Cost minimization of measurement devices under estimation accuracy constraints in the presence of Gaussian noise ☆

B. Dulek *, S. Gezici *

Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara TR-06800, Turkey

A R T I C L E   I N F O

A B S T R A C T

Novel convex measurement cost minimization problems are proposed based on various estimation accuracy constraints for a linear system subject to additive Gaussian noise. Closed form solutions are obtained in the case of an invertible system matrix. In addition, the effects of system matrix uncertainty are studied both from a generic perspective and by employing a specific uncertainty model. The results are extended to the Bayesian estimation framework by treating the unknown parameters as Gaussian distributed random variables. Numerical examples are presented to discuss the theoretical results in detail.

1. Introduction

In this paper, we propose measurement cost minimization problems under various constraints on estimation accuracy for a system characterized by a linear input–output relationship subject to Gaussian noise. For the measurement cost, we employ the recently proposed measurement device model in [1], and present a detailed treatment of the proposed measurement cost minimization problems. Although the statistical estimation problem in the presence of Gaussian noise is by far the most widely known and well-studied subject of estimation theory [2], approaches that consider the estimation performance jointly with system-resource constraints have become popular in recent years. Distributed detection and estimation problems took the first step by incorporating bandwidth and energy constraints due to data processing at the sensor nodes, and data transmission from sensor nodes to a fusion node in the context of wireless sensor networks (WSNs) [3–7]. Since then, the majority of the related studies have addressed the costs arising from similar system-level limitations with a relatively weak emphasis on the measurement costs due to amplitude resolution and dynamic range of the sensing apparatus. To begin with, we summarize the main aspects of the research that has been carried out in recent years to unfold the relationship between estimation capabilities and aforementioned costs of the sensing devices.

In [3], detection problems are examined under a constraint on the expected cost resulting from measurement and transmission stages. It is found out that optimal detection performance can be achieved by a randomized on–off transmission scheme of the acquired measurements at a suitable rate. The distributed mean-location parameter estimation problem is considered in [4] for WSNs based on quantized observations. It is shown that when the dynamic range of the estimated parameter is small or comparable with the noise variance, a class of maximum likelihood (ML) estimators exists with performance close to that of the sample mean estimator under stringent bandwidth constraint of one bit per sensor. When the dynamic range of the estimated parameter is comparable to or large than the noise variance, an optimum value for the quantization step results in the highest estimation accuracy possible for a given bandwidth constraint. In [5], a power scheduling strategy that minimizes the total energy consumption subject to a constraint on the worst mean-squared-error (MSE) distortion is derived for decentralized estimation in a heterogeneous sensing environment. Assuming an uncoded quadrature amplitude modulation (QAM) transmission scheme and uniform randomized quantization at the sensor nodes, it is stated that depending on the corresponding channel quality, a sensor is either on or off completely. When a sensor is active, the optimal values for transmission power and quantization level for the sensor can be determined analytically in terms of the channel path losses and local observation noise levels.

In [6], distributed estimation of an unknown parameter is discussed for the case of independent additive observation noises with possibly different variances at the sensors and over non-ideal fading wireless channels between the sensors and the fusion center. The concepts of estimation outage and estimation diversity are introduced. It is proven that the MSE distortion can be minimized under sum power constraints by turning off sensors transmitting over bad channels adaptively without degrading the diversity gain. In addition, performance decrease is reported when

individual power constraints are also imposed at each sensor. In [7], the distributed estimation of a deterministic parameter immersed in uncorrelated noise in a WSN is targeted under a total bit rate constraint. The number of active sensors is determined together with the quantization bit rate of each active sensor in order to minimize the MSE. The problem of estimating a spatially distributed, time-varying random field from noisy measurements collected by a WSN is investigated under bandwidth and energy constraints on the sensors in [8]. Using graph-theoretic techniques, it is shown that the energy consumption can be reduced by constructing reduced order Kalman–Bucy filters from only a subset of the sensors. In order to prevent degradation in the root-mean-squared (RMS) estimation error performance, efficient methods employing Pareto optimality criterion between the communication costs and RMS estimation error are presented. A power allocation problem for distributed parameter estimation is investigated under a total network power constraint for various topologies in [9]. It is shown that for the basic star topology, the optimal solution assumes either of the sensor selection, water-filling, or channel inversion forms depending on the measurement noise variance, and the corresponding analytical expressions are obtained. Asymptotically optimal power allocation strategies are derived for more complex branch, tree, and linear topologies assuming amplify-and-forward and estimate-and-forward transmission protocols. The decentralized WSN estimation is extended to incorporate the effects of imperfect data transmission from sensors to fusion center under stringent bandwidth constraints in [10].

Important results are also obtained for the sensor selection problem under various constraints on the system cost and estimation accuracy. The problem of choosing a set of k sensor measurements from a set of m available measurements so that the estimation error is minimized is addressed in [11] under a Gaussian assumption. It is shown that the combinatorial complexity of the solution can significantly be reduced without sacrificing much from the estimation accuracy by employing a heuristic based on convex optimization. In [12], a similar sensor selection problem is analyzed in a target detection framework when several classes of binary sensors with different discrimination performance and costs are available. Based on the conditional distributions of the observations at the fusion center, the performance of the corresponding optimal hypothesis tests is assessed using the symmetric Kullback–Leibler divergence. The solution of the resulting constrained maximization problem indicates that the sensor class with the best performance-to-cost ratio should be selected.

As outlined above, not much work has been performed, to the best of our knowledge, in the context of jointly designing the measurement stage from a cost-oriented perspective while performing estimation up to a predetermined level of accuracy. In other words, the trade-offs between measurement associated costs and estimation errors remain, to a large extent, undiscovered in the literature. On the other hand, if adopted, such an approach will inevitably require a general and reliable method of assessing the cost of measurements applicable to any real world phenomenon under consideration as well as an appropriate means of evaluating the best achievable estimation performance without reference to any specific estimator structure. For the fulfillment of the first requirement, a novel measurement device model is suggested in [1], where the cost of each measurement is determined by the number of amplitude levels that can reliably be distinguished. As a consequence, higher resolution (less noisy) measurements demand higher costs in accordance with the usual practice. Although the proposed model may lack in capturing the exact relationship between the cost and inner workings of any specific measurement hardware, it encompasses a sufficient amount of generality to remain useful under a multitude of circumstances. Based on this measurement model, an optimization problem is formulated in

[13] in order to calculate the optimal costs of measurement devices that maximize the average Fisher information for a scalar parameter estimation problem.

Although the optimal cost allocation problem is studied for the single parameter estimation case in [13], and the signal recovery based on linear minimum mean-squared-error (LMMSE) estimators is discussed under cost-constrained measurements using a linear system model in [1], no studies have analyzed the implications of the proposed measurement device model in a more general setting by considering both random and nonrandom parameter estimation under various estimation accuracy constraints and uncertainty in the linear system model. The main contributions of our study in this paper extend far beyond a multivariate analysis of the discussion in [13], and can be summarized as follows:

- Formulated new convex optimization problems for the minimization of the total measurement cost by employing constraints on various estimation accuracy criteria (i.e., different functionals of the eigenvalues of the Fisher information matrix (FIM)) assuming a linear system model[1] in the presence of Gaussian noise.
- Studied system matrix uncertainty both from a general perspective and by employing a specific uncertainty model.
- Obtained closed form solutions for two of the proposed convex optimization problems in the case of invertible system matrix.
- Extended the results to the Bayesian estimation framework by treating the unknown estimated parameters as Gaussian distributed random variables.

In addition to the items listed above, simulation results are presented to discuss the theoretical results. Namely, we compare the performance of various estimation quality metrics through numerical examples using optimal and suboptimal cost allocation schemes, and simulate the effects of system matrix uncertainty. We also examine the behavior of the optimal solutions returned by various estimation accuracy criteria under scaling of the system noise variances, and identify the most robust criterion to variations in the average system noise power via numerical examples. The relationship between the number of effective measurements and the quality of estimation is also investigated under scaling of the system noise variances.

The rest of this paper is organized as follows: In Section 2, we pose the optimal cost allocation problem as a convex optimization problem under various information criteria for nonrandom parameter vector estimation. In Section 3, we modify the proposed optimization problems to handle the worst-case scenarios under system matrix uncertainty. Next, we take a specific but nevertheless practical uncertainty model, and discuss how the optimization problems are altered while preserving convexity. In Section 4, we focus on two optimization problems proposed in Section 2, and simplify them to obtain closed form solutions in the case of invertible system matrix. In Section 5, we provide several numerical examples to illustrate the results presented in this paper. Extensions to Bayesian estimation with Gaussian priors are discussed in Section 6, and we conclude in Section 7.

2. Optimal cost allocation under estimation accuracy constraints

Consider a discrete-time system model as in Fig. 1 in which noisy measurements are obtained at the output of a linear system, and then the measurements are processed to estimate the value of a nonrandom parameter vector $\boldsymbol{\theta}$. The observation vector $\mathbf{x}$ at the

---

[1] Such linear models have a multitude of application areas, a few examples of which are channel equalization, wave propagation, compressed sensing, and Wiener filtering problems [14,15].
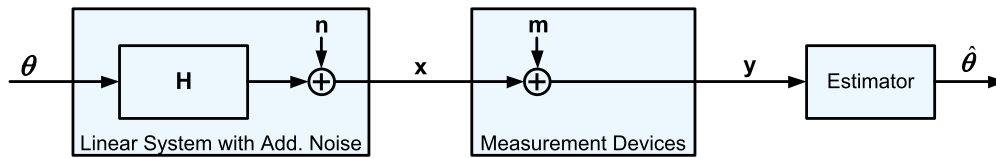
Fig. 1. Measurement and estimation systems model block diagram for a linear system with additive noise.

output of the linear system can be represented by $\mathbf{x} = \mathbf{H}^T \boldsymbol{\theta} + \mathbf{n}$, where $\boldsymbol{\theta} \in \mathbb{R}^L$ denotes a vector of parameters to estimate, $\mathbf{n} \in \mathbb{R}^K$ is the inherent random system noise, and $\mathbf{x} \in \mathbb{R}^K$ is the observation vector at the output of the linear system. The system noise $\mathbf{n}$ is assumed to be a Gaussian distributed random vector with zero-mean, independent but not necessarily identical components, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D_n})$, where $\mathbf{D_n} = \mathrm{diag}\{\sigma_{n_1}^2, \sigma_{n_2}^2, \ldots, \sigma_{n_K}^2\}$ is a diagonal covariance matrix, and $\mathbf{0}$ denotes the all-zeros vector of length K. We also assume that the number of observations is at least equal to the number of estimated parameters (i.e., $K \geqslant L$) and the system matrix $\mathbf{H}$ is an $L \times K$ matrix with full row rank L so that the columns of $\mathbf{H}$ span $\mathbb{R}^L$.

