CCIT Report #550
August 2005

Blind Minimax Estimation

Zvika Ben-Haim* Yonina C. Eldar

Abstract

We consider the linear regression problem of estimating an unknown, deterministic parameter vector based on measurements corrupted by colored Gaussian noise. We present and analyze blind minimax estimators (BMEs), which consist of a minimax estimator whose parameter set is itself estimated from measurements. This approach results in several extensions of the James-Stein estimator; however, unlike the James-Stein result, some of these extensions are non-shrinkage, making them applicable in a wider range of problem settings. We demonstrate analytically that the BMEs strictly dominate the least-squares estimator, i.e., they achieve lower mean-squared error for any value of the parameter vector. We also show through simulation that the BMEs generally outperform other extensions of the James-Stein technique.

EDICS Keyword: SSP-PARE

The authors are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: zvikabh@technion.ac.il, yonina@ee.technion.ac.il; phone +972-3-741-5291; fax +972-4-829-5757.
This work was supported by the 6th EU framework program, via the NEWCOM network of excellence, and by the Israel Science Foundation.

I. Introduction

The problem of estimating a parameter vector from noisy measurements has countless applications in science and engineering. Such estimation problems are typically modelled either in a Bayesian setting, in which a prior distribution on the parameter is assumed, or in a deterministic setting, in which no prior exists [1]. This paper examines the deterministic estimation problem. We further assume that the measurements $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ are linear combinations of the parameter vector $\mathbf{x}$, to which Gaussian noise $\mathbf{w}$ is added. Here the transformation matrix $\mathbf{H}$ and the noise covariance $E\{\mathbf{w}\mathbf{w}^*\}$ are assumed to be known. We seek an estimate $\hat{\mathbf{x}}$ which approximates $\mathbf{x}$ in the sense of minimal mean-squared error (MSE).

This ubiquitous problem was first addressed by Gauss [2], who proposed the classical *least-squares* (LS) estimator. Several lines of reasoning can be used to support the LS approach. Gauss' argument showed that the LS estimator minimizes the squared error between the measurements $\mathbf{y}$ and the transformed estimate $\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}}$. It is also well-known that the LS estimator is the maximum likelihood estimator for Gaussian noise. However, neither of these criteria are directly related to the MSE, or to any other measure of the distance between $\mathbf{x}$ and $\hat{\mathbf{x}}$. Another property of the LS estimator is that it is the linear, unbiased estimator achieving minimal MSE. Yet by removing the requirements of unbiasedness and linearity, estimators yielding lower MSE can be constructed. While linearity and unbiasedness may be intuitively appealing properties, they have no relation to the primary goal at hand, namely, achieving low estimation error.

Because the parameter vector $\mathbf{x}$ is deterministic, the MSE $E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\}$ is generally a function of $\mathbf{x}$. As a result, one estimator may be better than another for some values of $\mathbf{x}$, and worse for other values of $\mathbf{x}$. For instance, the trivial estimator $\hat{\mathbf{x}} = \mathbf{0}$ achieves optimal MSE when $\mathbf{x} = \mathbf{0}$, but it is a very poor estimator for other values of $\mathbf{x}$. Nonetheless, it is possible to impose a partial order among estimators [3], as follows. An estimator $\hat{\mathbf{x}}_1$ is said to *strictly dominate* a different estimator $\hat{\mathbf{x}}_2$ if the MSE of $\hat{\mathbf{x}}_1$ is lower than that of $\hat{\mathbf{x}}_2$, for all values of $\mathbf{x}$. If the MSE of $\hat{\mathbf{x}}_1$ is never higher than that of $\hat{\mathbf{x}}_2$, and is strictly lower for at least one parameter value, then $\hat{\mathbf{x}}_1$ is said to *dominate* $\hat{\mathbf{x}}_2$. An estimator is said to be *admissible* if it is not dominated by any other estimator. Surprisingly, the LS estimator turns out to be inadmissible, i.e., some estimators *always* achieve lower MSE [4]. Thus, it is of interest to characterize the class of admissible estimators, and to find estimators which dominate the LS estimator.

The study of admissibility is sometimes restricted to the set of linear estimators, which have the form $\hat{\mathbf{x}} = \mathbf{Gy}$. A linear admissible estimator is one which is not dominated by any other linear estimator. A simple rule characterizes the class of such estimators [5], and, given any linear inadmissible estimator, it is possible to construct a linear admissible alternative which dominates it [6]. However, the problem of admissibility is considerably more intricate when the linearity restriction is removed; generally, admissible estimators are either trivial (e.g., $\hat{\mathbf{x}} = \mathbf{0}$) or exceedingly complex [7], [8]. As a result, much research has focused on finding simple nonlinear techniques which dominate the LS estimator.

Early work on LS-dominating estimators considered the independent, identical-distribution (i.i.d.) case, for which $\mathbf{H} = \mathbf{I}$ and the noise is white. Among these, the James-Stein estimator [3], [9] is the best-known example; another is the Alam-Thompson estimator [10], [11]. Various extensions of the James-Stein estimator were later constructed for the general (non-i.i.d.) case [12]–[15]. Of these, Bock's estimator [13] is quoted most often [15], [16]. However, none of these approaches has become a standard alternative to the LS estimator, and they are rarely used in signal processing [15]. One reason for this is that the estimators are poorly justified and seem counterintuitive, and as such they are sometimes regarded with skepticism (see discussion following [17]). Another reason is that many of these approaches (including the James-Stein estimator and Bock's extension) result in shrinkage estimators, consisting of a gain factor multiplying the LS estimate. While this can certainly be used to reduce MSE, such estimators are inappropriate for some applications, in which a gain factor has no effect on final estimation quality.

In this paper, we provide a framework for generating a wide class of low-complexity, LS-dominating estimators, which are constructed from a simple, intuitive principle, called the blind minimax approach [18], [19]. This approach is used as a basis for selecting and generating estimators tailored for given estimation problems. Many blind minimax estimators (BMEs) reduce to Stein-type estimators in the i.i.d. case, and they continue to dominate the LS estimator in the general, non-i.i.d. case as well. Thus, we show analytically that the proposed estimators *always* achieve lower MSE than the LS estimator. Unlike Bock's estimator, BMEs may be constructed so that they are non-shrinkage, if this is required. Furthermore, extensive simulations show that BMEs considerably outperform Bock's estimator.

Blind minimax estimators are based on bounded parameter set minimax estimators [20], [21]. These are estimators designed for a slightly different problem, in which the parameter is known to lie within a given set. Such estimators have been studied extensively and a closed-form estimator is known for many types of parameter sets. In our case, however, no prior information about the parameter set is assumed. Instead, the blind minimax approach makes use of a two-stage estimation process. In the first stage, the parameter set is estimated from the measurements. In the second stage, a minimax estimator for this parameter set is used to estimate the parameter itself. The result may be viewed as a simple estimator, independent of this two-stage construction process. Indeed, our LS-dominance proofs are independent of the method by which the estimators are generated. In particular, the dominance results do not depend on the parameter actually lying within the estimated parameter set. However, the blind minimax technique provides a framework whereby many different estimators can be generated, and provides insight into the mechanism by which these estimators outperform the LS estimator.

