Probabilistically-constrained estimation of random parameters with unknown distribution

Sergiy A. Vorobyov
Communication Systems Group
Darmstadt University of Technology
Darmstadt 64283, Germany
Email: svor@nt.tu-darmstadt,de

Yonina C. Eldar
Dept. of Electrical Engineering
Technion, Israel Inst. of Technology
Haifa 32000, Israel
Email: yonina@ee.technion.ac.il

Alex B. Gershman
Communication Systems Group
Darmstadt University of Technology
Darmstadt 64283, Germany
Email: gershman@nt.tu-darmstadt.de

Abstract— The problem of estimating random unknown signal parameters in a noisy linear model is considered. It is assumed that the covariance matrices of the unknown signal parameter and noise vectors are known and that the noise is Gaussian, while the distribution of the random signal parameter vector is unknown. Instead of the traditional minimum mean squared error (MMSE) approach, where the average is taken over both the random signal parameters and noise realizations, we propose a linear estimator that minimizes the MSE which is averaged over the noise only. To make our design pragmatic, the minimization is performed for signal parameter realizations whose probability is sufficiently large, while "discarding" low-probability realizations. It is shown that the obtained linear estimator can be viewed as a generalization of the classical Wiener filter.

I. INTRODUCTION

Estimating a vector of unknown random or deterministic parameters in a noisy linear model represents a classic problem of estimation theory that has found numerous applications in signal processing, communications, radar, and other fields. Mathematically, this problem amounts to estimating an $n \times 1$ unknown signal parameter vector x in the linear model

$$y = Hx + w \qquad (1)$$

where y and w are $m \times 1$ observation and random noise vectors, respectively, and H is an $m \times n$ known transformation matrix. A linear estimate of the vector x can be written as

$$\hat{x} = Gy \qquad (2)$$

where G is some $n \times m$ matrix.

Different assumptions about the noise vector w and the unknown signal parameter vector x can be made depending on particular application. For example, it can be assumed that w and x are both zero-mean random stationary Gaussian with known positive definite covariance matrices $C_w \triangleq E\{ww^H\}$ and $C_x \triangleq E\{xx^H\}$, where $(\cdot)^H$ and $E\{\cdot\}$ denote the Hermitian transpose and the statistical expectation, respectively. In this case, a linear estimator that minimizes the mean squared error (MSE) averaged over both the signal parameters and noise realizations coincides with the Wiener filter [1], [2]. Mathematically, the Wiener filter finds its estimate $\hat{x}$ of x as

$$\arg_G \min_{\hat{x}=Gy} E_{w,x}\{\|\hat{x} - x\|^2\} \qquad (3)$$

where $\| \cdot \|^2$ denotes the Euclidian norm of a vector.

In other applications, the unknown parameter vector x is assumed to be deterministic with a known uncertainty region. In this case, an approach that is based on minimizing the worst-case MSE is frequently used [3]-[5]. Specifically, if the norm of x is bounded by some known constant U, then this estimation problem can be written as [5]

$$\arg_G \min_{\hat{x}=Gy} \max_{\|x\| \leq U} E_w\{\|\hat{x} - x\|^2\}. \qquad (4)$$

This estimator is known to be *robust* because it minimizes the MSE for the worst choice of x.

Robust estimators can also be obtained in the stochastic signal case, when the unknown parameter vector is random and its covariance matrix is known. In such a case, we can minimize the MSE (averaged over the noise

realizations only), while taking into account only realizations of $\boldsymbol{x}$ that occur with a sufficiently large probability, and discarding realizations of $\boldsymbol{x}$ whose probability is low. That way, the *outage probability*[1] of the so-obtained estimator of $\boldsymbol{x}$ can be maintained at an acceptably low level.

The latter estimation problem has been recently addressed in [6] for the case of a random unknown signal parameter vector with known (Gaussian) distribution. In this paper, we adopt the main framework of [6], but address a more general case when the distribution of the random signal parameter vector is *completely unknown*.