Noisy measurements of the observation vector $\mathbf{x}$ are made by K measurement devices at the output of the linear system, and then the measured values in vector $\mathbf{y} \in \mathbb{R}^K$ are processed to estimate the parameter vector $\boldsymbol{\theta}$. It is assumed that each measurement device is capable of sensing the value of a *scalar* physical quantity with some resolution in amplitude according to the measurement model $y_i = x_i + m_i$, where m_i denotes the measurement noise associated with the ith measurement device. In other words, measurement devices are modeled to introduce additive random measurement noise which can be expressed as $\mathbf{y} = \mathbf{x} + \mathbf{m}$. It is also reasonable to assume that measurement noise vector $\mathbf{m}$ is independent of the inherent system noise $\mathbf{n}$. In addition, the noise components introduced by the measurement devices (the elements of $\mathbf{m}$) are assumed to be zero-mean independent Gaussian random variables with possibly distinct variances,[2] i.e., $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \mathbf{D_m})$, where $\mathbf{D_m}$ is a diagonal covariance matrix given by $\mathbf{D_m} = \mathrm{diag}\{\sigma_{m_1}^2, \sigma_{m_2}^2, \ldots, \sigma_{m_K}^2\}$. Based on the outputs of the measurements devices, unknown parameter vector $\boldsymbol{\theta}$ is estimated.

In practical scenarios, a major issue is the cost of performing measurements. The cost of a measurement device is primarily assessed with its resolution, more specifically with the number of amplitude levels that the device can reliably discriminate. Intuitively, as the accuracy of a measurement device increases so does its cost. Therefore, it may not always be possible to make high resolution measurements with a limited budget. In a recent work [1], a novel measurement device model is proposed where the cost of each device is expressed quantitatively in terms of the number of amplitude levels that can be resolved reliably. In this model, the amplitude resolution of the measurement devices solely determines the cost of each measurement. The dynamic range or scaling of the input to the measurement device is assumed to have no effect on the cost as long as the number of resolvable levels stays the same. More explicitly, in [1], the cost associated with measuring the ith component of the observation vector $\mathbf{x}$ is given by $C_i = 0.5 \log_2(1 + \sigma_{x_i}^2 / \sigma_{m_i}^2)$, where $\sigma_{x_i}^2$ denotes the variance of the ith component of observation vector $\mathbf{x}$ (i.e., the variance of the input to the ith measurement device), and $\sigma_{m_i}^2$ is the variance of the ith component of $\mathbf{m}$ (i.e., the variance of the noise introduced by the ith measurement device).[3] Notice that $\sigma_{x_i}^2 = \sigma_{n_i}^2$,

$\forall i \in \{1, 2, \ldots, K\}$, since $\boldsymbol{\theta}$ is a deterministic parameter vector. Then, the overall cost of measuring all the components of the observation vector $\mathbf{x}$ is expressed as

$$C = \sum_{i=1}^{K} C_i = \sum_{i=1}^{K} \frac{1}{2} \log_2 \left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2} \right). \tag{1}$$

A closer look into (1) reveals that it is a nonnegative, monotonically decreasing and convex function of $\sigma_{m_i}^2$, $\forall \sigma_{n_i}^2 > 0$ and $\forall \sigma_{m_i}^2 > 0$. It is also noted that a measurement device has a higher cost if it can perform measurements with a lower measurement variance (i.e., with higher accuracy). Such an approach brings great flexibility by enabling to work with variable precision over the acquired measurements. After formulating the measurement device model as outlined above, our objective is to minimize the total cost of the measurement devices under a constraint on estimation accuracy. In other words, we are allowed to design the noise levels of the measurement devices such that the overall cost is minimized under a constraint on the minimum acceptable estimation performance.

In nonrandom parameter estimation problems, the Cramer–Rao bound (CRB) provides a lower bound on the mean-squared errors (MSEs) of unbiased estimators under some regularity conditions [16]. Specifically, the CRB on the estimation error for an arbitrary unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is expressed as

$$\mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} \succcurlyeq \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) \triangleq \mathrm{CRB}, \tag{2}$$

where $\mathbf{J}(\mathbf{y}, \boldsymbol{\theta})$ is the Fisher information matrix (FIM) of the measurement $\mathbf{y}$ relative to the parameter vector $\boldsymbol{\theta}$, which is defined as

$$\mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) \triangleq \int \frac{1}{p_{\mathbf{y}}^{\boldsymbol{\theta}}(\mathbf{y})} \left(\frac{\partial p_{\mathbf{y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial p_{\mathbf{y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \boldsymbol{\theta}} \right)^T d\mathbf{y}, \tag{3}$$

where $\partial / \partial \boldsymbol{\theta}$ denotes the gradient (i.e., a column vector of partial derivatives) with respect to parameters $\theta_1, \ldots, \theta_K$. Or, equivalently, the elements of the FIM can be calculated from [16]

$$J_{ij} = -\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}} \left\{ \frac{\partial^2 \log p_{\mathbf{y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_i \partial \theta_j} \right\}. \tag{4}$$

The symbol $\succcurlyeq$ between nonnegative definite matrices in (2) represents the inequality with respect to the positive semidefinite matrix cone. Specifically, it indicates that the difference matrix obtained by subtracting the right-hand side of the inequality from the left-hand side is nonnegative definite. Assuming independent Gaussian distributions for $\mathbf{n}$ and $\mathbf{m}$, it can be shown that the CRB is given as follows [17]

$$\mathrm{CRB} = \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) = \left(\mathbf{H} \mathrm{Cov}^{-1}(\mathbf{n} + \mathbf{m}) \mathbf{H}^T \right)^{-1}, \tag{5}$$

where $\mathrm{Cov}(\cdot)$ denotes the covariance matrix of the random vector $\mathbf{n} + \mathbf{m}$ and $\mathrm{Cov}(\mathbf{n} + \mathbf{m}) = \mathbf{D_n} + \mathbf{D_m} = \mathrm{diag}\{\sigma_{n_1}^2 + \sigma_{m_1}^2, \sigma_{n_2}^2 + \sigma_{m_2}^2, \ldots, \sigma_{n_K}^2 + \sigma_{m_K}^2\}$ due to independence. Then, $\mathbf{D} \triangleq \mathrm{Cov}^{-1}(\mathbf{n} + \mathbf{m}) = \mathrm{diag}\{1/(\sigma_{n_1}^2 + \sigma_{m_1}^2), 1/(\sigma_{n_2}^2 + \sigma_{m_2}^2), \ldots, 1/(\sigma_{n_K}^2 + \sigma_{m_K}^2)\}$, where $\mathrm{Cov}^{-1}(\cdot)$ represents the inverse of the covariance matrix. Notice that the CRB can actually be attained in this case by employing

---

[2] Since Gaussian distribution maximizes the differential entropy over all distributions with the same variance, the assumption that the errors introduced by the measurement devices are Gaussian distributed handles the worst-case scenario.

[3] For an in-depth discussion on the plausibility of this measurement device model and its relation to the number of distinguishable amplitude levels, we refer the reader to [1].

the maximum likelihood (ML) estimator (also the best linear unbiased estimator (BLUE) in this case), $\hat{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{HDH}^T)^{-1}\mathbf{HDy}$, where the efficiency of the estimator follows from linearity of the system and due to the assumption of Gaussian distributions [16]. Specifically, the covariance matrix of the estimator equals the inverse of the FIM, i.e., $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{y})) = (\mathbf{HDH}^T)^{-1}$.

Remark. When non-Gaussian distributions are assumed, we can utilize the preceding observation to obtain an upper bound on the CRB. To see this, a few preliminaries are needed. First, the FIM of a random vector $\mathbf{z}$ with respect to a translation parameter is defined as follows [17]

$$\mathbf{J}(\mathbf{z}) \triangleq \mathbf{J}(\boldsymbol{\theta} + \mathbf{z}, \boldsymbol{\theta}) = \int \frac{1}{p_{\mathbf{z}}(\mathbf{z})}\left(\frac{\partial p_{\mathbf{z}}(\mathbf{z})}{\partial \mathbf{z}}\right)\left(\frac{\partial p_{\mathbf{z}}(\mathbf{z})}{\partial \mathbf{z}}\right)^T d\mathbf{z}, \qquad (6)$$

where $p_{\mathbf{z}}(\mathbf{z})$ is the probability density function of $\mathbf{z}$ that is independent of $\boldsymbol{\theta}$. A well-known property of the FIM under translation is $\mathbf{J}(\mathbf{z}) \succcurlyeq \text{Cov}^{-1}(\mathbf{z})$ with equality if and only if $\mathbf{z}$ is Gaussian [17].

Based on these preliminaries, for linear models in the form of Fig. 1 but with arbitrary probability distributions for $\mathbf{n}$ and $\mathbf{m}$, it can be shown that $\mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) = \mathbf{HJ}(\mathbf{n}+\mathbf{m})\mathbf{H}^T$, where $\mathbf{J}(\mathbf{n}+\mathbf{m})$ indicates the FIM under a translation parameter of random vector $\mathbf{n}+\mathbf{m}$ [17]. In order to upper bound the CRB, it is first observed that $\mathbf{J}(\mathbf{n}+\mathbf{m}) \succcurlyeq \text{Cov}^{-1}(\mathbf{n}+\mathbf{m})$. Using the properties of nonnegative definite matrices, we have

$$\begin{aligned} \text{CRB} &= \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) = \left(\mathbf{HJ}(\mathbf{n}+\mathbf{m})\mathbf{H}^T\right)^{-1} \\ &\preccurlyeq \left(\mathbf{H}\,\text{Cov}^{-1}(\mathbf{n}+\mathbf{m})\mathbf{H}^T\right)^{-1}, \end{aligned} \qquad (7)$$

which naturally indicates that the difference matrix obtained by subtracting the CRB from the covariance matrix of the linear estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ must be nonnegative definite. Correspondingly, it is also possible to lower bound the CRB for independent random vectors $\mathbf{n}$ and $\mathbf{m}$. To that aim, we can revert to the Fisher Information Inequality (FII) [18]. FII states that $\mathbf{J}^{-1}(\mathbf{n}+\mathbf{m}) \succcurlyeq \mathbf{J}^{-1}(\mathbf{n}) + \mathbf{J}^{-1}(\mathbf{m})$ with equality if and only if $\mathbf{n}$ and $\mathbf{m}$ are Gaussian. Therefore,

$$\text{CRB} = \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) \succcurlyeq \left(\mathbf{H}\left(\mathbf{J}^{-1}(\mathbf{n}) + \mathbf{J}^{-1}(\mathbf{m})\right)^{-1}\mathbf{H}^T\right)^{-1}. \qquad (8)$$