Blind minimax estimators differ in the method by which the parameter set is estimated. In Section II, we study the case in which the estimated parameter set is a sphere; Section III derives estimators based on an ellipsoidal parameter set. The different estimators are compared in a numerical study in Section IV, and the paper concludes with a discussion in Section V.

Throughout this paper, vectors are denoted by lowercase boldface letters, while matrices are denoted by uppercase boldface letters. The ith component of a vector $\mathbf{v}$ is denoted v_i. The (unique) positive semidefinite square root of a positive semidefinite matrix $\mathbf{Q}$ is denoted $\mathbf{Q}^{1/2}$. The notation $\tilde{\mathbf{u}} \sim N_p(\mathbf{u}, \mathbf{Q})$ indicates that $\tilde{\mathbf{u}}$ is a random vector of length p, distributed normally with mean $\mathbf{u}$ and covariance $\mathbf{Q}$.

II. The Spherical Blind Minimax Estimator

Consider the problem of estimating an unknown deterministic parameter vector $\mathbf{x} \in \mathbb{C}^m$ from measurements $\mathbf{y} \in \mathbb{C}^n$ given by

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

where $\mathbf{H} \in \mathbb{C}^{n \times m}$ is a known matrix and $\mathbf{w}$ is a Gaussian random vector with zero mean and covariance $\mathbf{C_w}$. For simplicity, we assume that $\mathbf{H}$ is full-rank and that $\mathbf{C_w}$ is positive definite.

Suppose first that $\mathbf{x}$ is known to lie within a compact parameter set $\mathcal{S}$. In this case, a bounded parameter set minimax estimator may be constructed [20], [21]. This is the linear estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ minimizing the worst-case MSE among all possible values of $\mathbf{x}$ in $\mathcal{S}$,

$$\hat{\mathbf{x}}_{\mathrm{M}} = \arg \min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{x}\in\mathcal{S}} E\left\{ \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \right\}. \tag{2}$$

For example, when the set $\mathcal{S}$ is a sphere centered at the origin, $\mathcal{S} = \{\mathbf{x} : \|\mathbf{x}\|^2 \leq L^2\}$, the linear minimax estimator is [21]

$$\hat{\mathbf{x}}_{\mathrm{M}} = \frac{L^2}{L^2 + \epsilon_0}\hat{\mathbf{x}}_{\mathrm{LS}}, \tag{3}$$

where $\hat{\mathbf{x}}_{\mathrm{LS}}$ is the LS estimator,

$$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{y}, \tag{4}$$

ϵ_0 is the MSE of $\hat{\mathbf{x}}_{\mathrm{LS}}$, given by

$$\epsilon_0 = E\left\{\|\hat{\mathbf{x}}_{\mathrm{LS}} - \mathbf{x}\|^2\right\} = \mathrm{Tr}(\mathbf{Q}^{-1}), \tag{5}$$

and

$$\mathbf{Q} = \mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}. \tag{6}$$

It has recently been shown that any linear minimax estimator achieves lower MSE than that of the LS estimator, for all values of $\mathbf{x}$ in $\mathcal{S}$ [18]. Thus, as long as *some* bounded parameter set is known to contain $\mathbf{x}$, minimax estimators outperform the LS estimator.

The blind minimax approach uses minimax estimators when no parameter set is known. This is done in a two-stage process:

1) A parameter set $\mathcal{S}$ is estimated from the measurements;
2) A minimax estimator designed for $\mathcal{S}$ is used to estimate the parameter vector $\mathbf{x}$.

Blind minimax estimators differ in the method by which the parameter set is estimated. In this section, we use a spherical parameter set, and estimate the sphere radius from the measurements. We assume for now that the sphere is centered on the origin. The resulting *spherical BME* (SBME) will have the form (3), where L^2 is estimated from the measurements.

As an estimate of L^2, we seek a value as close as possible to $\|\mathbf{x}\|^2$: a smaller value would exclude the true vector $\mathbf{x}$ from the parameter set, while a larger value would yield an overly conservative estimator. Since $\mathbf{x}$ is unknown, a natural alternative is to use $\hat{\mathbf{x}}_{\mathrm{LS}}$ instead; for instance, one may estimate L^2 as $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$. This results in the *direct SBME*

$$\hat{\mathbf{x}}_{\mathrm{DSBM}} = \frac{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2}{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2 + \epsilon_0}\hat{\mathbf{x}}_{\mathrm{LS}}. \tag{7}$$

In fact, $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$ tends to be an overestimate of $\|\mathbf{x}\|^2$, since $E\{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2\} = \|\mathbf{x}\|^2 + \epsilon_0$. To correct for this effect, we may alternatively estimate L^2 as $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2 - \epsilon_0$. Substituting this value into (3), the *balanced SBME* is given by

$$\hat{\mathbf{x}}_{\mathrm{BSBM}} = \left(1 - \frac{\epsilon_0}{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2}\right)\hat{\mathbf{x}}_{\mathrm{LS}}. \tag{8}$$

Both SBMEs reduce to well-known results in the i.i.d. case: the balanced version $\hat{\mathbf{x}}_{\mathrm{BSBM}}$ reduces to Stein's approach [4], and the direct version $\hat{\mathbf{x}}_{\mathrm{DSBM}}$ reduces to a variant of the Alam-Thompson estimator [10], [11]. Under suitable conditions, the techniques proposed by Stein and by Alam and Thompson are known to strictly dominate the LS estimator, meaning that they achieve lower MSE for all values of $\mathbf{x}$. However, the SBMEs are equally well-defined for the non-i.i.d. case. Furthermore, as we shall see, both versions of the SBME strictly dominate the LS estimator in the non-i.i.d. case.

Up to this point, we have arbitrarily chosen the parameter set to be centered on the origin. The result was a weighted average between the LS estimate and the origin. The weight given to the LS estimate may be viewed as a restraint, which lessens the effect of measurement noise. As we shall see, the proposed BMEs outperform the LS estimator, illustrating the fact that the LS estimator is an overestimate. However, the choice of a parameter set centered on the origin is completely arbitrary; BMEs may be constructed around any constant center point $\mathbf{x}_0$. This would result in a weighted average between the LS estimator and $\mathbf{x}_0$, which may be useful if the parameter vector is expected to lie near a particular point. For example, the "off-center" balanced SBME is given by

$$\hat{\mathbf{x}} = \left(1 - \frac{\epsilon_0}{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2}\right)\hat{\mathbf{x}}_{\mathrm{LS}} + \left(\frac{\epsilon_0}{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2}\right)\mathbf{x}_0. \tag{9}$$

All dominance results continue to hold for the off-center estimators as well. In the sequel, we assume $\mathbf{x}_0 = \mathbf{0}$ merely for the sake of notational simplicity.

The following theorem demonstrates that the SBMEs are guaranteed to outperform the LS estimator in terms of MSE.

Theorem 1: Suppose $\epsilon_0/\epsilon_{\max} > 4$, where ϵ_0 is given by (5), $\epsilon_{\max}$ is the largest eigenvalue of $\mathbf{Q}^{-1}$, and $\mathbf{Q}$ is given by (6). Then, both the direct SBME and the balanced SBME strictly dominate the LS estimator.

