MINIMAX ESTIMATORS DOMINATING THE LEAST-SQUARES ESTIMATOR

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 present several analytical and numerical results demonstrating the superiority of minimax estimators over least-squares (LS) estimation. We show that, for any bounded parameter set, a linear minimax estimator achieves lower mean-squared error than the LS estimator, over the entire parameter set. When a parameter set is unknown, we propose to estimate the parameter set from the data, and show that in many cases, the obtained *blind minimax* estimator still dominates the LS estimator. The results are related to and compared with other LS-dominating estimators, such as the James-Stein estimator.

1. INTRODUCTION

Consider the system of observations $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{x}$ is a deterministic parameter vector, $\mathbf{H}$ is a known transformation matrix, and $\mathbf{w}$ is Gaussian random noise with zero mean and known covariance $\mathbf{C}_\mathbf{w}$. An estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ is a function of $\mathbf{y}$ designed to be close to $\mathbf{x}$ in some sense, for example, in terms of the mean-squared error (MSE) between $\mathbf{x}$ and $\hat{\mathbf{x}}$. The standard estimation technique in this case seeks the linear unbiased estimator minimizing the MSE; this approach leads to the well-known least-squares (LS) estimator [1]. However, by allowing the use of biased and nonlinear estimators, significant reduction in MSE may be achieved.

The goal of this paper is to develop estimators which dominate the LS estimator. An estimator is said to dominate the LS estimator if its MSE is never higher than that of the LS estimator, while being strictly lower for some values of $\mathbf{x}$ [2]. Thus, if an estimator dominates the LS estimator, it is always preferable in terms of MSE performance.

Our design is based on the use of minimax MSE estimators [3, 4]. These are estimators designed to minimize the worst-case estimation error, for all parameters $\mathbf{x}$ in a specified set $\mathcal{U}$. In Section 2, we show that for any bounded parameter set $\mathcal{U}$, the MSE obtained by the minimax estimator is lower than the MSE of the LS estimator, for *all* $\mathbf{x}$ in $\mathcal{U}$ [5].

We seek to apply this result to the general case, in which no parameter set $\mathcal{U}$ is known. In Section 3, we consider "blind minimax" estimators, which are minimax estimators whose parameter set is itself estimated from measurements. The result is a nonlinear estimator whose performance is superior to the LS estimator in many cases. We discuss two types of blind minimax estimators, and show that for many scenarios, both estimators dominate the LS estimator. This is done by showing that the blind minimax estimator is a modified version of estimators known to dominate the LS estimator [6–9]. For analytical tractability, the dominance proof is limited to the case $\mathbf{H} = \mathbf{I}$ and $\mathbf{C}_\mathbf{w} = \mathbf{I}$. In Section 4, the improved performance of the blind minimax estimator is illustrated numerically for more general cases. The results are summarized and discussed in Section 5.

2. MINIMAX ESTIMATION

Consider the system of measurements $\mathbf{y} \in \mathbb{C}^m$,

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

where $\mathbf{x} \in \mathbb{C}^n$ is an unknown deterministic vector, $\mathbf{H} \in \mathbb{C}^{m\times n}$ is a known full-rank matrix, and $\mathbf{w}$ is a zero-mean Gaussian random vector with known positive definite covariance $\mathbf{C}_\mathbf{w}$. We wish to construct an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$, such that the mean-squared error (MSE) is minimal. In this section we limit our discussion to linear estimators, $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ [1]. For such estimators, the MSE is given by

$$\epsilon(\hat{\mathbf{x}}, \mathbf{x}) = E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} = v(\hat{\mathbf{x}}) + \|\mathbf{b}(\hat{\mathbf{x}})\|^2, \tag{2}$$

where the variance $v(\hat{\mathbf{x}})$ is given by

$$v(\hat{\mathbf{x}}) = E\{\|\hat{\mathbf{x}} - E\hat{\mathbf{x}}\|^2\} = \mathrm{Tr}(\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^*), \tag{3}$$

and the bias $\mathbf{b}(\hat{\mathbf{x}})$ equals

$$\mathbf{b}(\hat{\mathbf{x}}) = E\{\mathbf{x} - \hat{\mathbf{x}}\} = (\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}. \tag{4}$$