As a result, a lower bound on the CRB can also be obtained in terms of the FIMs under translation parameters (6) of random vectors $\mathbf{n}$ and $\mathbf{m}$ with arbitrary probability distributions. □

Returning to our case of independent Gaussian system noise and measurement noise, the CRB is equal to the covariance matrix (i.e., estimation error covariance) of the ML estimator $\hat{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{HDH}^T)^{-1}\mathbf{HDy}$ as mentioned in the paragraph following (5). Furthermore, when the system and measurement noise distributions are not restricted to Gaussian, the covariance matrix of the linear estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ can also be used as an upper bound to the CRB as shown in (7). For this reason, in the following analysis we employ several performance metrics based on the CRB given in (5) in order to assess the quality of estimation. In other words, we propose measurement cost minimization formulations under various estimation accuracy constraints based on the CRB expression in (5). However, before that analysis, we first express the CRB in a more familiar form in the optimization theoretic sense

$$\text{CRB} = \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) = \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2}\mathbf{h}_i\mathbf{h}_i^T\right)^{-1}, \qquad (9)$$

and the corresponding ML estimator that achieves this bound becomes

$$\begin{aligned} \hat{\boldsymbol{\theta}}(\mathbf{y}) &= \left(\mathbf{HDH}^T\right)^{-1}\mathbf{HDy} \\ &= \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2}\mathbf{h}_i\mathbf{h}_i^T\right)^{-1}\sum_{i=1}^{K}\frac{y_i}{\sigma_{n_i}^2 + \sigma_{m_i}^2}\mathbf{h}_i. \end{aligned} \qquad (10)$$

2.1. Average mean-squared error

The diagonal components of the CRB provide a lower bound on the MSE while estimating the components of parameter $\boldsymbol{\theta}$. Specifically,

$$\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\left\{\|\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\|_2^2\right\} \geqslant \text{tr}\left\{\mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta})\right\},$$

where $\text{tr}\{\cdot\}$ denotes the trace operator [16]. In other words, the harmonic average of the eigenvalues of the FIM is taken as the performance metric. Based on this metric, the following measurement cost minimization problem is proposed:

$$\begin{aligned} &\min_{\{\sigma_{m_i}^2\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\ &\text{subject to} \quad \text{tr}\left\{\left(\sum_{i=1}^{K}\frac{1}{\sigma_{n_i}^2+\sigma_{m_i}^2}\mathbf{h}_i\mathbf{h}_i^T\right)^{-1}\right\} \leqslant \text{E}, \end{aligned} \qquad (11)$$

where E denotes a constraint on the maximum allowable average estimation error. Due to the inevitable intrinsic system noise, the design criterion E must satisfy $\text{E} > \text{tr}\{(\mathbf{HD}_{\mathbf{n}}^{-1}\mathbf{H}^T)^{-1}\} = \text{tr}\{(\sum_{i=1}^{K}\frac{\mathbf{h}_i\mathbf{h}_i^T}{\sigma_{n_i}^2})^{-1}\}$. Substituting $\mu_i = 1/(\sigma_{n_i}^2 + \sigma_{m_i}^2)$, (11) becomes

$$\begin{aligned} &\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 - \sigma_{n_i}^2\mu_i\right) \\ &\text{subject to} \quad \text{tr}\left\{\left(\sum_{i=1}^{K}\mu_i\mathbf{h}_i\mathbf{h}_i^T\right)^{-1}\right\} \leqslant \text{E}. \end{aligned} \qquad (12)$$

It is noted that the objective function is smooth and concave for $\forall \mu_i \in [0, 1/\sigma_{n_i}^2)$. Since the constraint is also a convex function of μ_i's for $\forall \mu_i \geqslant 0$, this is a convex optimization problem [19, Section 7.5.2]. Consequently, it can be efficiently solved in polynomial time using interior point methods and the numerical convergence is assured. It is also possible to express this optimization problem using linear matrix inequalities (LMIs) as follows:

$$\begin{aligned} &\max_{\{z_i\}_{i=1}^L,\,\{\mu_i\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 - \sigma_{n_i}^2\mu_i\right) \\ &\text{subject to} \quad \begin{bmatrix} \sum_{i=1}^{K}\mu_i\mathbf{h}_i\mathbf{h}_i^T & \mathbf{e}_j \\ \mathbf{e}_j^T & z_i \end{bmatrix} \succcurlyeq 0, \\ &\qquad j = 1, \ldots, L, \ \sum_{i=1}^{K}z_i \leqslant \text{E}, \end{aligned} \qquad (13)$$

where $\mathbf{e}_j$ denotes the column vector of length L with a 1 in the jth coordinate and 0's elsewhere. Or equivalently,

$$\begin{aligned} &\max_{\mathbf{Z} \in \mathbf{S}_L,\,\{\mu_i\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^{K}\log_2\left(1 - \sigma_{n_i}^2\mu_i\right) \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{Z} & \mathbf{I} \\ \mathbf{I} & \sum_{i=1}^{K}\mu_i\mathbf{h}_i\mathbf{h}_i^T \end{bmatrix} \succcurlyeq 0, \quad \text{tr}(\mathbf{Z}) \leqslant \text{E}, \end{aligned} \qquad (14)$$

where $\mathbf{S}_L$ denotes the set of symmetric $L \times L$ matrices.

2.2. Shannon information

An alternative measure of the estimation accuracy considers the Shannon (mutual) information content between the unknown parameter vector $\boldsymbol{\theta}$ and the measurement vector $\mathbf{y}$. More explicitly, the interest is to place a constraint on the log volume of the η-confidence ellipsoid which is defined as the minimum ellipsoid that contains the estimation error with probability η [19, Section 7.5.2]. As shown in [11], the η-confidence ellipsoid is given by

$$\varepsilon_\alpha = \left\{ \mathbf{z} \mid \mathbf{z}^T \mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) \mathbf{z} \leqslant \alpha \right\}, \tag{15}$$

where $\alpha = F_{\chi_K^2}^{-1}(\eta)$ is obtained from the cumulative distribution function of a chi-squared random variable with K degrees of freedom. Then, the log volume of the η-confidence ellipsoid is obtained as[4]

$$\log \mathbf{vol}(\varepsilon_\alpha) = \beta - \frac{1}{2} \log \det\left(\sum_{i=1}^K \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \mathbf{h}_i^T \right),$$

$$\text{where } \beta = \frac{n}{2} \log(\alpha \pi) - \log\left(\Gamma\left(\frac{n}{2} + 1\right) \right), \tag{16}$$

with Γ denoting the Gamma function. Notice that the design criterion is related to the geometric mean of the eigenvalues of the FIM. Based on this metric, the following measurement cost optimization problem can be obtained:

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \log \det\left(\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \right) \geqslant 2(\beta - \mathrm{S}), \tag{17}$$

where μ_i is as defined in (12) and S is a constraint on the log volume of η-confidence ellipsoid satisfying $\mathrm{S} > \beta - 0.5 \log \det$ $(\mathbf{H}\mathbf{D}_{\mathbf{n}}^{-1}\mathbf{H}^T) = \beta - 0.5 \log \det(\sum_{i=1}^K \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2})$. Since $\log \det(\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T)$ is a smooth concave function of μ_i for $\mu_i \geqslant 0$, the resulting optimization problem is convex [19, Section 3.1.5]. The smoothness property of the problem is also very helpful for obtaining the solution via numerical methods.

By introducing a lower triangular nonsingular matrix $\mathbf{L}$ and utilizing Cholesky decomposition of positive definite matrices, it is possible to rewrite the constraint in terms of a lower bound. To that aim, let $\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \succcurlyeq \mathbf{L}\mathbf{L}^T$. Then, the optimization problem can be expressed equivalently as

$$\max_{\mathbf{L} \in \mathbf{U}_L, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{I} & \mathbf{L}^T \\ \mathbf{L} & \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \end{bmatrix} \succcurlyeq 0, \quad \sum_{i=1}^L \log L_{i,i} \geqslant (\beta - \mathrm{S}), \tag{18}$$

where $\mathbf{U}_L$ denotes the set of lower triangular nonsingular $L \times L$ square matrices, $L_{i,i}$ represents the ith diagonal coefficient of $\mathbf{L}$, and L is the dimension of $\mathbf{L}$.

2.3. Worst-case error variance

When the primary concern shifts from accuracy requirements towards robust behavior, it may be more desirable to have a constraint on the worst-case variance of the estimation error, which

---

is associated with the maximum (minimum) eigenvalue of the CRB (FIM) [11,20–22]. The corresponding optimization problem is stated as follows:

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \lambda_{\min}\left\{ \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \right\} \geqslant \Lambda, \tag{19}$$

where $\lambda_{\min}\{\cdot\}$ represents the minimum eigenvalue of its argument, and Λ is a predetermined lower bound on the minimum eigenvalue of the FIM satisfying $\Lambda < \lambda_{\min}\{\mathbf{H}\mathbf{D}_{\mathbf{n}}^{-1}\mathbf{H}^T\} = \lambda_{\min}\{\sum_{i=1}^K \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2}\}$. Since the constraint can be represented in the form of an LMI, this problem can equivalently be expressed as

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \succcurlyeq \Lambda \mathbf{I}, \tag{20}$$

where $\mathbf{I}$ is the $L \times L$ identity matrix. The resulting problem is also convex [19, Section 7.5.2].

2.4. Worst-case coordinate error variance

Another variation of the worst-case error criteria can be obtained by placing a constraint on the maximum error variance among all the individual estimator components, i.e., restricting the largest diagonal entry of the CRB. Using this performance criterion, we have the following optimization problem

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \max_{j=1,\dots,K} \left(\left(\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \right)^{-1} \right)_{j,j} \leqslant \varrho, \tag{21}$$

where ϱ is a constraint on the maximum allowable diagonal entry of the CRB (estimation error covariance matrix) satisfying $\varrho > \max_{j=1,\dots,K}((\mathbf{H}\mathbf{D}_{\mathbf{n}}^{-1}\mathbf{H}^T)^{-1})_{j,j} = \max_{j=1,\dots,K}((\sum_{i=1}^K \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2})^{-1})_{j,j}$. This problem can equivalently be expressed as

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \varrho & \mathbf{e}_j^T \\ \mathbf{e}_j & \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \end{bmatrix} \succcurlyeq 0, \quad j = 1, \dots, L, \tag{22}$$

where $\mathbf{e}_j$ denotes the column vector of length L with a 1 in the jth coordinate and 0's elsewhere. This is also a convex optimization problem [19, Section 7.5.2].