The value $\epsilon_0/\epsilon_{\max}$ is known as the effective dimension, and may be roughly described as the number of independent measurements in the system. In the i.i.d. case, for example, the effective dimension simply equals the length of the measurement vector.

Note that both SBMEs can be written as

$$\hat{\mathbf{x}}_{\mathrm{SBM}} = \left(1 - \frac{\epsilon_0}{b + \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}}\right)\hat{\mathbf{x}}_{\mathrm{LS}}, \tag{10}$$

where $b = \epsilon_0$ for the direct SBME and $b = 0$ for the balanced SBME. Thus, rather than proving Theorem 1, we prove the following, more general proposition.

Proposition 1: Suppose $\epsilon_0/\epsilon_{\max} > 4$, where ϵ_0 is given by (5), $\epsilon_{\max}$ is the largest eigenvalue of $\mathbf{Q}^{-1}$, and $\mathbf{Q}$ is given by (6). Then, the estimator (10) dominates the LS estimator, for any $b \geq 0$.

The proof of Proposition 1 makes use of the following result, known as Stein's lemma [3, Theorem 1.5.15].

Lemma 1 (Stein): Let $\hat{\mathbf{v}} \sim N_p(\mathbf{v}, \mathbf{I})$, and let $g(\hat{\mathbf{v}})$ be a differentiable function such that $E\left\{\left|\frac{\partial g(\hat{\mathbf{v}})}{\partial \hat{v}_i}\right|\right\} < \infty$ for all i. Then,

$$E\left\{\frac{\partial g(\hat{\mathbf{v}})}{\partial \hat{v}_i}\right\} = -E\{g(\hat{\mathbf{v}})(v_i - \hat{v}_i)\}\,. \tag{11}$$

Proof of Proposition 1. We will prove that $\hat{\mathbf{x}}_{\mathrm{SBM}}$ strictly dominates $\hat{\mathbf{x}}_{\mathrm{LS}}$ for any $b \geq 0$. First, note that the MSE of $\hat{\mathbf{x}}_{\mathrm{SBM}}$ is given by

$$R(\hat{\mathbf{x}}_{\mathrm{SBM}}) = E\left\{\|\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{SBM}}\|^2\right\} = \epsilon_0 + E\left\{\frac{\epsilon_0^2 \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}}{(b + \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}})^2}\right\} + 2E\left\{\frac{\epsilon_0}{b + \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}} \hat{\mathbf{x}}_{\mathrm{LS}}^* (\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{LS}})\right\}. \tag{12}$$

Let $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$ be the eigenvalue decomposition of $\mathbf{Q}$, such that $\mathbf{V}$ is unitary and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots \lambda_m)$. Define $\hat{\mathbf{v}} = \mathbf{V}^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\mathrm{LS}}$ and $\mathbf{v} = \mathbf{V}^* \mathbf{Q}^{1/2} \mathbf{x}$. Note the following relations between $\hat{\mathbf{v}}$ and $\hat{\mathbf{x}}_{\mathrm{LS}}$:

$$\begin{aligned}
\hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \hat{\mathbf{v}} &= \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}, \\
\hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \mathbf{v} &= \hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{x}, \\
\hat{\mathbf{v}}^* \mathbf{\Lambda}^{-2} \hat{\mathbf{v}} &= \hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{Q}^{-1} \hat{\mathbf{x}}_{\mathrm{LS}}.
\end{aligned} \tag{13}$$

Using these properties, we now evaluate the third term in (12), obtaining

$$\begin{aligned}
E\left\{\frac{\epsilon_0}{b + \hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}} \hat{\mathbf{x}}_{\mathrm{LS}}^* (\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{LS}})\right\} &= E\left\{\frac{\epsilon_0}{b + \hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \hat{\mathbf{v}}} \hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} (\mathbf{v} - \hat{\mathbf{v}})\right\} \\
&= E\left\{\frac{\epsilon_0}{b + \hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \hat{\mathbf{v}}} \sum_{i=1}^{p} \lambda_i^{-1} \hat{v}_i (v_i - \hat{v}_i)\right\} \\
&= \epsilon_0 \sum_{i=1}^{p} \lambda_i^{-1} E\left\{\frac{\hat{v}_i (v_i - \hat{v}_i)}{b + \hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \hat{\mathbf{v}}}\right\}.
\end{aligned} \tag{14}$$

Let

$$g_i(\hat{\mathbf{v}}) \triangleq \frac{\hat{v}_i}{b + \hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1} \hat{\mathbf{v}}}, \tag{15}$$

and note that $\hat{\mathbf{v}}$ is distributed normally with mean $\mathbf{v}$ and covariance $\mathbf{I}$. We can thus apply

Lemma 1 to obtain

$$E\left\{\frac{\epsilon_0}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\hat{\mathbf{x}}_{\text{LS}}^*(\mathbf{x}-\hat{\mathbf{x}}_{\text{LS}})\right\} = -\epsilon_0\sum_i \lambda_i^{-1}E\left\{\frac{1}{b+\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}} - 2\frac{\lambda_i^{-1}\hat{v}_i^2}{(b+\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}})^2}\right\}$$
$$= -\epsilon_0 E\left\{\frac{\text{Tr}(\mathbf{\Lambda}^{-1})}{b+\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}}\right\} + 2\epsilon_0 E\left\{\frac{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-2}\hat{\mathbf{v}}}{(b+\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}})^2}\right\}$$
$$= -\epsilon_0 E\left\{\frac{\text{Tr}(\mathbf{Q}^{-1})}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\right\} + 2\epsilon_0 E\left\{\frac{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\text{LS}}}{(b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}})^2}\right\}. \quad (16)$$

Substituting this result back into (12), we have

$$R(\hat{\mathbf{x}}_{\text{SBM}}) = \epsilon_0 + E\left\{\frac{\epsilon_0}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\left(\epsilon_0\frac{\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}} - 2\epsilon_0 + 4\frac{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\text{LS}}}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\right)\right\}. \quad (17)$$

Since $b \geq 0$,

$$R(\hat{\mathbf{x}}_{\text{SBM}}) \leq \epsilon_0 + E\left\{\frac{\epsilon_0}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\left(\frac{\epsilon_0\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}{\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}} - 2\epsilon_0 + 4\frac{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\text{LS}}}{\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}\right)\right\}$$
$$\leq \epsilon_0 + E\left\{\frac{\epsilon_0}{b+\hat{\mathbf{x}}_{\text{LS}}^*\hat{\mathbf{x}}_{\text{LS}}}(-\epsilon_0 + 4\epsilon_{\max})\right\}. \quad (18)$$

If $\epsilon_0 > 4\epsilon_{\max}$, then the expectation is taken over a strictly negative range, and hence $R(\hat{\mathbf{x}}_{\text{SBM}})$ is always lower than ϵ_0, so that $\hat{\mathbf{x}}_{\text{SBM}}$ strictly dominates $\hat{\mathbf{x}}_{\text{LS}}$. ■

As we have shown, in terms of MSE, both direct and balanced SBMEs outperform the LS estimator, providing us with a first example of the power of blind minimax estimation.