Since $\mathbf{b}(\hat{\mathbf{x}})$ depends on the unknown value of $\mathbf{x}$, direct minimization of the MSE is not possible. A common approach is to limit discussion to unbiased estimators, in which case the MSE no longer depends on $\mathbf{x}$, and then seek the linear estimator that minimizes the MSE. This results in the least-squares (LS) estimator,

$$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{y}. \tag{5}$$

The MSE of the LS estimator is constant for all $\mathbf{x}$, and is given by

$$\epsilon_0 = \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}\right). \tag{6}$$

While the LS estimator is the best linear *unbiased* estimator, in many cases biased estimators exist which outperform the LS estimator in terms of MSE. In particular, if the parameter $\mathbf{x}$ is known to lie within a given parameter set $\mathcal{U}$, a linear minimax MSE estimator may be constructed, which minimizes the worst-case MSE within the parameter set $\mathcal{U}$ [3, 4]. Formally, a linear estimator $\hat{\mathbf{x}}_{\mathrm{M}}$ is a linear minimax MSE estimator for a parameter set $\mathcal{U}$ if, for any other linear estimator $\hat{\mathbf{x}}$,

$$\sup_{\mathbf{x}\in\mathcal{U}} E\{\|\hat{\mathbf{x}}_{\mathrm{M}} - \mathbf{x}\|^2\} \le \sup_{\mathbf{x}\in\mathcal{U}} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}. \tag{7}$$

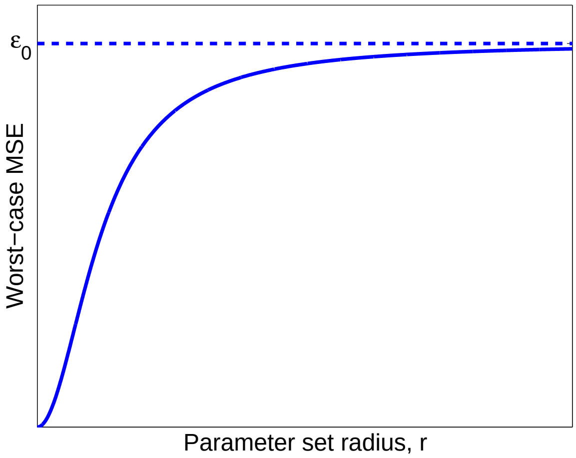

Fig. 1. Worst-case error for various minimax estimators.

The linear minimax MSE estimator for the spherical parameter set $\{\mathbf{x} : \|\mathbf{x}\| \le r\}$ is given by [4]

$$\hat{\mathbf{x}}_r = \frac{r^2}{r^2+\epsilon_0}(\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{y}. \tag{8}$$

As the parameter set radius r increases, the minimax estimator approaches the LS estimator. Indeed, the LS estimator is the linear minimax MSE estimator for the parameter set $\mathbb{C}^n$.

The following theorem demonstrates that the minimax estimator for any bounded set $\mathcal{U}$ outperforms the LS estimator, in that it provides lower MSE than the LS estimator for any $\mathbf{x}$ in $\mathcal{U}$.

Theorem 1. *Let $\mathcal{U}$ be a bounded parameter set, and let $\hat{\mathbf{x}}_\mathrm{M}$ be a linear minimax MSE estimator for the parameter set $\mathcal{U}$. Then, the MSE of $\hat{\mathbf{x}}_\mathrm{M}$ is lower than the MSE of the LS estimator* (5), *for all $\mathbf{x} \in \mathcal{U}$.*

Proof. For any bounded $\mathcal{U}$, there exists a finite r such that $\mathcal{U}$ is bounded within the sphere $\{\mathbf{x} : \|\mathbf{x}\| \le r\}$. The linear minimax MSE estimator for this sphere is given by (8). We now show that $\hat{\mathbf{x}}_r$ achieves a lower MSE than the LS estimator for all $\mathbf{x} \in \mathcal{U}$. The bias of $\hat{\mathbf{x}}_r$ is given by

$$\mathbf{b}(\hat{\mathbf{x}}_r) = E\{\hat{\mathbf{x}}_r - \mathbf{x}\} = (\beta - 1)\mathbf{x}, \tag{9}$$

where $\beta = \frac{r^2}{r^2+\epsilon_0}$. It follows from (3) that the variance of $\hat{\mathbf{x}}_r$ is

$$v(\hat{\mathbf{x}}_r) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^*) = \beta^2\epsilon_0. \tag{10}$$

