MSE-RATIO REGRET ESTIMATION WITH BOUNDED DATA UNCERTAINTIES

Yonina C. Eldar

Department of Electrical Engineering
Technion–Israel Institute of Technology, Haifa, Israel
Email: yonina@ee.technion.ac.il.

ABSTRACT

We consider the problem of robust estimation of a deterministic bounded parameter vector $\mathbf{x}$ in a linear model. While in an earlier work, we proposed a minimax estimation approach in which we seek the estimator that minimizes the worst–case mean-squared error (MSE) *difference regret* over all bounded vectors $\mathbf{x}$, here we consider an alternative approach, in which we seek the estimator that minimizes the worst–case MSE *ratio regret*, namely, the worst–case *ratio* between the MSE attainable using a linear estimator ignorant of $\mathbf{x}$, and the minimum MSE attainable using a linear estimator that knows $\mathbf{x}$. The rational behind this approach is that the value of the difference regret may not adequately reflect the estimator performance, since even a large regret should be considered insignificant if the value of the optimal MSE is relatively large.

1. INTRODUCTION

The classical least-squares estimator for estimating an unknown parameter vector $\mathbf{x}$ in a linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known matrix and $\mathbf{w}$ is a noise vector, is well-known to result, in many cases, in a large residual mean-squared error (MSE). This observation has motivated the search for alternative linear estimators of $\mathbf{x}$, when no prior statistics on $\mathbf{x}$ are available.

If $\mathbf{x}$ is deterministic, then the MSE of an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ will in general depend explicitly on $\mathbf{x}$, and therefore cannot be minimized directly. A possible design approach in this case, is in the spirit of the minimax MSE approach initiated by Huber [1], in which the estimator is chosen to minimize the worst-case MSE over all values of $\mathbf{x}$, in the region of uncertainty [2, 3]. However, this approach is pessimistic in nature, since it optimizes the performance for the worst-possible choice of parameters, which may in turn result in a loss of performance for all other cases.

To improve the performance over the minimax MSE approach, in an earlier work [4], we considered the case in which $\mathbf{x}$ is known to satisfy a (possibly weighted) norm constraint, and developed a competitive minimax MSE estimator that minimizes the worst-case *difference regret*, which is the difference between the MSE of the linear estimator ignorant of $\mathbf{x}$, and the smallest attainable MSE with a linear estimator that knows $\mathbf{x}$. The motivation behind this estimator is that such an estimator performs uniformly as close as possible to the optimal linear estimator, in the region of uncertainty.

A possible drawback of the minimax difference regret (MDR) estimator is that the value of the regret may not adequately reflect the estimator performance, since even a large regret should be considered insignificant if the value of the optimal MSE is relatively large. On the other hand, if the optimal MSE is small, then even a small regret should be considered significant. Therefore, in this paper, instead of considering the worst-case difference regret, we suggest a minimax ratio regret (MRR) estimator that minimizes the worst-case *ratio* between the MSE of a linear estimator that does not know $\mathbf{x}$, and the best possible MSE.

In Section 2, we show that the MRR estimator can be described by m parameters, which are the solution to a convex optimization problem. We then specialize the results, in Section 3, to two special choices of the weighting matrix. In the first choice, $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$, which may be of interest, for example, when an unknown signal $\mathbf{x}$ is sent through a known channel, and the output signal-to-noise ratio is bounded. In the second choice, $\mathbf{T} = \mathbf{I}$, which may be of interest when an unknown signal $\mathbf{x}$ is sent through a known channel, and the power of $\mathbf{x}$ is bounded. In these special cases, we show that the MRR estimator can be derived as the solution to explicit, simple, and computationally tractable convex optimization problems. In Section 4, we present an alternative derivation of the MRR estimator, and show, that in some cases, it leads to an (almost) closed form solution for the optimal estimator. Section 5 presents several examples illustrating the performance advantage of the MRR estimator over the MDR estimator, and other conventional linear estimators.

Proofs of theorems, which are omitted here for brevity, can be found in [5].