Mathematically, our problem amounts to the design of a linear estimator that minimizes the MSE (averaged over the noise realizations only) with a certain selected probability under the assumption that the vector $\boldsymbol{x}$ is random with known covariance matrix $\boldsymbol{C_x}$ and unknown probability density function (pdf). Such an assumption can be motivated by the fact that for non-Gaussian distributions, the problem of estimating the covariance matrix is much simpler than that of estimating the whole pdf.

II. PROBLEM FORMULATION

Using (1) and (2), the MSE for any given $\boldsymbol{x}$ can be expressed as

$$\mathrm{E}_{\boldsymbol{w}}\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\} = \mathrm{Tr}\{\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^H\} + \boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} \qquad (5)$$

where $\boldsymbol{\Lambda} \triangleq (\boldsymbol{GH} - \boldsymbol{I})^H(\boldsymbol{GH} - \boldsymbol{I})$, and $\mathrm{Tr}\{\cdot\}$ and $\boldsymbol{I}$ stand for the trace operator and the identity matrix, respectively.

Our objective is to design a robust estimator by considering only the realizations of $\boldsymbol{x}$ which occur with a sufficiently high probability, while discarding the realizations whose probability is low. Mathematically, this is equivalent to replacing the expectation over $\boldsymbol{x}$ in (3) by the probability operator, and constraining this probability to be greater than or equal to a certain (preliminary selected) threshold value $p \in (0, 1)$. Thus, using (5), our estimator is the solution to

$$\min_{\boldsymbol{G},t}\{t : \mathrm{Pr}_{\boldsymbol{x}}\{\mathrm{Tr}\{\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^H\} + \boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} \geq t\} \leq 1 - p\}$$
$$(6)$$

or, equivalently,

$$\min_{\boldsymbol{G},t}\{t : \mathrm{Pr}_{\boldsymbol{x}}\{\mathrm{Tr}\{\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^H\} + \boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} < t\} \geq p\} \quad (7)$$

[1]Note that our definition of (estimator) outage is somewhat different from that used in communication theory. In the latter context, outage is defined as the situation when the channel is so poor that no scheme can communicate reliably at a certain target data rate [9].

where $\mathrm{Pr}_{\boldsymbol{x}}\{\cdot\}$ denotes the probability operator. This problem belongs to the class of *probability-constrained stochastic programming* problems [7], [8].

As the pdf of $\boldsymbol{x}$ is unknown, our approach is to solve the problem (7) for the *worst-case* pdf. To this end we define

$$\mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\} \triangleq \inf\{\mathrm{Pr}_{\boldsymbol{x}}\{\boldsymbol{x} \in \mathcal{A}\} \mid \boldsymbol{C_x}\} \qquad (8)$$

where the set $\mathcal{A}$ is given by

$$\mathcal{A} \triangleq \{\boldsymbol{x} \in \mathbb{C}^n \mid \boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} - t' < 0\} \qquad (9)$$

with

$$t' \triangleq t - \mathrm{Tr}\{\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^H\} \qquad (10)$$

and the infimum is over all probability distributions of $\boldsymbol{x} \in \mathbb{C}^n$ with covariance $\boldsymbol{C_x}$. Equation (8) defines the worst case corresponding to the least favorable distribution of $\boldsymbol{x}$. We then modify the (7) as

$$\min_{\boldsymbol{G},t}\{t : \mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\} \geq p\}. \qquad (11)$$

To develop a solution to the stochastic programming problem (11), we first convert the constraint to a simpler deterministic form. This derivation is given in Section 3. A solution to the problem (11) is then given in Section 4.

III. DETERMINISTIC FORM OF THE PROBABILITY OPERATOR

In Theorem 1 below we show that $\mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\}$ can be bounded above and below by solutions of a semidefinite program (SDP). We then show that these bounds are tight, and solve the corresponding SDPs.