Both SBMEs are shrinkage estimators, i.e., they consist of the LS estimator multiplied by a gain factor smaller than one. The SBMEs thus illustrate the fact that the LS estimator tends to be an overestimate, and shrinkage can improve its performance. However, in some applications, such as image reconstruction, a gain factor has no effect on the end result. In the next section, we use the blind minimax approach to develop a non-shrinkage estimator, which also dominates the LS estimator.

III. The Ellipsoidal Blind Minimax Estimator

Occasionally, one must use the MSE in place of a more accurate error measure, which may be overly complex or even subjective. For example, in communication systems, one is interested in minimizing the bit error rate (BER). However, BER minimization techniques are generally too computationally expensive to be practical. In many situations, one chooses instead to minimize the MSE between the transmitted and reconstructed symbols [22]. This is done in the hope that low MSE provides low BER. However, MSE and BER are not always directly related. In the previous section, it was shown that SBMEs achieve lower MSE than the LS estimator. Yet in a binary communication system, only the signs of the estimated elements are used to evaluate the received bits. Since the SBMEs are shrinkage estimators, they yield the same sign as the LS estimator. Thus, in a binary communication system, the BER obtained by SBMEs is identical to the BER of the LS estimator.

Shrinkage estimators are therefore not applicable to *all* estimation problems. In this section, we develop two non-shrinkage estimators by considering ellipsoidal, rather than spherical, parameter sets as the basis for the blind minimax estimation technique. In some situations, these estimators also outperform the SBMEs in terms of MSE.

Not all elements of the least-squares estimate $\hat{\mathbf{x}}_{\mathrm{LS}}$ are equally trustworthy. Rather, $\hat{\mathbf{x}}_{\mathrm{LS}}$ is a Gaussian random vector with mean $\mathbf{x}$ and covariance $\mathbf{Q}^{-1} = \left(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1}$. Thus, some elements in $\hat{\mathbf{x}}_{\mathrm{LS}}$ have lower variance than others. In this sense, the scalar shrinkage factor of the SBME (10) and other extended Stein estimators [13] seems inadequate. Indeed, several researchers have proposed shrinking each measurement separately according to its variance [12], [14]. Ironically, however, there has been disagreement as to whether high-variance components should be shrunk more [12] or less [14], and little justification has been given to either choice.

The blind minimax approach provides a natural framework for resolving these disputes. To see this, note that $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{x} + \mathbf{u}$, where $\mathbf{u} \sim N_m(\mathbf{0}, \mathbf{Q}^{-1})$. The SBME was constructed by using $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$ as an estimate for $\|\mathbf{x}\|^2$. However, since the noise $\mathbf{u}$ is colored, it is sensible to first whiten the noise by writing

$$\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{Q}^{1/2}\mathbf{x} + \tilde{\mathbf{u}}, \tag{19}$$

where $\tilde{\mathbf{u}} \sim N_m(\mathbf{0}, \mathbf{I})$. One may then estimate $\|\mathbf{Q}^{1/2}\mathbf{x}\|^2$ using $\|\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$, obtaining the *ellipsoidal BME* (EBME). Such an estimate can be readily incorporated into the blind minimax framework by using an ellipsoidal parameter set, $\mathcal{S} = \{\mathbf{x} : \mathbf{x}^*\mathbf{Q}\mathbf{x} \le L^2\}$, rather than the spherical parameter set of the SBME. The ellipsoidal parameter set is elongated in directions of low noise, resulting in lower shrinkage for those directions. In the i.i.d. case, $\mathbf{Q} = \mathbf{I}$, and the estimator reduces to the SBME.

As with the construction of the SBME, one may estimate L^2 in two ways. The *direct EBME* is obtained by substituting $L^2 = \hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}$. The *balanced EBME* is obtained by observing that $E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}\} = \mathbf{x}^*\mathbf{Q}\mathbf{x} + m$, and hence substituting $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} - m$ for L^2.

A closed form for minimax estimators of an ellipsoidal parameter set was developed in [21].

By substituting the value of L^2 into this closed form, we obtain the following expression for the EBMEs.

Proposition 2 (Closed-Form EBME): Let $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^*$ be the eigenvalue decomposition of $\mathbf{Q}$, so that $\mathbf{V}$ is unitary, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots \lambda_m)$, and $\lambda_1 \geq \cdots \geq \lambda_m$. The direct and balanced EBMEs are then given by

$$\hat{\mathbf{x}}_{\mathrm{EBM}} = \mathbf{V}\,\mathrm{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})\mathbf{V}^* \left(\mathbf{I} - \alpha \mathbf{Q}^{1/2}\right) \hat{\mathbf{x}}_{\mathrm{LS}}, \tag{20}$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m} \lambda_i^{-1/2}}{\hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{Q} \hat{\mathbf{x}}_{\mathrm{LS}} + b - k}. \tag{21}$$

Here, k is the smallest integer $0 \leq k \leq m-1$ such that $\alpha < \lambda_{k+1}^{-1/2}$, and b is a constant which equals 0 for the balanced EBME and m for the direct EBME.

Proof: Follows from Proposition 1 of [21]. ■

While the closed form of the EBMEs appears somewhat more intimidating than that of the SBMEs, the computational complexities of all estimators are comparable. The major difference is the calculation of the value k, for which m divisions are required. Like the SBMEs, the EBMEs also dominate the LS estimator under suitable conditions, as shown in the following theorem.

Theorem 2: Suppose $\mathrm{Tr}(\mathbf{Q}^{-1/2}) > 4\epsilon_{\max}^{1/2}$, where $\epsilon_{\max}^{1/2}$ is the largest eigenvalue of $\mathbf{Q}^{-1/2}$. Then, both the balanced and the direct EBME strictly dominate $\hat{\mathbf{x}}_{\mathrm{LS}}$.

The proof of Theorem 2 is based on an analogy between the diagonal matrix $\mathrm{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})$ in (20) and Baranchik's positive-part modification [3], [23] of the James-Stein estimator. Baranchik proposed using a shrinkage factor of 0 whenever the James-Stein estimator uses negative shrinkage, and showed that the resulting *positive-part estimator* dominates the James-Stein estimator. Although the EBME is not a shrinkage estimator, it resembles Baranchik's modification. To see this, consider the estimator $\hat{\mathbf{x}}_0$ obtained by removing the term $\mathrm{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})$ from (20),

$$\begin{aligned} \hat{\mathbf{x}}_0 &= (\mathbf{I} - \alpha \mathbf{Q}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}} \\ &= \mathbf{V}\,\mathrm{diag}\left(1 - \alpha\lambda_1^{1/2}, \cdots 1 - \alpha\lambda_m^{1/2}\right) \mathbf{V}^* \hat{\mathbf{x}}_{\mathrm{LS}}. \end{aligned} \tag{22}$$

Since $\alpha \geq \lambda_i^{-1/2}$ for all $i \leq k$, this would introduce negative componentwise shrinkage for the first k eigenvectors of $\mathbf{V}$. The following proposition, which is a generalization of Baranchik's result, shows that the MSE can be reduced by eliminating this negative shrinkage.

