

PROBABILITY-CONSTRAINED APPROACH TO ESTIMATION OF RANDOM GAUSSIAN PARAMETERS

Sergiy A. Vorobyov* Yonina C. Eldar[†] Arkadi Nemirovski[‡] Alex B. Gershman*

* Communication Systems Group, Darmstadt University of Technology, Merckstr. 25, 64283 Darmstadt, Germany

[†] Dept. of Electrical Engineering, Technion - Israel Institute of Technology 32000 Haifa, Israel

[‡] Minerva Optimization Center, Technion - Israel Institute of Technology 32000 Haifa, Israel

ABSTRACT

The problem of estimating a random signal vector \mathbf{x} observed through a linear transformation \mathbf{H} and corrupted by an additive noise is considered. A linear estimator that minimizes the mean squared error (MSE) with a certain selected probability is derived under the assumption that both the additive noise and random signal vectors are zero mean Gaussian with known covariance matrices. Our approach can be viewed as a robust generalization of the Wiener filter. It simplifies to the recently proposed robust minimax estimator in some special cases.

1. INTRODUCTION

The theory of parameter estimation in linear models has been studied extensively in the second half of the past century, following the classical works of Wiener and Kolmogorov [1], [2]. A fundamental problem addressed in these papers can be briefly described as that of estimating unknown parameter in the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (1)$$

where \mathbf{x} and \mathbf{w} is an $n \times 1$ zero-mean vector of unknown stochastic parameters with known covariance matrix $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{C}_x$, \mathbf{w} is an $m \times 1$ zero-mean random noise vector with known covariance matrix \mathbf{C}_w , \mathbf{y} is the $m \times 1$ observation vector, \mathbf{H} is an $m \times n$ known transformation matrix, $(\cdot)^H$ stands for the Hermitian transpose, and $E\{\cdot\}$ denotes the expectation operator. Both \mathbf{C}_x and \mathbf{C}_w are assumed to be positive definite matrices.

Let the parameter vector \mathbf{x} be estimated by a linear estimator

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{y} \quad (2)$$

where \mathbf{G} is some $n \times m$ matrix. The Wiener-Kolmogorov estimator finds the estimate of \mathbf{x} as

$$\arg_G \min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} E_{w,x} \{ \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \} \quad (3)$$

where the expectation is taken with respect to both the additive noise \mathbf{w} and the unknown vector \mathbf{x} , which are assumed to have Gaussian distribution, and $\|\cdot\|^2$ denotes the Euclidian norm of a vector. It is also assumed that the exact statistical information about \mathbf{w} and \mathbf{x} is available. Moreover, the basic assumption used

The work of Y. C. Eldar was supported in part by the EU 6th framework programme, via the NEWCOM network of excellence.

in the estimator (3) is the ergodic assumption meaning that the observation time is so long as to reveal the long-term ergodic properties of the processes \mathbf{w} and \mathbf{x} . However, these two assumptions may not be satisfied in practice.

If the statistical information about \mathbf{w} and \mathbf{x} deviates from the assumed, then the performance of the Wiener-Kolmogorov estimator can degrade substantially. The modifications of the Wiener-Kolmogorov estimator, which are robust against imprecise knowledge of the statistical information about \mathbf{w} and \mathbf{x} , have been developed in [3], [4]. Another approach for designing robust estimators is based on minimizing the worst-case MSE in the uncertainty region for the parameter \mathbf{x} , where \mathbf{x} is assumed to be deterministic unknown [5]-[7]. Specifically, if the norm of \mathbf{x} is bounded by some known constant U defining an uncertainty region, then the estimation problem can be written as [7]

$$\arg_G \min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\| \leq U} E_w \{ \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \}. \quad (4)$$

The estimator (4) is called *robust* because it minimizes the MSE for the worst choice of \mathbf{x} .