Theorem 1: The probability operator in (11) can be upper- and lower-bounded by the solutions of the following SDP problems:
Upper bound SDP:

$$\min_{\boldsymbol{Z}, \lambda} \quad 1 - \lambda \;\geq\; \mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\}$$
$$\text{subject to } \mathrm{Tr}\{\boldsymbol{\Lambda}\boldsymbol{Z}\} - t'\lambda \geq 0$$
$$\boldsymbol{0} \preceq \boldsymbol{Z} \preceq \boldsymbol{C_x}, \quad 0 \leq \lambda \leq 1 \qquad (12)$$

where $\boldsymbol{Z}$ is an $n \times n$ Hermitian matrix and λ is a scalar.

Lower bound SDP:

$$\max_{\boldsymbol{P}, \tau} \quad 1 - \mathrm{Tr}\{\boldsymbol{C_x}\boldsymbol{P}\} \;\leq\; \mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\}$$
$$\text{subject to } \boldsymbol{P} \succeq \tau\boldsymbol{\Lambda}, \quad t'\tau - 1 \geq 0$$
$$\tau \geq 0 \qquad (13)$$

where $\boldsymbol{P}$ is an $n \times n$ Hermitian matrix and τ is a scalar.

Proof: We start with the proof of the upper bound property (12). We will prove that if $\boldsymbol{Z}$ and λ satisfy

the constraints in the upper bound SDP (12), then there exists a random vector $\boldsymbol{x}$ with

$$E\{\boldsymbol{xx}^H\} = \boldsymbol{C_x}, \ \ \mathrm{Pr}_{\boldsymbol{x}}\{\boldsymbol{x} \in \mathcal{A}\} \leq 1 - \lambda. \quad (14)$$

Hence, it will be shown that $1 - \lambda$ is an upper bound for $\mathrm{Pr}_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C_x}\}$.

To show this, the following lemma will be needed.

Lemma 1 [10]: If the inequality

$$\mathrm{Tr}\{\boldsymbol{\Lambda Z}\} - d \geq 0 \quad (15)$$

is satisfied for some $n \times n$ Hermitian matrix $\boldsymbol{Z} \succeq \boldsymbol{0}$, then there exist $n \times 1$ vectors $\boldsymbol{v}_i$ and scalars $\alpha_i \geq 0$, $i = 1, \ldots, K$, with $K \leq 2n$, such that

$$\boldsymbol{v}_i^H \boldsymbol{\Lambda v}_i - d \geq 0, \ \ i = 1, \ldots, K \quad (16)$$

$$\sum_{i=1}^{K} \alpha_i = 1, \ \sum_{i=1}^{K} \alpha_i \boldsymbol{v}_i = 0, \ \sum_{i=1}^{K} \alpha_i \boldsymbol{v}_i \boldsymbol{v}_i^H \preceq \boldsymbol{Z}. \ (17)$$

An important fact following from this lemma is that, if $\boldsymbol{s}$ is an $n \times 1$ random zero-mean vector with covariance matrix $\boldsymbol{Z} = E\{\boldsymbol{ss}^H\} \succeq \boldsymbol{0}$ satisfying

$$E\{\boldsymbol{s}^H \boldsymbol{\Lambda s} - d\} = \mathrm{Tr}\{\boldsymbol{\Lambda Z}\} - d \geq 0 \quad (18)$$

then there exists a discrete random zero-mean vector $\boldsymbol{x}$ with $K \leq 2n$ possible values, that satisfies

$$\boldsymbol{x}^H \boldsymbol{\Lambda x} - d \geq 0, \ \ \ E\{\boldsymbol{xx}^H\} \preceq E\{\boldsymbol{ss}^H\} = \boldsymbol{Z}. \quad (19)$$

Proof: The proof of a rather general form of this lemma can be found in [10]. In what follows, the proof of our (more specific) formulation of this lemma is given for a zero-mean $\boldsymbol{x}$, positive semi-definite $\boldsymbol{\Lambda}$, and $d > 0$.