Proposition 3: Let $\hat{\mathbf{x}}$ be any estimator of the form $\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^*\hat{\mathbf{x}}_{\mathrm{LS}}$, where $\mathbf{D}$ is a diagonal matrix, whose diagonal elements d_i are functions of the random variable $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}$. Suppose at least one of the elements d_i is negative with nonzero probability. Then, $\hat{\mathbf{x}}$ is dominated by the (generalized) positive-part estimator

$$\hat{\mathbf{x}}_+ = \mathbf{V}\mathbf{D}_+\mathbf{V}^*\hat{\mathbf{x}}_{\text{LS}}, \tag{23}$$

where $\mathbf{D}_+$ is a diagonal matrix with diagonal elements $d_{i+} = \max(0, d_i)$.

Proof: We will show that $\text{MSE}(\hat{\mathbf{x}}) - \text{MSE}(\hat{\mathbf{x}}_+)$ is nonnegative for all $\mathbf{x}$, and positive for any value of $\mathbf{x}$ whose elements are all nonzero.

$$\begin{aligned}\text{MSE}(\hat{\mathbf{x}}) - \text{MSE}(\hat{\mathbf{x}}_+) &= E\left\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right\} - E\left\{\|\hat{\mathbf{x}}_+ - \mathbf{x}\|^2\right\} \\ &= E\left\{\|\hat{\mathbf{x}}\|^2 - \|\hat{\mathbf{x}}_+\|^2\right\} - 2E\{\hat{\mathbf{x}}^*\mathbf{x} - \hat{\mathbf{x}}_+^*\mathbf{x}\} \\ &= E\left\{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{V}(\mathbf{D}^2 - \mathbf{D}_+^2)\mathbf{V}^*\hat{\mathbf{x}}_{\text{LS}}\right\} \\ &\quad - 2E\{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}.\end{aligned} \tag{24}$$

Since $d_i^2 - d_{i+}^2 \geq 0$ for all i, the first term in (24) is nonnegative. Hence, to prove the proposition, it suffices to show that $E\{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}$ is nonpositive for all $\mathbf{x}$, and negative for values $\mathbf{x}$ with nonzero elements.

To this end, define $\mathbf{z} = \mathbf{V}^*\mathbf{x}$ and $\hat{\mathbf{z}} = \mathbf{V}^*\hat{\mathbf{x}}_{\text{LS}}$. We note that $\hat{\mathbf{z}} \sim N_m(\mathbf{z}, \mathbf{\Lambda}^{-1})$, so that the elements of $\hat{\mathbf{z}}$ are statistically independent. To calculate $E\{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}$, we condition on $\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\text{LS}}$, obtaining

$$E\{\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\} = E\{E\{\hat{\mathbf{z}}^*(\mathbf{D} - \mathbf{D}_+)\mathbf{z}|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}\}\} = E\left\{\sum_{i=1}^{m}(d_i - d_{i+})E\{\hat{z}_i z_i|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}\}\right\}, \tag{25}$$

where we used the fact that $\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\text{LS}} = \hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}$, and that d_i and d_{i+} are deterministic when conditioned on $\hat{\mathbf{x}}_{\text{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\text{LS}}$.

We now define

$$r_i(\hat{\mathbf{z}}) \triangleq \sqrt{\frac{\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}} - \sum_{j\neq i}\lambda_j\hat{z}_j^2}{\lambda_i}}, \tag{26}$$

and note that $\hat{z}_i = \text{sgn}(\hat{z}_i)r_i(\hat{\mathbf{z}})$. For each i, we further condition on all values $\{\hat{z}_j\}_{j\neq i}$, to obtain

$$E\{\hat{z}_i z_i|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}, \{\hat{z}_j\}_{j\neq i}\} = z_i r_i(\hat{\mathbf{z}})E\{\text{sgn}(\hat{z}_i)|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}, \{\hat{z}_j\}_{j\neq i}\}. \tag{27}$$

Since $\hat{z}_i$ is independent of $\{\hat{z}_j\}_{j\neq i}$, it follows that $\text{sgn}(\hat{z}_i)$ is jointly independent of $\{\hat{z}_j\}_{j\neq i}$ and $\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}$. Thus

$$E\{\text{sgn}(\hat{z}_i)|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}, \{\hat{z}_j\}_{j\neq i}\} = E\{\text{sgn}(\hat{z}_i)\}. \tag{28}$$

Combining this result with (25) and (27), we obtain

$$E\{\hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{V}(\mathbf{D}-\mathbf{D}_+)\mathbf{V}^*\mathbf{x}\} = E\left\{\sum_{i=1}^{m}(d_i - d_{i+})r_i(\hat{\mathbf{z}})|z_i|\operatorname{sgn}(z_i)E\{\operatorname{sgn}(\hat{z}_i)\}\right\}. \tag{29}$$

When $z_i = 0$, the ith term in the sum above equals 0. In all other cases, we use the fact that $\hat{z}_i$ is Gaussian with mean z_i to obtain

$$\Pr\{\operatorname{sgn}(\hat{z}_i) = \operatorname{sgn}(z_i)\} > \Pr\{\operatorname{sgn}(\hat{z}_i) \neq \operatorname{sgn}(z_i)\}. \tag{30}$$

Thus, $|z_i|\operatorname{sgn}(z_i)E\{\operatorname{sgn}(\hat{z}_i)\}$ is positive if $z_i \neq 0$, and equals zero if $z_i = 0$. It follows that all terms in (29) are nonnegative, except for the term $(d_i - d_{i+})$, which is nonpositive. As a result, (29) (and hence (24)) is nonpositive for all $\mathbf{x}$, so that the MSE of $\hat{\mathbf{x}}_+$ is never higher than that of $\hat{\mathbf{x}}$.

We must also show that for some $\mathbf{x}$, (29) is strictly negative. To this end we choose $\mathbf{x}$ for which all elements are nonzero; as a result, all terms in (29) are strictly positive, except for $(d_i - d_{i+})$. This last term is negative when $d_i < 0$ and zero otherwise. Since d_i is negative with nonzero probability for at least one value of i, we conclude that for the chosen value of $\mathbf{x}$, (29) is strictly negative, completing the proof of Proposition 3. ■

This generalization of the concept of a positive part estimator is now used to prove Theorem 2.

Proof of Theorem 2. We show that $\hat{\mathbf{x}}_0$ of (22) strictly dominates the LS estimator. The result then follows from Proposition 3, since $\hat{\mathbf{x}}_{\mathrm{EBM}}$ is the positive part of $\hat{\mathbf{x}}_0$.

