left(-\left\|\mathbf{z}-\mathbf{H}\mathbf{x}\right\|^2_{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}\right) d\mathbf{x} \tag{11}$$

and the corresponding MMSE (conditioned on $\mathbf{z}$) is given by the conditional variance

$$\begin{aligned} var(\hat{\mathbf{x}}_{MMSE}) &= \frac{1}{c}\int_{\mathbf{X}} \tilde{\mathbf{x}}_{MMSE}\; \tilde{\mathbf{x}}'_{MMSE} \\ &\quad exp\left(-\left\|\mathbf{z}-\mathbf{H}\mathbf{x}\right\|^2_{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}\right) d\mathbf{x} \end{aligned} \tag{12}$$

In this paper, however, the attention will be focused on *Maximum A-Posteriori (MAP)* estimation for which some interesting results will be derived in the sequel, although MMSE estimation is clearly a possible alternative for hard-constrained estimation exploiting, for instance, *Monte Carlo methods* [7], [8].

The MAP estimate of $\mathbf{x}$, under the assumption of a uniform prior distribution, is given by

$$\hat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} f(\mathbf{x}|\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbf{X}} \left\|\mathbf{z}-\mathbf{H}\mathbf{x}\right\|^2_{\mathbf{\Sigma}_{\mathbf{w}}^{-1}} \tag{13}$$

Hence the MAP estimate coincides with the *constrained ML* (or equivalently *constrained weighted LS*) estimate

$$\hat{\mathbf{x}}_c = \arg\max_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x}|\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbf{X}} \left\|\mathbf{z}-\mathbf{H}\mathbf{x}\right\|^2_{\mathbf{\Sigma}_{\mathbf{w}}^{-1}} \tag{14}$$

In order to quantify the benefits provided by the knowledge of $\mathbf{X}$, it would be helpful to evaluate the *bias* $E\left[\tilde{\mathbf{x}}_c\right]$ and the *MSE* $E\left[\tilde{\mathbf{x}}_c\tilde{\mathbf{x}}'_c\right]$ of the constrained estimator (14). Unfortunately this is a prohibitive task for a general membership set $\mathbf{X}$.

A. Single bounded parameter

Conversely, interesting analytical formulas can be obtained in the special case of a single bounded parameter, i.e.

$$\mathbf{X} = \left\{\mathbf{x} = [x_1, \mathbf{x}'_2]' : x_{1,min} \le x_1 \le x_{1,max}, \mathbf{x}_2 \in \mathbb{R}^{n-1}\right\} \tag{15}$$

Such formulas are given by the following theorem.

Theorem 1 - *Let us consider the constrained estimation problem (14) with $\mathbf{X}$ as in (15). Then:*
(i) *the resulting estimator turns out to be*

$$\hat{x}_{1,c} = \begin{cases} x_{1,min}, & \text{if } \hat{x}_1 < x_{1,min} \\ \hat{x}_1, & \text{if } x_{1,min} \le \hat{x}_1 \le x_{1,max} \\ x_{1,max}, & \text{if } \hat{x}_1 > x_{1,max} \end{cases} \tag{16}$$

$$\hat{\mathbf{x}}_{2,c} = \mathbf{\Sigma}_{2|1}\mathbf{H}'_2\mathbf{\Sigma}_{\mathbf{w}}^{-1}\left(\mathbf{z}-\mathbf{h}_1\hat{x}_{1,c}\right) \tag{17}$$

where $\hat{\mathbf{x}} = [\hat{x}_1, \hat{\mathbf{x}}'_2]'$ is the unconstrained estimate provided by (2) and $\mathbf{\Sigma}_{2|1}$ is given by (5).
(ii) *The constrained estimate $\hat{x}_{1,c}$ provided by (16) has pdf*

$$\begin{aligned} f_c(\xi_1) &= F(x_{1,min})\;\delta(\xi_1 - x_{1,min}) \\ &+ (1 - F(x_{1,max}))\;\delta(\xi_1 - x_{1,max}) \\ &+ f(\xi_1)\;\mathcal{I}_{[x_{1,min},x_{1,max}]}(\xi_1) \end{aligned} \tag{18}$$

where $f(\cdot)$ is the Gaussian pdf of $\hat{x}_1$ and $F(\cdot)$ the corresponding distribution function.
Further, let

$$\alpha \triangleq \frac{x_{1,max} - x_{1,min}}{2\;\sigma_1}, \qquad \beta \triangleq \frac{x_1 - x_{1,min}}{x_{1,max} - x_{1,min}} \tag{19}$$

then:
(iii) *the estimator (16)-(17) is biased with bias*

$$E\left[\tilde{x}_{1,c}\right] = \delta(\alpha,\beta)\;\sigma_1 \tag{20}$$

$$E\left[\tilde{\mathbf{x}}_{2,c}\right] = -\;\mathbf{L}\;\delta(\alpha,\beta)\;\sigma_1 \tag{21}$$

where the matrix $\mathbf{L}$ is defined in (7) and

$$\delta(\alpha,\beta) = \Delta(\alpha\beta) - \Delta(\alpha(1-\beta)) \tag{22}$$

$$\Delta(y) \triangleq \frac{1}{\sqrt{2\pi}}\;\exp\left(-2y^2\right) - y\;\left(1 - erf\left(\sqrt{2}\;y\right)\right); \tag{23}$$

(iv) *the estimator (16)-(17) is characterized by the following MSE*

$$\sigma^2_{1,c} \triangleq E\left[\tilde{x}^2_{1,c}\right] = (1-\gamma(\alpha,\beta))\;\sigma^2_1 \tag{24}$$

$$\mathbf{\Sigma}_{2,c} \triangleq E\left[\tilde{\mathbf{x}}_{2,c}\tilde{\mathbf{x}}'_{2,c}\right] = \mathbf{\Sigma}_2 - \mathbf{L}\mathbf{L}'\;\gamma(\alpha,\beta)\;\sigma^2_1 \tag{25}$$

where

$$\gamma(\alpha,\beta) = \Gamma(\alpha\beta) + \Gamma(\alpha(1-\beta)) \tag{26}$$

$$\begin{aligned} \Gamma(y) &\triangleq \frac{1}{2}\left(1 - erf\left(\sqrt{2}\;y\right)\right) + \sqrt{\frac{2}{\pi}}\;y\;\;\exp\left(-2y^2\right) \\ &- 2y^2\left(1 - erf\left(\sqrt{2}\;y\right)\right). \end{aligned} \tag{27}$$

Proof - *See the Appendix.*

Remarks on Theorem 1

- In the case of a single bounded parameter, the estimator becomes straightforward and requires no optimization. In fact, it first determines $\hat{x}_{1,c}$ by the simple check (16) on the unconstrained estimate $\hat{x}_1$ and then determines $\hat{\mathbf{x}}_{2,c}$ by the conditioning over $x_1 = \hat{x}_{1,c}$ in (17). Conversely, in the general case, the estimator (14) involves the solution of a *nonlinear programming* problem, e.g. a *quadratic programming* problem whenever $\mathbf{X}$ is a polyhedron (i.e. a set defined by linear inequalities).
- The estimation performance (bias and MSE reduction) is completely characterized by the variables α and β defined in (19). The variable $\alpha \ge 0$ is the width of the bounding semi-interval of x_1 normalized by the standard deviation of the unconstrained estimate of x_1. Notice that $\alpha \to \infty$ corresponds to the unconstrained estimation (2) while $\alpha = 0$ corresponds to the estimation for known

x_1 (4). Conversely $\beta \in [0,1]$ expresses the location of the true parameter x_1 within the bounding interval $[x_{1,min}, x_{1,max}]$ i.e. $\beta = 0$, $\beta = 1$, $\beta = \frac{1}{2}$ correspond to $x_1 = x_{1,min}$, $x_1 = x_{1,max}$,
$x_1 = x_{1,m} \stackrel{\triangle}{=} (x_{1,min} + x_{1,max})/2$ respectively.