Using (2), we have that for any $\mathbf{x} \in \mathcal{U}$,

$$\begin{aligned} \mathrm{MSE}(\hat{\mathbf{x}}_r) &= \beta^2\epsilon_0 + (1-\beta)^2\|\mathbf{x}\|^2 \\ &\le \beta^2\epsilon_0 + (1-\beta)^2 r^2 \\ &= \left(\frac{r^2}{r^2+\epsilon_0}\right)\epsilon_0 \\ &< \epsilon_0. \end{aligned} \tag{11}$$

Hence, for all $\mathbf{x} \in \mathcal{U}$, the MSE using $\hat{\mathbf{x}}_r$ is lower than the MSE for an unbiased estimator. From (7), it follows that

$$\mathrm{MSE}(\hat{\mathbf{x}}_\mathrm{M}) \le \mathrm{MSE}(\hat{\mathbf{x}}_r) < \epsilon_0 \quad \text{for all } \mathbf{x} \in \mathcal{U}, \tag{12}$$

which completes the proof. □

From Theorem 1, it follows that if the parameter $\mathbf{x}$ is known to lie in some bounded parameter set $\mathcal{U}$, then the minimax estimator is provably better than the LS estimator. This is illustrated in Figure 1, which plots the worst-case MSE obtained by minimax estimators using spherical parameter sets with various radii, for the case $\mathbf{H} = \mathbf{I}$ and $\mathbf{C}_\mathbf{w} = \mathbf{I}$; similar results are obtained for other settings. The worst-case MSE increases with the parameter set radius, and approaches the MSE ϵ_0 of the LS estimator as $r \to \infty$.

In the following section, we extend the use of minimax estimators to the case in which the parameter set is unknown and must be estimated from the measurements.

3. BLIND MINIMAX ESTIMATION

3.1. Definitions

Minimax estimators outperform the LS estimator because they are designed for a subset of all possible parameters; this subset is the parameter set $\mathcal{U}$. In many cases, however, no parameter set is known. In these situations, we propose to use a spherical parameter set, centered at the origin, whose radius r is itself estimated from the measurements. Thus we obtain a two-stage process:

1. Estimate the parameter set radius r from the measurements.
2. Estimate the parameter $\mathbf{x}$ using a minimax estimator whose parameter set is $\mathcal{U} = \{\mathbf{x} : \|\mathbf{x}\| \le r\}$.

As we shall see, this *blind minimax estimation* approach results in estimators which improve substantially upon the LS estimator, while using exactly the same input.

The parameter set radius indicates the degree of certainty of the value of $\mathbf{x}$: as r increases, less knowledge is assumed about $\mathbf{x}$. Thus we would like to choose r to be as small as possible, while still including the true value of $\mathbf{x}$; optimally, we would like to choose $r = \|\mathbf{x}\|$. Since $\mathbf{x}$ is unknown, we can instead use the LS estimate $\hat{\mathbf{x}}_\mathrm{LS}$, and select

$$r_\mathrm{d} = \|\hat{\mathbf{x}}_\mathrm{LS}\|. \tag{13}$$

Substituting r_d into the linear minimax MSE estimator (8), we obtain the *direct blind minimax estimator* (DBME)

$$\hat{\mathbf{x}}_\mathrm{d} = \frac{\|\hat{\mathbf{x}}_\mathrm{LS}\|^2}{\|\hat{\mathbf{x}}_\mathrm{LS}\|^2+\epsilon_0}\hat{\mathbf{x}}_\mathrm{LS}. \tag{14}$$

It turns out that r_d^2 tends to be an overestimate of $\|\mathbf{x}\|^2$ [10]. To see this, consider the expectation of r_d^2. Denoting $\mathbf{G}_\mathrm{LS} = (\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_\mathbf{w}^{-1}$, we have

$$\begin{aligned} E\{r_\mathrm{d}^2\} &= E\{\|\mathbf{G}_\mathrm{LS}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\} \\ &= E\{\mathrm{Tr}\left((\mathbf{x}+\mathbf{G}_\mathrm{LS}\mathbf{w})(\mathbf{x}+\mathbf{G}_\mathrm{LS}\mathbf{w})^*\right)\} \\ &= \mathrm{Tr}(\mathbf{x}\mathbf{x}^*) + E\{\mathrm{Tr}\left(\mathbf{G}_\mathrm{LS}\mathbf{w}\mathbf{w}^*\mathbf{G}_\mathrm{LS}^*\right)\} \\ &= \|\mathbf{x}\|^2 + \epsilon_0, \end{aligned} \tag{15}$$