Denoting $s = \sum_{i=k+1}^{m} \lambda_i^{-1/2}$, the MSE of $\hat{\mathbf{x}}_0$ is given by

$$\begin{aligned}\mathrm{MSE} &= E\left\{\left\|\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{LS}} + \frac{s\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} + b - k}\right\|^2\right\} \\ &= \epsilon_0 + E\left\{\frac{s^2\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}}{(\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} + b - k)^2}\right\} \\ &\quad + 2E\left\{\frac{s(\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{LS}})^*\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} + b - k}\right\}.\end{aligned} \tag{31}$$

We now define $\hat{\mathbf{v}} = \mathbf{V}^*\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}$ and $\mathbf{v} = \mathbf{V}^*\mathbf{Q}^{1/2}\mathbf{x}$. Using this notation, the third term in (31) may be written as

$$A_3 \triangleq E\left\{ \frac{s(\mathbf{x}-\hat{\mathbf{x}}_{\mathrm{LS}})^*\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k} \right\}$$
$$= E\left\{ \frac{s(\mathbf{v}-\hat{\mathbf{v}})^*\mathbf{\Lambda}^{-1/2}\hat{\mathbf{v}}}{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k} \right\}$$
$$= \sum_{i=1}^{m} \lambda_i^{-1/2} E\left\{ \frac{s(v_i-\hat{v}_i)\hat{v}_i}{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k} \right\}, \tag{32}$$

where we have used the fact that $\hat{\mathbf{v}} \sim N_m(\mathbf{v}, \mathbf{I})$. Let

$$g_i(\hat{\mathbf{v}}) = \frac{s\hat{v}_i}{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k}, \tag{33}$$

noting that k is implicitly dependent on $\hat{\mathbf{v}}$, and that s is implicitly dependent on k. Thus, $g_i(\hat{\mathbf{v}})$ is discontinuous for some values of $\hat{\mathbf{v}}$, namely, those values for which $\alpha = \lambda_i^{-1/2}$. However, these values of $\hat{\mathbf{v}}$ occur with probability zero; for all other values, k (and hence s) are constant for sufficiently small changes in $\hat{\mathbf{v}}$. Thus,

$$\frac{\partial g_i(\hat{\mathbf{v}})}{\partial \hat{v}_i} = s\frac{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k-2\hat{v}_i^2}{(\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k)^2} \quad \text{with probability 1}, \tag{34}$$

and $E\left\{ \left| \frac{\partial g_i(\hat{\mathbf{v}})}{\partial \hat{v}_j} \right| \right\} < \infty$ for all j. Using Lemma 1, we have

$$E\left\{ \frac{s(v_i-\hat{v}_i)\hat{v}_i}{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k} \right\} = -E\left\{ s\frac{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k-2\hat{v}_i^2}{(\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k)^2} \right\}. \tag{35}$$

Combining (35) with (32), we obtain

$$A_3 = -\sum_{i=1}^{m} \lambda_i^{-1/2} E\left\{ s\frac{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k-2\hat{v}_i^2}{(\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k)^2} \right\}$$
$$= E\left\{ -\frac{s\,\mathrm{Tr}(\mathbf{Q}^{-1/2})}{\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k} + 2s\frac{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1/2}\hat{\mathbf{v}}}{(\hat{\mathbf{v}}^*\hat{\mathbf{v}}+b-k)^2} \right\}$$
$$= E\left\{ -\frac{s\,\mathrm{Tr}(\mathbf{Q}^{-1/2})}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k} + 2s\frac{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}}{(\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k)^2} \right\}. \tag{36}$$

We note that k is chosen in Proposition 2 in a manner which ensures that $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k \geq 0$. Hence

$$A_3 \leq E\left\{ \frac{s}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k} \left(-\mathrm{Tr}(\mathbf{Q}^{-1/2}) + 2\frac{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}} \right) \right\}$$
$$\leq E\left\{ \frac{s}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}+b-k} \left(-\mathrm{Tr}(\mathbf{Q}^{-1/2}) + 2\epsilon_{\max}^{1/2} \right) \right\}, \tag{37}$$

where $\epsilon_{\max}^{1/2}$ is the largest eigenvalue of $\mathbf{Q}^{-1/2}$. Substituting this result back into (31), and using the fact that $s \leq \mathrm{Tr}(\mathbf{Q}^{-1/2})$, yields

$$\mathrm{MSE} \leq \epsilon_0 + E\left\{ \frac{s\left(-\mathrm{Tr}(\mathbf{Q}^{-1/2}) + 4\epsilon_{\max}^{1/2}\right)}{\hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{Q} \hat{\mathbf{x}}_{\mathrm{LS}} + b - k} \right\}. \tag{38}$$

If $\mathrm{Tr}(\mathbf{Q}^{-1/2}) > 4\epsilon_{\max}^{1/2}$, then the expectation above is negative, so that $\hat{\mathbf{x}}_0$ (and hence $\hat{\mathbf{x}}_{\mathrm{EBM}}$) strictly dominate the LS estimator. ■

As we have seen, both the EBME and the SBME achieve lower MSE than the least-squares estimator. These results pose several further questions: Do BMEs significantly improve the MSE? How do BMEs compare with other extended Stein estimators? Is there a substantial difference between the performance of the spherical and ellipsoidal estimators? These questions will be answered in the numerical study in the next section.

IV. Numerical Results

Estimator performance generally depends on a number of operating conditions, including the effective dimension, the signal-to-noise ratio (SNR), the eigenvalues $\lambda_1, \ldots \lambda_m$ of $\mathbf{Q}$, and the value of the unknown parameter vector $\mathbf{x}$. Several computer simulations were implemented to test the effect of these conditions on performance. The simulations were also used to compare the BMEs with Bock's estimator [13], which is the most commonly-used extended Stein estimator [15], [16]. Like Stein's results, Bock's approach consists of a shrinkage estimator, given by

$$\hat{\mathbf{x}}_{\mathrm{Bock}} = \left(1 - \frac{\epsilon_0/\epsilon_{\max} - 2}{\hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{Q} \hat{\mathbf{x}}_{\mathrm{LS}}}\right) \hat{\mathbf{x}}_{\mathrm{LS}}. \tag{39}$$

The theorems of Sections II and III ensure that the BMEs achieve lower MSE than the LS estimator, but do not guarantee that this improvement is substantial. To measure this performance gain, we first chose a typical scenario, in which the number of parameters m and the number of measurements n were both 15. The system matrix $\mathbf{H}$ was chosen as $\mathrm{diag}(1, 1, 1, .5, .3, .2, .2, .2, .2, .1, .1, .1, .1, .05, .05)$, and the parameter vector was chosen randomly as a zero-mean, unit-variance i.i.d. Gaussian vector. The noise was i.i.d. with variance σ^2 chosen to achieve the desired SNR. Estimates of the MSE were calculated for a range of SNR values by generating 3000 random realizations of noise and parameter vectors per SNR value. The results are plotted in Fig. 1.