- The formulas (20)-(23) give the bias of the estimator as a function of α and β. Notice that the estimator is unbiased for $\alpha = 0$ (known x_1) and for $\alpha \to \infty$ (unconstrained estimation), since $\Delta(y) \to 0$ for $y \to \infty$. The estimator is also unbiased for $\beta = 1/2$, i.e. whenever the true parameter x_1 coincides with the midpoint $x_{1,m}$ of the bounding interval. Further, since $\Delta(y)$ is monotonically decreasing for $y \geq 0$, the bias is positive (negative) if and only if $\beta < 1/2$ or equivalently $x_1 < x_{1,m}$ (respectively $\beta > 1/2$ or $x_1 > x_{1,m}$). This fact has a simple interpretation: whenever the true parameter x_1 is located unsymmetrically with respect to the bounds, the closer bound pushes the estimate towards the other (farther) bound. Also notice that for a given α the bias is anti-symmetric w.r.t. $\beta = 1/2$ and is maximum (in absolute value) for $\beta = 0$ and $\beta = 1$. Finally, since $\Delta(y)$ is decreasing for $y \geq 0$ and $\Delta(0) = 1/\sqrt{2\pi}$, the bias never exceeds the threshold $\sigma_1/\sqrt{2\pi}$. However, the bias is not a good measure of the estimator's quality. In fact, a little amount of bias can be tolerated provided that this implies a MSE reduction.
- The formulas (24)-(27) provide the MSE as a function of α and β. From (24), it is clear that $\gamma(\alpha,\beta) \in [0,1]$ represents the relative MSE reduction on x_1 implied by the a-priori information $x_1 \in [x_{1,min}, x_{1,max}]$, i.e.

$$\gamma(\alpha,\beta) \stackrel{\triangle}{=} \frac{\sigma_{1,c}^2 - \sigma_1^2}{\sigma_1^2}.$$

Moreover, (25) states that the implied MSE reduction on $\mathbf{x}_2$ is obtained via multiplication of the MSE reduction on x_1 by the gain matrix $\mathbf{L}'\mathbf{L}$ (like in the case of x_1 known). It is interesting to analyze the dependence of the MSE reduction on the amount of a-priori information available on x_1 (variable α) as well as on the position of the true parameter x_1 (variable β). Notice that for $\alpha \to 0$ (tight constraints), $\gamma(\alpha,\beta) \to 1$ (100% MSE reduction); in particular for $\alpha = 0$ (24) reduces, as expected, to the formula (6) relative to the case in which x_1 is known. Conversely, since $\Gamma(y) \to 0$ for $y \to \infty$, $\gamma(\alpha,\beta) \to 0$ (no reduction) for $\alpha \to \infty$ (loose constraints), as expected. As far as the dependence on β is concerned, notice that $\gamma(\alpha,\beta)$ is symmetric w.r.t. $\beta = 1/2$. Further useful results concerning the dependence of $\gamma(\alpha,\beta)$ on β are given in the subsequent corollary.

Corollary 1 - *Let us consider the normalized MSE $\sigma_{1,c}^2/\sigma_1^2 = 1 - \gamma(\alpha,\beta)$, where $\gamma(\alpha,\beta)$ is defined in (26)-(27). Then there exists $\alpha_0 > 0$ such that:*

(i) *for $\alpha \leq \alpha_0$, the MSE has an unique minimum for $\beta = 1/2$;*

(ii) *for $\alpha > \alpha_0$, the MSE has two minima symmetric w.r.t. $\beta = 1/2$, which tend to the endpoints of the interval ($\beta = 0$ and $\beta = 1$) for increasing α.*

Proof - *See the Appendix.*

The above analysis is illustrated in fig. 1 which plots the relative MSE $1 - \gamma(\alpha,\beta)$ and displays the two types of behaviour: (1) unique minimum in the center of the interval for $\alpha \leq \alpha_0$; (2) two minima located symmetrically w.r.t. the center of the interval for $\alpha > \alpha_0$. Notice that the threshold value is $\alpha_0 \approx 0.75$. This means that for an interval $[x_{1,min}, x_{1,max}]$ sufficiently small, i.e. $x_{1,max} - x_{1,min} \leq 1.5\ \sigma_1$, a maximum MSE reduction is obtained whenever the true parameter is exactly in the middle of the interval, i.e. $x_1 = (x_{1,max} + x_{1,min})/2$. Conversely, for an increasing interval width, the % MSE reduction decreases and for $x_{1,max} - x_{1,min} > 1.5\ \sigma_1$ the minimum splits up into two minima symmetrically moving towards the interval's extreme points. Please notice that, also for very large α (i.e. $\alpha >> \alpha_0$) a significant MSE reduction is obtained whenever the true parameter is located close to the interval's boundaries.

B. Multiple bounded parameters

Let us now consider the general case of multiple bounded parameters wherein the parameter membership set $\mathbf{X}$ is an hyper-rectangle in $\mathbb{R}^n$ as defined in (8). Now the dependence of $\hat{x}_{1,c}$ from $\hat{x}_1$ is no longer provided by (16) and may involve $\hat{\mathbf{x}}_2$. The relationship (16) holds if $\hat{x}_1 \in [x_{1,min}, x_{1,max}]$ but it may happen, for instance, that if $\hat{x}_i > x_{i,max}$ for some $i \neq 1$ and $\hat{x}_1 > x_{1,max}$, nevertheless $\hat{x}_{1,c} < x_{1,max}$. This is related to the ellipsoidal shape of the level surfaces of the LS cost functional $(\mathbf{x} - \hat{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \hat{\mathbf{x}})$. In the case of a single bounded parameter it was possible to derive the pdf $f_c(\cdot)$, see (18), as the sum of three terms one for each possible set of active constraints: a continuous term $f(\cdot)$ with support the interval $[x_{1,min}, x_{1,max}]$ and two weighted singular (impulsive) terms centered in $x_{1,min}$ and $x_{1,max}$. The weights of the impulsive terms correspond to the probabilities, under the unconstrained Gaussian pdf $f(\cdot)$, of the regions $x_1 < x_{1,min}$ and $x_1 > x_{1,max}$ which, in this simple case, represent the probabilities that the active constraint is $x_1 = x_{1,min}$ or, respectively, $x_1 = x_{1,max}$. In the general case of n bounded parameters, there are clearly 3^n possible active sets and the situation becomes by far more complicated. Let $a \in \mathcal{A}$ denote a generic active set, $\mathcal{A}$ being the set of all possible active sets including the empty set $\emptyset$. Then a induces a partitioning of $\mathbf{x}$ into *active* components $\mathbf{x}_a \in \mathbb{R}^{n_a}$, subject to the equality constraint $\mathbf{x}_a = \check{\mathbf{x}}_a$ where $\check{\mathbf{x}}_a$ is a vector of bounds depending on a, and *free* components $\mathbf{x}_f \in \mathbb{R}^{n-n_a}$. Notice that the problem (14) is equivalent to

$$\hat{\mathbf{x}}_c = \arg\min_{\mathbf{x} \in \mathbf{X}} \ \|\mathbf{x} - \hat{\mathbf{x}}\|_{\boldsymbol{\Sigma}^{-1}}^2 \tag{28}$$

which can be regarded as a *multiparametric Quadratic Programming (mpQP)* problem where $\hat{\mathbf{x}}$ is a parameter of the optimization problem and the vector $\mathbf{x}$ must be optimized for all values of $\hat{\mathbf{x}}$. By the theory of mpQP [9], [10], for each $a \in \mathcal{A}$ there exists a polyhedral set $\widehat{\mathbf{X}}_a$ such that

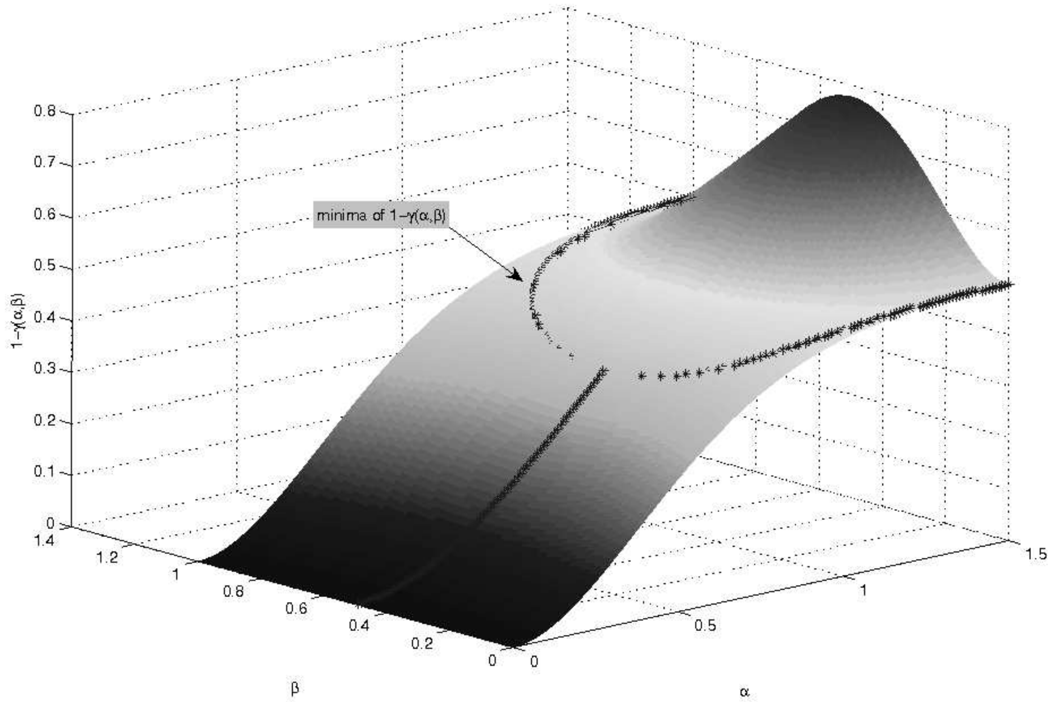

Fig. 1. Constrained-to-unconstrained MSE ratio $\sigma^2_{1,c}/\sigma^2_1$ vs. α and β.

