BLIND MINIMAX ESTIMATORS: IMPROVING ON LEAST-SQUARES ESTIMATION

Zvika Ben-Haim *Yonina C. Eldar*

zvikabh@technion.ac.il, yonina@ee.technion.ac.il
Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, Israel

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 estimators based on the blind minimax approach, a technique whereby a parameter set is estimated from measurements and then used to construct a minimax estimator. We demonstrate analytically that the obtained estimators strictly dominate the least-squares estimator (LSE), i.e., they achieve lower mean-squared error for any value of the parameter vector. Simulations show that these estimators outperform Bock's estimator, which also dominates the LSE.

1. INTRODUCTION

We consider the classical problem of estimating a deterministic, unknown parameter vector $\mathbf{x}$ from a measurement vector $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{H}$ is a known matrix and $\mathbf{w}$ represents Gaussian noise with known covariance $\mathbf{C_w}$ [1]. An estimator $\hat{\mathbf{x}}$ is a function of $\mathbf{y}$ intended to be close to $\mathbf{x}$, typically in terms of mean-squared error (MSE). However, the MSE of an estimator generally depends on the unknown value of $\mathbf{x}$, and no estimator minimizes the MSE for *all* values of $\mathbf{x}$. Hence, a standard approach is to limit discussion to the set of linear unbiased estimators, for which the MSE does not depend on $\mathbf{x}$. The well-known least-squares estimator (LSE) minimizes the MSE among all linear unbiased estimators.

Yet biased and nonlinear estimators may outperform the LSE in terms of MSE; this is known as the Stein phenomenon [2, 3]. One example is the James-Stein estimator [4], an estimator designed for the i.i.d. case, in which $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \sigma^2\mathbf{I}$. Under simple regularity conditions, the James-Stein estimator strictly dominates the LSE, meaning that it achieves lower MSE for *any* value of $\mathbf{x}$.

Several extended Stein estimators attempt to apply these results to the more interesting case in which $\mathbf{H}$ and $\mathbf{C_w}$ are not identity matrices [5–7]. However, none of these methods has become a standard alternative to the LSE. One reason is that Stein estimators are considered counterintuitive (see the discussion following [8]). Another reason is that many extended Stein estimators are shrinkage estimators, rendering them inappropriate for applications (such as image enhancement) in which a gain factor does not affect performance.

A somewhat different estimation problem occurs when the parameter $\mathbf{x}$ is unknown, but lies within a known set $\mathcal{S}$. In this case, a linear minimax estimator may be designed which minimizes the worst-case MSE among all possible values of $\mathbf{x}$ in $\mathcal{S}$ [9, 10]. It has recently been shown that minimax estimators achieve lower MSE than the LSE, as long as $\mathbf{x}$ does indeed lie within the set $\mathcal{S}$ [11, 12].

We seek to apply the success of minimax estimators to the general estimation problem, in which no information on $\mathbf{x}$ is available. To do so, we propose a two-stage estimation process. First, the measurements are used to estimate a parameter set $\mathcal{S}$ likely to contain the true value of $\mathbf{x}$. Next, a minimax estimator is constructed for the set $\mathcal{S}$, to obtain the final estimate of $\mathbf{x}$. We refer to the resulting estimator as a blind minimax estimator (BME).

BMEs are successful because a parameter set can be estimated far more accurately than the actual value of $\mathbf{x}$. Subsequently, in many cases the obtained estimator outperforms the LSE. In a recent paper [11], BMEs were examined for the i.i.d. case, and shown to strictly dominate the LSE under simple regularity conditions. Furthermore, in the i.i.d. case, BMEs were shown to closely resemble the James-Stein estimator.

One advantage of BMEs is that the two-stage blind minimax technique described above extends naturally to the general (non-i.i.d.) setting. This paper focuses on such an extension. We provide closed forms for the BMEs, and show that they strictly dominate the LSE and also outperform other extended Stein estimators.

We use the blind minimax approach to construct two different estimators. The spherical BME (Section 2) is a shrinkage estimator which can be implemented very efficiently, while the ellipsoidal BME (Section 3) is a non-shrinkage estimator which is slightly more computationally complex. Both estimators dominate the LSE and outperform other extended Stein estimators, and both estimators reduce to the blind minimax estimator of [11] in the i.i.d. case. The proposed estimators are compared with existing estimators in an empirical study in Section 4, and the results are summarized in Section 5.

2. 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

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

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 $\mathbf{C_w}$.