It is evident from this figure that substantial improvement in MSE can be achieved by using BMEs in place of the LS estimator: in some cases the MSE of the LS estimator is more than

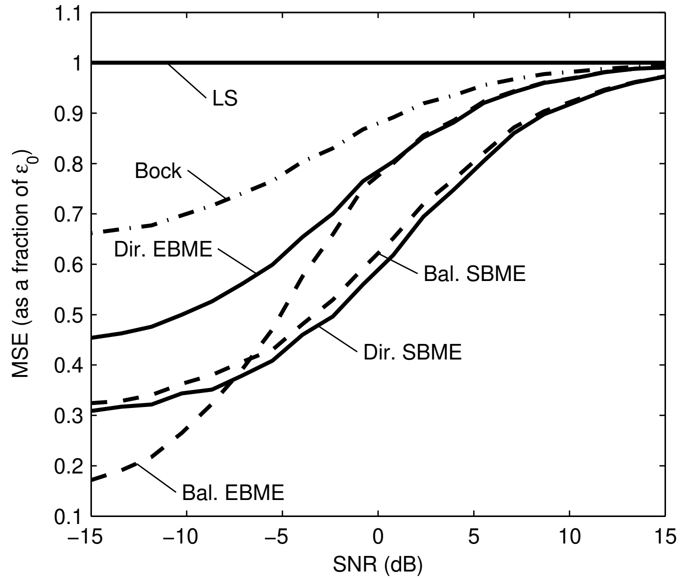

Fig. 1. MSE vs. SNR for a typical operating condition: effective dimension 5.1, $m = n = 15$.

five times the MSE of the BMEs. The performance gain is particularly noticeable at low and moderate SNR. At infinite SNR, the LS estimator is known to be optimal [1], and all other estimators converge to the value of the LS estimate; as a result, performance gain is smaller at high SNR.

Different settings call for the use of different estimators, as no single estimator is optimal under all operating conditions. To demonstrate this, estimator MSE was measured for various SNRs and effective dimensions. The parameter vector for this simulation was randomly chosen from an i.i.d. normal distribution. The number of measurements and the number of parameters were both equal to 10. The system matrix $\mathbf{H}$ was equal to $\mathbf{I}$, and the noise covariance matrix $\mathbf{C_w}$ was diagonal, with diagonal elements $c_1 = 1$, $c_2 = 0.5$, and $c_3 = \cdots = c_{10} = t$, where t was chosen to obtain the desired effective dimension. The simulation was repeated for 20 different effective dimensions in the range 1.5 to 10, and for 20 different SNR values in the range –15 dB to 15 dB. For each operating condition, the MSE of each estimator was calculated. Fig. 2 displays the estimator achieving lowest MSE (among the estimators tested); when two or more estimators achieve MSE within 5% of the lowest value, this is indicated by their combined pattern. The LS estimator is outperformed by all estimators in this simulation, so it is not displayed in the figure.

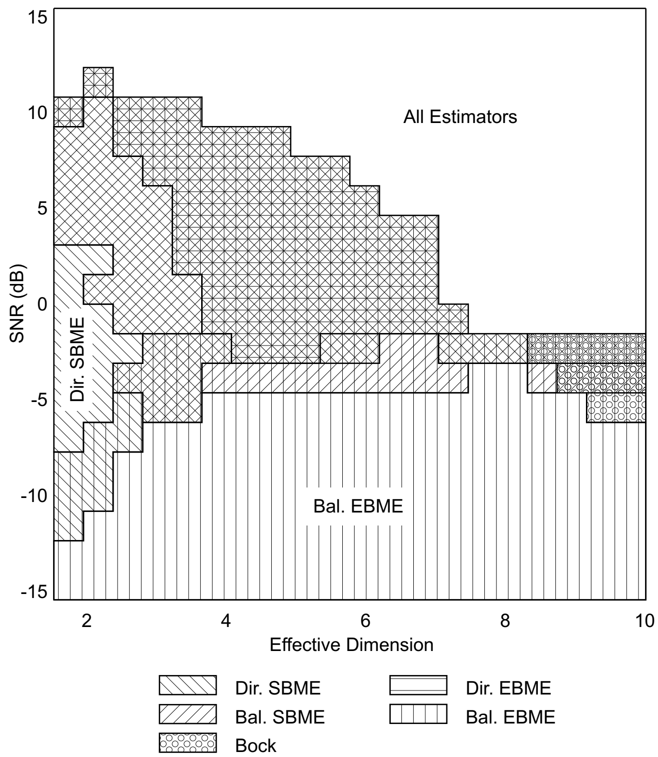

Fig. 2. Estimators achieving lowest MSE, among the five estimators tested ($m = n = 10$)

Fig. 2 demonstrates that the BMEs significantly outperform Bock's estimator under a very wide range of operating conditions. It is notable that the BMEs continue to outperform Bock's estimator and the LS estimator at effective dimensions of 2–4; the dominance results of Sections II and III only apply to effective dimensions above 4.

Fig. 2 also indicates that, for most operating conditions, the balanced EBME is the optimal estimator among those tested; its improvement is significant particularly at moderate and low SNR. However, at low effective dimensions, the SBMEs significantly outperform the EBMEs. This is indicative of a more subtle limitation of the EBMEs, namely, their sensitivity to the condition number[1] of $\mathbf{Q}$; this limitation also appears in Bock's estimator. The EBME makes use of an ellipsoid of the form $\{\mathbf{x} : \mathbf{x}^*\mathbf{Q}\mathbf{x} \leq L^2\}$, which becomes eccentric when the condition number of $\mathbf{Q}$ is large. As a result, a slight increase in the measurements along a narrow axis of the ellipse greatly increases the radius in the wide axes. This causes the corresponding parameters to be estimated with negligible shrinkage, thus reducing the improvement over

[1]The condition number of a matrix is defined as the ratio between its largest and smallest eigenvalues.

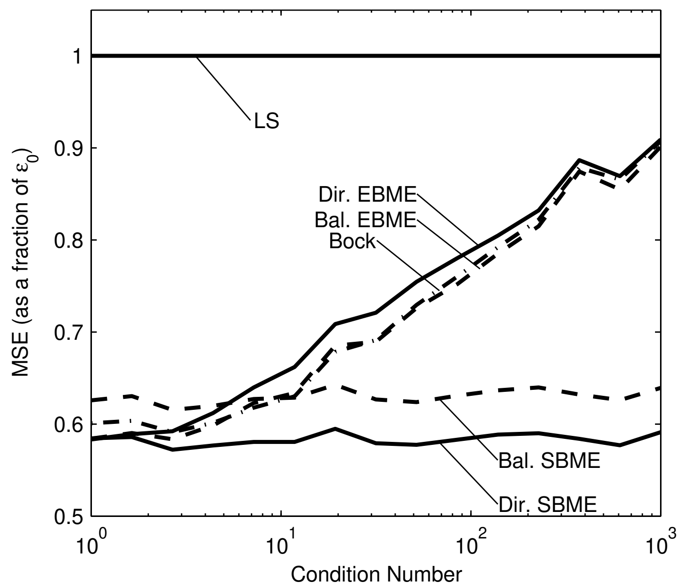

Fig. 3. Estimator MSE vs. condition number; $m = n = 10$, SNR 0 dB

the LS estimator. Bock's estimator suffers from the same effect, since its shrinkage factor is also a function of $\hat{\mathbf{x}}_{\mathrm{LS}}^* \mathbf{Q} \hat{\mathbf{x}}_{\mathrm{LS}}$. Only the SBMEs, whose parameter set estimates are based on the value $\hat{\mathbf{x}}_{\mathrm{LS}}^* \hat{\mathbf{x}}_{\mathrm{LS}}$, continue to perform well for high condition numbers.