Similarly, the performance of the Wiener-Kolmogorov estimator can degrade substantially if the ergodic assumption is not satisfied. However, in practice, this assumption is, indeed, often violated. Especially, it might be the case with respect to the process \mathbf{x} . Motivated by this fact, an appealing idea would be to develop an estimator which relaxes the ergodic assumption by considering only the scenarios (realizations of \mathbf{x}) which occur with large probabilities, while discarding the scenarios with small probabilities. It is equivalent to replacing the expectation taken with respect to \mathbf{x} in (3) by the probability operator, and bounding such probability.

In the next section we give the mathematical formulation for a new estimator, which we call hereafter as a *probabilistically-constrained estimator*, develop the convex approximation for such estimator, and show some interesting connections to the worst-case estimator (4). The simulation results are presented in Section 3. Section 4 contains our concluding remarks.

2. PROBABILISTICALLY-CONSTRAINED ESTIMATOR

In this section, we develop a new approach to estimation of random parameters that is based on minimization of the MSE with a certain selected probability.

2.1. Problem formulation

Mathematically, we seek the solution to the following optimization problem

$$\begin{aligned} & \arg_{\mathbf{G}} \min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}, t'} t' \\ & \text{subject to } \Pr_{\mathbf{x}}\{E_w\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} \leq t'\} \geq p \end{aligned} \quad (5)$$

where p is a selected probability value, and $\Pr_{\mathbf{x}}\{\cdot\}$ denotes the probability operator whose form is assumed to be known and that applies only to \mathbf{x} . This problem belongs to a class of *chance- or probability-constrained stochastic programming* problems [8].

Using (1) and (2), the MSE is given by

$$\begin{aligned} E_w\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} &= \text{Tr}\{\mathbf{G}\mathbf{C}_w\mathbf{G}\} \\ &+ \mathbf{x}^H(\mathbf{G}\mathbf{H} - \mathbf{I})^H(\mathbf{G}\mathbf{H} - \mathbf{I})\mathbf{x} \end{aligned} \quad (6)$$

where $\text{Tr}\{\cdot\}$ and \mathbf{I} stand for the trace operator and the identity matrix, respectively. Introducing new notations $t \triangleq t' - \text{Tr}\{\mathbf{G}\mathbf{C}_w\mathbf{G}\}$, $\delta \triangleq 1 - p$, and $\mathbf{\Lambda} \triangleq (\mathbf{G}\mathbf{H} - \mathbf{I})^H(\mathbf{G}\mathbf{H} - \mathbf{I})$ (where $\mathbf{\Lambda}$ is a Hermitian positive semi-definite matrix), and using (6), problem (5) can be equivalently written as

$$\begin{aligned} & \arg_{\mathbf{G}, t} \min \{\text{Tr}\{\mathbf{G}\mathbf{C}_w\mathbf{G}\} + t\} \\ & \text{subject to } \Pr_{\mathbf{x}}\{\mathbf{x}^H\mathbf{\Lambda}\mathbf{x} - t > 0\} \leq \delta. \end{aligned} \quad (7)$$

The problem (7) is mathematically intractable in general. However, if \mathbf{x} is Gaussian distributed, then, using convex approximation of probability constraints [9], (7) can be approximated by a tractable convex problem and can be efficiently solved.

2.2. Convex upper bound on the probability constraint

Introducing a new random variable $\boldsymbol{\xi}$ with zero mean and identity covariance matrix, such that $\mathbf{x} = \mathbf{C}_x^{1/2}\boldsymbol{\xi}$, the constraint of (7) can be rewritten as:

$$\Pr_{\mathbf{x}}\{\mathbf{x}^H\mathbf{\Lambda}\mathbf{x} - t > 0\} = \Pr_{\boldsymbol{\xi}}\{\boldsymbol{\xi}^H\mathbf{C}_x^{1/2}\mathbf{\Lambda}\mathbf{C}_x^{1/2}\boldsymbol{\xi} - t > 0\} \leq \delta. \quad (8)$$

Let λ_i ($i = 1, \dots, n$) be the eigenvalues¹ of the matrix $\mathbf{C}_x^{1/2}\mathbf{\Lambda}\mathbf{C}_x^{1/2}$. Then, (8) can be equivalently written as

$$p(t, \boldsymbol{\lambda}) \triangleq \Pr_{\boldsymbol{\eta}}\left\{\sum_{i=1}^n \lambda_i \eta_i^2 - t > 0\right\} \leq \delta \quad (9)$$

where η_i , ($i = 1, \dots, n$) are independent zero-mean random variables with unit variance, $\boldsymbol{\eta} \triangleq [\eta_1, \dots, \eta_n]^T$, $\boldsymbol{\lambda} \triangleq [\lambda_1, \dots, \lambda_n]^T$, and $[\cdot]^T$ stands for the transpose. The following theorem suggests a convex approximation of the constraint (9).