whenever $\hat{\mathbf{x}} \in \widehat{\mathbf{X}}_a$ the inequality-constrained estimate $\hat{\mathbf{x}}_c$ in (28) coincides with the equality-constrained estimate under the active constraints $\mathbf{x}_a = \check{\mathbf{x}}_a$, namely

$$\widehat{\mathbf{X}}_a = \left\{ \hat{\mathbf{x}} \in \mathbb{R}^n : \hat{\mathbf{x}}_c = \arg \min_{\mathbf{x}_a = \check{\mathbf{x}}_a} \|\mathbf{x} - \hat{\mathbf{x}}\|^2_{\mathbf{\Sigma}^{-1}} \right\} \quad (29)$$

Notice that obviously $\widehat{\mathbf{X}}_a = \mathbf{X}$ for $a = \emptyset$ (no active constraints). Conversely, for $a \in \mathcal{A} \backslash \emptyset$, the determination of $\widehat{\mathbf{X}}_a$ is not obvious but can be faced, at the price of a high computational burden, with the mpQP solvers available in the literature [9], [10]. Assume that the sets $\widehat{\mathbf{X}}_a$ are given, the pdf $f_c(\cdot)$ of the constrained estimate can be formally expressed as follows

$$\begin{aligned} f_c(\boldsymbol{\xi}) &= f(\boldsymbol{\xi})\mathcal{I}_{\mathbf{X}} + \sum_{a \in \mathcal{A}_v} \delta\left(\boldsymbol{\xi} - \check{\mathbf{x}}_a\right) \int_{\widehat{\mathbf{X}}_a} f(\boldsymbol{\chi}) d\boldsymbol{\chi} \\ &+ \sum_{a \in \mathcal{A}_0} \delta\left(\boldsymbol{\xi}_a - \check{\mathbf{x}}_a\right) f_a(\boldsymbol{\xi}_f) \end{aligned} \quad (30)$$

where $\mathcal{A}_v$ is the set of active sets of cardinality $n_a = n$ (vertices of $\mathbf{X}$), $\mathcal{A}_0$ is the set of active sets of cardinality $0 < n_a < n$ (hyperfaces of dimension $1, \ldots, n-1$ of $\mathbf{X}$) and $f_a(\boldsymbol{\xi}_f)$ represent the pdf's, on such hyperfaces, of the free variables. The expression (30) is a generalization of (18) to $n \geq 1$ bounded parameters. Its evaluation requires:

1) the determination of the regions $\widehat{\mathbf{X}}_a$ via mpQP techniques;
2) the calculation of the n-dimensional integrals of the unconstrained Gaussian pdf $f(\cdot)$ over $\widehat{\mathbf{X}}_a$ for all $a \in \mathbf{A}_v$;
3) the determination of the functions $f_a(\cdot)$ for all $a \in \mathcal{A}_0$.

Notice that in particular the above task 3, which is not needed in the case $n = 1$, seems analytically untractable. To see the difficulty, let us consider the case of $n = 2$ bounded parameters. In this case, $f_c(\xi_1, \xi_2)$ consists of $3^2 = 9$ terms: a continuous term $f(\xi_1, \xi_2)$ with support the rectangle $\mathbf{X}$; 4 singular terms with support the edges of $\mathbf{X}$; 4 singular terms with support the vertices of $\mathbf{X}$. This is illustrated in fig. 2 wherein two possible unconstrained estimates $\hat{\mathbf{x}}_a$, $\hat{\mathbf{x}}_b$ and the ellipsoidal level curves of the LS cost function centered around them are displayed; it can be seen that in case a the constrained estimate $\hat{\mathbf{x}}_{\mathbf{ca}}$ is in the vertex $[x_{1,max}, x_{2,max}]'$ while in case b the constrained estimate $\hat{\mathbf{x}}_{\mathbf{cb}}$ is on the segment $x_2 = x_{2,max}$, $x_1 \in [x_{1,min}, x_{1,max}]$. However the weights of the above-mentioned components of $f_c(\cdot)$, which correspond to the probabilities, under the Gaussian pdf $f(\cdot)$, of nine regions of the parameter space, cannot be computed simply by evaluating the distribution function $F(\cdot)$ as in (18). For instance, the term associated to the vertex $\mathbf{v} = [x_{1,max}, x_{2,max}]'$

$$\delta(\xi_1 - x_{1,max})\ \delta(\xi_2 - x_{2,max}) \int_{\widehat{\mathbf{X}}_{\mathbf{v}}} f(\chi_1, \chi_2)\ d\chi_1\ d\chi_2$$

requires the determination of the set $\widehat{\mathbf{X}}_{\mathbf{v}}$ of unconstrained estimates $\hat{\mathbf{x}}$ for which the constrained estimate $\hat{\mathbf{x}}_c$ coincides with $\mathbf{v}$ and then the evaluation of the probability of the event $\hat{\mathbf{x}} \in \widehat{\mathbf{X}}_{\mathbf{v}}$ under the distribution $f(\cdot)$. Clearly, this probability can only be evaluated in a numerical way. The situation is even more complicated for the edges. For instance, the term associated to the edge

$e = \{\mathbf{x} : x_{1,min} \le x_1 \le x_{1,max}, x_2 = x_{2,max}\}$ requires the determination of the pdf $f_e(\xi_1)$ over the edge, which seems to be related to the unconstrained Gaussian distribution in an overly too complicated fashion. From the above arguments, it is clear that the computation of the statistics of the hard-constrained estimator $\hat{\mathbf{x}}_\mathbf{c}$ is prohibitive, if not impossible, for multiple bounded parameters. Computations become easier if we consider a sub-optimal constrained estimator operating as follows.

Sub-Optimal Constrained Estimator

$$\begin{aligned}
&\text{for } i = 1 \text{ to } n\\
&\hat{\mathbf{x}}_{i:n} = \left(\mathbf{H}'_{i:n}\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}_{i:n}\right)^{-1}\ \mathbf{H}'_{i:n}\mathbf{\Sigma}_\mathbf{w}^{-1}\left(\mathbf{z} - \mathbf{H}_{1:i-1}\hat{\mathbf{x}}_{1:i-1,s}\right)\\
&\hat{x}_{i,s} = \begin{cases} x_{i,min}, & \text{if } \hat{x}_i < x_{i,max} \\ \hat{x}_i, & \text{if } x_{i,min} \le \hat{x}_i \le x_{i,max} \\ x_{i,max}, & \text{if } \hat{x}_i > x_{i,max} \end{cases}
\end{aligned} \tag{31}$$

The above estimator provides the estimate vector $\mathbf{x}_s = [\hat{x}_{1,s}, \hat{x}_{2,s}, \ldots, \hat{x}_{n,s}]'$ sequentially as follows: it first determines the estimate $\hat{x}_{1,s}$ of x_1 based on $x_1 \in [x_{1,min}, x_{1,max}]$, then it determines the estimate $\hat{x}_{2,s}$ of x_2 based on $x_1 = \hat{x}_{1,s}$ and $x_2 \in [x_{2,min}, x_{2,max}]$ and so forth. At stage i, the estimator gets $\hat{x}_{i,s}$ based on $x_{1:i-1} = \hat{x}_{1:i-1,s}$ and $x_i \in [x_{i,min}, x_{i,max}]$. In general, $\hat{\mathbf{x}}_s \neq \hat{\mathbf{x}}_c$ and, therefore, $\hat{\mathbf{x}}_s$ is sub-optimal in the sense of (14). Further, $\hat{\mathbf{x}}_s$ clearly depends on the ordering of the entries in the parameter vector. Since (31) actually involves the sequential application of the hard-constrained estimator for a single bounded parameter, it is possible to use the analytical results of Theorem 1 in order to evaluate, in a conservative way, the MSE reduction that can be achieved exploiting bounds on multiple parameters. In fact, the MSE $\sigma^2_{i,s}$ of the estimators $\hat{x}_{i,s}$, provided by (31), can be computed by the following recursion.

1) Initialization

$$\mathbf{\Sigma}_s = \mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{1,s} & * & * & * \\ * & \ddots & * & * \\ * & * & \sigma^2_{i,s} & * \\ * & * & * & \mathbf{\Sigma}_{i+1:n,s} \end{bmatrix}$$

2) for $i = 1$ to n

$$\begin{aligned}
\alpha_i &= \frac{x_{i,max} - x_{i,min}}{2\ \sigma_{i,s}}\\
\beta_i &= \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}}\\
\mathbf{L}_i &= \left(\mathbf{H}'_{i+1:n}\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H}_{i+1:n}\right)^{-1}\mathbf{H}'_{i+1:n}\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{h}_i\\
\mathbf{\Sigma}_{i+1:n,s} &= \mathbf{\Sigma}_{i+1:n,s} - \mathbf{L}_i\mathbf{L}'_i\ \gamma(\alpha_i,\beta_i)\ \sigma^2_{i,s}\\
\sigma^2_{i,s} &= \sigma^2_{i,s}\ (1 - \gamma(\alpha_i,\beta_i))
\end{aligned} \tag{32}$$

Please notice that overwriting has been used in (32) and $*$ denote don't care blocks at the i-th iteration. At the end of the recursion, the quantity

$$\gamma_{n,s}(\beta) = \frac{\sigma_n^2 - \sigma^2_{n,s}}{\sigma_n^2}$$

provides, for the parameter x_n and for a given location of the true parameter specified by the vector $\beta = [\beta_1, \ldots, \beta_n]'$, a conservative estimate (lower bound) for the relative MSE reduction achievable by the use of constraints. Such an estimate can be obtained for all the parameters x_i via (32) by a suitable re-ordering of the components of $\mathbf{x}$ which moves x_i in the bottom position of the re-ordered vector.