Suppose $\mathbf{x}$ is known to lie within a compact parameter set $\mathcal{S}$. In this case, a *linear minimax* estimator may be constructed [9, 10]. This is the linear estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ minimizing the worst-case MSE

This work was supported by the 6th EU framework program, via the NEWCOM network of excellence, and by the Israel Science Foundation.

among all possible values of $\mathbf{x}$ in $\mathcal{S}$,

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

For example, when the set $\mathcal{S}$ is a sphere centered on the origin, $\mathcal{S} = \{\mathbf{x} : \|\mathbf{x}\|^2 \le L^2\}$, the linear minimax estimator is [10]

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

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

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

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

$$\epsilon_0 = E\{\|\hat{\mathbf{x}}_{\mathrm{LS}}-\mathbf{x}\|^2\} = \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}\right). \tag{5}$$

It has recently been shown that any linear minimax estimator achieves lower MSE than that of the LSE, for all values of $\mathbf{x}$ in $\mathcal{S}$ [11,12]. Thus, as long as some parameter set $\mathcal{S}$ is known, minimax estimators outperform the LSE.

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 centered on the origin, and estimate the sphere radius from the measurements. The resulting *spherical blind minimax estimator* (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 $\|\mathbf{x}\|^2$ is unknown, a natural alternative is to use the LSE to estimate $\|\mathbf{x}\|^2$; for instance, one may estimate L^2 as $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$. Closer observation reveals that $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2$ is, in fact, an overestimate of $\|\mathbf{x}\|^2$, since

$$E\{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2\} = \|\mathbf{x}\|^2 + \epsilon_0. \tag{6}$$

To correct for this effect, we estimate L^2 as $\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2 - \epsilon_0$. Substituting this value into (3), the SBME is given by

$$\hat{\mathbf{x}}_{\mathrm{SBM}} = \left(1 - \frac{\epsilon_0}{\|\hat{\mathbf{x}}_{\mathrm{LS}}\|^2}\right)\hat{\mathbf{x}}_{\mathrm{LS}}. \tag{7}$$

Thus, the SBME is a shrinkage estimator: it consists of multiplying the LSE by a scaling factor. The scaling factor is a "restraint" which lessens the effect of random fluctuations in the measurements.

It is remarkable that for the i.i.d. case, the above estimator reduces to the original Stein estimator [2], which was derived in a different manner and later shown to strictly dominate the LSE [4]. An estimator is said to strictly dominate the LSE if it achieves lower MSE for all values of $\mathbf{x}$. By comparison, an estimator dominates the LSE if its MSE is at least as low as that of the LSE for all values of $\mathbf{x}$, and is strictly lower for at least one value of $\mathbf{x}$.

While the SBME reduces to Stein's estimator in the i.i.d. case, it is equally well-defined for the non-i.i.d. case. Furthermore, as the following theorem shows, the SBME strictly dominates the LSE in the non-i.i.d. case.

Theorem 1. *Suppose $\epsilon_0/\lambda_{\max} > 4$, where $\lambda_{\max}$ is the largest eigenvalue of $(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}$. Then, the SBME* (7) *strictly dominates the LSE* (4)*.*

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

The proof of Theorem 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 $E\left|\frac{\partial g(\hat{\mathbf{v}})}{\partial \hat{v}_i}\right| < \infty$ for all i. Then,*

$$E\left\{\frac{\partial g(\hat{\mathbf{v}})}{\partial \hat{v}_i}\right\} = -E\{g(\hat{\mathbf{v}})(v_i - \hat{v}_i)\}\,. \tag{8}$$

Proof of Theorem 1. The MSE of $\hat{\mathbf{x}}_{\mathrm{SBM}}$ is given by

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{SBM}}) &= E\left\{\left\|\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{LS}} + \frac{\epsilon_0}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\hat{\mathbf{x}}_{\mathrm{LS}}\right\|^2\right\} \\ &= \epsilon_0 + E\left\{\frac{\epsilon_0^2}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\right\} + 2E\left\{\frac{\epsilon_0\hat{\mathbf{x}}_{\mathrm{LS}}^*(\mathbf{x}-\hat{\mathbf{x}}_{\mathrm{LS}})}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\right\}.\end{aligned} \tag{9}$$