Theorem 1: For $0 \leq \lambda_i < 1/2$ ($i = 1, \dots, n$), the inequality

$$\ln p(t, \boldsymbol{\lambda}) \leq \ln\left(E_{\boldsymbol{\eta}}\left\{\exp\left\{\sum_{i=1}^n \lambda_i \eta_i^2 - t\right\}\right\}\right) \quad (10)$$

$$= -\frac{1}{2} \sum_{i=1}^n \ln(1 - 2\lambda_i) - t \triangleq P(t, \boldsymbol{\lambda}) \quad (11)$$

¹Note that the eigenvalues λ_i ($i = 1, \dots, n$) are real nonnegative because $\mathbf{C}_x^{1/2}\mathbf{\Lambda}\mathbf{C}_x^{1/2}$ is Hermitian positive semi-definite.

holds, where $P(t, \boldsymbol{\lambda})$ is a logarithmically convex function² in variables t and $\boldsymbol{\lambda}$.

Proof: Let \mathcal{E} be the following set

$$\mathcal{E} = \left\{ \boldsymbol{\eta} \mid \sum_{i=1}^n \lambda_i \eta_i^2 - t > 0 \right\}$$

and let ψ be a *real nonnegative nondecreasing convex* function satisfying the following properties:

$$\psi(t, \boldsymbol{\lambda}) > \psi(0, \mathbf{0}) \geq 1. \quad (12)$$

Then, using (12) we can write

$$p(t, \boldsymbol{\lambda}) = \Pr_{\boldsymbol{\eta}}\{\boldsymbol{\eta} \in \mathcal{E}\} = E_{\boldsymbol{\eta}}\{\mathbf{1}_{\mathcal{E}}(\boldsymbol{\eta})\} \leq E_{\boldsymbol{\eta}}\{\psi(t, \boldsymbol{\lambda})\} \quad (13)$$

where $\mathbf{1}_{\mathcal{E}}(\boldsymbol{\eta})$ denotes the indicator function of the set \mathcal{E} , i.e., $\mathbf{1}_{\mathcal{E}}(\boldsymbol{\eta}) = 1$ if $\boldsymbol{\eta} \in \mathcal{E}$ and $\mathbf{1}_{\mathcal{E}}(\boldsymbol{\eta}) = 0$ if $\boldsymbol{\eta} \notin \mathcal{E}$.

The exponential function $\exp(t, \boldsymbol{\lambda}) = \exp\{\sum_{i=1}^n \lambda_i \eta_i^2 - t\}$ is nonnegative, nondecreasing, convex, and satisfies the property (12). Thus, substituting $\exp(t, \boldsymbol{\lambda})$ instead of $\psi(t, \boldsymbol{\lambda})$ in (13) and taking the logarithm (that is also nonnegative, nondecreasing, convex, and monotonic function), we obtain (10). Note that the inequality (10) is similar to the one used in the Chernoff bound [10].

Next, using the fact that η_i^2 ($i = 1, \dots, n$) are central χ^2 distributed, and using the characteristic function [10] of the central χ^2 distribution, we obtain

$$\begin{aligned} P(t, \boldsymbol{\lambda}) &= \ln\left(e^{-t} E_{\boldsymbol{\eta}}\left\{\exp\left\{\sum_{i=1}^n \lambda_i \eta_i^2\right\}\right\}\right) \\ &= \ln\left(\frac{e^{-t}}{\prod_{i=1}^n (1 - 2\lambda_i)^{1/2}}\right) = -\frac{1}{2} \sum_{i=1}^n \ln(1 - 2\lambda_i) - t. \end{aligned}$$