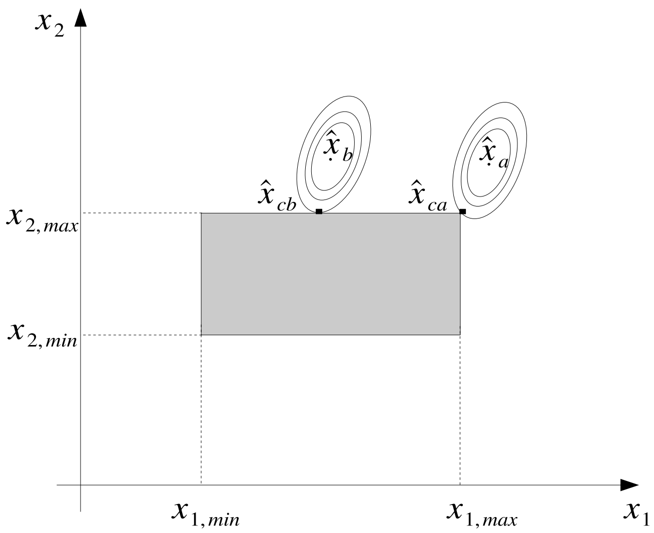

Fig. 2. Hard-constrained solution for the case of two bounded parameters

IV. Soft-constrained estimation

An alternative way of incorporating a-priori information in parameter estimation is to introduce *soft constraints*, i.e. a penalization in the cost functional of the deviation from some prior parameter value. This can be interpreted, in a Bayesian framework, as the adoption of a (normal) Gaussian prior distribution on the parameters. In fact, assuming that $\mathbf{x} \sim \mathcal{N}\left(\overline{\mathbf{x}}, \overline{\mathbf{\Sigma}}\right)$, the MAP criterion (10) becomes [2]

$$f(\mathbf{x}|\mathbf{z}) = \frac{1}{c}\exp\left[-\left(\|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{\Sigma}_\mathbf{w}^{-1}} + \|\mathbf{x} - \overline{\mathbf{x}}\|^2_{\overline{\mathbf{\Sigma}}^{-1}}\right)\right] \tag{33}$$

which involves an additional weighted penalization of $\mathbf{x} - \overline{\mathbf{x}}$. Accordingly the MAP estimate, which maximizes (33), is given by [2], [3]

$$\hat{\mathbf{x}}_{sc} = \mathbf{\Sigma}_{sc}\left(\mathbf{H}'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{z} + \overline{\mathbf{\Sigma}}^{-1}\overline{\mathbf{x}}\right) \tag{34}$$

where

$$\mathbf{\Sigma}_{sc} \triangleq \left(\mathbf{H}'\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{H} + \overline{\mathbf{\Sigma}}^{-1}\right)^{-1} = \left(\mathbf{\Sigma}^{-1} + \overline{\mathbf{\Sigma}}^{-1}\right)^{-1} \tag{35}$$

The estimator (34), that will be referred to hereafter as the *soft-constrained estimator*, has bias

$$E\left[\tilde{\mathbf{x}}_{sc}\right] = \mathbf{\Sigma}_{sc}\ \overline{\mathbf{\Sigma}}^{-1}\ \left(\overline{\mathbf{x}} - E[\mathbf{x}]\right) \tag{36}$$

and MSE

$$E\left[\tilde{\mathbf{x}}_{sc}\tilde{\mathbf{x}}_{sc}'\right] = \boldsymbol{\Sigma}_{sc}\,\overline{\boldsymbol{\Sigma}}^{-1}\,E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\right]\,\overline{\boldsymbol{\Sigma}}^{-1}\,\boldsymbol{\Sigma}_{sc} + \boldsymbol{\Sigma}_{sc}\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Sigma}_{sc} \quad (37)$$

If indeed the assumption $\mathbf{x} \sim \mathcal{N}\left(\overline{\mathbf{x}}, \overline{\boldsymbol{\Sigma}}\right)$ holds true, the soft-constrained estimator (34) is unbiased and its MSE in (37) turns out to be equal to $\boldsymbol{\Sigma}_{sc}$. In this case, (34) is clearly optimal also in the MMSE sense. Conversely, if the available a-priori information is the hard constraint $\mathbf{x} \in \mathbf{X}$, $\overline{\mathbf{x}}$ and $\overline{\boldsymbol{\Sigma}}$ can be regarded as design parameters of the soft-constrained estimator (34) to be chosen so as to possibly minimize, in some sense, the MSE in (37). In particular, the following two settings can be considered.

Stochastic Setting - The parameter $\mathbf{x}$ is regarded as a random variable uniformly distributed over the set $\mathbf{X}$ defined in (8). In this case, it is natural to choose $\overline{\mathbf{x}}$ as the center of $\mathbf{X}$ i.e.

$$\overline{\mathbf{x}} = E[\mathbf{x}] = \mathbf{x}_m \stackrel{\triangle}{=} \frac{1}{2}\,(\mathbf{x}_{min} + \mathbf{x}_{max}) \quad (38)$$

where $\mathbf{x}_{min}$ and $\mathbf{x}_{max}$ are the vectors of entries $x_{i,min}$ and, respectively, $x_{i,max}$. Moreover,

$$E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\right] = diag\left\{\frac{(x_{i,max}-x_{i,min})^2}{12}\right\}_{i=1,\ldots,n} \quad (39)$$

and the MSE in (37) becomes a function of $\overline{\boldsymbol{\Sigma}}$ only. For instance, one can minimize the MSE's trace i.e.

$$\min_{\overline{\boldsymbol{\Sigma}}} E\left[\tilde{\mathbf{x}}_{sc}'\tilde{\mathbf{x}}_{sc}\right] = \min_{\overline{\boldsymbol{\Sigma}}} tr\left\{\left(\boldsymbol{\Sigma}^{-1}+\overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\left[\overline{\boldsymbol{\Sigma}}^{-1}E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\right]\overline{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}\right]\left(\boldsymbol{\Sigma}^{-1}+\overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\right\} \quad (40)$$

Deterministic Setting - The parameter $\mathbf{x}$ is regarded as a deterministic variable in $\mathbf{X}$. Also in this case, $\overline{\mathbf{x}}$ is naturally chosen as the center of $\mathbf{X}$. Further, $E\left[(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\right] = (\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'$ and the MSE in (37) depends, therefore, on the unknown true parameter $\mathbf{x}$, besides $\overline{\boldsymbol{\Sigma}}$. To avoid the dependence on $\mathbf{x}$, one can perform the *min-max* optimization of the MSE's trace

$$\min_{\overline{\boldsymbol{\Sigma}}}\max_{\mathbf{x}\in\mathbf{X}} E\left[\tilde{\mathbf{x}}_{sc}'\tilde{\mathbf{x}}_{sc}\right] = \min_{\overline{\boldsymbol{\Sigma}}}\max_{\mathbf{x}\in\mathbf{X}} tr\left\{\left(\boldsymbol{\Sigma}^{-1}+\overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\left[\overline{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\overline{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}^{-1}\right]\left(\boldsymbol{\Sigma}^{-1}+\overline{\boldsymbol{\Sigma}}^{-1}\right)^{-1}\right\} \quad (41)$$

It can be noticed that in both of the above settings, $\overline{\boldsymbol{\Sigma}}$ becomes a design parameter which can be chosen in order to obtain a good performance for the estimator (34). A possible design approach is to take $\overline{\boldsymbol{\Sigma}} = \mathbf{V}\;diag\{\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_n\}\;\mathbf{V}'$, where $\mathbf{V}$ is the eigenvector matrix of $\boldsymbol{\Sigma}$ i.e. $\boldsymbol{\Sigma} = \mathbf{V}\;diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}\;\mathbf{V}'$, and to minimize the trace of the MSE via (40) (MSE averaged over $\mathbf{x} \sim \mathcal{U}(\mathbf{X})$) or via (41) (worst case MSE w.r.t. $\mathbf{x} \in \mathbf{X}$). The optimal solution turns out to be [11]

$$\overline{\lambda}_i = \frac{(x_{i,max}-x_{i,min})^2}{\kappa^2} \quad (42)$$

where

$$\kappa = \begin{cases} \sqrt{12}, & \text{for the stochastic setting} \\ 2, & \text{for the deterministic setting} \end{cases}. \quad (43)$$

The resulting soft-constrained estimator (34) coincides with the well known *ridge* estimator [11]. An alternative design approach, which does not impose any structure on $\overline{\boldsymbol{\Sigma}}$, has been proposed in [12]. The approach in [12] is based on the deterministic setting and consists of minimizing $tr(\overline{\boldsymbol{\Sigma}})$ subject to the constraint that the MSE (37) of the soft-constrained estimator does not exceed, for any possible $\mathbf{x} \in \mathbf{X}$, the MSE $\boldsymbol{\Sigma}$ of the unconstrained estimator.