Let us denote $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}$, and 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 $\mathbf{v} = \mathbf{V}\mathbf{Q}^{1/2}\mathbf{x}$. The third term in (9) may now be written as

$$\begin{aligned}E_3 &\triangleq E\left\{\frac{\epsilon_0\hat{\mathbf{x}}_{\mathrm{LS}}^*(\mathbf{x}-\hat{\mathbf{x}}_{\mathrm{LS}})}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\right\} \\ &= E\left\{\frac{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}(\mathbf{v}-\hat{\mathbf{v}})}{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}}\right\} \\ &= \sum_{i=1}^{m}\lambda_i^{-1}E\left\{\frac{\hat{v}_i(v_i-\hat{v}_i)}{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}}\right\},\end{aligned} \tag{10}$$

where in the last step we used the fact that $\hat{\mathbf{v}} \sim N_m(\mathbf{v}, \mathbf{I})$. Defining $g_i(\hat{\mathbf{v}}) = \frac{\hat{v}_i}{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}}$, we apply Stein's lemma to obtain

$$\begin{aligned}E_3 &= -\sum_i \lambda_i^{-1}E\left\{\frac{1}{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}} - 2\frac{\lambda_i^{-1}\hat{v}_i^2}{(\hat{\mathbf{v}}^*\mathbf{\Lambda}\hat{\mathbf{v}})^2}\right\} \\ &= -E\left\{\frac{\mathrm{Tr}(\mathbf{\Lambda}^{-1})}{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}}}\right\} + 2E\left\{\frac{\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-2}\hat{\mathbf{v}}}{(\hat{\mathbf{v}}^*\mathbf{\Lambda}^{-1}\hat{\mathbf{v}})^2}\right\} \\ &= -E\left\{\frac{\epsilon_0}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\right\} + 2E\left\{\frac{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\mathrm{LS}}}{(\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}})^2}\right\}.\end{aligned} \tag{11}$$

Substituting this result back into (9) yields

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{SBM}}) &= \epsilon_0 + E\left\{\frac{\epsilon_0}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\left(4\frac{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}^{-1}\hat{\mathbf{x}}_{\mathrm{LS}}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}\right) - \epsilon_0\right\} \\ &\le \epsilon_0 + E\left\{\frac{\epsilon_0}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\hat{\mathbf{x}}_{\mathrm{LS}}}(-\epsilon_0 + 4\lambda_{\max})\right\}.\end{aligned} \tag{12}$$

If $\epsilon_0 > 4\lambda_{\max}$, then the expectation is taken over a strictly negative range, and hence $\hat{\mathbf{x}}_{\mathrm{SBM}}$ always has strictly lower MSE than $\hat{\mathbf{x}}_{\mathrm{LS}}$, which proves the theorem. □

As we have shown, in terms of MSE, the SBME is a better estimator than the LSE. This is particularly notable in light of the simple mechanism used to generate the SBME. However, the SBME is a shrinkage estimator, i.e., it consists of the LSE multiplied by a gain factor. 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 LSE.

3. THE ELLIPSOIDAL BLIND MINIMAX ESTIMATOR

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 $\mathbf{Q}^{-1} = (\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}$. Thus, some elements in $\hat{\mathbf{x}}_{\mathrm{LS}}$ have lower variance than others. In this sense, the scalar shrinkage factor of the SBME (7) and other extended Stein estimators [6] seems inadequate. Indeed, several researchers have proposed shrinking each element according to its variance [5, 7]. Ironically, however, there has been disagreement as to whether high-variance components should be shrunk more [5] or less [7], and little justification has been given to shrinkage factor choice.

The blind minimax approach provides a natural framework for solving these disputes. To see this, denote $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{x} + \mathbf{u}$, 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

$$\mathbf{Q}^{1/2}\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{Q}^{1/2}\mathbf{x} + \tilde{\mathbf{u}}, \tag{13}$$

where $\tilde{\mathbf{u}} \sim N_m(\mathbf{0}, \mathbf{I})$. A better approach is then to estimate $\mathbf{x}^*\mathbf{Q}\mathbf{x}$ using $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}$. Such an estimate can be readily incorporated into the blind minimax framework by using an ellipsoidal parameter set, $\{\mathbf{x} : \mathbf{x}^*\mathbf{Q}\mathbf{x} \le L^2\}$, rather than the spherical parameter set of the SBME. This 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, we observe that

$$E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}\} = \mathbf{x}^*\mathbf{Q}\mathbf{x} + m, \tag{14}$$

so that $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}$ is an overestimate of $\mathbf{x}^*\mathbf{Q}\mathbf{x}$. We therefore use $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} - m$ to estimate the ellipsoid radius.