where ϵ_0 is the MSE of the LS estimator (6). Thus, in the average case, r_d^2 overestimates $\|\mathbf{x}\|^2$ by ϵ_0. To see whether this results in performance degradation, we define the unbiased estimator

$$r_\mathrm{u}^2 = \|\hat{\mathbf{x}}_\mathrm{LS}\|^2 - \epsilon_0. \tag{16}$$

Note that r_u^2 may become negative when $\hat{\mathbf{x}}_\mathrm{LS}$ is small. Substituting r_u^2 into the linear minimax MSE estimator (8) yields

$$\hat{\mathbf{x}}_\mathrm{u} = \frac{\|\hat{\mathbf{x}}_\mathrm{LS}\|^2-\epsilon_0}{\|\hat{\mathbf{x}}_\mathrm{LS}\|^2}\hat{\mathbf{x}}_\mathrm{LS}. \tag{17}$$

We refer to this estimator as the *unbiased blind minimax estimator* (UBME). Although the estimator itself is biased, it is based on an unbiased estimate of $\|\mathbf{x}\|^2$.

3.2. Comparison with LS Estimation

We now compare the MSE obtained by the blind minimax estimators with the MSE of the LS estimator. For analytical tractability, we limit our discussion to the special case in which $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \mathbf{I}$. Our goal is to show that the blind minimax estimators *dominate* the LS estimator. An estimator $\hat{\mathbf{x}}_1$ is said to dominate $\hat{\mathbf{x}}_2$ if $\mathrm{MSE}(\hat{\mathbf{x}}_1) \le \mathrm{MSE}(\hat{\mathbf{x}}_2)$ for all $\mathbf{x}$, with strict inequality for some value of $\mathbf{x}$ [2]. The following theorems show that this is indeed the case.

Theorem 2. *Let $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \mathbf{I}$, and assume that $n \ge 4$. Then, the DBME* (14) *dominates the LS estimator* (5).

The proof of Theorem 2 is based on the following general result, which is due to Strawderman [8].

Lemma 1. *Let $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \mathbf{I}$, and assume that $n \ge 3$. Let $\rho(\lambda)$ be a non-decreasing function satisfying $0 \le \rho(\lambda) \le 2$. Then, the estimator*

$$\hat{\mathbf{x}} = \left(1 - \rho\left(\tfrac{1}{2}\|\mathbf{y}\|^2\right) \frac{n-2}{\|\mathbf{y}\|^2}\right) \mathbf{y} \tag{18}$$

dominates the LS estimator (5).

Proof of Theorem 2. Since $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \mathbf{I}$, we have $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{y}$ and $\epsilon_0 = n$. Substituting these values into (14) yields

$$\hat{\mathbf{x}}_{\mathrm{d}} = \left(\frac{\|\mathbf{y}\|^2}{\|\mathbf{y}\|^2 + n}\right) \mathbf{y}. \tag{19}$$

This is a modified version [11] of an estimator first proposed by Alam and Thompson [7],

$$\hat{\mathbf{x}}_{\mathrm{AT}} = \left(\frac{\|\mathbf{y}\|^2}{\|\mathbf{y}\|^2 + c}\right) \mathbf{y}, \quad 0 \le c \le n-2. \tag{20}$$

The Alam-Thompson estimator (20) is known to dominate the LS estimator [9]. We now use Lemma 1 to show that the modified estimator (19) also dominates the LS estimator. Consider the function

$$\rho(\lambda) = \left(\frac{n}{n-2}\right)\left(\frac{2\lambda}{2\lambda + n}\right). \tag{21}$$

The function $\rho(\lambda)$ is nondecreasing in λ, and for $n \ge 4$, we have $0 \le \rho(\lambda) \le 2$. Therefore, by Lemma 1, the estimator (18) dominates the LS estimator $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{y}$. Substituting (21) for $\rho(\cdot)$ yields the required estimator (19). ☐