Like in the previous sections on the unconstrained and hard-constrained estimation, let us now consider the case in which a-priori information is available on the parameter x_1 only. In the soft-constrained setting, this amounts to assuming that

$$\overline{\mathbf{x}} = \begin{bmatrix} x_{1,m} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \dfrac{x_{1,min}+x_{1,max}}{2} \\ \mathbf{0} \end{bmatrix}, \qquad \overline{\boldsymbol{\Sigma}}^{-1} = \begin{bmatrix} \dfrac{1}{\overline{\sigma}_1^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (44)$$

The following result suggests how to choose $\overline{\sigma}_1$ in an optimal way.

Theorem 2 - *The optimal solution of (40) (for the stochastic setting) or (41) (for the deterministic setting) under (15) and (44) is given by*

$$\overline{\sigma}_1 = \frac{x_{1,max}-x_{1,min}}{\kappa} \quad (45)$$

where κ is defined in (43).

Proof - *See the Appendix.*

Remark - The estimators obtained in the deterministic and stochastic setting have not equivalent properties. The former guarantees $E\left[\tilde{\mathbf{x}}_{sc}'\tilde{\mathbf{x}}_{sc}\right] \leq E\left[\tilde{\mathbf{x}}'\tilde{\mathbf{x}}\right] = tr\;\Sigma$ for any $\mathbf{x} \in \mathbf{X}$ while for the latter the above inequality holds in the *mean*, over the uniform distribution of $\mathbf{x}$, but not necessarily for any value of $\mathbf{x}$ in $\mathbf{X}$. As a consequence, the stochastic setting is characterized by a lower covariance (uncertainty) than the deterministic setting; this suggests that the deterministic MAP$_s$ estimator is overly conservative.

V. Hard-constrained vs. soft-constrained: numerical example

In this section, the aim is to compare *hard-constrained* and *soft-constrained* estimation on a simple *line fitting* case-study. The problem is to fit position measurements of a target obtained at different time instants via a line (affine function of time). The measured position $z(i)$ at sample time i is modeled by

$$z(i) \;=\; x_1 \;+\; iT\, x_2 \;+\; w(i) \tag{46}$$

where: $i \in \{0, 1, \ldots, N-1\}$; $N > 2$ is the number of measurements; $T > 0$ is the sampling interval; the initial position x_1 and velocity x_2 are the two parameters to be estimated; $w(i) \sim \mathcal{N}(0, \sigma_w^2)$ is the measurement noise. This is a special, 2-dimensional, case of the linear regression (1) with

$$\mathbf{z} = \begin{bmatrix} z(0) \\ z(1) \\ \vdots \\ z(N-1) \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & T \\ \vdots & \vdots \\ 1 & (N-1)\,T \end{bmatrix} \tag{47}$$

Specializing (3) to this case, one gets

$$\begin{aligned} \boldsymbol{\Sigma} &\triangleq \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \\ &= \sigma_w^2 \begin{bmatrix} \dfrac{2(2N-1)}{N(N+1)} & -\dfrac{6}{TN(N+1)} \\ -\dfrac{6}{TN(N+1)} & \dfrac{12}{T^2 N(N^2-1)} \end{bmatrix} \end{aligned} \tag{48}$$

which shows how the MSE of the unconstrained estimator decreases by increasing the size N of the data sample and the sampling rate $1/T$ or by decreasing the variance σ_w^2 of the measurement noise.

According to the available a-priori information and the way in which it is processed, three cases will be distinguished:

- information on a single parameter;
- information on both parameters (optimal approach);
- information on both parameters (sub-optimal approach).

For ease of reference, these cases will be referred to in the sequel as **SP (Single Parameter)**, **MP (Multiple Parameters)** and **SMP (Sequential Multiple Parameters)**. Remind that in the sub-optimal approach the information on both parameters is processed in a sequential (sub-optimal) way operating as follows: to compute the estimate $\hat{x}_2$ of x_2, the information on x_1 is used to estimate x_1 and then this estimate is used, along with the information on x_2, to compute $\hat{x}_2$; a reverse order is adopted to obtain the estimate $\hat{x}_1$ of x_1. It is worth pointing out that the sub-optimal approach can be applied in the same way to both hard-constrained and soft-constrained estimation, but turns out to be advantageous only for hard-constrained estimation where it avoids the solution of a QP problem. Whenever the information is available on a single parameter, the equations (24)-(27) and (37) with (44) provide the MSE for the hard- and, respectively, soft-constrained estimators. To compare the two estimators in this case it is useful, like in (24)-(27), to express the relationship (37) in terms of α. In fact, this allows to evaluate the performance of the constrained estimators versus the ratio between the constraint width and the standard deviation of the unconstrained estimator and, thus, quantify independently of the values of N, T, σ_w the effectiveness of constraints for MSE reduction. From (19) and (45), it follows that a fair comparison *"soft vs. hard"* can be performed by choosing $\overline{\sigma}_1 = 2\alpha\, \sigma_1/\kappa$ for the soft-constrained estimator.

Let us first consider the SP case. Fig. 3 shows, for both the constrained parameter x_1 and the free one x_2 as well as for both choices of κ (deterministic or stochastic setting), the MSE[1], averaged on β, of the unconstrained, soft-constrained (MAP$_s$) and hard-constrained (MAP$_h$) estimators for varying α. From these plots it can be seen that for low α, wherein the benefits of the constrained estimation are more evident, soft constraints yield better performance than hard constraints whereas the converse holds for high α. Fig. 3 also displays the MSE, for the constrained variable x_1, obtained by using a-priori information only (i.e. ignoring measurements); this amounts to taking the center $x_{1,m}$ as estimate of x_1 with associated normalized MSE $\alpha^2/3$. Notice that the estimator based on prior information only, gives a lower MSE than MAP$_h$ for $\alpha < 1.35$. Conversely, it gives a MSE lower than the deterministic setting of MAP$_s$ for $\alpha < 1$ and never lower than the stochastic setting of MAP$_s$. Also notice that, the stochastic MAP$_s$ estimator always gives a better performance on average than its deterministic counterpart. Fig. 4 displays the MSE behaviour of the constrained parameter x_1 versus β (which characterizes the position of the true parameter in the interval $[x_{1,min}, x_{1,max}]$), for different values of α. It can be noticed that the concavity of the MSE curve for the hard-constrained estimator changes with α, as previously remarked after Corollary 1. It can also be seen that, for $\alpha = 5$, the MAP$_s$ estimator designed under the stochastic setting provides a higher MSE than the unconstrained estimator whenever the true parameter is close to the border of the constraint($\beta \approx 0$ or $\beta \approx 1$), while under the deterministic setting the MAP$_s$ performance is never worse than in the unconstrained case (even if on average the MSE in the stochastic setting will be lower than in the deterministic setting). Summing up, the deterministic MAP$_s$ estimator is better when the true value is close to the extreme points, while the stochastic MAP$_s$ estimator is better on average. Since the latter represents the majority of cases, whereas a worst-case design (such as the deterministic one) will be worse on average but slightly better in some cases of low probability, it can be concluded that the stochastic MAP$_s$ estimator seems the most valuable.

Let us now examine the MP case. In this case, there is no analytical (exact) formula to evaluate the MSE and resort has been made to Monte Carlo simulations wherein both the noise realization and the true parameter location have been randomly varied. As in the SP case, the MSE is reported as a function of α, defined as the ratio between the constraint semi-width and the standard deviation of the unconstrained estimate for both parameters, i.e.

$$\alpha = \frac{x_{1,max} - x_{1,min}}{2\sigma_1} = \frac{x_{2,max} - x_{2,min}}{2\sigma_2}$$

For fixed α, the matrix $\overline{\boldsymbol{\Sigma}}$ of the MAP$_s$ estimator has been taken with the same eigenvectors of $\boldsymbol{\Sigma}$ and with eigenvalues

[1]Different measures, such as posterior entropy, could certainly be used but the concern in this paper is on MSE, for which we have been able to derive analytical formulas and an approximate evaluation algorithm in presence of hard bounds on the parameters.

optimally chosen as in (42)-(43). The simulation results are shown in Table I which reports, for different values of α, the MSE, averaged over β, of the MAP$_s$ (deterministic setting) and MAP$_h$ estimators in the SP, MP and SMP cases. Notice that the MSE has been normalized w.r.t. the MSE of the unconstrained estimator and that the SP case is relative to x_1 constrained (the results relative to x_2 constrained can be obtained by interchanging the columns x_1 and x_2 in the table). It is worth pointing out that the same MSE values within less than 1% of accuracy have been obtained with Monte Carlo simulations and with the analytical formulas given in the previous section whenever available (i.e. except for the MAP$_h$ estimator in the MP case); this confirms the validity of the theoretical results of Theorem 1. Examination of the table reveals that, in the case of both bounded parameters, hard-constrained estimation yields better performance starting from lower values of α. This difference w.r.t. the case of a single bounded parameter is due to the fact that in the 2-dimensional space the special choice of $\overline{\mathbf{\Sigma}}$, aligned with the eigenvectors of $\mathbf{\Sigma}$, is not able to effectively capture the a-priori information provided by hard constraints. A different design procedure of the parameters $\overline{\mathbf{x}}$ and $\overline{\mathbf{\Sigma}}$ of the soft-constrained estimator, much more effective for multiple parameters and general polytopic constraints, can be found in [12]. Finally, it can be noticed that the sub-optimal approach (SMP) gives a performance which is much better than the optimal single parameter approach (SP) and very close to the optimal two-parameter approach (MP). This means not only that the sub-optimal hard-constrained estimator (31) provides a good trade-off among performance and computational cost but also that the sequential MSE analysis (32) gives a very good approximation of the MSE obtained with the optimal hard-constrained estimator (14).