The *ellipsoidal blind minimax estimator* (EBME) can thus be defined as follows. First, calculate the value $L^2 = \hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} - m$. Next, substitute L^2 into the formula [10] for the minimax estimator designed for the parameter set $\{\mathbf{x} : \mathbf{x}^*\mathbf{Q}\mathbf{x} \le L^2\}$. Formally, let $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$ be the eigenvalue decomposition of $\mathbf{Q}$, so that $\mathbf{V}$ is unitary, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots \lambda_m)$, and $\lambda_1 \ge \cdots \ge \lambda_m$. We then have the following closed form for the EBME.

Proposition 1 (Closed-Form EBME). *The EBME is given by*

$$\hat{\mathbf{x}}_{\mathrm{EBM}} = \mathbf{V}\,\mathrm{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})\mathbf{V}^*\left(\mathbf{I} - \alpha\mathbf{Q}^{1/2}\right)\hat{\mathbf{x}}_{\mathrm{LS}}, \tag{15}$$

where

$$\alpha = \frac{\sum_{i=k+1}^{m}\lambda_i^{-1/2}}{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}} - k}, \tag{16}$$

and k is the smallest integer $0 \le k \le m-1$ such that $\alpha < \lambda_{k+1}^{-1/2}$.

Proof. Follows from Proposition 1 of [10]. □

While the closed form of the EBME appears somewhat more intimidating than that of the SBME, their computational complexities are comparable. The major difference is the calculation of the value k, for which m divisions are required. Like the SBME, the EBME also dominates the LSE under suitable conditions.

Theorem 2. *Suppose* $\mathrm{Tr}(\mathbf{Q}^{-1/2}) > 4\lambda_{\max}^{1/2}$*, where* $\lambda_{\max}^{1/2}$ *is the largest eigenvalue of* $\mathbf{Q}^{-1/2}$*. Then, the EBME* (15) *strictly dominates the LSE* (4).

The proof of Theorem 2 is based on an analogy between the diagonal matrix $\mathrm{diag}(\mathbf{0}_k, \mathbf{1}_{m-k})$ in (15) and Baranchik's positive-part modification [3, 13] 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 (15),

$$\begin{aligned}\hat{\mathbf{x}}_0 &= (\mathbf{I} - \alpha\mathbf{Q}^{1/2})\hat{\mathbf{x}}_{\mathrm{LS}}\\ &= \mathbf{V}\,\mathrm{diag}\left(1 - \alpha\lambda_1^{1/2}, \cdots 1 - \alpha\lambda_m^{1/2}\right)\mathbf{V}^*\hat{\mathbf{x}}_{\mathrm{LS}}.\end{aligned} \tag{17}$$

Since $\alpha \ge \lambda_i^{-1/2}$ for all $i \le k$, this would introduce negative shrinkage for the first k eigenvectors of $\mathbf{V}$. As the following proposition shows, the MSE can be reduced by eliminating this negative shrinkage.

Proposition 2 (Generalized Positive-Part Estimator). *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 may be 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*

$$\hat{\mathbf{x}}_+ = \mathbf{V}\mathbf{D}_+\mathbf{V}^*\hat{\mathbf{x}}_{\mathrm{LS}}, \tag{18}$$

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

Proof. The estimator $\hat{\mathbf{x}}_+$ is said to dominate $\hat{\mathbf{x}}$ if $\mathrm{MSE}(\hat{\mathbf{x}}) \ge \mathrm{MSE}(\hat{\mathbf{x}}_+)$ for all $\mathbf{x}$, with strict inequality for at least one value of $\mathbf{x}$. We will show that $\mathrm{MSE}(\hat{\mathbf{x}}) - \mathrm{MSE}(\hat{\mathbf{x}}_+)$ is nonnegative for all $\mathbf{x}$, and positive for any value of $\mathbf{x}$ whose elements are all nonzero.

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{x}}) - \mathrm{MSE}(\hat{\mathbf{x}}_+) &= E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} - E\{\|\hat{\mathbf{x}}_+ - \mathbf{x}\|^2\}\\ &= E\{\|\hat{\mathbf{x}}\|^2 - \|\hat{\mathbf{x}}_+\|^2\} - 2E\{\hat{\mathbf{x}}^*\mathbf{x} - \hat{\mathbf{x}}_+^*\mathbf{x}\}\\ &= E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{V}(\mathbf{D}^2 - \mathbf{D}_+^2)\mathbf{V}^*\hat{\mathbf{x}}_{\mathrm{LS}}\}\\ &\quad - 2E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}.\end{aligned} \tag{19}$$