2. THE MINIMAX RATIO ESTIMATOR

Consider the problem of estimating the unknown deterministic parameter vector $\mathbf{x}$ in the linear model

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

where $\mathbf{H}$ is a known $n \times m$ matrix with full rank m, and $\mathbf{w}$ is a zero-mean random vector with known covariance $\mathbf{C}_w$. We assume that $\mathbf{x}$ satisfies the weighted norm constraint $M \leq \|\mathbf{x}\|_{\mathbf{T}} \leq L$ for some positive definite matrix $\mathbf{T}$ and scalars $0 < M \leq L$, where $\|\mathbf{x}\|^2_{\mathbf{T}} = \mathbf{x}^*\mathbf{T}\mathbf{x}$ and $(\cdot)^*$ denotes the Hermitian conjugate.

We estimate $\mathbf{x}$ using a linear estimator so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m \times n$ matrix $\mathbf{G}$. We would like to design an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ to minimize the MSE, which is given by

$$E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}. \tag{2}$$

Since the MSE depends explicitly on the unknown $\mathbf{x}$, we cannot choose an estimate to directly minimize the MSE (2).

To develop a competitive estimator, we consider a minimax ratio criterion, in which the estimator is obtained by minimizing the worst-case *ratio* between the MSE of a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ of $\mathbf{x}$, and the smallest possible nonzero

MSE attainable using any estimator of the form $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$ where $\mathbf{x}$ is assumed to be known, so that $\mathbf{G}$ can depend explicitly on $\mathbf{x}$. Since we are restricting ourselves to estimators of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, even in the case in which $\mathbf{x}$ is known, we cannot in general achieve zero MSE (unless $\mathbf{x} = \mathbf{0}$).

It was shown in [4] that for known $\mathbf{x}$ the optimal estimator is given by $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{x})\mathbf{y}$, where

$$\mathbf{G}(\mathbf{x}) = \frac{1}{1+\mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{x}}\mathbf{x}\mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{3}$$

and the smallest possible MSE is

$$\text{MSE}^o = \frac{\mathbf{x}^*\mathbf{x}}{1+\mathbf{x}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{x}}. \tag{4}$$

Thus, we seek the matrix $\mathbf{G}$ that is the solution to the problem

$$\min_{\mathbf{G}} \max_{M^2 \le \mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2} \frac{E(\|\mathbf{G}\mathbf{y}-\mathbf{x}\|^2)}{\text{MSE}^o}, \tag{5}$$

where MSE^o is given by (4).

For analytical tractability, we restrict our attention to weighting matrices $\mathbf{T}$ such that $\mathbf{T}$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix. Thus, if $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ has an eigendecomposition $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and Σ is a diagonal matrix, then $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$ for some diagonal matrix Λ. Theorem 1 below establishes the general form of the solution to (5) for any such $\mathbf{T}$.

Theorem 1 *Let $\mathbf{x}$ denote the unknown deterministic vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and Σ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_i > 0$ and let $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$ where Λ is an $m \times m$ diagonal matrix with diagonal elements $\lambda_i > 0$. Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{M \le \|\mathbf{x}\|_{\mathbf{T}} \le L} \frac{E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}$$

has the form

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i which are the solution to the convex optimization problem

$$(\Gamma): \min_{\gamma, d_i}\left\{\gamma : \max_{\mathbf{s}\in\mathscr{S}}\left\{\sum_{i=1}^m (1-d_i)^2 s_i - \gamma\frac{\sum_{i=1}^m s_i}{1+\sum_{i=1}^m \sigma_i s_i}\right\} + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0\right\}, \tag{6}$$

with

$$\mathscr{S} = \left\{\mathbf{s}\in\mathscr{R}^m, \mathbf{s}\ge 0 \,\middle|\, \sum_{i=1}^m \lambda_i s_i = M^2 \text{ or } \sum_{i=1}^m \lambda_i s_i = L^2\right\}. \tag{7}$$

Theorem 1 reduces the problem of minimizing the ratio regret to the simpler optimization problem (6). As we show in Sections 3.1 and 3.2, for certain choices of $\mathbf{T}$, the problem can be further simplified. In Section 4 we consider a general method for solving (6) that exploits its connection with a related convex optimization problem. We then demonstrate, in Section 4.1, that in some cases this approach can lead to further insight into the MRR estimator.