Let us use the fact that any positive semi-definite Hermitian matrix can be decomposed as a sum of rank-one matrices as

$$\boldsymbol{Z} = \sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{w}_i^H. \quad (20)$$

We will refer to such a decomposition as a *dyadic decomposition* taking into account that all the matrices $\boldsymbol{w}_i \boldsymbol{w}_i^H \ (i = 1, \ldots, n)$ are dyadic. Inserting (20) into (15),

$$0 \leq \mathrm{Tr}\{\boldsymbol{\Lambda Z}\} - d = \sum_{i=1}^{n} \boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i - d. \quad (21)$$

The terms $\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i$ in (21) are all non-negative because $\boldsymbol{\Lambda} \succeq \boldsymbol{0}$. Let $r \ (0 < r \leq n)$ denote the number of non-zero terms. Then (21) can be rewritten as

$$\sum_{i=1}^{r} \boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i \geq d. \quad (22)$$

Let β_i and β_{r+i} $(i = 1, \ldots, r)$ be the positive and negative roots of the equations

$$\beta_i^2 \boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i = d, \ \ \ i = 1, \ldots, r. \quad (23)$$

These roots are given by

$$\beta_i = \sqrt{\frac{d}{\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i}}, \ \ \ \beta_{r+i} = -\sqrt{\frac{d}{\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i}}. \quad (24)$$

Let $K = 2r$ and

$$\boldsymbol{v}_i = \beta_i \boldsymbol{w}_i, \ \ \ \ \alpha_i = \frac{\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i}{2 \sum_{i=1}^{r} \boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i},$$

$$\boldsymbol{v}_{i+r} = \beta_{i+r} \boldsymbol{w}_i, \ \ \ \alpha_{i+r} = \alpha_i. \quad (25)$$

It can be readily seen that the conditions (16) are satisfied for the vectors $\boldsymbol{v}_i$. As $\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i > 0$, we have that $\alpha_i > 0$ and $\alpha_{i+r} > 0$. Moreover,

$$\sum_{i=1}^{2r} \alpha_i = \sum_{i=1}^{r} 2\alpha_i = \sum_{i=1}^{r} \frac{\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i}{\sum_{k=1}^{r} \boldsymbol{w}_k^H \boldsymbol{\Lambda w}_k} = 1 \quad (26)$$

and

$$\sum_{i=1}^{2r} \alpha_i \boldsymbol{v}_i = \sum_{i=1}^{r} \alpha_i (\beta_i + \beta_{r+i}) \boldsymbol{w}_i = 0 \quad (27)$$

because $\beta_i + \beta_{r+i} = 0$. Finally, using (22) and (24),

$$\begin{aligned} \sum_{i=1}^{2r} \alpha_i \boldsymbol{v}_i \boldsymbol{v}_i^H &= \sum_{i=1}^{r} \alpha_i (\beta_i^2 + \beta_{i+r}^2) \boldsymbol{w}_i \boldsymbol{w}_i^H \\ &= \sum_{i=1}^{r} \frac{2d\alpha_i}{\boldsymbol{w}_i^H \boldsymbol{\Lambda w}_i} \boldsymbol{w}_i \boldsymbol{w}_i^H \\ &= \frac{d}{\sum_{k=1}^{r} \boldsymbol{w}_k^H \boldsymbol{\Lambda w}_k} \sum_{i=1}^{r} \boldsymbol{w}_i \boldsymbol{w}_i^H \\ &\preceq \sum_{i=1}^{r} \boldsymbol{w}_i \boldsymbol{w}_i^H \preceq \boldsymbol{Z}. \end{aligned} \quad (28)$$

Thus, the conditions (17) are satisfied for the choice of $\boldsymbol{v}_i$ and α_i $(i = 1, \ldots, K)$ of (25). This completes the proof. $\square$

Returning to the proof of the property (12), let us consider the case $0 < \lambda < 1$. Note that the case $\lambda = 0$ is trivial because $\mathrm{Pr}_{\boldsymbol{x}}\{\boldsymbol{x} \in \mathcal{A}\} \leq 1$ always holds. The case $\lambda = 1$ will be considered at the end of the proof.