Theorem 3. *Let $\mathbf{H} = \mathbf{I}$ and $\mathbf{C_w} = \mathbf{I}$. For $n > 4$, the UBME* (17) *achieves lower MSE than the LS estimator, for all $\mathbf{x}$. For $n = 4$, the UBME achieves the same MSE as the LS estimator, for all $\mathbf{x}$.*

Proof. Substituting $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{y}$ and $\epsilon_0 = n$ into (17), we obtain

$$\hat{\mathbf{x}}_{\mathrm{u}} = \left(1 - \frac{n}{\|\mathbf{y}\|^2}\right) \mathbf{y}. \tag{22}$$

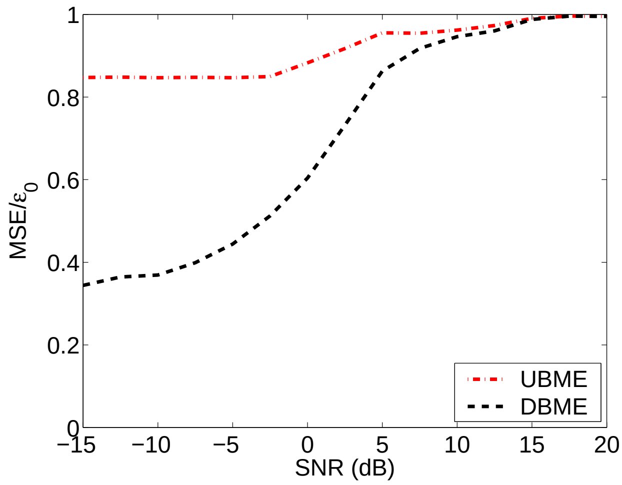

Fig. 2. MSE of blind minimax estimators (as a fraction of ϵ_0, the MSE of the LS estimator) for varying SNR levels.

This estimator was first proposed by Stein [12], and was further analyzed by James and Stein [6], who showed that its MSE is given by

$$\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{u}}) = n + n(4-n)E\left\{\frac{1}{n-2+2K}\right\}, \tag{23}$$

where K is a Poisson random variable with mean $\frac{1}{2}\|\mathbf{x}\|^2$. The expectation $E\{1/(n-2+2K)\}$ contains only positive terms and is therefore always positive. For $n > 4$, the expression $n\,(4-n)$ is negative, so that $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{u}}) < n$. For $n = 4$, we have $n(4-n) = 0$, so $\mathrm{MSE}(\hat{\mathbf{x}}_{\mathrm{u}}) = n$, completing the proof. ☐

As we have seen in Theorems 2 and 3, the proposed blind minimax estimators outperform the LS estimator for all values of the parameter $\mathbf{x}$, in the analytically simple case $\mathbf{H} = \mathbf{I}$, $\mathbf{C_w} = \mathbf{I}$. A numerical simulation of the general case follows in Section 4.

4. NUMERICAL RESULTS

We have seen in Section 3 that blind minimax estimators dominate the LS estimator for the special case $\mathbf{H} = \mathbf{I}$, $\mathbf{C_w} = \mathbf{I}$. In this section we provide numerical evidence demonstrating that this occurs for many cases of correlated measurements as well.

To test estimation performance, a setup with $m = n = 5$ was used. The parameter $\mathbf{x}$ was chosen as an independent, identically distributed Gaussian random variable with zero mean and unit variance. Correlation between measurements can equivalently arise either from $\mathbf{H} \ne \mathbf{I}$ or from $\mathbf{C_w} \ne \mathbf{I}$. We arbitrarily chose $\mathbf{H} = \mathbf{I}$ and

$$\mathbf{C_w} = \begin{pmatrix} b & \alpha b & \cdots & \alpha b \\ \alpha b & b & \cdots & \alpha b \\ \vdots & \vdots & \ddots & \vdots \\ \alpha b & \alpha b & \cdots & b \end{pmatrix}, \tag{24}$$

where α controls the noise correlation, and b controls the signal-to-noise ratio (SNR), defined as

$$\mathrm{SNR} = \frac{E\{\|\mathbf{x}\|^2\}}{\mathrm{Tr}(\mathbf{C_w})} = \frac{1}{b}. \tag{25}$$