3. MRR ESTIMATOR FOR SOME CHOICES OF $\mathbf{T}$

3.1 MRR Estimator For $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$

We first consider the case in which $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$, so that the eigenvalues λ_i of $\mathbf{T}$ are equal to the eigenvalues σ_i of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. The MRR estimator in this case is given by the following theorem.

Theorem 2 *Let $\mathbf{x}$ denote the unknown deterministic vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and Σ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_1 \ge \ldots \ge \sigma_m > 0$. Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{M \le \|\mathbf{x}\|_{\mathbf{T}} \le L} \frac{E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}$$

with $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ is given by

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i that are the solution to

$$\min_{\gamma, d_i, y, z}\left\{\gamma : \begin{array}{c} L^2 y + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0 \\ \frac{1}{\sigma_i}\left((1-d_i)^2 - \frac{\gamma}{1+L^2}\right) \le y, \quad 1 \le i \le m \\ L^2 y + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0 \\ \frac{1}{\sigma_i}\left((1-d_i)^2 - \frac{\gamma}{1+L^2}\right) \le y, \quad 1 \le i \le m \\ M^2 z + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0 \\ \frac{1}{\sigma_i}\left((1-d_i)^2 - \frac{\gamma}{1+M^2}\right) \le z, \quad 1 \le i \le m \end{array}\right\}.$$

3.2 MRR Estimator For $\mathbf{T} = \mathbf{I}$

Theorem 3 below considers the MRR estimator for $\mathbf{T} = \mathbf{I}$.

Theorem 3 *Let $\mathbf{x}$ denote the unknown deterministic vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and Σ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_1 \ge \ldots \ge \sigma_m > 0$. Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{M \le \|\mathbf{x}\| \le L} \frac{E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}$$

has the form

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i that are the solution to

$$\min_{\gamma, d_i, \mu, \lambda, \zeta, \eta}\left\{\gamma : \begin{array}{c} -2L\sqrt{\mu\gamma} + \mu + L^2\lambda + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0 \\ (1-d_i)^2 + \sigma_i\mu \le \lambda, \quad 1 \le i \le m \\ \mu \ge 0 \\ -2M\sqrt{\zeta\gamma} + \zeta + M^2\eta + \sum_{i=1}^m \frac{d_i^2}{\sigma_i} \le 0 \\ (1-d_i)^2 + \sigma_i\zeta \le \eta, \quad 1 \le i \le m \\ \zeta \ge 0. \end{array}\right\}.$$

If in addition $M = L$, then the elements d_i are given by

$$d_i = \begin{cases} 1 - \sqrt{\lambda - \sigma_i \mu}, & i \le k; \\ 0, & i \ge k+1, \end{cases}$$

with $k = \arg\min \gamma_i$, $\mu = \mu_k$ and $\lambda = \lambda_k$. Here $\gamma_i, \mu_i, \lambda_i, 1 \le i \le m$ are the optimal solutions to the problem (Γ_i) given by

$$(\Gamma_i): \min_{\gamma,\mu,\lambda} \left\{ \gamma : \begin{array}{c} -2L\sqrt{\mu}\gamma + \mu + L^2\lambda + \sum_{j=1}^{i} \frac{\left(1-\sqrt{\lambda-\sigma_j\mu}\right)^2}{\sigma_j} \le 0 \\ \sigma_1\mu \le \lambda \le 1 + \sigma_i\mu \\ \mu \ge 0 \\ \text{if } i < m \text{ then } \lambda \ge 1 + \sigma_{i+1}\mu \end{array} \right\}.$$