Using the result of Lemma 1 together with the first constraint in (12), the constraint $\boldsymbol{Z} \succeq \boldsymbol{0}$, and the condition $0 < \lambda < 1$, we can define a random vector $\boldsymbol{x}_1$ that satisfies the following inequalities

$$\boldsymbol{x}_1^H \boldsymbol{\Lambda x}_1 - t' \geq 0, \ \ \ E\{\boldsymbol{x}_1 \boldsymbol{x}_1^H\} \preceq \boldsymbol{Z}/\lambda. \quad (29)$$

Then, using the second constraint in (12) and (29), we have

$$\lambda E\{\boldsymbol{x}_1\boldsymbol{x}_1^H\} \preceq \boldsymbol{Z} \preceq \boldsymbol{C}_{\boldsymbol{x}}. \tag{30}$$

According to (30), the matrix

$$\boldsymbol{C}_0 \triangleq \frac{1}{1-\lambda}\left(\boldsymbol{C}_{\boldsymbol{x}} - \lambda E\{\boldsymbol{x}_1\boldsymbol{x}_1^H\}\right) \tag{31}$$

is positive semi-definite. Using a dyadic decomposition of $\boldsymbol{C}_0$ in the form

$$\boldsymbol{C}_0 = \sum_{i=1}^{r} \boldsymbol{c}_i\boldsymbol{c}_i^H \tag{32}$$

with $r \leq n$, we can build a discrete random vector $\boldsymbol{x}_0$ with the covariance matrix $E\{\boldsymbol{x}_0\boldsymbol{x}_0^H\} = \boldsymbol{C}_0$ as follows. If $r = 0$, we choose $\boldsymbol{x}_0 = \boldsymbol{0}$. If $r > 0$, then

$$\boldsymbol{x}_0 = \begin{cases} \sqrt{r}\boldsymbol{c}_i, & \text{with probability } 1/(2r) \\ -\sqrt{r}\boldsymbol{c}_i, & \text{with probability } 1/(2r) \end{cases} \tag{33}$$

is chosen.

Summarizing, we have defined two independent random vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ that satisfy (29) and

$$(1-\lambda)E\{\boldsymbol{x}_0\boldsymbol{x}_0^H\} + \lambda E\{\boldsymbol{x}_1\boldsymbol{x}_1^H\} = \boldsymbol{C}_{\boldsymbol{x}}. \tag{34}$$

Considering the random vector with a mixture distribution

$$\boldsymbol{x} = \begin{cases} \boldsymbol{x}_0, & \text{with probability } 1-\lambda \\ \boldsymbol{x}_1, & \text{with probability } \lambda \end{cases} \tag{35}$$

and using (34), we can verify that $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{C}_{\boldsymbol{x}}$. Moreover, since $\boldsymbol{x}_1 \notin \mathcal{A}$, then $\Pr_{\boldsymbol{x}}\{\boldsymbol{x} \in \mathcal{A}\} \leq 1 - \lambda$, and hence $1 - \lambda$ is an upper bound for $\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\}$.

If $\lambda = 1$, we can define $\tilde{\boldsymbol{Z}} \triangleq (1-\varepsilon)\boldsymbol{Z}$, and $\tilde{\lambda} \triangleq (1-\varepsilon)\lambda = 1-\varepsilon$, where $0 < \varepsilon < 1$. Similarly to the case $0 < \lambda < 1$, we can construct a random variable with $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{C}_{\boldsymbol{x}}$ and $\Pr_{\boldsymbol{x}}\{\boldsymbol{x} \in \mathcal{A}\} \leq 1 - \tilde{\lambda} = \varepsilon$ for any $\varepsilon, 0 < \varepsilon < 1$. Therefore, if $\lambda = 1$ then $\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\} = 0$. This completes the proof of the upper bound property.