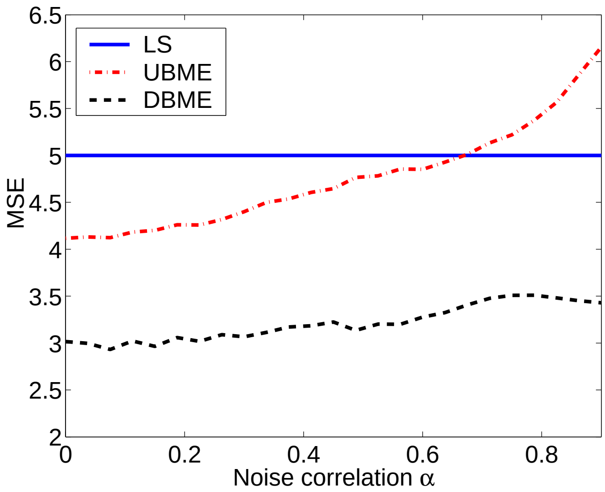

Fig. 3. Estimator performance for varying noise correlation levels.

The simulation was performed at a variety of SNR levels, with correlation $\alpha = \frac{1}{3}$. The results clearly show dominance of both blind minimax estimators over the LS estimator, for moderate correlation levels (see Figure 2). Both estimators converge to the LS estimator for high SNR.

Next, the effect of measurement correlation on estimator performance was examined. To this end, the simulation was repeated with various correlation values α, at a constant SNR of 0 dB (see Figure 3). These results indicate that the DBME outperforms the LS estimator for all correlation levels tested, while the UBME fails for high noise correlation levels.

5. SUMMARY AND DISCUSSION

The purpose of this paper is to illustrate the power of minimax estimators, and to analytically show that these estimators are superior to the commonly-used LS estimator. We began by showing that, for any bounded parameter set $\mathcal{U}$, linear minimax estimators achieve lower MSE than the LS estimator, for *any* $\mathbf{x}$ in $\mathcal{U}$. Being linear, these estimators require no more computational resources than the LS estimator. Thus, when a parameter set is known, minimax estimators should definitely be preferred over LS estimators.

Even when a parameter set is unknown, *blind* minimax estimators can still be utilized by first estimating a parameter set radius, then using a minimax estimator for the estimated parameter set. Because of the two-stage estimation process, the obtained estimators are nonlinear, but like the LS estimator, these estimators require only $O(nm)$ multiplications per estimation. We propose two methods for estimating the parameter set radius: directly estimating the radius with the LS estimator (DBME), or using an unbiased estimator for the radius (UBME). These estimators were compared with the LS estimator, both analytically and numerically. The analytical comparison was limited to the case $\mathbf{H} = \mathbf{I}$, $\mathbf{C_w} = \mathbf{I}$, and showed that when the number of variables n is greater than 4, both estimators dominate the LS estimator. Numerical comparison indicates that blind minimax estimators continue to dominate the LS estimator for other cases as well. The conditions required for dominance of the LS estimator in the general case are currently one of our research topics.

The parameter set radius, as estimated by the DBME, was shown to be an overestimate of the true value $\|\mathbf{x}\|$. This led to the development of the UBME, for which the radius estimate is unbiased. It is therefore somewhat surprising that the DBME outperforms the UBME in most cases (Figures 2 and 3). A possible explanation for this result is that prudence takes preference over accuracy: a conservative radius estimate yields a parameter set which almost always contains the parameter, while a tight parameter set does not. This result is all the more surprising considering that the UBME is related to the well-known James-Stein estimator [6], while the DBME is related to the seldom-used Alam-Thompson estimator [7]. The DBME does not dominate the UBME, as there are some extreme cases in which the UBME outperforms the DBME. However, our results indicate that the DBME may outperform the UBME in practical situations. Analytical comparison of these estimators is another topic for further research.

6. REFERENCES

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

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

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

[4] 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.

[5] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Processing*, 2005, to appear.

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

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

[8] W. E. Strawderman, "Proper Bayes minimax estimators of the multivariate normal mean," *Ann. Math. Statist.*, vol. 42, no. 1, pp. 385–388, Feb. 1971.

[9] 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.

[10] T. Dvorkind, 2004, personal communication.

[11] J. R. Thompson, "Some shrinkage techniques for estimating the mean," *J. Amer. Statist. Assoc.*, vol. 63, no. 321, pp. 113–122, Mar. 1968.

[12] 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.