3. Extensions to cases with system matrix uncertainty – robust measurement

It may also be the case that there exists some uncertainty concerning the elements in the system matrix $\mathbf{H}$ [11]. Suppose that the system matrix $\mathbf{H}$ can take values from a given finite set $\mathcal{H}$. In the robust measurement problem, we consider the optimization over the worst-case scenario. Specifically, we choose the matrix

---

[4] We use 'log' without a subscript to denote the natural logarithm.

from the family of system matrices $\mathcal{H}$ resulting in the worst estimation accuracy constraint, and perform the optimization accordingly. Recalling that the infimum (supremum) preserves concavity (convexity), it is possible to restate the measurement cost optimization problems given in Section 2, and still maintain convex optimization problems. Then, the resulting optimization problems with respect to each criterion are expressed as follows.

3.1. Average mean-squared error

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \sup_{\mathbf{H}\in\mathcal{H}} \text{tr}\left\{\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T\right\}^{-1} \leqslant E, \tag{23}$$

or equivalently,

$$\max_{\mathbf{Z}\in\mathbf{S}_L,\, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{Z} & \mathbf{I} \\ \mathbf{I} & \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \end{bmatrix} \succcurlyeq 0 \quad \text{for all } \mathbf{H}\in\mathcal{H},$$

$$\text{tr}(\mathbf{Z}) \leqslant E. \tag{24}$$

3.2. Shannon information

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \inf_{\mathbf{H}\in\mathcal{H}} \log\det\left\{\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T\right\} \geqslant 2(\beta - S), \tag{25}$$

or equivalently,

$$\max_{\mathbf{L}\in\mathbf{U}_L,\, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{I} & \mathbf{L}^T \\ \mathbf{L} & \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \end{bmatrix} \succcurlyeq 0 \quad \text{for all } \mathbf{H}\in\mathcal{H},$$

$$\sum_{i=1}^L \log L_{i,i} \geqslant (\beta - S). \tag{26}$$

3.3. Worst-case error variance

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T \succcurlyeq \Lambda \mathbf{I} \quad \text{for all } \mathbf{H}\in\mathcal{H}. \tag{27}$$

3.4. Worst-case coordinate error variance

$$\max_{\{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \sup_{\mathbf{H}\in\mathcal{H}} \max_{j=1,\dots,K}\left(\left\{\sum_{i=1}^K \mu_i \mathbf{h}_i \mathbf{h}_i^T\right\}^{-1}\right)_{j,j} \leqslant \varrho. \tag{28}$$

When the set $\mathcal{H}$ is finite, the problem can be solved using standard arguments from convex optimization. However, the set

$\mathcal{H}$ is in general not finite, and the solutions of the above optimization problems require general techniques from semi-infinite convex optimization such as those explained in [23,24]. In the following, a specific uncertainty model is considered where it is possible to further simplify the optimization problems given in (26) and (27) by expressing the constraints as LMIs. To that aim, let $\mathbf{H}\in\mathcal{H} = \{\bar{\mathbf{H}} + \boldsymbol{\Delta}:\, \|\boldsymbol{\Delta}^T\|_2 \leqslant \epsilon\}$, where $\|\cdot\|_2$ denotes the spectral norm (i.e., the square root of the largest eigenvalue of the positive semidefinite matrix $\boldsymbol{\Delta}\boldsymbol{\Delta}^T$). It is possible to express this constraint as an LMI, $\boldsymbol{\Delta}\boldsymbol{\Delta}^T \preccurlyeq \epsilon^2 \mathbf{I}$. Suppose also that $\boldsymbol{\mu}$ is defined as the following diagonal matrix $\boldsymbol{\mu} \triangleq \text{diag}\{\mu_1, \mu_2, \dots, \mu_K\}$, and $\mathbf{W} \triangleq \mathbf{L}\mathbf{L}^T$ is a symmetric positive definite matrix. Then, the constraint in (26) can be expressed in terms of $\bar{\mathbf{H}}$ and $\boldsymbol{\Delta}$ as

$$\mathbf{W} \preccurlyeq \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \bar{\mathbf{H}}\boldsymbol{\mu}\boldsymbol{\Delta}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\boldsymbol{\Delta}^T,$$

$$\text{for all } \boldsymbol{\Delta}\boldsymbol{\Delta}^T \preccurlyeq \epsilon^2 \mathbf{I}. \tag{29}$$

Similarly, the constraint in (27) is given by

$$\Lambda\mathbf{I} \preccurlyeq \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \bar{\mathbf{H}}\boldsymbol{\mu}\boldsymbol{\Delta}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\bar{\mathbf{H}}^T + \boldsymbol{\Delta}\boldsymbol{\mu}\boldsymbol{\Delta}^T,$$

$$\text{for all } \boldsymbol{\Delta}\boldsymbol{\Delta}^T \preccurlyeq \epsilon^2 \mathbf{I}. \tag{30}$$

In [25, Theorem 3.3], a necessary and sufficient condition is derived for quadratic matrix inequalities in the form of (29) and (30) to be true. In the light of this theorem, (29) holds if and only if there exists $t \geqslant 0$ such that

$$\begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - \mathbf{W} - t\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succcurlyeq 0, \tag{31}$$

and (30) holds if and only if there exists $t \geqslant 0$ such that

$$\begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - (\Lambda + t)\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succcurlyeq 0. \tag{32}$$

Notice that (31) and (32) are both linear in $\boldsymbol{\mu}, \mathbf{W}$ and t. Hence, under this specific uncertainty model, we can express the optimization problem in (26) as

$$\max_{t,\mathbf{W}\in\mathbf{s}_{++}^L,\, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - \mathbf{W} - t\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succcurlyeq 0,$$

$$\log\det(\mathbf{W}) \geqslant 2(\beta - S), \quad t \geqslant 0, \tag{33}$$

where $\mathbf{s}_{++}^L$ denotes symmetric positive-definite $L \times L$ matrices. Similarly, it is possible to write the optimization problem in (27) as

$$\max_{t,\, \{\mu_i\}_{i=1}^K} \frac{1}{2} \sum_{i=1}^K \log_2\left(1 - \sigma_{n_i}^2 \mu_i\right)$$

$$\text{subject to} \quad \begin{bmatrix} \bar{\mathbf{H}}\boldsymbol{\mu}\bar{\mathbf{H}}^T - (\Lambda + t)\mathbf{I} & \bar{\mathbf{H}}\boldsymbol{\mu} \\ \boldsymbol{\mu}\bar{\mathbf{H}}^T & \boldsymbol{\mu} + \frac{t}{\epsilon^2}\mathbf{I} \end{bmatrix} \succcurlyeq 0, \quad t \geqslant 0. \tag{34}$$

4. Special case – invertible system matrix H

When the system matrix $\mathbf{H}$ is a $K \times K$ invertible matrix meaning that the number of unknown parameters is equal to the number of observations, it is possible to obtain closed-form solutions of the optimization problems stated in (11) and (17). Moreover, for the solution of (11), it is not necessary to assume that the components of the system noise $\mathbf{n}$ are independent; it is sufficient to have $\mathbf{n}$ as a Gaussian distributed random vector with zero-mean and

arbitrary covariance matrix (possibly colored), i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{n})$ with $\{\sigma_{n_1}^2, \sigma_{n_2}^2, \ldots, \sigma_{n_K}^2\}$ constituting the diagonal components of $\boldsymbol{\Sigma}_\mathbf{n}$, and $\mathbf{0}$ denoting the all-zeros vector of length K as before. To that aim, assuming independent Gaussian distributions for $\mathbf{n}$ and $\mathbf{m}$, and square $\mathbf{H}$ with full-rank (invertible), it is observed that

$$\begin{aligned}
\text{CRB} = \mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta}) &= \left(\mathbf{H}\,\text{Cov}^{-1}(\mathbf{n}+\mathbf{m})\mathbf{H}^T\right)^{-1} \\
&= \left(\mathbf{H}^{-1}\right)^T \text{Cov}(\mathbf{n}+\mathbf{m})\mathbf{H}^{-1} \\
&= \left(\mathbf{H}^{-1}\right)^T \boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}^{-1} + \left(\mathbf{H}^{-1}\right)^T \mathbf{D}_\mathbf{m}\mathbf{H}^{-1},
\end{aligned} \tag{35}$$

where the first part of the CRB, $(\mathbf{H}^{-1})^T \boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}^{-1}$ is a known quantity, and the second part $(\mathbf{H}^{-1})^T \mathbf{D}_\mathbf{m}\mathbf{H}^{-1}$ will be subject to design while assessing the quality of the estimation. Similar to the previous discussion, CRB can be achieved in this case by employing the corresponding linear unbiased estimator which turns out simply to be a multiplication of the measurement vector with the inverse of the system matrix, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{y}) = (\mathbf{H}^{-1})^T\mathbf{y}$. Returning to two commonly used performance metrics introduced in Section 2, we next examine the closed-form solutions of the corresponding cost minimization problems.

4.1. Average mean-squared error

Due to the CRB, it is known that the average MSE while estimating the components of the parameter $\boldsymbol{\theta}$ is bounded from below as

$$\begin{aligned}
\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\left\{\left\|\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\right\|_2^2\right\} &\geqslant \text{tr}\left\{\mathbf{J}^{-1}(\mathbf{y}, \boldsymbol{\theta})\right\} \\
&= \text{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}^{-1}\right\} + \text{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \mathbf{D}_\mathbf{m}\mathbf{H}^{-1}\right\},
\end{aligned}$$

where the last equality follows from the linearity of the trace operator and the invertibility of $\mathbf{H}$. Since $(\mathbf{H}^{-1})^T \boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}^{-1}$ is known, let $t = \text{tr}\{(\mathbf{H}^{-1})^T \boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}^{-1}\}$. When the aim is to minimize the measurement cost subject to a constraint on the lower bound for the average MSE (achievable in the case of Gaussian distributions), the optimization problem can be expressed similarly to (11) as follows:

$$\begin{aligned}
&\min_{\{\sigma_{m_i}^2\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^K \log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\
&\text{subject to} \quad \text{tr}\left\{\left(\mathbf{H}^{-1}\right)^T \mathbf{D}_\mathbf{m}\mathbf{H}^{-1}\right\} \leqslant \text{E} - t,
\end{aligned} \tag{36}$$

where E denotes a constraint for the overall average estimation error suggested by the CRB (achievable in this case), and t represents the unavoidable estimation error due to intrinsic system noise $\mathbf{n}$. Notice that for consistency, the design parameter E should be selected as $\text{E} > t$.