We now continue with the proof of the lower bound property (13). We will show that, if the constraints in (13) are satisfied, then

$$1 - \text{Tr}\{\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{P}\} \leq \Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\}. \tag{36}$$

The first and second constraints in (13) can be combined as

$$\begin{bmatrix} \boldsymbol{P} - \tau\boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0}^T & t'\tau - 1 \end{bmatrix} \succeq \boldsymbol{0}. \tag{37}$$

Positive semi-definite property of the matrix in (37) implies that for all $\boldsymbol{x}$

$$\boldsymbol{x}^H\boldsymbol{P}\boldsymbol{x} \geq 1 + \tau(\boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} - t'). \tag{38}$$

Similarly, because of the first and third constraints in (13) the matrix $\boldsymbol{P}$ is positive semi-definite. It implies that for all $\boldsymbol{x}$

$$\boldsymbol{x}^H\boldsymbol{P}\boldsymbol{x} \geq 0. \tag{39}$$

We observe that if $\boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} - t' < 0$ (that is, $\boldsymbol{x} \in \mathcal{A}$), then the right-hand side of (38) is smaller than one. Similarly, if $\boldsymbol{x}^H\boldsymbol{\Lambda}\boldsymbol{x} - t' \geq 0$ (that is, $\boldsymbol{x} \notin \mathcal{A}$), then the right-hand side in (38) is larger than one. Using this observation and the constraint (39), we have

$$\boldsymbol{x}^H\boldsymbol{P}\boldsymbol{x} \geq 1 - \mathbf{1}_{\mathcal{A}}(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{x} \notin \mathcal{A} \\ 0, & \boldsymbol{x} \in \mathcal{A} \end{cases} \tag{40}$$

where $\mathbf{1}_{\mathcal{A}}(\boldsymbol{x})$ denotes the indicator function of the set $\mathcal{A}$, i.e., $\mathbf{1}_{\mathcal{A}}(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in \mathcal{A}$ and $\mathbf{1}_{\mathcal{A}}(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \notin \mathcal{A}$. Taking the expectation of the right- and the left-hand sides of (40) and using the facts that $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{C}_{\boldsymbol{x}}$ and that $E\{\mathbf{1}_{\mathcal{A}}(\boldsymbol{x})\} = \Pr\{\boldsymbol{x} \in \mathcal{A}\}$, we obtain

$$\begin{aligned} E\{\boldsymbol{x}^H\boldsymbol{P}\boldsymbol{x}\} &= \text{Tr}\{\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{P}\} \\ &\geq 1 - E\{\mathbf{1}_{\mathcal{A}}(\boldsymbol{x})\} \\ &= 1 - \Pr\{\boldsymbol{x} \in \mathcal{A}\}. \end{aligned} \tag{41}$$

Therefore, $1 - \text{Tr}\{\boldsymbol{C}_{\boldsymbol{x}}\boldsymbol{P}\}$ is a lower bound for $\Pr\{\boldsymbol{x} \in \mathcal{A}\}$, and (36) holds. This completes the proof of the lower bound property and Theorem 1. $\square$

We now show that the bounds in Theorem 1 are tight. It can be easily seen that the objective function in (12) is minimized if $\boldsymbol{Z} = \boldsymbol{C}_{\boldsymbol{x}}$ and $\lambda = \text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\}/t'$ provided that $\text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\} < t'$, which holds in our problem. Then, the upper bound on the probability operator in the constraint of (11) is given by the optimal solution of (12) and can be written as

$$\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\} \leq 1 - \text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\}/t'. \tag{42}$$

Furthermore, provided that $t' > 0$, the optimal solution of (13) is given by $\tau = 1/t'$ and $\boldsymbol{P} = \boldsymbol{\Lambda}/t'$. Inserting this optimal value of $\boldsymbol{P}$ into the objective function of (13), we can express the lower bound for $\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\}$ in the constraint of (11) as

$$\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\} \geq 1 - \text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\}/t'. \tag{43}$$