Since $d_i^2 - d_{i+}^2 \ge 0$ for all i, the first term in (19) is nonnegative. Hence, to prove the proposition, it suffices to show that $E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}$ 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 $\hat{\mathbf{z}} = \mathbf{V}^*\hat{\mathbf{x}}_{\mathrm{LS}}$. 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 $E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\}$, we condition on $\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{Q}\hat{\mathbf{x}}_{\mathrm{LS}}$, obtaining

$$\begin{aligned}E\{\hat{\mathbf{x}}_{\mathrm{LS}}^*\mathbf{V}(\mathbf{D} - \mathbf{D}_+)\mathbf{V}^*\mathbf{x}\} &= E\{E\{\hat{\mathbf{z}}^*(\mathbf{D} - \mathbf{D}_+)\mathbf{z}|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}\}\}\\ &= E\left\{\sum_{i=1}^{m}(d_i - d_{i+})E\{\hat{z}_i z_i|\hat{\mathbf{z}}^*\mathbf{\Lambda}\hat{\mathbf{z}}\}\right\},\end{aligned} \tag{20}$$

where we used the fact that d_i and d_{i+} are deterministic when conditioned on $\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS}$. We now define

$$r_i(\hat{\mathbf{z}}) \triangleq \sqrt{\frac{\hat{\mathbf{z}}^* \mathbf{\Lambda} \hat{\mathbf{z}} - \sum_{j \neq i} \lambda_j \hat{z}_j^2}{\lambda_i}}, \quad (21)$$

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

$$E\{\hat{z}_i z_i | \hat{\mathbf{z}}^* \mathbf{\Lambda} \hat{\mathbf{z}}, \{\hat{z}_j\}_{j \neq i}\} = z_i r_i(\hat{\mathbf{z}}) E\{\operatorname{sgn}(\hat{z}_i) | \hat{\mathbf{z}}^* \mathbf{\Lambda} \hat{\mathbf{z}}, \{\hat{z}_j\}_{j \neq i}\}. \quad (22)$$

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

$$E\{\operatorname{sgn}(\hat{z}_i) | \hat{\mathbf{z}}^* \mathbf{\Lambda} \hat{\mathbf{z}}, \{\hat{z}_j\}_{j \neq i}\} = E\{\operatorname{sgn}(\hat{z}_i)\}. \quad (23)$$

Combining this result with (20) and (22), we obtain

$$\begin{aligned} &E\{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{V}(\mathbf{D} - \mathbf{D}_+) \mathbf{V}^* \mathbf{x}\} \\ &= E\left\{ \sum_{i=1}^m (d_i - d_{i+}) r_i(\hat{\mathbf{z}}) |z_i| \operatorname{sgn}(z_i) E\{\operatorname{sgn}(\hat{z}_i)\} \right\}. \quad (24) \end{aligned}$$

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

$$\Pr\{\operatorname{sgn}(\hat{z}_i) = \operatorname{sgn}(z_i)\} > \Pr\{\operatorname{sgn}(\hat{z}_i) \neq \operatorname{sgn}(z_i)\}. \quad (25)$$

Thus, $\operatorname{sgn}(z_i) E\{\operatorname{sgn}(\hat{z}_i)\}$ is positive if $z_i \neq 0$, and equals zero if $z_i = 0$. It follows that all terms in (24) are nonnegative, except for the term $(d_i - d_{i+})$, which is nonpositive. As a result, (24) (and hence (19)) 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 $\mathbf{x}$, (24) is strictly negative. To this end we choose $\mathbf{x}$ for which all elements are nonzero; as a result, all terms in (24) are strictly positive, except for $(d_i - d_{i+})$. This last term is negative when $d_i < 0$ and zero otherwise. Since, for at least one value of i, d_i is negative with nonzero probability, we conclude that for the chosen value of $\mathbf{x}$, (24) is strictly negative, completing the proof of Proposition 2. □

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 (17) strictly dominates the LSE. The result follows since $\hat{\mathbf{x}}_{\rm EBM}$ 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

$$\begin{aligned} \text{MSE} &= E\left\{ \left\| \mathbf{x} - \hat{\mathbf{x}}_{\rm LS} + \frac{s \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \right\|^2 \right\} \\ &= \epsilon_0 + E\left\{ \frac{s^2 \hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS}}{(\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k)^2} \right\} \\ &\quad + 2E\left\{ \frac{s(\mathbf{x} - \hat{\mathbf{x}}_{\rm LS})^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \right\}. \quad (26) \end{aligned}$$