This effect is demonstrated in Fig. 3. Here, the settings are identical to those of Fig. 2, except that the SNR is constant and equals 0 dB, and the noise covariance matrix contains nine eigenvalues equal to 1, and an additional small eigenvalue whose value is modified to control the condition number. Thus, the condition number is changed with little influence on the effective dimension. As expected, the performance of the EBMEs and Bock's estimator deteriorates when high condition numbers are used, while the SBMEs are hardly affected by the change. Fortunately, both the effective dimension and the condition number depend only on the system matrix $\mathbf{H}$ and the noise covariance $\mathbf{C_w}$, so that an informed choice may be made for any given estimation problem.

V. Discussion

The blind minimax approach is a general technique for using minimax estimators in situations for which no parameter set is known. We considered an application of this concept to the Gaussian linear regression model. Four novel estimators were proposed: estimators based on the spherical and ellipsoidal parameter sets, and which make use of the direct and balanced estimation techniques. In Sections II and III, all of these estimators were shown to dominate the LS estimator. Thus, in *any* application which makes use of a LS estimator, the MSE performance can be improved by using a BME instead.

It can readily be shown that the dominance condition of the SBMEs (Theorem 1) is weaker than the dominance condition of the EBMEs (Theorem 2), i.e., the conditions for SBME dominance hold whenever the conditions for EBME dominance hold. The dominance condition of Bock's estimator [13] is still weaker[2]. This would seem to indicate that Bock's estimator is superior to the proposed estimators. Yet the results of Section IV demonstrate that the opposite is true: the BMEs almost always outperform Bock's estimator — even in cases where their performance is not guaranteed by the dominance theorems. Thus, while dominance theorems are useful in providing sufficient conditions for improving on the LS estimator, they are ill-suited for comparing LS-dominating estimators. This conclusion is significant since many researchers have justified estimator choices by maximizing the range of conditions for which dominance is guaranteed. It seems that other analytical tools are required for comparing LS-dominating estimators. For example, it may be possible to prove that BMEs dominate Bock's estimator, for some problem settings.

The choice between the different BMEs is application-dependent. As explained in Section IV, the balanced EBME outperforms all other estimators at low SNR, while the direct SBME is sometimes better at moderate SNR, particularly for low effective dimensions. Also, the EBMEs (as well as Bock's estimator) perform poorly when the condition number of $\mathbf{Q}$ is large, while the SBMEs remain effective for any condition number. These results may be used to select an estimator depending on the problem setting, in particular since the effective dimension and condition number may be calculated in advance.

A more fundamental difference is that the SBMEs are shrinkage estimators, while the EBMEs are not. Thus, in applications where the only goal is minimization of the MSE, the direct SBME may be preferred for its robustness and simplicity. For example, the SBME is an excellent estimator of system parameters, such as autoregression (AR) coefficients. However, in certain applications, MSE minimization is only a nominal goal which approximates some other error

[2]A simple change to the SBME (adding -2 to the numerator) changes its dominance condition to that of Bock's estimator, without significantly affecting its performance. However, we have been unable to derive this modification using the blind minimax approach, and thus prefer the simpler form of the SBME used in the paper.

criterion. In some of these cases, a shrinkage estimator has no impact on the actual objective. For example, if the vector $\mathbf{x}$ is an image which is to be reconstructed, its subjective quality is not affected by multiplying the entire estimate by a scalar. Likewise, in a binary receiver, the sign of $\mathbf{x}$ must be determined, but the sign does not change when the estimate is shrunk. In such applications, SBMEs (and Bock's estimator) have no effect on the final result, whereas the balanced EBME can be used to improve performance.

In this paper, we have explored the idea of blind minimax estimation, whereby one uses linear minimax estimators whose parameter set is itself estimated from measurements. This simple concept was examined in the setting of a linear system of measurements with colored Gaussian noise, where we have shown that the BMEs dominate the LS estimator. Hence, in any such problem, the proposed estimators can be used in place of the LS estimator, with a guaranteed performance gain. Apart from being useful in and of themselves, the proposed estimators support the underlying concept of blind minimax estimation. This concept can be applied to many other estimation problems, such as estimation with uncertain system matrices, estimation with non-Gaussian noise, and sequential estimation. Use of the blind minimax approach in such problems remains a topic for further study.

Stein's discovery of LS-dominating estimators, half a century ago, shocked the statistics community, and LS-dominating estimators are still rarely used in practice. It is our hope that the blind minimax concept will provide additional support for such estimators, both by supplying an intuitive understanding of Stein's phenomenon, and by providing a wide class of powerful new estimators.

References

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[2] K. F. Gauss, "Theoria combinationis obsercationunt erronbus minimis obnoxiae," 1821.

[3] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer, 1998.

[4] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate distribution," in *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 1, 1956, pp. 197–206.

[5] A. Cohen, "All admissible linear estimators of the mean vector," *Ann. Math. Statist.*, vol. 37, no. 2, pp. 458–463, Apr. 1966.

[6] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Trans. Signal Processing*, to appear.

[7] Y. Maruyama, "A unified and broadened class of admissible minimax estimators of a multivariate normal mean," *J. Multivariate Analysis*, vol. 64, pp. 196–205, 1998.

[8] ——, "Minimax admissible estimation of a multivariate normal mean and improvement upon the James-Stein estimator," Ph.D. dissertation, University of Tokyo, 2000.

[9] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, vol. 1, 1961, pp. 311–319.

[10] K. Alam and J. R. Thompson, "Estimation of the mean of a multivariate normal distribution," Indiana Univ., Tech. Rep., 1964.

[11] A. J. Baranchik, "A family of minimax estimators of the mean of a multivariate normal distribution," *Ann. Math. Statist.*, vol. 41, no. 2, Apr. 1970.

[12] B. Efron and C. Morris, "Stein's estimation rule and its competitors: an empirical Bayes approach," *J. Amer. Statist. Assoc.*, vol. 68, pp. 117–130, 1973.

[13] M. E. Bock, "Minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 3, no. 1, pp. 209–218, Jan. 1975.

[14] J. O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss," *Ann. Statist.*, vol. 4, no. 1, pp. 223–226, Jan. 1976.

[15] J. H. Manton, V. Krishnamurthy, and H. V. Poor, "James-Stein state filtering algorithms," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2431–2447, Sept. 1998.

[16] E. Greenberg and C. E. Webster, Jr., *Advanced Econometrics*, 2nd ed. New York: Wiley, 1983.

[17] B. Efron and C. Morris, "Combining possibly related estimation problems," *J. Roy. Statist. Soc. B*, vol. 35, no. 3, pp. 379–421, 1973.

[18] Z. Ben-Haim and Y. C. Eldar, "Minimax estimators dominating the least-squares estimator," in *Proc. Int. Conf. Acoust., Speech and Signal Processing (ICASSP 2005)*, vol. IV, Philadelphia, PA, Mar. 2005, pp. 53–56.

[19] ——, "Blind minimax estimators: Improving on least-squares estimation," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP 2005)*, Bordeaux, France, July 2005.

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

[21] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation with bounded data uncertainties," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[22] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[23] A. J. Baranchik, "Multiple regression and estimation of the mean of a multivariate normal distribution," Department of Statistics, Stanford University, Stanford, CA, Tech. Rep. #51, 1964.