From the independence of the measurement noise components, $\mathbf{D}_\mathbf{m} = \text{diag}\{\sigma_{m_1}^2, \sigma_{m_2}^2, \ldots, \sigma_{m_K}^2\}$ is a diagonal covariance matrix with $\sigma_{m_i}^2 > 0$, $\forall i \in \{1, 2, \ldots, K\}$. In the view of this observation, it is possible to simplify the objective function further by defining $\mathbf{F} \triangleq (\mathbf{H}^{-1})^T = [\mathbf{f}_1 \ \mathbf{f}_2 \ \ldots \ \mathbf{f}_K]$, where $\mathbf{f}_i$ represents the ith row of the inverse of the system matrix $\mathbf{H}$. Let $f_i \triangleq \|\mathbf{f}_i\|_2^2$ denote the square of the Euclidean norm of the vector $\mathbf{f}_i$, that is, the sum of squares of the elements in $\mathbf{f}_i$. It is noted that f_i is always positive for invertible $\mathbf{H}$, and is constant for fixed $\mathbf{H}$. Then the optimization problem in (36) can be expressed as follows:

$$\begin{aligned}
&\min_{\{\sigma_{m_i}^2\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^K \log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\
&\text{subject to} \quad \sum_{i=1}^K f_i\sigma_{m_i}^2 \leqslant \text{E} - t, \quad \sigma_{m_i}^2 \geqslant 0, \ \forall i \in \{1, 2, \ldots, K\}.
\end{aligned} \tag{37}$$

From (37), it is noted that the constraint function is linear in $\sigma_{m_i}^2$'s, the objective function is convex, and both functions are continuously differentiable which altogether indicate that Slater's condition holds. Therefore, Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality. Then, the optimal measurement noise variances can be calculated from

$$\sigma_{m_i}^2 = -\frac{\sigma_{n_i}^2}{2} + \sqrt{\frac{\sigma_{n_i}^4}{4} + \gamma\,\frac{\sigma_{n_i}^2}{f_i}}, \tag{38}$$

where $\gamma > 0$ is obtained by substituting (38) into the average MSE constraint, that is $\sum_{i=1}^K f_i\sigma_{m_i}^2 = \text{E} - t$.

Special case: When the inverse of the system matrix has normalized rows, i.e., $f_i = 1$, and the components of the system noise are independent zero-mean Gaussian random variables, the optimal measurement noise variances should satisfy $\sum_{i=1}^K \sigma_{m_i}^2 = \text{E} - \sum_{i=1}^K \sigma_{n_i}^2$. If identical system noise components are assumed as well, i.e., $\sigma_{n_i}^2 = \sigma_n^2$, $i = 1, \ldots, K$, then the optimal solution results in $\sigma_{m_i}^2 = \sigma_m^2$, $i = 1, \ldots, K$, where $\sigma_m^2 = \text{E}/K - \sigma_n^2$ is obtained from the average MSE constraint. The corresponding optimal cost is given by $(K/2)\log_2(\text{E}/(\text{E} - K\sigma_n^2))$. This is an increasing function of K for fixed E. Furthermore, the derivatives of all orders with respect to K exist, and are positive for $K < \text{E}/\sigma_n^2$. Therefore, estimating more parameters under an average error constraint based on the CRB requires even more accurate measurement devices with higher costs as long as $K < \text{E}/\sigma_n^2$ is satisfied.

4.2. Shannon information

Another measure of estimation accuracy that results in a closed form solution in the case of invertible system matrix $\mathbf{H}$ is the Shannon information criterion. Using this metric as the constraint function, we are effectively restricting the log volume of the η-confidence ellipsoid to stay below a predetermined value S. Using similar arguments to Section 2.2 and the invertibility of $\mathbf{H}$,

$$\begin{aligned}
&\log\det\left(\mathbf{H}\,\text{Cov}^{-1}(\mathbf{n}+\mathbf{m})\mathbf{H}^T\right) \\
&= \log\left(\det\mathbf{H}\cdot\det\left(\text{Cov}^{-1}(\mathbf{n}+\mathbf{m})\right)\cdot\det\mathbf{H}^T\right) \\
&= 2\log|\det\mathbf{H}| - \sum_{i=1}^K \log\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right),
\end{aligned} \tag{39}$$

where the second equality follows the properties of the determinant and logarithm, i.e., $\det\mathbf{H} = \det\mathbf{H}^T$, $\det(\text{Cov}^{-1}(\mathbf{n}+\mathbf{m})) = 1/\det(\text{Cov}(\mathbf{n}+\mathbf{m}))$, and $\text{Cov}(\mathbf{n}+\mathbf{m}) = \mathbf{D}_\mathbf{n} + \mathbf{D}_\mathbf{m} = \text{diag}\{\sigma_{n_1}^2 + \sigma_{m_1}^2, \sigma_{n_2}^2 + \sigma_{m_2}^2, \ldots, \sigma_{n_K}^2 + \sigma_{m_K}^2\}$ due to Gaussian distributed independent system and measurement noises with independent components. Since the system matrix $\mathbf{H}$ is known, let $\alpha \triangleq \log|\det\mathbf{H}|$. Under these conditions, the optimization problem in (17) can be stated as

$$\begin{aligned}
&\min_{\{\sigma_{m_i}^2\}_{i=1}^K} \frac{1}{2}\sum_{i=1}^K \log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right) \\
&\text{subject to} \quad \sum_{i=1}^K \log\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right) \leqslant 2(\text{S} + \alpha - \beta),
\end{aligned} \tag{40}$$

where S and β are as defined in (17).

Notice that although the objective in (40) is a convex function of $\sigma_{m_i}^2$'s, the constraint is not a convex set. In fact, the constraint set is what is left after the convex set

$$\mathcal{C} = \left\{\boldsymbol{\sigma}_\mathbf{m}^2 \succcurlyeq 0 : \sum_{i=1}^K \log\left(\sigma_{n_i}^2 + \sigma_{m_i}^2\right) > 2(\text{S} + \alpha - \beta)\right\}$$

is subtracted from $\{\sigma_{\mathbf{m}}^2 \succcurlyeq 0\}$. Since the global minimum of the unconstrained objective function is achieved for $\sigma_{\mathbf{m}}^2 = \infty$ which is contained in set $\mathcal{C}$ and the objective function is convex, it is concluded that the minimum of the objective function has to occur at the boundary, i.e., $\sum_{i=1}^{K} \log(\sigma_{n_i}^2 + \sigma_{m_i}^2) = 2(S + \alpha - \beta)$ must be satisfied [26]. Therefore, we can take the constraint as equality in (40). This is a standard optimization problem that can be solved using Lagrange multipliers. Hence, by defining $\varrho \triangleq 2(S + \alpha - \beta)$, we can write the Lagrange functional as

$$J(\sigma_{m_1}^2, \ldots, \sigma_{m_K}^2) = \frac{1}{2} \sum_{i=1}^{K} \log_2\left(1 + \frac{\sigma_{n_i}^2}{\sigma_{m_i}^2}\right)$$
$$+ \lambda\left(\sum_{i=1}^{K} \log(\sigma_{n_i}^2 + \sigma_{m_i}^2) - \varrho\right), \quad (41)$$

and differentiating with respect to $\sigma_{m_i}^2$, we have the following assignment of the noise variances to the measurement devices

$$\sigma_{m_i}^2 = (\gamma^{1/K} - 1)\sigma_{n_i}^2, \quad \text{where } \gamma = \frac{2^\varrho}{\prod_{j=1}^{K} \sigma_{n_j}^2}. \quad (42)$$

For consistency, the design parameter S should be selected as $\varrho = 2(S + \alpha - \beta) > \sum_{i=1}^{K} \log(\sigma_{n_i}^2)$ since the intrinsic system noise puts a lower bound on the minimum attainable volume of the confidence ellipsoid. Some properties of the obtained solution can be summarized as follows:

- For given ϱ, K and $\sigma_{n_i}^2$'s, the minimum achievable cost is $(K/2) \log_2\left(\frac{\gamma^{1/K}}{\gamma^{1/K} - 1}\right)$, where γ is computed as in (42).
- For a fixed value of K (available number of observations), relaxing the constraint on the volume of the η-confidence ellipsoid (increasing the value of ϱ) results in smaller measurement device costs with a limiting value of 0, as expected.
- If the observation variances are equal; that is, $\sigma_{n_i}^2 = \sigma_n^2$, $i = 1, \ldots, K$, employing identical measurement devices for all the observations; that is, $\sigma_{m_i}^2 = \sigma_m^2$, $i = 1, \ldots, K$, is the optimal strategy. From (42), the optimal value of the measurement noise variances is calculated as $\sigma_{m,\text{opt}}^2 = e^{\varrho/K} - \sigma_n^2$, and the corresponding minimum total measurement cost is given as $\varrho/(2\log 2) - (K/2)\log_2(e^{\varrho/K} - \sigma_n^2)$ which is an increasing function of K for $\varrho > K \log \sigma_n^2$. Intuitively, this result as well indicates that estimating more parameters under a fixed constraint on the volume of the ellipsoid containing the estimation errors requires a higher total measurement device cost.

5. Numerical results

In this section, we present an example that illustrates several theoretical results developed in the previous section. To that aim, a discrete-time linear system as depicted in Fig. 1 is considered

$$\mathbf{y} = \mathbf{H}^T \boldsymbol{\theta} + \mathbf{n} + \mathbf{m}, \quad (43)$$

where $\boldsymbol{\theta}$ is a length-20 vector containing the unknown parameters to be estimated, $\mathbf{H}$ is a 20×100 system matrix with full row rank, the intrinsic system noise $\mathbf{n}$ and the measurement noise $\mathbf{m}$ are length-100 Gaussian distributed random vectors with independent components. The entries of the system matrix $\mathbf{H}$ are generated from a process of i.i.d. uniform random variables in the interval $[-0.1, 0.1]$. Also, the components of the system noise vector $\mathbf{n}$ are independently Gaussian distributed with zero mean, and it is assumed that their variances come from a uniform distribution defined in the interval $[0.05, 1]$. The implication of this assumption is that the observations at the output of the linear system possess uniformly varying degrees of accuracy. In other words,

it is assured that observations corrupted by weak, moderate and strong levels of Gaussian noise are available with similar proportions for the estimation stage. In the following, we look into the problem of optimally assigning costs to measurement devices under various estimation accuracy constraints when the variances of the intrinsic system noise components are uniformly distributed as explained above. Note that our results obtained in the previous section are still valid for Gaussian system noise processes with arbitrary diagonal covariance matrices (i.e., the nonzero components of the diagonal covariance matrix need not be uniformly distributed as in this example). In obtaining the optimal solutions for the convex optimization problems stated above, *fmincon* method from MATLAB's Optimization Toolbox and the CVX software [27] are used.