VI. Conclusions

The paper has investigated two alternative ways of exploiting a-priori knowledge on the estimated parameters in the estimation process. The *hard-constrained* estimation approach, wherein the a-priori knowledge is specified in terms of a membership set to which the unknown parameters certainly belong, requires constrained optimization and is, therefore, computationally more expensive than *soft-constrained estimation*, wherein the knowledge is expressed in terms of a Gaussian prior distribution with known mean and covariance of the unknown parameters. Exact analytical formulas for the bias and MSE reduction of the hard-constrained estimator have been derived for a linear regression model and interval constraints on a single parameter. These formulas express the performance improvement in terms of the key variable α, defined as the ratio between the width of the constraint semi-interval and the standard deviation of the unconstrained estimator. It has also been shown how these formulas can be employed to evaluate in an approximate way the performance improvement in the general case in which bounds are available on all parameters. Results on the choice of the prior covariance of the soft-constrained estimator for optimal MSE performance have also been given. A case-study concerning a line fitting estimation problem has confirmed the validity of the presented theoretical results and has also provided an interesting comparison between the hard-constrained and soft-constrained approaches. From this comparison, the following considerations can be drawn.

- The exploitation of a-priori information, when available, is always beneficial, the more beneficial the lower is α i.e. for tighter constraints, shorter data samples and higher noise variance.
- For low α, the soft-constrained estimator is preferable while, for α greater than a suitable threshold value, the hard-constrained estimator yields better performance.
- The above mentioned threshold decreases by increasing the number of bounded parameters.
- Even for multiple bounded parameters, there seems to be an appropriate choice of the prior covariance of the soft-constrained estimator [12] for which its performance is comparable to that of the hard-constrained estimator.

A perspective for future work is the generalization of the CRLB to inequality-constrained estimation. The fact that the inequality-constrained estimation problem is converted, with a given probability, into one of several possible equality-constrained problems can probably be exploited in order to derive a sort of CRLB using the results in [1].

Appendix

Proof of Theorem 1

Part (i) - If $\hat{x}_1 \in [x_{1,min}, x_{1,max}]$, the unconstrained estimate (2) satisfies the constraints and, therefore, provides the solution of (14)-(15). Otherwise, the solution of (14)-(15) has an active constraint, either $\hat{x}_{1,c} = x_{1,min}$ or $\hat{x}_{1,c} = x_{1,max}$ according to whether $\hat{x}_1 < x_{1,min}$ or $\hat{x}_1 > x_{1,max}$. Hence $\hat{x}_{1,c}$ is given by (16). Moreover, since $\mathbf{x}_2$ is unconstrained, $\hat{\mathbf{x}}_{2,c}$ is nothing but the unconstrained estimate of $\mathbf{x}_2$ for known $x_1 = \hat{x}_{1,c}$ and is, therefore, given by (17).

Part (ii) - From (16), the distribution function $F_c(\cdot)$ of the constrained estimate $\hat{x}_{1,c}$ turns out to be

$$F_c(\xi_1) = \begin{cases} 0, & \text{if } \xi_1 < x_{1,min} \\ F(\xi_1), & \text{if } x_{1,min} \leq \xi_1 < x_{1,max} \\ 1, & \text{if } \xi_1 \geq x_{1,max} \end{cases} \tag{49}$$

where $F(\cdot)$ is the Gaussian distribution function of the unconstrained estimate $\hat{x}_1$. Hence, the pdf $f_c(\cdot)$ of $\hat{x}_{1,c}$ turns out to be given by (18) where $f(\cdot)$ is the Gaussian pdf of $\hat{x}_1$.

Part (iii) - To compute the bias of the estimate $\hat{x}_{1,c}$, let us consider the definition

$$E\left[\hat{x}_{1,c}\right] = \int_{-\infty}^{\infty} \xi_1 \ f_c(\xi_1) \ d\xi_1 \tag{50}$$

Exploiting (18) and the relationships

$$\int_{-\infty}^{\infty} \xi_1 \ f(\xi_1) \ d\xi_1 = E\left[\hat{x}_1\right] = x_1 \qquad \text{(unbiasedness of } \hat{x}_1) \tag{51}$$

$$F(y) = \int_{-\infty}^{y} f(\xi_1) d\xi_1 \tag{52}$$

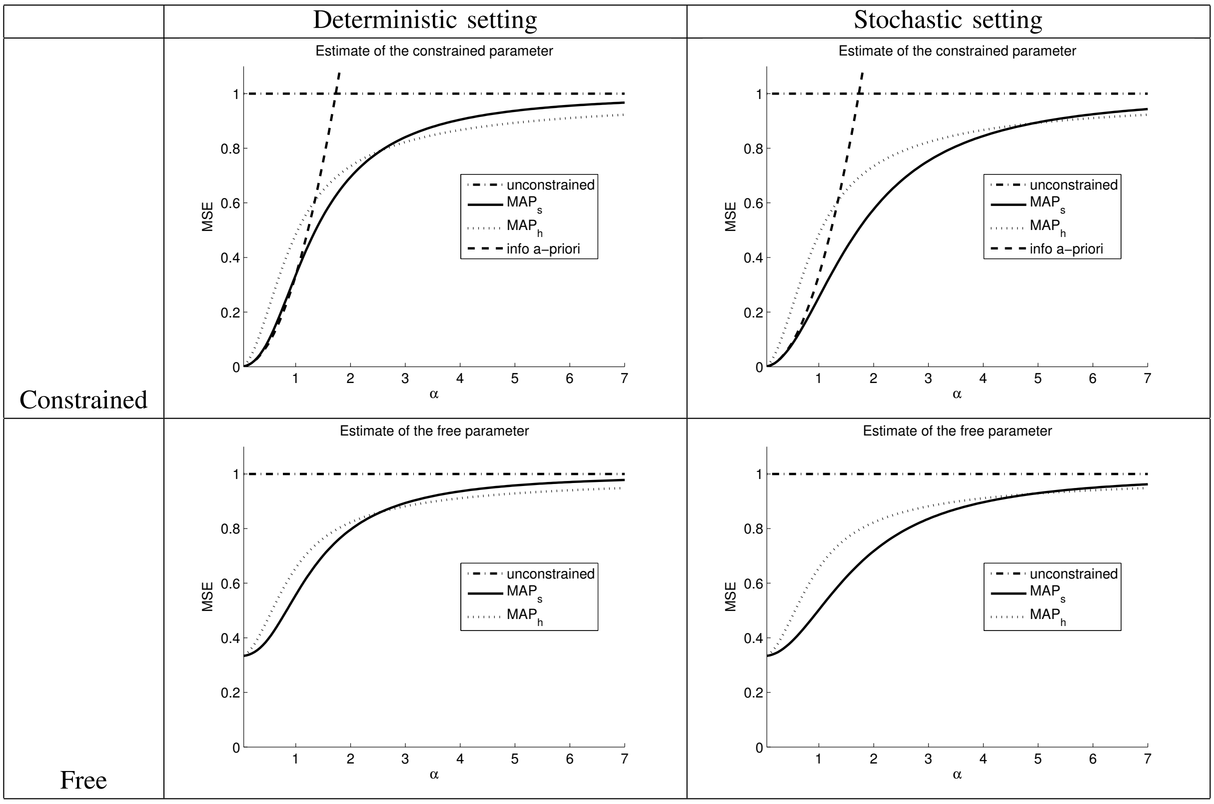

Fig. 3. MSE (averaged over β) vs. α for the unconstrained (dash-dotted line), hard-constrained (dotted line) and soft-constrained (solid line) estimators. The dashed line in the two upper figures indicates the MSE $\alpha^2/3$ of the estimate based on prior information only.

TABLE I

NORMALIZED % MSE (MSE = 100% FOR UNCONSTRAINED ESTIMATION) FOR THE CASES SP, MP AND SMP. THE MAP$_s$ ESTIMATOR IN THE SMP CASE IS NOT CONSIDERED.