The latter expression is the same as (11), and it is convex in variables t and $\boldsymbol{\lambda}$ if $0 \leq \lambda_i < 1/2$ ($i = 1, \dots, n$). The convexity follows from the general fact that a weighted, with nonnegative weights, sum of logarithmically convex functions, e.g., that of $\exp(t, \boldsymbol{\lambda})$, is itself logarithmically convex. \square

Note that Theorem 1 provides a convex approximation for $\ln p(t, \boldsymbol{\lambda})$ only if $0 \leq \lambda_i < 1/2$ ($i = 1, \dots, n$). However, these conditions may not be satisfied in practice. Thus, the inequality (10) should be generalized for arbitrary real positive λ_i ($i = 1, \dots, n$) before it can be used to approximate the probability constraint in (7). Toward this end, let us introduce a new optimization variable $\gamma > 0$ which serves as a normalization coefficient such that $\lambda_i/\gamma < 1/2$ ($i = 1, \dots, n$). Then, it can be checked that the function

$$\Phi(\gamma, t, \boldsymbol{\lambda}) \triangleq \gamma P(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda}) = -\frac{\gamma}{2} \sum_{i=1}^n \ln(1 - 2\lambda_i/\gamma) - t$$

is convex in γ , t , and $\boldsymbol{\lambda}$. We can now formulate the following theorem on the constraint (9).

Theorem 2: If $\delta \in (0, 1)$, then the condition

$$\exists \gamma > 0 \quad \text{such that} \quad \Phi(\gamma, t, \boldsymbol{\lambda}) - \gamma \ln \delta \leq 0 \quad (14)$$

is a sufficient condition for the validity of the inequality (9).

²Note that a function $f(z)$ is logarithmically convex on the interval $[a, b]$ if $f(z) > 0$ and $\ln f(z)$ is convex on $[a, b]$.

Proof: Based on Theorem 1, the function $P(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda})$ is an upper bound on the function $\ln p(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda})$. Moreover, it is easy to see that $\ln p(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda}) = \ln p(t, \boldsymbol{\lambda})$. Hence, the function $P(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda})$ is also an upper bound on $\ln p(t, \boldsymbol{\lambda})$, that is, $\ln p(t, \boldsymbol{\lambda}) \leq P(\gamma^{-1}t, \gamma^{-1}\boldsymbol{\lambda})$ that must hold for some $\gamma > 0$. Thus, if there exists $\gamma > 0$ such that $\Phi(\gamma, t, \boldsymbol{\lambda}) \leq \gamma \ln \delta$, then it implies also that $\ln p(t, \boldsymbol{\lambda}) \leq \ln \delta$ or equivalently $p(t, \boldsymbol{\lambda}) \leq \delta$. \square

The condition (14) provides a ‘‘safe approximation’’ of the constraint (9). That is, if a given pair $(t, \boldsymbol{\lambda})$ with $\boldsymbol{\Lambda} \succeq \mathbf{0}$ can be extended by a properly chosen γ to a solution of (14), then the constraint (9) holds true. However, the condition (14) can not yet be applied to the problem (7), because it is not convex with respect to \mathbf{G} . Indeed, the eigendecomposition is a non-convex function in \mathbf{G} . This difficulty is resolved in the next subsection where the approximate probabilistically-constrained estimator is given.

2.3. Approximate probabilistically-constrained estimator

A convex in \mathbf{G} approximation of the probabilistically-constrained estimator (7) is given in the following theorem.