5.1. Performance of various estimation quality metrics under perfect system state information

First, we investigate the cost assignment problem under perfect information on the system matrix and intrinsic noise variances. Recall that four different performance constraints are proposed for that purpose in Section 2. In the following four experiments, we analyze the behavior of the total measurement cost while each constraint metric is varied between its extreme values. The total cost is measured in bits by taking logarithms with respect to base 2. The constraint metric is expressed as the ratio of its current value to the value it attains for the limiting case when zero measurement noise variances are assumed. As an example, for average mean-squared-error criterion, the total measurement cost C will be tabulated versus $E/\operatorname{tr}\{(\mathbf{H}\mathbf{D_n}^{-1}\mathbf{H}^T)^{-1}\}$.

In addition to the optimal cost allocation scheme proposed in this paper, we also consider two suboptimal cost allocation strategies:

- *Equal cost to all measurement devices:* In this strategy, it is assumed that a single set of measurement devices with identical costs is employed for all observations so that $C_i = C$, $i = 1, 2, \ldots, K$. This, in turn, implies that the ratio of the measurement noise variance to the intrinsic system noise variance, $x \triangleq \sigma_{m_i}^2/\sigma_{n_i}^2$, is constant for all measurement devices. Then, the total cost can be expressed in terms of x as $C = 0.5K \log_2(1 + 1/x)$, and similarly the FIM becomes $\mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) = \frac{\mathbf{H}\mathbf{D_n}^{-1}\mathbf{H}^T}{x+1} = \frac{1}{x+1} \sum_{i=1}^{K} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2}$. Using this observation, the constraint functions provided for different performance metrics in the optimization problems (11), (17), (19), and (21) can be algebraically solved for equality to determine the value of x without applying any convex optimization techniques, and the corresponding measurement variances and cost assignments can be obtained.
- *Equal measurement noise variances:* In this case, measurement devices are assumed to introduce random errors with equal noise variances, that is, $\sigma_{m_i}^2 = \sigma_m^2$, $i = 1, 2, \ldots, K$. In other words, all observations are assumed to be corrupted with identical noise processes, and the best measurement noise variance value that minimizes the overall measurement cost while satisfying the estimation accuracy constraint is selected. Accordingly, the objective function in the proposed optimization problems simplifies to $C = 0.5 \sum_{i=1}^{K} \log_2(1 + \sigma_{n_i}^2/\sigma_m^2)$ and the FIM employed in the constraint functions takes the form $\mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{K} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\sigma_{n_i}^2 + \sigma_m^2}$. By substituting these expressions into the various optimization approaches provided in Section 2, these problems can be solved rapidly over a *single* parameter σ_m^2 using the tools of convex analysis, and the optimal cost allocations can be obtained for the case of equal measurement noise variances.

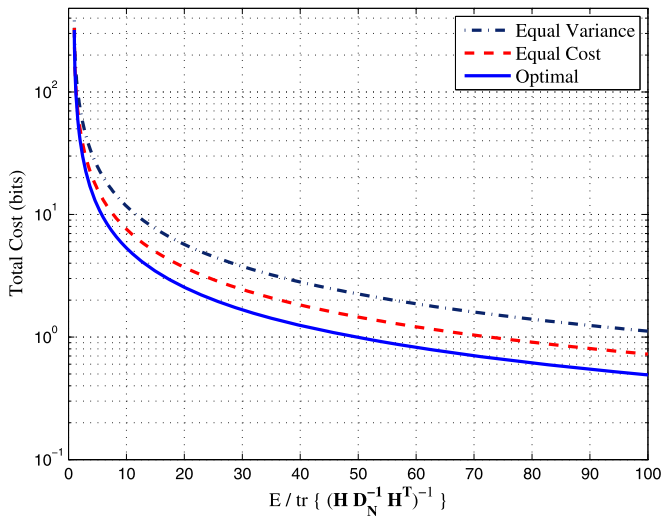**Fig. 2.** Total cost versus normalized average MSE constraint.

Fig. 3. Total cost versus normalized Shannon information constraint.

Fig. 4. Total cost versus normalized worst-case error variance constraint.

5.1.1. Average mean-squared-error criterion

In this experiment, we study the effects of the average MSE constraint on the total measurement device cost. Starting from the minimum achievable value for the average MSE due to intrinsic system noise (i.e., $\mathrm{tr}\{(\mathbf{HD_n^{-1}H}^T)^{-1}\}$), we increase the constraint up to 100 times this minimal value, as depicted in Fig. 2. Three curves are presented corresponding to the optimal cost allocation strategy and two suboptimal strategies, one employing equal cost and the other employing equal noise variance among the measurement devices. It is noted that the optimal strategy results in the minimum cost for all values of the MSE constraint as expected. Its performance is followed by the equal cost assignment scheme, and the worst performing strategy is the one that assigns equal measurement noise variances to all the devices. When the average MSE criterion is stringent (for smaller values of E), all the strategies require increasingly more accurate measurements (hence higher costs) to satisfy the constraint. As the MSE constraint is relaxed (i.e., for larger values of E), the measurement costs of three different strategies start to drop down to zero but become less responsive as they move along.

5.1.2. Shannon information criterion

This experiment aims to discover the relationship between Shannon information constraint and total measurement device cost. Since the constraint is expressed as a 'greater than' inequality, we begin with the maximum attainable value of $\log\det(\mathbf{HD_n^{-1}H}^T)$ and loosen the constraint by decreasing towards the negative multiples of this quantity as shown in Fig. 3. When the constraint is very restrictive (corresponding to high values of $2(\beta - S)$), the differences among the performances of optimal and suboptimal strategies disappear. As the constraint is relaxed away from the maximum attainable value, it is observed that the decrease in the total cost is less responsive with respect to the average MSE. However, as the relaxation continues we see that the drop in the total cost for the Shannon information criterion maintains its pace for a longer time while the drop in the average MSE criterion seems to saturate. Again similar to the previous case, the performance of the optimal strategy is superior to the equal measurement device cost strategy, and the worst performance belongs to the equal measurement variance scheme.

5.1.3. Worst-case error variance

In this experiment, we investigate the effects of the worst-case error variance criterion on the total measurement device cost un-
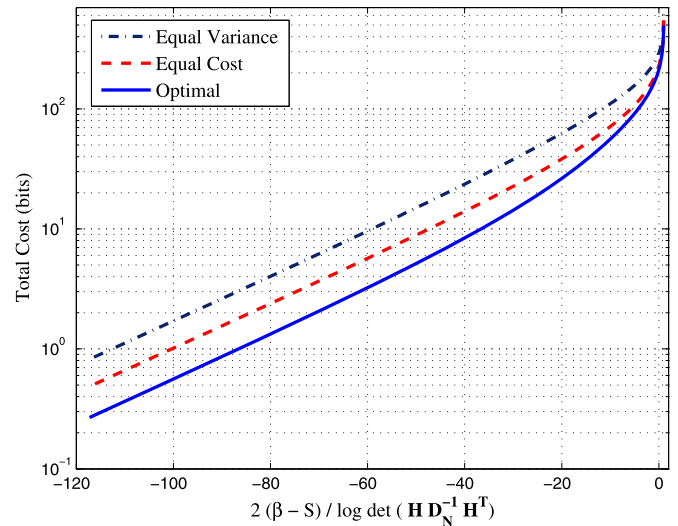
der different cost allocation strategies. Similar conclusions to the previous experiments can be drawn by examining Fig. 4.
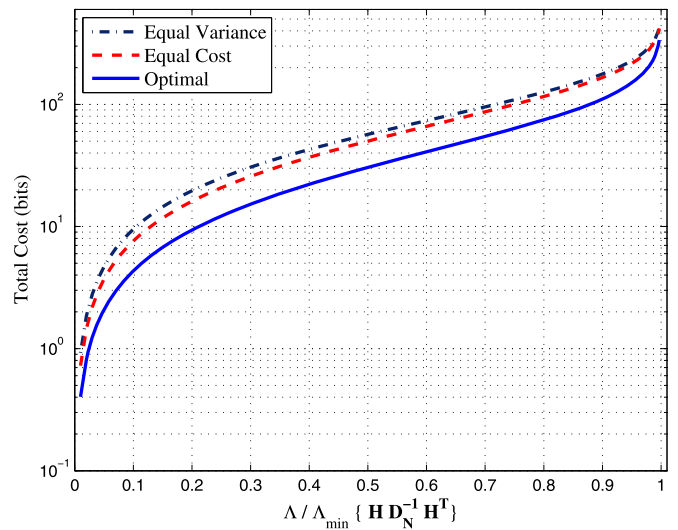
5.1.4. Worst-case coordinate error variance

This experiment focuses on the relationship between the constraint on the largest diagonal entry of the CRB and the total measurement device costs achievable via different cost allocation strategies. The results are illustrated in Fig. 5. It is noted that the plots depicted in Fig. 2 embody a large degree of resemblance to those given in Fig. 5. This similarity is anticipated and can be attributed to the fact that the former criterion puts a constraint on the average of the diagonal entries of the CRB whereas the latter places a similar constraint on their maximum.

Finally, we can stress a few more points. It is necessary that the intrinsic system noise variances and the system matrix are jointly evaluated to compute the optimal measurement noise variances and the corresponding cost allocations. In other words, in order to assign more cost to a specific observation, it is not sufficient to just know that the particular observation is reliable (i.e., has smaller variance) but we also need to know its intrinsic combinations with the other observations due to linear system matrix. Furthermore, the performance figures are quite useful in the sense that they
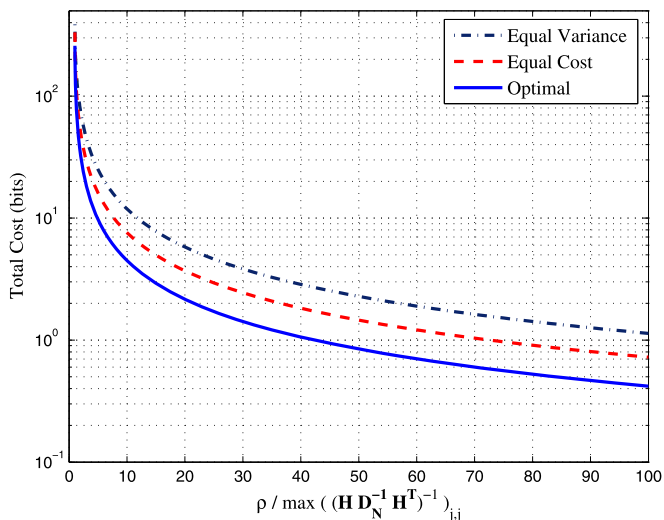
Fig. 5. Total cost versus normalized worst-case coordinate error variance constraint.

Fig. 6. The performance of various optimal cost allocation strategies under scaling of the system noise variances. All costs are equal for $c = 0.5$.

provide the minimum cost necessary to obtain a desired level of estimation accuracy.