	SP			MP			SMP		
	MSE	x_1	x_2	**MSE**	x_1	x_2	**MSE**	x_1	x_2
$\alpha = 0.5$	MAP$_s$	10	40	MAP$_s$	10	22	MAP$_s$	–	–
	MAP$_h$	21	47	MAP$_h$	17	17	MAP$_h$	18	18
	MSE	x_1	x_2	**MSE**	x_1	x_2	**MSE**	x_1	x_2
$\alpha = 1$	MAP$_s$	34	56	MAP$_s$	34	48	MAP$_s$	–	–
	MAP$_h$	48	65	MAP$_h$	37	37	MAP$_h$	38	38
	MSE	x_1	x_2	**MSE**	x_1	x_2	**MSE**	x_1	x_2
$\alpha = 1.5$	MAP$_s$	56	71	MAP$_s$	55	66	MAP$_s$	–	–
	MAP$_h$	64	76	MAP$_h$	51	51	MAP$_h$	52	52

from (50) we get

$$E\left[\tilde{x}_{1,c}\right] = \int_{-\infty}^{x_{1,min}} (x_{1,min} - \xi_1)\ f(\xi_1)\ d\xi_1 + \int_{x_{1,max}}^{\infty} (x_{1,max} - \xi_1)\ f(\xi_1)\ d\xi_1 \tag{53}$$

Notice that the bias (53) clearly depends on the true parameter value x_1 via $f(\xi_1)$. The first term of (53) provides a positive bias while the second provides a negative bias. Exploiting the Gaussianity of $f(\cdot)$ i.e.

$$f(\xi_1) = \frac{1}{\sqrt{2\pi}\ \sigma_1} \exp\left(-\ \frac{(\xi_1 - x_1)^2}{2\ \sigma_1^2}\right) \tag{54}$$

and evaluating the integrals in (53) we get, after lengthy calculations which are not reported due to lack of space, the expression provided by (20) and (22)-(23). Next step is to compute the bias of $\hat{\mathbf{x}}_{2,c}$. Substitution of (2) into (17) yields

$$\hat{\mathbf{x}}_{2,c} = \mathbf{x}_2 - \mathbf{L}\tilde{x}_{1,c} + \mathbf{\Sigma}_{2|1}\mathbf{H}_2'\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{w}$$

where $\mathbf{L}$ is defined in (7). Consequently, $\tilde{\mathbf{x}}_{2,c}$ is conditionally Gaussian to $\hat{x}_{1,c}$ with mean $-\mathbf{L}\tilde{x}_{1,c}$ and covariance $\mathbf{\Sigma}_{2|1}$, i.e.

$$\tilde{\mathbf{x}}_{2,c} \sim \mathcal{N}\left(-\mathbf{L}\tilde{x}_{1,c}, \mathbf{\Sigma}_{2|1}\right). \tag{55}$$

Hence

$$E\left[\tilde{\mathbf{x}}_{2,c}\right] = E\left[\ E\left[\ \tilde{\mathbf{x}}_{2,c}|\hat{x}_{1,c}\ \right]\ \right] = -\ \mathbf{L}\ E\left[\tilde{x}_{1,c}\right]$$

which, by (20), yields (21).

Part (iv) - Let us now evaluate the MSE of $\hat{x}_{1,c}$ defined as

$$\sigma_{1,c}^2 \triangleq E\left[\tilde{x}_{1,c}^2\right] = \int_{-\infty}^{\infty} (\xi_1 - x_1)^2\ f_c(\xi_1)\ d\xi_1.$$

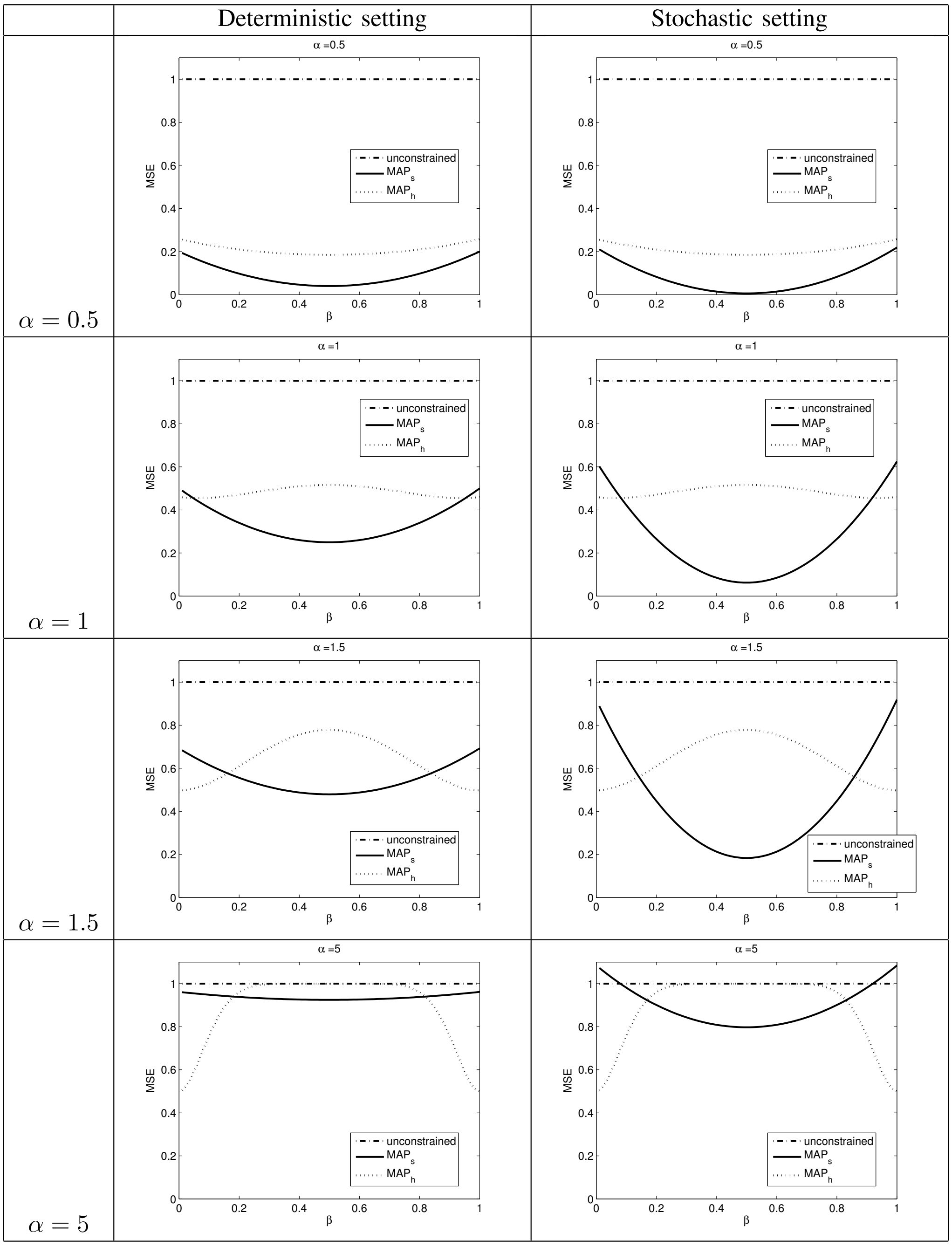

Fig. 4. MSE vs. β for different values of α

Exploiting (18), (52) and

$$\int_{-\infty}^{\infty} (\xi_1 - x_1)^2 f(\xi_1)\, d\xi_1 \;=\; \sigma_1^2$$

we get

$$\sigma_{1,c}^2 \;=\; \sigma_1^2 - \varepsilon$$

where

$$\begin{aligned} \varepsilon \;=\; & \int_{-\infty}^{x_{1,min}} \left[(x_1 - \xi_1)^2 - (x_1 - x_{1,min})^2 \right] \; f(\xi_1)\; d\xi_1 \\ + \; & \int_{x_{1,max}}^{\infty} \left[(\xi_1 - x_1)^2 - (x_{1,max} - x_1)^2 \right] \; f(\xi_1)\; d\xi_1 \end{aligned} \tag{56}$$

Notice that $\varepsilon > 0$ is the absolute MSE reduction; in particular, the two terms in (56) represent the reduction due to the a-priori information provided by the constraints $x_1 \geq x_{1,min}$ and, respectively, $x_2 \leq x_{2,max}$. Again, exploiting (54) and evaluating the integrals in (56) with the use of the *erf* function, after tedious calculations omitted due to space considerations, we get for $E[\tilde{x}_{1,c}^2]$ the expression in (24) and (26)-(27).
The final step of the proof is to compute the MSE of $\hat{\mathbf{x}}_{2,c}$.

From (55), using (6) and (24),

$$\begin{aligned} E\left[\tilde{\mathbf{x}}_{2,c}\tilde{\mathbf{x}}'_{2,c}\right] &= E\left[\ E\left[\ \tilde{\mathbf{x}}_{2,c}\tilde{\mathbf{x}}'_{2,c}\ \mid\ \hat{x}_{1,c}\ \right]\ \right] \\ &= E\left[\ \mathbf{L}\mathbf{L}'\tilde{x}^2_{1,c} + \mathbf{\Sigma}_{2|1}\ \right] \\ &= \mathbf{\Sigma}_2 - \mathbf{L}\mathbf{L}'\sigma_1^2 + \mathbf{L}\mathbf{L}'E\left[\tilde{x}^2_{1,c}\right] \\ &= \mathbf{\Sigma}_2\ -\ \mathbf{L}\mathbf{L}'\ \gamma(\alpha,\beta)\ \sigma_1^2 \end{aligned}$$

which is just (25).