Combining (42) and (43), we conclude that if $\text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\} < t'$, then the upper and lower bounds coincide and, therefore,

$$\Pr_{\boldsymbol{x}}\{\mathcal{A}, \boldsymbol{C}_{\boldsymbol{x}}\} = 1 - \text{Tr}\{\boldsymbol{\Lambda}\boldsymbol{C}_{\boldsymbol{x}}\}/t'. \tag{44}$$

IV. Probabilistically-constrained estimator

Inserting (44) into (11), we obtain the following equivalent deterministic optimization problem:

$$\min_{\boldsymbol{G},t}\{\ t:\ \mathrm{Tr}\{\boldsymbol{\Lambda C_x}\} < \delta(t - \mathrm{Tr}\{\boldsymbol{GC_wG}^H\})\} \quad (45)$$

where $\delta \triangleq 1 - p$. Note that according to the constraint in (45), the condition $\mathrm{Tr}\{\boldsymbol{\Lambda C_x}\} < t'$ is always satisfied. The problem (45) can be equivalently rewritten as

$$\min_{\boldsymbol{G}} \mathrm{Tr}\{\boldsymbol{GC_wG}^H + \delta^{-1}(\boldsymbol{GH}-\boldsymbol{I})\boldsymbol{C_x}(\boldsymbol{GH}-\boldsymbol{I})^H\}. \tag{46}$$

Since the objective function is strictly convex in $\boldsymbol{G}$, the optimal solution can be found be setting the derivative to 0 which results in the following probabilistically-constrained estimator of $\boldsymbol{x}$:

$$\hat{\boldsymbol{x}} = \boldsymbol{C_x H}^H \left(\delta \boldsymbol{C_w} + \boldsymbol{HC_x H}^H\right)^{-1} \boldsymbol{y}. \tag{47}$$

Interestingly, the linear estimator (47) can be viewed as a generalization of the Wiener filter. In particular, it coincides with the Wiener filter when $\delta \to 1$ ($p \to 0$), that is, when the constraint in (7) is satisfied independently of the realization of $\boldsymbol{x}$. More generally, it is equal to a Wiener filter matched to a weighted noise covariance $\delta \boldsymbol{C_w}$. Thus, our approach effectively modifies the noise covariance by a factor of δ.

V. Conclusions

A linear estimator has been proposed that, in contrast to the traditional MMSE approach, minimizes the MSE averaged over the noise realizations only. The minimization takes into account the signal realizations whose probability is sufficiently large, while discarding the low-probability realizations. It has been shown that the proposed linear estimator can be viewed as a generalization of the classical Wiener filter providing an improved robustness against unknown signal distribution and "bad" realizations.

References

[1] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York: Wiley, 1949.

[2] A. Kolmogorov, "Interpolation and extrapolation," *Bull. Acad. Sci., USSR, Ser. Math.*, vol. 5, pp. 314, 1941.

[3] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," *Problems Inform. Trans.*, vol. 16, pp. 120-133, 1980.

[4] V. L. Girko and N. Christopeit, "Minimax estimators for linear models with nonrandom disturbances," *Random Operators and Stochastic Equations*, vol. 3, no. 4, pp. 361-377, 1995.

[5] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168-181, Jan. 2005.

[6] S. A. Vorobyov, Y. C. Eldar, A. Nemirovski, and A. B. Gershman, "Probability-constrained approach to estimation of random Gaussian parameters," in *Proc. 1st IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Puerto Vallarta, Mexico, Dec. 2005, pp. 101-104.

[7] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Springer-Verlag, 1997.

[8] A. Prékopa, *Stochastic Programming*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.

[9] J. G. Proakis, *Digital Communications*. 4th edition, McGraw-Hill, Inc., 2001.

[10] L. Vandenberghe, S. Boyd, and K. Comanor, "Generalized Chebyshev bounds via semidefinite programming," *SIAM Review*, to appear in 2006.