We now define $\hat{\mathbf{v}} = \mathbf{V}^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}$ and $\mathbf{v} = \mathbf{V}^* \mathbf{Q}^{1/2} \mathbf{x}$. Using this notation, the third term in (26) may be written as

$$\begin{aligned} E_3 &\triangleq E\left\{ \frac{s(\mathbf{x} - \hat{\mathbf{x}}_{\rm LS})^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \right\} \\ &= E\left\{ \frac{s(\mathbf{v} - \hat{\mathbf{v}})^* \mathbf{\Lambda}^{-1/2} \hat{\mathbf{v}}}{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k} \right\} \\ &= \sum_{i=1}^m \lambda_i^{-1/2} E\left\{ \frac{s(v_i - \hat{v}_i) \hat{v}_i}{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k} \right\}, \quad (27) \end{aligned}$$

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

Let $g_i(\hat{\mathbf{v}}) = \frac{s \hat{v}_i}{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k}$, 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,

$$\frac{\partial g_i(\hat{\mathbf{v}})}{\partial \hat{v}_i} = s \frac{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k - 2\hat{v}_i^2}{(\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k)^2} \quad \text{with probability } 1, \quad (28)$$

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

$$E\left\{ \frac{s(v_i - \hat{v}_i) \hat{v}_i}{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k} \right\} = -E\left\{ s \frac{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k - 2\hat{v}_i^2}{(\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k)^2} \right\}. \quad (29)$$

Combining with (27), we obtain

$$\begin{aligned} E_3 &= -\sum_{i=1}^m \lambda_i^{-1/2} E\left\{ s \frac{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k - 2\hat{v}_i^2}{(\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k)^2} \right\} \\ &= E\left\{ -\frac{s \operatorname{Tr}(\mathbf{Q}^{-1/2})}{\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k} + 2s \frac{\hat{\mathbf{v}}^* \mathbf{\Lambda}^{-1/2} \hat{\mathbf{v}}}{(\hat{\mathbf{v}}^* \hat{\mathbf{v}} - k)^2} \right\} \\ &= E\left\{ -\frac{s \operatorname{Tr}(\mathbf{Q}^{-1/2})}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} + 2s \frac{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}}{(\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k)^2} \right\}. \quad (30) \end{aligned}$$

We note that k is chosen in Proposition 1 in a manner which ensures that $\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k \geq 0$. Hence

$$\begin{aligned} E_3 &\leq E\left\{ \frac{s}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \left(-\operatorname{Tr}(\mathbf{Q}^{-1/2}) + 2 \frac{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q}^{1/2} \hat{\mathbf{x}}_{\rm LS}}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS}} \right) \right\} \\ &\leq E\left\{ \frac{s}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \left(-\operatorname{Tr}(\mathbf{Q}^{-1/2}) + 2\lambda_{\max}^{1/2} \right) \right\}, \quad (31) \end{aligned}$$

where $\lambda_{\max}^{1/2}$ is the largest eigenvalue of $\mathbf{Q}^{-1/2}$. Substituting this result back into (26), and using the fact that $s \leq \operatorname{Tr}(\mathbf{Q}^{-1/2})$, yields

$$\text{MSE} \leq \epsilon_0 + E\left\{ \frac{s\left(-\operatorname{Tr}(\mathbf{Q}^{-1/2}) + 4\lambda_{\max}^{1/2} \right)}{\hat{\mathbf{x}}_{\rm LS}^* \mathbf{Q} \hat{\mathbf{x}}_{\rm LS} - k} \right\}. \quad (32)$$

If $\operatorname{Tr}(\mathbf{Q}^{-1/2}) > 4\lambda_{\max}^{1/2}$, then the expectation above is negative, so that $\hat{\mathbf{x}}_0$ (and hence $\hat{\mathbf{x}}_{\rm EBM}$) 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 spherical and ellipsoidal estimators? These questions will be answered in the numerical study in the next section.