5.2. Performance comparison of estimation quality metrics under scaling of the system noise variances

In this section, we devise a new experiment in order to jointly assess the performance of the proposed optimal cost assignment strategies under different estimation quality metrics. Using the same set of system noise variances employed in the previous experiments, we scale them with a factor c that varies inside the interval [0.1, 1] with 0.01 increments. Specifically, $\hat{\sigma}_{n_i}^2 = c\sigma_{n_i}^2$, $i = 1, \ldots, K$, where $c \in \{0.1 : 0.01 : 1\}$. For such a comparison to make sense, the constraints on the estimation quality metrics are selected so that the optimal total measurement costs returned by the various approaches are equal for a certain value of the scale parameter c. Then, using the same value as the constraint, we evaluate the performance of each optimal cost allocation strategy for the rest of the scale parameter values.

Two examples are constructed for this case. In the first one, the performances of the optimal schemes under four different performance metrics are equated for $c = 0.5$, producing an optimal total cost of 40.11. The corresponding constraint function values are E = 23.1371 for the average MSE criterion, $2(\beta - S) = 1.9389$ for Shannon information criterion, $\Lambda = 0.4364$ for the worst-case error variance criterion and $\varrho = 1.3646$ for the worst-case coordinate error variance criterion. The results are illustrated in Fig. 6. Intuitively, as the intrinsic system noise variances are increased, more reliable measurements (higher costs) are required to satisfy the same level of accuracy. Comparing the performances in Fig. 6, where all the costs are equated for $c = 0.5$, we observe that the average MSE criterion results in the least (i.e., the best) optimal cost score for increasing values of the scale parameter c. Its performance is followed by the Shannon information criterion, next by the worst-case coordinate error variance criterion, and finally by the worst-case error variance criterion. In other words, the effects of increasing system noise variances are much more pronounced for the worst-case error variance criterion, which operates by setting a constraint on the minimum eigenvalue of the FIM, than the remaining criteria. If the noise scale parameter c is decreased below 0.5, it is observed that the Shannon information criterion produces the lowest measurement cost followed by the worst-case coordinate error variance criterion, worst-case error variance criterion, and finally average MSE criterion in the order
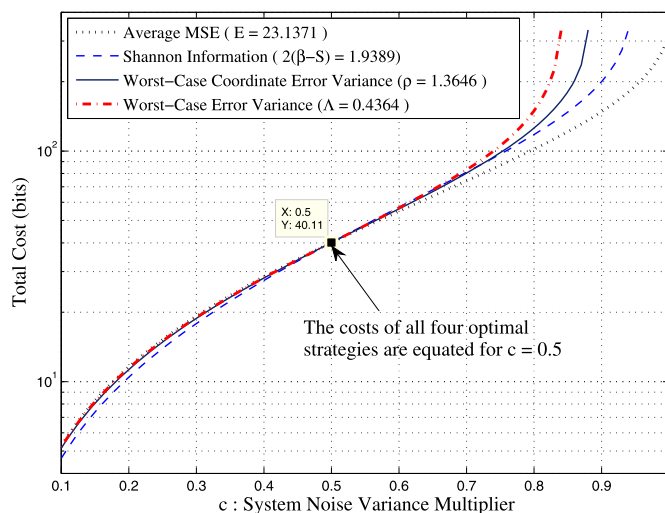
of increasing costs. It is noted that, except for the average MSE criterion, the performance of the remaining three metrics stays in the same order for values of c above and below 0.5. Another important observation is that among the four estimation quality metrics, the performance of the MSE criterion is the one that is least susceptible to changes in the system noise variance. That is, as c is increased beyond 0.5 and decreased below 0.5, the least varying performance metric corresponds to the average MSE criterion. Therefore, in applications where the level of the system noise variance are likely to fluctuate around a nominal value and a predetermined value of the estimation accuracy has to be satisfied, the average MSE criterion provides the most robust alternative in terms of the measurement device selection. However, even in this case, a small change of order 0.01 in the value of the scale parameter disturbs the total cost by more than 1 bit for the average MSE metric.

In the second example, the performances of the estimation quality metrics are equated for $c = 1$, resulting in a total cost score of 320.8. We employ the same constraint value (E = 23.1371) for the average MSE criterion, and the adjustments are applied to the remaining metrics. The corresponding constraint function values are calculated as $2(\beta - S) = 0.66$ for the Shannon information criterion, $\Lambda = 0.3664$ for the worst-case error variance criterion, and $\varrho = 1.5519$ for the worst-case coordinate error variance criterion. The results are illustrated in Fig. 7. In accordance with the observations for high values of c in the previous example, the worst-case error variance metric quickly responds to the drop in the level of the system noise variance values. Hence, the lowest cost is provided by the worst-case error variance criterion for $c < 1$. On the other hand, the optimal cost value for the average MSE criterion exhibits the slowest descent for decreasing values of c. Also noted from the figure is that the performance curve for the Shannon information criterion down-crosses the curve corresponding to the worst-case coordinate error variance criterion at around $c = 0.21$.

5.3. The relationship between the number of effective measurements and the quality of estimation under scaling of the system noise variances

In this experiment we discuss the relationship between the number of effective measurements K_{eff} and various estimation quality metrics under scaling of the system noise variances. A measurement is assessed as effective whenever the cost of that measurement exceeds a certain fraction of the optimal value of the
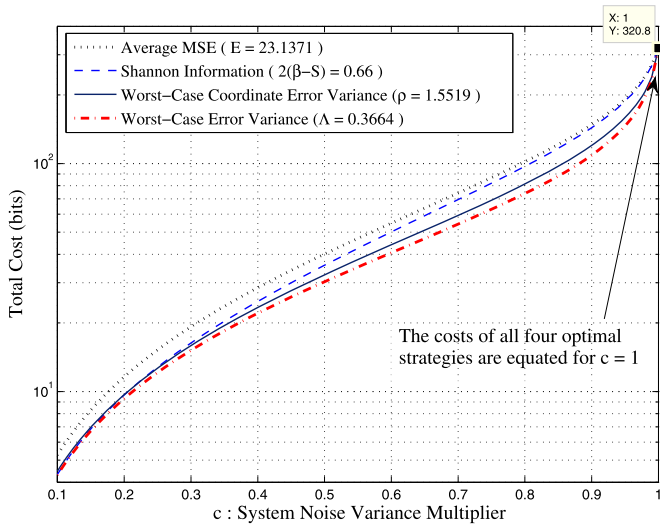
Fig. 7. The performance of various optimal cost allocation strategies under scaling of the system noise variances. All costs are equal for $c = 1$.
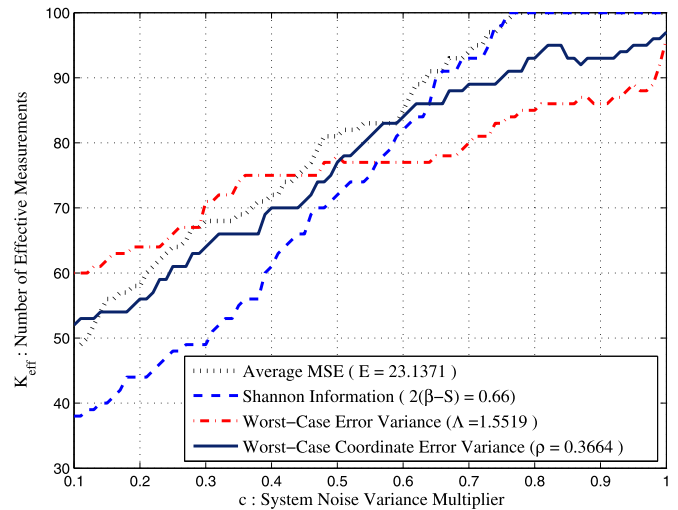

Fig. 8. Number of effective measurements under the scaling of the system noise variances for various estimation accuracy metrics.

total measurement cost. More specifically, we require that $C_i > p(C/K)$ where K represents the total number of measurements. With this construction, it is assured that the total cost of the effective measurements is greater than $(1 - p)C$, from which a suitable value for p can be determined [1]. For small values of p, we can safely assume that the remaining measurements do not cause a significant change on the total cost or provide any significant contribution to the estimation accuracy. Similar to the study in [1], $p = 0.125$ is selected. The same constraint values as in Fig. 7 are employed for the estimation accuracy metrics. Since the performances of all four estimation accuracy criteria are fixed to a high cost score of 320.8 for $c = 1$, it is noted from Fig. 8 that most of the observations are utilized at this value of the scale parameter in order to satisfy the strict constraints. As the average system noise power is reduced by assigning smaller values to the system noise variance multiplier c, the number of effective measurements decreases for all the four cases in accordance with decreasing measurement costs. In other words, lower noise variances result in looser constraints which can be achieved by using fewer number of high resolution (costly) measurements. For small values of c, the worst-case error variance requires the largest number of measurements followed by the average MSE criterion, the worst-case coordinate error variance criterion, and finally the Shannon information criterion. For higher values of c, the situation is reversed apart from the average MSE criterion which requires the largest number of effective measurements. When $c \leqslant 0.56$, a relatively small number of accurate measurements is sufficient to conduct a reliable estimation using the Shannon information criterion with respect to the remaining criteria.

5.4. Effects of system matrix uncertainty

So far, we have assumed that the system matrix is known perfectly at the measurement stage. In this experiment, we consider the case in which the measurement system can only have partial knowledge about the system matrix according to the specific uncertainty model introduced in Section 3. That is, the system matrix is represented as the sum of a known matrix plus a random disturbance matrix $\mathbf{H} \in \mathcal{H} = \{ \bar{\mathbf{H}} + \mathbf{\Delta}: \ \|\mathbf{\Delta}^T\|_2 \leqslant \epsilon \}$, where the degree of uncertainty is controlled with the spectral norm of the disturbance matrix $\mathbf{\Delta}$. Below, we present the results concerning the effects of system uncertainty on the optimal cost allocation problem for
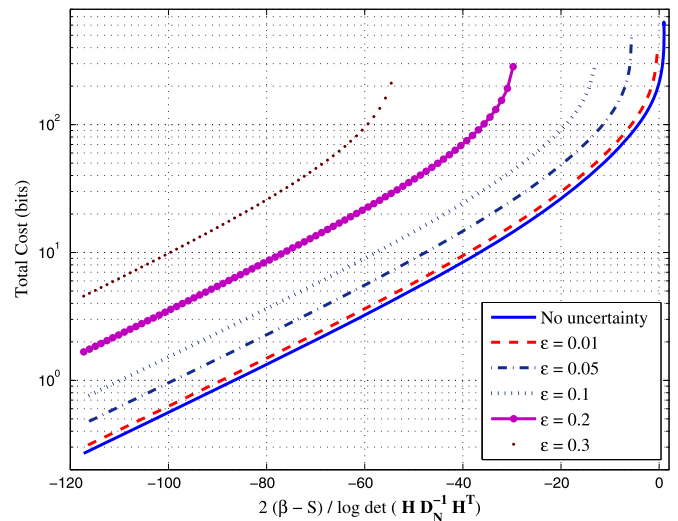

Fig. 9. Effects of system matrix uncertainty on the total measurement cost for Shannon information criterion.

the Shannon information and the worst-case error variance criteria in Fig. 9 and Fig. 10, respectively. For both cases, it is observed that the total cost increases as the amount of uncertainty in the system matrix increases for a given value of the constraint. The increase in the system matrix uncertainty also leads to smaller values of the maximum attainable estimation accuracy measures (the asymptotes where the total cost increases unboundedly).