Theorem 3: The convex in variables $\gamma, t, \mathbf{G}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n$ and $\boldsymbol{\Omega}$ problem

$$\begin{aligned} & \arg_{\mathbf{G}} \min_{\gamma, t, \mathbf{G}, \boldsymbol{\mu}, \boldsymbol{\Omega}} \text{Tr}\{\mathbf{G}\mathbf{C}_w\mathbf{G}\} + t \\ & \text{subject to} \quad -t - \frac{\gamma}{2} \sum_{i=1}^n \ln \left(1 - 2\frac{\mu_i}{\gamma} \right) - \gamma \ln \delta \leq 0 \\ & \quad \mu_1 \geq \mu_2 \geq \dots \geq \mu_n \geq 0 \\ & \quad 2\mu_1 < \gamma \\ & \quad S_k \{\boldsymbol{\Lambda}\mathbf{C}_x\} \leq \sum_{i=1}^k \mu_i, \quad k = 1, \dots, n-1 \\ & \quad \text{Tr}\{\boldsymbol{\Omega}\mathbf{C}_x\} \leq \sum_{i=1}^n \mu_i \\ & \quad \boldsymbol{\Omega} \preceq \boldsymbol{\Lambda} \end{aligned} \quad (15)$$

is a safe approximation of the problem (7), where the last constraint is convex by means of Schur complement. Here $\boldsymbol{\mu} \triangleq [\mu_1, \dots, \mu_n]^T$, $\boldsymbol{\Lambda} = (\mathbf{G}\mathbf{H} - \mathbf{I})^H(\mathbf{G}\mathbf{H} - \mathbf{I})$ and $S_k\{\mathbf{A}\}$ denotes the sum of k largest eigenvalues of a Hermitian matrix \mathbf{A} .

Proof: It is easy to check that the objective function and the constraints in (15) are all convex. Moreover, the last constraint is satisfied with equality at the optimum and, thus,

$$\text{Tr}\{\boldsymbol{\Omega}\mathbf{C}_x\} = \text{Tr}\{\boldsymbol{\Lambda}\mathbf{C}_x\} = \sum_{i=1}^n \mu_i. \quad (16)$$

Indeed, we observe that both the true probability $\Pr_{\mathbf{x}}\{\mathbf{x}^H\boldsymbol{\Lambda}\mathbf{x} > t\}$ and the constraints in (15) are monotone in $\boldsymbol{\Lambda} \succeq \mathbf{0}$. The latter means that for some $\boldsymbol{\Omega}_1 \succeq \boldsymbol{\Lambda} \succeq \mathbf{0}$, both the true probability and the constraints in (15) will be greater or equal to similar quantities corresponding to $\boldsymbol{\Lambda}$. Thus, it follows that at the optimum $\boldsymbol{\Omega} = \boldsymbol{\Lambda}$, and the last two constraints (15) can be substituted by (16).

To prove that the problem (15) is a safe approximation of the problem (7), we need to prove that the constraints in (15) are equivalent to the constraint (14). The latter equivalence means that a triple $(\gamma, t, \boldsymbol{\lambda})$ with $\boldsymbol{\Lambda} \succeq \mathbf{0}$ is feasible for (14) if and only if it can

be extended by properly chosen $\boldsymbol{\mu}$ to a feasible solution to the system given by the constraints in (15), where the last two constraints are substituted by (16).

Indeed, if $(\gamma, t, \boldsymbol{\lambda})$ is feasible for (14), then setting $\mu_i = \lambda_i$, where the eigenvalues are arranged in the non-ascending order, we extend $(\gamma, t, \boldsymbol{\lambda})$ to a feasible solution of to the system given by the constraints in (15).

Next, we prove that if $(\gamma, t, \mathbf{G}, \boldsymbol{\mu}, \boldsymbol{\Omega})$ is feasible for the system given by the constraints in (15), then $(\gamma, t, \boldsymbol{\lambda})$ is feasible for (14). This proof is readily given by the *Majorization Principle*³ which states the following:

Given two $n \times 1$ real valued vectors \mathbf{e} and \mathbf{f} , a necessary and sufficient condition for \mathbf{f} to belong to the convex hull of all permutations of \mathbf{e} is $s_k(\mathbf{e}) \geq s_k(\mathbf{f})$, for $k = 1, \dots, n-1$, and $s_n(\mathbf{e}) = s_n(\mathbf{f})$, where $s_k(\mathbf{a})$, $1 \leq k \leq n$, stands for the sum of k largest entries in \mathbf{a} .