4. ALTERNATIVE DERIVATION

In this section, we develop further insight into the MRR estimator, by developing an alternative formulation of the estimator. In particular, we show that the MRR estimator of Theorem 1 with d_i given as the solution to the problem (Γ) of (6), can be determined by first solving the simpler problem

$$(\Phi): \min_{t,d_i} \left\{ t : \max_{\mathbf{s}\in\mathscr{S}} \left\{ \sum_{i=1}^{m} (1-d_i)^2 s_i - \gamma \frac{\sum_{i=1}^{m} s_i}{1+\sum_{i=1}^{m}\sigma_i s_i} \right\} + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \le t \right\}, \quad (8)$$

where $\mathscr{S}$ is given by (7) and $\gamma \ge 1$ is fixed. Note, that (Φ) is equivalent to

$$\min_{d_i} \left\{ \max_{\mathbf{s}\in\mathscr{S}} \left\{ \sum_{i=1}^{m} (1-d_i)^2 s_i - \gamma \frac{\sum_{i=1}^{m} s_i}{1+\sum_{i=1}^{m}\sigma_i s_i} \right\} + \sum_{i=1}^{m} \frac{d_i^2}{\sigma_i} \right\}, \quad (9)$$

which has one less variable than the problem (Γ) of (6).

Let $\hat{t}(\gamma)$ denote the optimal value of t in the problem (Φ) of (8), and let $\hat{\gamma}$ be the unique value of $\gamma \ge 1$ such that $\hat{t}(\gamma) = 0$ (as we show below in Proposition 1, such a γ always exists, and is unique). Then, denoting by $\hat{d}_i$ the optimal value of d_i in the problem (Φ) with $\gamma = \hat{\gamma}$, we now show that $\hat{d}_i$ and $\hat{\gamma}$ are the optimal solutions to the problem (Γ) of (6): Since $\hat{d}_i$ and $\hat{\gamma}$ are feasible for (Φ) with $t = 0$, they are also feasible for (Γ). Now suppose, conversely, that there exists feasible d_i and $\gamma < \hat{\gamma}$ for (Γ). It then follows that $\hat{t}(\gamma) \le 0$. But since $\hat{t}(\gamma)$ is decreasing in γ and $\gamma < \hat{\gamma}$, we have that $\hat{t}(\gamma) \ge \hat{t}(\hat{\gamma}) = 0$, from which we conclude that $\hat{t}(\gamma) = 0$, which is a contradiction since $\hat{\gamma}$ is the *unique* value for which $\hat{t}(\gamma) = 0$. Therefore, to solve (Γ) we may first solve the simpler problem (Φ), and then find $\hat{\gamma}$ by a simple line search, for example using bisection. Specifically, we may start by choosing $\gamma = 1$. For each choice of γ we compute $\hat{t}(\gamma)$. If $\hat{t}(\gamma) > 0$, then we increase γ, and if $\hat{t}(\gamma) < 0$, then we decrease γ, continuing until $\hat{t}(\gamma) = 0$. Due to the continuity and monotonicity properties of $\hat{t}(\gamma)$, established in Proposition 1 below, the algorithm is guaranteed to converge.

Proposition 1 *Let $\hat{t}(\gamma)$ denote the optimal value of t in the problem (Φ) of (8). Then*

1. *$\hat{t}(\gamma)$ is continuous in γ;*
2. *$\hat{t}(\gamma)$ is strictly decreasing in γ;*
3. *there is a unique value of γ for which $\hat{t}(\gamma) = 0$.*

Thus, instead of solving the problem (Γ) of (6), we may solve the problem (Φ) of (8), which in some cases may provide more insight into the solution. To illustrate the possible advantage of this approach, in the next section we consider the case in which $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $L = M$, and show that this approach leads to new insight into the optimal solution.

4.1 Alternative Derivation For $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$

By exploiting the connection between problems (6) and (8) for $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $L = M$, we now show that the MRR estimator can be expressed in terms of two parameters, which can be found using an inner and outer line search algorithm.