4. NUMERICAL RESULTS

Estimator performance generally depends on a number of operating conditions, including the effective dimension, the signal-to-noise ratio (SNR), the distribution of eigenvalues $\lambda_1, \ldots \lambda_m$, and

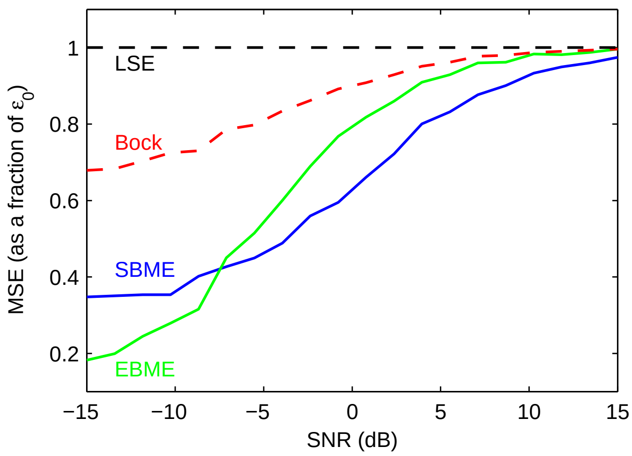

Fig. 1. MSE vs. SNR for a typical operating condition: $\mathbf{x}$ distributed normally i.i.d., effective dimension 4.8, $m = n = 15$.

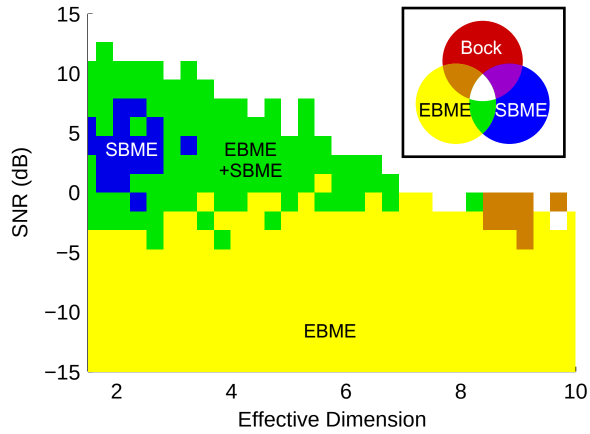

Fig. 2. Optimal estimator(s) for various effective dimensions ($\mathbf{x}$ distributed normally i.i.d.; $m = n = 10$)

the value of the unknown parameter vector $\mathbf{x}$. A computer simulation was used to test the effect of these conditions on estimator performance. The simulations were also used to compare the BMEs with Bock's estimator [6]

$$\hat{\mathbf{x}}_{\text{Bock}} = \left(1 - \frac{\epsilon_0/\lambda_{\max} - 2}{\hat{\mathbf{x}}_{\text{LS}}^* \mathbf{H}^* \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{H} \hat{\mathbf{x}}_{\text{LS}}}\right) \hat{\mathbf{x}}_{\text{LS}}, \qquad (33)$$

which is the most commonly-used extended Stein estimator [14].

Typical simulation results are plotted in Fig. 1. In this simulation, both the number of measurements and the number of parameters is 15, and the effective dimension is 4.8. The parameter vector $\mathbf{x}$ is chosen from a zero-mean i.i.d. normal distribution, whose variance is chosen to yield the required SNR, defined as $E\{\|\mathbf{x}\|^2\} \,/\, \text{Tr}((\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1})$. The plot displays the average MSE among 50 random realizations of $\mathbf{x}$; for each value of $\mathbf{x}$ the MSE was calculated as the average error obtained from 30 noise realizations.

The BMEs clearly outperform both the LSE and Bock's estimator in the case of Fig. 1. However, it turns out that there exist operating conditions for which *each* of the estimators tested (with the exception of the LSE) outperforms all other estimators, if only by a small margin. A good estimator is therefore one which is rarely dominated, and then only by a small margin.

To test which estimators satisfy this requirement, the MSE of the various estimators was calculated under many different operating conditions. Some of the results of these simulations appear in Figs. 2 and 3. In these figures, the optimal estimator under each operating condition is indicated by color[1], as defined in the legend in Fig. 2. (The LSE was outperformed in all operating conditions tested, so it is not assigned a color.) When two or more estimators achieve MSE performance within 5% of the optimal, this is indicated by their combined color. For example, a green region indicates operating conditions for which the performance of the EBME (yellow) and the SBME (blue) was nearly identical. Such a plot allows one to effectively compare estimators under a wide range of operating conditions.