Now let $(\gamma, t, \mathbf{G}, \tilde{\boldsymbol{\mu}}, \boldsymbol{\Omega})$ be feasible for the system given by the constraints in (15), and let $\boldsymbol{\lambda}$ be the vector of eigenvalues of $\mathbf{C}_x^{1/2}\boldsymbol{\Lambda}\mathbf{C}_x^{1/2}$. Then, from the second and fourth constraints in (15) and the constraint (16) it follows that $s_k(\boldsymbol{\lambda}) \leq s_k(\tilde{\boldsymbol{\mu}})$, with equality for $k = n$. Using the Majorization Principle, we conclude that $\boldsymbol{\lambda}$ is a convex combination of permutations of $\tilde{\boldsymbol{\mu}}$. However, the first inequality in (15) is valid only when $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$ and, thus, it is valid when $\boldsymbol{\mu}$ is a permutation of $\tilde{\boldsymbol{\mu}}$. Since the left hand side of this inequality is convex in $\boldsymbol{\mu}$, we conclude that this inequality is valid if $\boldsymbol{\mu}$ is a convex combination of permutations of $\tilde{\boldsymbol{\mu}}$. In particular, this inequality is valid when $\boldsymbol{\mu} = \boldsymbol{\lambda}$. The latter means that $(\gamma, t, \boldsymbol{\lambda})$ is feasible for (14). \square

Since the problem (15) is convex, it can be efficiently solved using interior-point methods [12]. The complexity of solving the problem (15) is equivalent to that of semi-definite programming (SDP) problem complexity because of the last three constraints. However, the first constraint of (15) is logarithmically convex that makes the problem non-SDP.

For better understanding of the estimator (15), a special case which relates it to the minimax estimator of [7] is considered.

2.4. Special case

In this case, $\mathbf{x} = x$ is a scalar, $\mathbf{H} = \mathbf{h}$ is an $n \times 1$ vector, and the variance of x is denoted as σ_x^2 .

The estimator \hat{x} is given by $\hat{x} = \mathbf{g}^H \mathbf{y}$ for some $n \times 1$ vector \mathbf{g} . Then, the problem (7) can be written as

$$\begin{aligned} & \arg_{\mathbf{g}, t} \min \left\{ \mathbf{g}^H \mathbf{C}_w \mathbf{g} + t \right\} \\ & \text{subject to} \quad \Pr_{\mathbf{x}} \{ x^2 (\mathbf{g}^H \mathbf{h} - 1)^2 - t > 0 \} \leq \delta \end{aligned}$$

and the problem (15) can be rewritten as

$$\arg_{\mathbf{g}, \gamma} \min \left\{ \mathbf{g}^H \mathbf{C}_w \mathbf{g} + \gamma \ln \frac{1}{\delta} - \frac{\gamma}{2} \ln \left(1 - \frac{2\sigma_x^2 (\mathbf{g}^H \mathbf{h} - 1)^2}{\gamma} \right) \right\}. \quad (17)$$

Next, it can be shown that

$$\begin{aligned} & \min_{\gamma > 0} \left\{ \gamma \ln(1/\delta) - \frac{\gamma}{2} \ln \left(1 - \frac{2\sigma_x^2 (\mathbf{g}^H \mathbf{h} - 1)^2}{\gamma} \right) \right\} \\ & \leq (1 + 2 \ln(1/\delta) + 2\sqrt{\ln(1/\delta)}) \sigma_x^2 (\mathbf{g}^H \mathbf{h} - 1)^2. \end{aligned} \quad (18)$$

³For proof see [11, pp.147-149].

Let us denote $U^2 \triangleq (1 + 2 \ln(1/\delta) + 2\sqrt{\ln(1/\delta)})\sigma_x^2$. Then, based on the approximation (18), the problem (17) can be rewritten as

$$\arg_g \min_g \left\{ g^H C_w g + U^2 (g^H h - 1)^2 \right\}.$$

The solution of this problem

$$g = \frac{U^2}{1 + U^2 h^H C_w^{-1} h} C_w^{-1} h$$

is equivalent to the minimax MSE estimator developed in [7] for estimating the scalar x in the model $y = xh + w$ with $|x| \leq U$, where U is chosen to be equal to $(1 + 2 \ln(1/\delta) + 2\sqrt{\ln(1/\delta)})\sigma_x^2$.