Proof of Corollary 1

From (56), it follows that

$$\begin{aligned} &\gamma(\alpha,\beta) = \frac{\varepsilon}{\sigma_1^2} \\ &= \frac{1}{\sigma_1^2}\left\{\int_{-\infty}^{x_{1,min}} \left[(\xi_1 - x_1)^2 - (x_{1,min} - x_1)^2\right]\ f(\xi_1)\ d\xi_1 \right. \\ &\left. + \int_{x_{1,max}}^{\infty} \left[(\xi_1 - x_1)^2 - (x_{1,max} - x_1)^2\right]\ f(\xi_1)\ d\xi_1 \right\} \end{aligned} \tag{57}$$

Differentiating $\gamma(\alpha,\beta)$ w.r.t. x_1 one gets

$$\begin{aligned} \frac{\partial\gamma(\alpha,\beta)}{\partial x_1} &= \int_{-\infty}^{x_{1,min}} \left[-2(\xi_1 - x_1) + 2(x_{1,min} - x_1) \right. \\ &+ \left. \frac{(\xi_1 - x_1)^3}{\sigma_1^2} - (x_{1,min} - x_1)\frac{(\xi_1 - x_1)}{\sigma_1^2}\right] \frac{f(\xi_1)}{\sigma_1^2}\ d\xi_1 \\ &+ \int_{x_{1,max}}^{\infty} \left[-2(\xi_1 - x_1) + 2(x_{1,max} - x_1) \right. \\ &+ \left. \frac{(\xi_1 - x_1)^3}{\sigma_1^2} - (x_{1,max} - x_1)\frac{(\xi_1 - x_1)}{\sigma_1^2}\right] \frac{f(\xi_1)}{\sigma_1^2}\ d\xi_1 \end{aligned} \tag{58}$$

Exploiting the relationship $\dfrac{\partial f(\xi_1)}{\partial \xi_1} = -\dfrac{(\xi_1 - x_1)}{\sigma_1^2} f(\xi_1)$ due to the gaussianity of $f(\cdot)$, after straightforward calculations, (58) can be simplified to

$$\begin{aligned} \sigma_1^2\ \frac{\partial\gamma(\alpha,\beta)}{\partial x_1} &= 2\ (x_{1,max} - x_1)\int_{x_{1,max}}^{\infty} f(\xi_1)\ d\xi_1 \\ &- 2\ (x_1 - x_{1,min})\int_{-\infty}^{x_{1,min}} f(\xi_1)\ d\xi_1 \end{aligned} \tag{59}$$

Differentiating (59) w.r.t. x_1

$$\begin{aligned} \sigma_1^2\ \frac{\partial^2\gamma(\alpha,\beta)}{\partial x_1^2} &= -\ 2\int_{-\infty}^{x_{1,min}} f(\xi_1)\ d\xi_1 \\ &+\ 2\ (x_1 - x_{1,min})\ f(x_{1,min}) \\ &-\ 2\int_{x_{1,max}}^{\infty} f(\xi_1)\ d\xi_1 \\ &+\ 2\ (x_{1,max} - x_1)\ f(x_{1,max}) \end{aligned} \tag{60}$$

Finally, exploiting the definition of the error function, from (59)-(60) and taking into account that $dx_1/d\beta = 2\alpha\sigma_1$, one gets

$$\frac{\partial\gamma(\alpha,\beta)}{\partial\beta} = \Gamma_1(\alpha\beta) - \Gamma_1(\alpha(1-\beta)) \tag{61}$$

$$\frac{\partial^2\gamma(\alpha,\beta)}{\partial\beta^2} = \Gamma_2(\alpha\beta) + \Gamma_2(\alpha(1-\beta)) \tag{62}$$

where

$$\begin{aligned} \Gamma_1(y) &= -\frac{4\alpha y}{\sigma_1}\ \left[1 - \text{erf}\left(\sqrt{2}\ y\right)\right] \\ \Gamma_2(y) &= 4\alpha^2\ \left[\text{erf}\left(\sqrt{2}\ y\right) - 1 + \frac{4y}{\sqrt{2\pi}}\ \exp\left(-2y^2\right)\right] \end{aligned} \tag{63}$$

Studying the nulls of the first derivative and the corresponding signs of the second derivative, one proves the existence of $\alpha_0 > 0$ for which the results (i) and (ii) hold. ■

Proof of Theorem 2

The matrix $\mathbf{\Sigma}^{-1}$ is partitioned as follows:

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} a & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \tag{64}$$

where a is scalar. Replacing (64) in (37) and taking $\overline{\mathbf{\Sigma}}^{-1}$ as in (44), one gets:

$$\begin{aligned} \text{MSE} \triangleq E\left[\tilde{\mathbf{x}}_{sc}\tilde{\mathbf{x}}'_{sc}\right] &= \begin{bmatrix} a + \frac{1}{\overline{\sigma}_1^2} & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix}^{-1} \\ & \begin{bmatrix} a + \frac{\chi^2}{\overline{\sigma}_1^4} & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \\ & \begin{bmatrix} -\mathbf{b}'\mathbf{C}^{-1}\mathbf{b}a + \frac{1}{\overline{\sigma}_1^2} & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix}^{-1} \end{aligned} \tag{65}$$

where

$$\chi = \frac{x_{1,max} - x_{1,min}}{\kappa}$$

In order to carry out the matrix product in (65), one can exploit the block matrix inversion

$$\begin{aligned} & \begin{bmatrix} a + \frac{1}{\overline{\sigma}_1^2} & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} S_{\mathbf{C}}^{-1} & -S_{\mathbf{C}}^{-1}\mathbf{b}'\mathbf{C}^{-1} \\ -S_{\mathbf{C}}^{-1}\mathbf{C}^{-1}\mathbf{b} & \mathbf{C}^{-1} + S_{\mathbf{C}}^{-1}\mathbf{C}^{-1}\mathbf{b}\mathbf{b}'\mathbf{C}^{-1} \end{bmatrix} \end{aligned} \tag{66}$$

where $S_{\mathbf{C}} = a + \frac{1}{\overline{\sigma}_1^2} - \mathbf{b}'\mathbf{C}^{-1}\mathbf{b}$ is the *Schur complement* of the matrix w.r.t. $\mathbf{C}$. By straightforward calculations it follows that the trace of the MSE (65) is:

$$\begin{aligned} tr\ \text{MSE} &= \frac{a - \mathbf{b}'\mathbf{C}^{-1}\mathbf{b} + \dfrac{\chi^2}{\overline{\sigma}_1^4}}{\left(a - \mathbf{b}'\mathbf{C}^{-1}\mathbf{b} + \dfrac{1}{\overline{\sigma}_1^2}\right)^2}\ \left(1 - \mathbf{b}'\mathbf{C}^{-2}\mathbf{b}\right) \\ &+\ tr\ \mathbf{C}^{-1} \end{aligned} \tag{67}$$

Finally, studying the derivative w.r.t. $\overline{\sigma}_1$, one can see that (67) has a minimum at $\overline{\sigma}_1 = \chi$ ■

ACKNOWLEDGMENT

The authors would like to remember their dear colleague and friend Giovanni Zappa, inspirator of this work and prematurely deceased on August 23, 2004.

REFERENCES

[1] T.L. Marzetta, "A simple derivation of the constrained multiple parameter Cramer-Rao bound", *IEEE Trans. on Signal Processing*, vol. 41, pp. 2247-2249, 1993.

[2] S.M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice Hall, 1993.

[3] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, *Estimation with applications to tracking and navigation*, John Wiley & Sons, 2001.

[4] D. Simon, T. Chia, "Kalman filtering with state equality constraints", *IEEE Trans. on Aerospace and Electronic Systems*, vol. 39, pp. 128-136, 2002.

[5] A. Farina, G. Golino, L. Ferranti, "Constrained tracking filters for A-SMGCS", *Proc. 6th International Conference on Information Fusion*, Cairns, Australia, 2003.

[6] J. Garcia Herrero, J.A. Besada Portas, J.R. Casar Corredera, "Use of map information for tracking targets on airport surface", *IEEE Trans. on Aerospace and Electronic Systems*, vol. 39, pp. 675-692, 2003.

[7] B. Ristic, S. Arulampalam, N. Gordon, *Beyond the Kalman filter: particle filters for tracking applications*, Artech House, 2004.

[8] S. Challa, S. Arulampalam, N. Bergman, "Target tracking incorporating flight envelope information", *Proc. 3rd International Conference on Information Fusion*, CD-ROM, ThC2-4, Paris, France, 2000.

[9] A. Bemporad, M. Morari, V. Dua, E.N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems", *Automatica*, vol. 38, pp. 3-20, 2002.

[10] P. Tondel, T.A. Johansen, A. Bemporad, "Further results on multiparametric quadratic programming", *Proc. 42nd IEEE Conf. on Decision and Control*, vol. 3, pp. 3173-3178, Maui, USA, 2003.

[11] A.E. Hoerl, R.W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems", *Technometrics*, vol. 12, pp. 55-67, 1970.

[12] A. Benavoli, L. Chisci, A. Farina, "Estimation of constrained parameters with guaranteed MSE improvement", submitted to *IEEE Trans. on Signal Processing*, July 2005. Available: *http://www.dsi.unifi.it/users/chisci/recent_publ.htm*.