[1]A color version of this manuscript is available at `http://www.technion.ac.il/~zvikabh/published/bme_imp_lse.pdf`

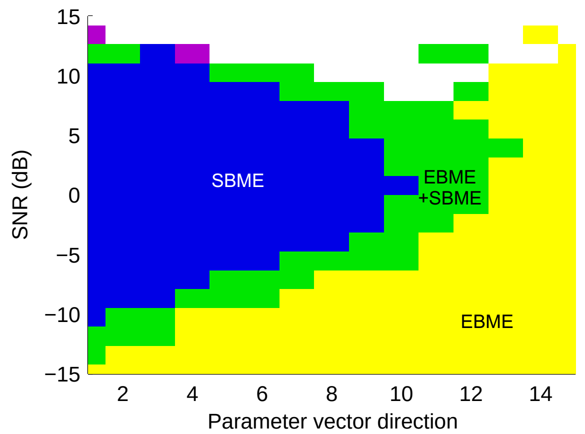

Fig. 3. Optimal estimator(s) for various values of $\mathbf{x}$ (effective dimension 5; $m = n = 15$)

In Fig. 2, the two most influential operating conditions were examined: these are the effective dimension and the SNR. The parameter vector for this simulation was randomly chosen from an i.i.d. normal distribution, as in Fig. 1. In this simulation, the number of measurements and the number of parameters were both equal to 10. The eigenvalues of $(\mathbf{H}^*\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H})^{-1}$ were 1, 0.5, and eight repetitions of an identical eigenvalue t, whose value was modified to obtain the desired effective dimension.

It is evident from Fig. 2 that all estimators perform comparably at high SNR. This is a result of the fact that, at high SNR, all estimators converge to the LSE, which is optimal for infinite SNR. However, for moderate and low SNR (below 5–10 dB), the BMEs perform significantly better than Bock's estimator, with the EBME dominating the SBME at low SNR. It is notable that the BMEs continue to outperform Bock's estimator and the LSE at effective dimensions of 2–4; the dominance results of Sections 2 and 3 only apply to effective dimensions above 4.

The fact that other operating conditions also affect estimator performance is evident in Fig. 3. Here, $m = n = 15$, and the parameter vector $\mathbf{x}$ is chosen in the direction of different eigenvectors of $(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}$, with the minimum eigenvector denoted by 1 and the maximum eigenvector denoted by 15. This has the effect of choosing vectors $\mathbf{x}$ which lie in the direction of minimal and maximal noise, respectively. The eigenvalues of $(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}$ are chosen as a geometric sequence $\{q^i\}_{i=1}^{15}$, where q is selected so that the effective dimension is 5; thus, each eigenvector is associated with a different eigenvalue. In all of these cases, the BMEs outperform Bock's estimator. However, the ellipsoidal and spherical versions are better suited for different values of the parameter vector.

Extensive measurements of the performance of the BMEs thus demonstrate that they are rarely dominated by Bock's estimator, and then only by a small margin. These results provide strong evidence in favor of the blind minimax approach.

5. 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. Two estimators were considered, the spherical and ellipsoidal BMEs. In Sections 2 and 3, both estimators were shown to dominate the LSE under simple regularity conditions. Thus, in *any* application which makes use of a LSE, the MSE performance can be improved by using either BME instead.

In Section 4, the BMEs were empirically shown to significantly outperform Bock's estimator. This is, perhaps, due to the fact that the proposed estimators are designed using a systematic technique, whereas no justification is given to the form of Bock's estimator beyond the fact that it dominates the LSE.

The choice between the ellipsoidal and spherical BMEs is application-dependent. The simulations performed indicate that the EBME outperforms the SBME at low SNR, while the SBME is often better at moderate SNR. More importantly, however, the SBME is a shrinkage estimator, while the EBME is not. In applications where the only goal is minimization of the MSE, the SBME may be preferred for its simplicity. Thus, 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 criterion. In some of these cases, a shrinkage estimator has no impact on the actual objective. For example, if the vector $\mathbf{x}$ is a reconstructed image, its subjective quality is hardly 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, the EBME must be used to improve performance.

The blind minimax approach was initially applied to the i.i.d. case, in which the noise is white [11]. In this paper, we have shown that the results can be generalized to the case of colored Gaussian noise and arbitrary transformation matrices. The analytical and empirical results we presented serve as a figure of merit for the proposed estimators in and of themselves. More importantly, they support the underlying concept of blind minimax estimation, which can be generalized to many other estimation problems, such as estimation with uncertain system matrices, estimation with non-Gaussian noise, and sequential estimation. Application of the blind minimax approach to these problems remains a topic for further study.

6. REFERENCES

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

[2] 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.

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

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

[5] 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.

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] ——, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Processing*, 2005, to appear. [Online]. Available: http://www.technion.ac.il/~zvikabh/published/max_set_est.pdf

[13] 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.

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