3. SIMULATIONS

In our simulations, the proposed estimator (15) is compared with the Wiener-Kolmogorov estimator. We assume that the dimensions of the vectors y and x of the linear model (1) are $m = 5$ and $n = 3$, respectively. The covariance matrix C_x is equal to $\sigma_x^2 \mathbf{I}$, with the variance $\sigma_x^2 = 3$. The noise realizations are taken from zero-mean Gaussian distribution with the covariance matrix $C_w = \sigma_w^2 \mathbf{I}$, where σ_w^2 denotes the noise variance. The elements of H are drawn from a Gaussian distribution with zero mean and unit variance and remain fixed for each run. The probability p for the proposed estimator is equal to 0.97. Each point in simulations is obtained by averaging over 100 independent runs.

Note that the value of MSE depends on the choice of x . The best-case (lowest) MSE corresponds to $x = \mathbf{0}$. Indeed, in this case, the second term of (6) is equal to zero and the MSE that is defined only by the first term of (6) has the lowest value. The worst-case (highest) MSE corresponds to the case when x is a vector in the direction of the principle eigenvector of Λ with the norm equal to the its variance σ_x^2 . In this case the second term of (6) has the largest value that gives the largest value for the MSE. Note that the worst-case x depends on G , and is different for different estimators. To find the worst-case x we generate a vector which lies in the direction of the principle eigenvector of Λ , and then normalize it to ensure that $\|x\|^2 = \text{Tr}\{E_x\{xx^H\}\} = \text{Tr}\{C_x\} = n\sigma_x^2$. Finally, the common-case MSE lies between the two bounds given by the best and worst-case MSEs. Thus, it corresponds, for example, to any x taken from Gaussian distribution with zero mean and variance equal to σ_x^2 .

Fig. 1 shows the best-case, worst-case and common-case MSEs of aforementioned estimators versus the inverse noise power given by $-10 \log \sigma_w^2$. The proposed estimator has similar performance to the Wiener-Kolmogorov estimator for the best-case x , and outperforms it for the worst-case and common-case x . Therefore, the proposed estimator has better overall performance than the Wiener-Kolmogorov estimator. However, for the sake of fairness we would like to note here that the MSE in our simulations is calculated using (6) where only the expectation with respect to w is taken, while the MSE for the Wiener-Kolmogorov estimator uses the averaging with respect to both w and x .

4. CONCLUSIONS

An approximation to a linear estimator that minimizes the MSE with a certain selected probability is derived under the assumption

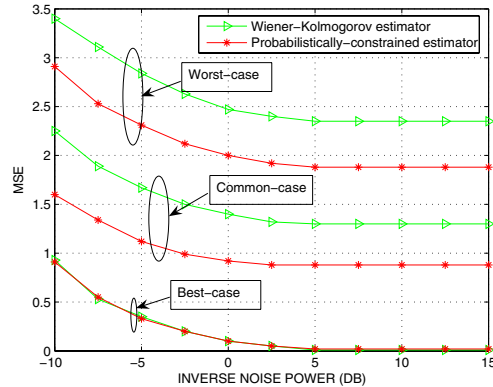


Fig. 1. MSE versus the inverse noise power.

that both the additive observation noise and the unknown random vector are zero mean Gaussian and their covariance matrices are known. Such linear estimator can be viewed as a robust generalization of the Wiener filter, and in some special cases can be simplified to the well-known minimax estimator.

5. 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] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316327, Mar. 1984.
- [4] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Processing*, vol. 52, pp. 1931-1946, July, 2004.
- [5] M. S. Pinsker, "Optimal filtering of square-integrable signals in Gaussian noise," *Problems Inform. Trans.*, vol. 16, pp. 120-133, 1980.
- [6] 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.
- [7] 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.
- [8] A. Prékopa, *Stochastic Programming*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.
- [9] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, (submitted 2004).
- [10] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Inc., 3rd Edition, 1991.
- [11] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. ser. MPS-SIAM Series on Optimization, 2001.
- [12] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.