6. Extension to Bayesian framework

In Section 2, parameter $\boldsymbol{\theta}$ is modeled as a deterministic unknown parameter. Whenever prior information is available about the distribution of the unknown parameter, this additional information can be utilized at the estimation stage. As a result, a more refined metric to assess the quality of the estimator performance is employed which is commonly known as the Bayesian CRB (BCRB) and expressed as follows:

$$\mathbb{E}\big\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\big\} \geqslant (\mathbf{J_D} + \mathbf{J_P})^{-1} \triangleq \text{BCRB}, \qquad (44)$$

where $\mathbf{J_D}$ represents data information matrix and $\mathbf{J_P}$ represents prior information matrix, whose elements are [16]
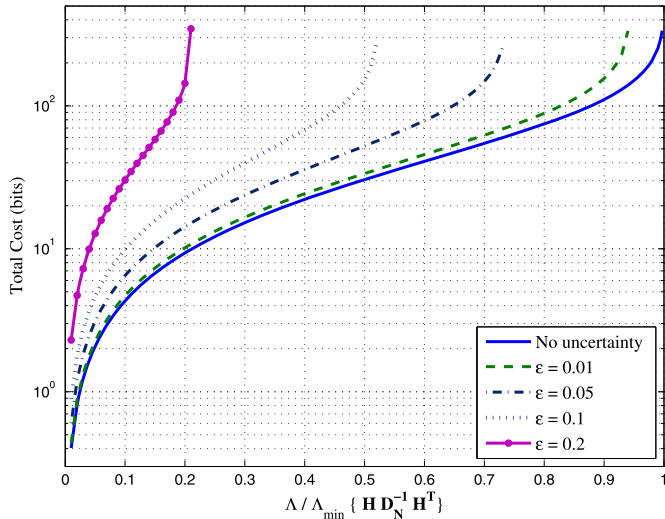
Fig. 10. Effects of system matrix uncertainty on the total measurement cost for worst-case error variance criterion.

$$J_{D_{ij}} = -\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}} \left\{ \frac{\partial^2 \log p_{\mathbf{y}}^{\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_i \partial \theta_j} \right\} = \mathbb{E}_{\boldsymbol{\theta}} \left\{ \mathbf{J}(\mathbf{y}, \boldsymbol{\theta}) \right\} \quad \text{and}$$

$$J_{P_{ij}} = -\mathbb{E}_{\boldsymbol{\theta}} \left\{ \frac{\partial^2 \log w(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}, \tag{45}$$

where $\mathbf{J}(\mathbf{y}, \boldsymbol{\theta})$ is the standard Fisher information matrix defined in (3).

When the prior probability of the parameter is Gaussian with $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, under the same assumptions regarding the independence of $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D_n})$ and $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \mathbf{D_m})$, the BCRB for the linear system given in Fig. 1 can be obtained as

$$\text{BCRB} = \left(\sum_{i=1}^{K} \frac{1}{\sigma_{n_i}^2 + \sigma_{m_i}^2} \mathbf{h}_i \mathbf{h}_i^T + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \right)^{-1}. \tag{46}$$

Correspondingly, the total cost function should be restated to incorporate the change in the variance of the input to each measurement noise device as follows:

$$C = \sum_{i=1}^{K} C_i = \sum_{i=1}^{K} \frac{1}{2} \log_2 \left(1 + \frac{\sigma_{x_i}^2}{\sigma_{m_i}^2} \right), \tag{47}$$

where $\sigma_{x_i}^2$ is the ith diagonal entry of the observation covariance matrix $\text{Cov}(\mathbf{x}) = \mathbf{H}^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{H} + \mathbf{D_n}$.

Based on these expressions, all the proposed cost minimization formulations in Section 2 can be modified accordingly to obtain the optimal cost assignment strategies in the presence of prior information. Specifically, the CRB is replaced with the BCRB, and the cost function stated in (47) is substituted as the objective function inside the optimization problems given in (14), (18), (20), and (22). However, the modified optimization problems are not necessarily convex. It is also noted that the problem formulation constructed by employing the LMMSE estimator in [1] is equivalent to the dual of the Bayesian estimation case under the average MSE criterion given in (11) when Gaussian priors are assumed.

7. Conclusion

In this paper, we have studied the measurement cost minimization problem for a linear system in the presence of Gaussian noise based on the measurement device model introduced in [1]. By considering the nonrandom parameter estimation case, novel convex optimization problems have been obtained under various

estimation accuracy constraints. Uncertainty in the system matrix has been modeled both under general terms and by using a specific uncertainty model. It has been indicated that the convexity properties of the proposed optimization problems are preserved under uncertainty. When the system matrix is invertible, closed form expressions have been presented for two different estimation accuracy metrics which enable a quick assessment of the corresponding cost allocation strategies analytically or via simpler numerical techniques. It has been shown that the prior information can be incorporated into the optimization problems but the resulting problems need no longer be convex. Through numerical examples, the relationships among various criteria have been analyzed in depth.

References

[1] A. Ozcelikkale, H.M. Ozaktas, E. Arikan, Signal recovery with cost-constrained measurements, IEEE Trans. Signal Process. 58 (2010) 3607–3617.

[2] H.V. Poor, An Introduction to Signal Detection and Estimation, Springer-Verlag, New York, 1994.

[3] S. Appadwedula, V.V. Veeravalli, D.L. Jones, Energy-efficient detection in sensor networks, IEEE J. Sel. Areas Commun. 23 (2005) 693–702.

[4] A. Ribeiro, G.B. Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian case, IEEE Trans. Signal Process. 54 (2006) 1131–1143.

[5] J.-J. Xiao, S. Cui, Z.-Q. Luo, A.J. Goldsmith, Power scheduling of universal decentralized estimation in sensor networks, IEEE Trans. Signal Process. 54 (2006) 413–422.

[6] S. Cui, J.-J. Xiao, A.J. Goldsmith, Z.-Q. Luo, H.V. Poor, Estimation diversity and energy efficiency in distributed sensing, IEEE Trans. Signal Process. 55 (2007) 4683–4695.

[7] J. Li, G. AlRegib, Rate-constrained distributed estimation in wireless sensor networks, IEEE Trans. Signal Process. 55 (2007) 1634–1643.

[8] H. Zhang, J. Moura, B. Krogh, Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost, IEEE Trans. Signal Process. 57 (2009) 2383–2395.

[9] G. Thatte, U. Mitra, Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology, IEEE Trans. Signal Process. 56 (2008) 2649–2661.

[10] T.C. Aysal, K.E. Barner, Constrained decentralized estimation over noisy channels for sensor networks, IEEE Trans. Signal Process. 56 (2008) 1398–1410.

[11] S. Joshi, S. Boyd, Sensor selection via convex optimization, IEEE Trans. Signal Process. 57 (2009) 451–462.

[12] M. Lazaro, M. Sanchez-Fernandez, A. Artes-Rodriguez, Optimal sensor selection in binary heterogeneous sensor networks, IEEE Trans. Signal Process. 57 (2009) 1577–1587.

[13] B. Dulek, S. Gezici, Average Fisher information maximisation in presence of cost-constrained measurements, Electron. Lett. 47 (2011) 654–656.

[14] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, Upper Saddle River, NJ, 1993.

[15] M. Hayes, Statistical Digital Signal Processing and Modeling, John Wiley & Sons, 1996.

[16] H.L.V. Trees, Detection, Estimation, and Modulation Theory: Part I, 2nd ed., John Wiley & Sons, New York, NY, 2001.

[17] R. Zamir, A proof of the Fisher information inequality via a data processing argument, IEEE Trans. Inform. Theory 44 (1998) 1246–1250.

[18] A. Dembo, T.M. Cover, J.A. Thomas, Information theoretic inequalities, IEEE Trans. Inform. Theory 37 (1991) 1501–1518.

[19] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.

[20] Y. Eldar, A. Ben-Tal, A. Nemirovski, Robust mean-squared error estimation in the presence of model uncertainties, IEEE Trans. Signal Process. 53 (2005) 168–181.

[21] Z. Ben-Haim, Y.C. Eldar, Maximum set estimators with bounded estimation error, IEEE Trans. Signal Process. 53 (2005) 3172–3182.

[22] A. Das, D. Kempe, Sensor selection for minimizing worst-case prediction error, in: Int. Conf. Inform. Process. Sensor Networks (IPSN'08), pp. 97–108.

[23] R. Hettich, K. Kortanek, Semi-infinite programming: Theory, methods, and applications, SIAM Rev. 35 (1993) 380–429.

[24] A. Mutapcic, S. Boyd, Cutting-set methods for robust convex optimization with pessimizing oracles, Optim. Methods Softw. 24 (2009) 381–406.

[25] Z. Quan Luo, J.F. Sturm, S. Zhang, Multivariate nonnegative quadratic mappings, SIAM J. Optim. 14 (2002) 1140–1162.

[26] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1968.

[27] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 1.21, http://cvxr.com/cvx, 2011.

Berkan Dulek received the B.S. and M.S. degrees with high honors in electrical engineering from Bilkent University, Turkey, in 2003 and 2006, respectively. He is currently studying toward the Ph.D. degree at Bilkent University. His research interests are in statistical signal processing and communications with emphasis on stochastic signaling, randomized detection and estimation under cost constraints.

Sinan Gezici received the B.S. degree from Bilkent University, Turkey in 2001, and the Ph.D. degree in electrical engineering from Princeton University in 2006. From 2006 to 2007, he worked at Mitsubishi Electric Research Laboratories, Cambridge, MA. Since February 2007, he has been an Assistant Professor in the Department of Electrical and Electronics Engineering at Bilkent University. Dr. Gezici's research interests are in the areas of detection and estimation theory, wireless communications, and localization systems. Among his publications in these areas is the book *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols* (Cambridge University Press, 2008).