Theorem 4 *Let $\mathbf{x}$ denote the unknown deterministic vector in the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w$. Let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\mathbf{V}$ is a unitary matrix and Σ is an $m \times m$ diagonal matrix with diagonal elements $\sigma_1 \ge \ldots \ge \sigma_m > 0$. Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}\|_{\mathbf{T}}=L} \frac{E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}$$

with $\mathbf{T} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ is given by

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ is an $m \times m$ diagonal matrix with diagonal elements

$$d_i = \begin{cases} 0, & i \le k(\rho_*); \\ 1 - \sqrt{\rho_*\sigma_i + \gamma/(L^2+1)}, & i \ge k(\rho_*) + 1. \end{cases}$$

Here ρ_* is the unique positive root of $\phi(\rho) = \sum_{i=1}^{m} \sigma_i \eta_i(\rho) - L^2$, with $\eta_i(\rho)$ given by

$$\eta_i(\rho) = \max\left[\frac{1}{\sigma_i}\left(\frac{1}{\sqrt{\rho\sigma_i + \gamma/(L^2+1)}} - 1 \right), 0 \right],$$

$$k(\rho_*) = \begin{cases} 0, & \eta_1(\rho_*) > 0; \\ \max\{k : \eta_k(\rho_*) = 0\}, & \text{otherwise}, \end{cases}$$

and γ is chosen such that $\hat{t}(\gamma) = 0$, where

$$\hat{t}(\gamma) = -\frac{\gamma}{1+L^2} \sum_{i=1}^{m} \hat{\eta}_i + \sum_{i=1}^{m} \frac{\hat{\eta}_i}{1+\sigma_i\hat{\eta}_i},$$

with $\hat{\eta}_i = \eta_i(\rho_*)$.

From Theorem 4, it follows that the MRR estimator can be found by using two bisection algorithms. In the first, γ is fixed and bisection is used to find the optimal ρ_*. In the second, bisection is used to find the optimal γ satisfying $\hat{t}(\gamma) = 0$.

5. EXAMPLES

To illustrate the MRR estimator, we consider the problem of estimating a 2D image from noisy observations, which are obtained by blurring the image with a 2D filter, and adding random Gaussian noise. Specifically, we generate an image $x(z_1, z_2)$ which is the sum of m harmonic oscillations:

$$x(z_1, z_2) = \sum_{\ell=1}^{m} a_\ell \cos(\omega_{\ell,1} z_1 + \omega_{\ell,2} z_2 + \phi_\ell), \quad (10)$$

where $\omega_{\ell,i} = 2\pi k_{\ell,i}/n$, and $k_{\ell,i} \in \mathbb{Z}^2$ are given parameters. Clearly, the image $x(z_1, z_2)$ is periodic with period n. Therefore, we can represent the image by a length-n^2 vector $\mathbf{x}$, with components $\{x(z_1, z_2) : 0 \le z_1, z_2 \le n-1\}$.

$(k_{\ell,1}, k_{\ell,2})$	a_ℓ	ϕ_ℓ
(1,1)	1.0681	2.1438
(2,1)	0.8704	3.3557
(1,2)	1.2027	4.5686
(2,2)	1.0466	1.9433
(3,2)	0.9449	5.2684

Table 1: Simulation parameters.

Estimator	Relative Error
LS	5.0e8
MMX	1.000
MDR	0.843
MRR	0.120

Table 2: Relative error for the data of Table 1.

The observed image $y(z_1,z_2)$ is given by

$$y(z_1,z_2) = \sum_{\tau_1,\tau_2} H(\tau_1,\tau_2) x(z_1-\tau_1-d_1, z_2-\tau_2-d_2) + \sigma w(z_1,z_2),$$

where $H(z_1,z_2)$ is a blurring filter defined by

$$H(z_1,z_2) = \max\left(1 - \frac{\sqrt{z_1^2+z_2^2}}{\rho}, 0\right), \qquad (11)$$

for some parameter ρ, d_1 and d_2 are randomly chosen shifts, and $w(z_1,z_2)$ is an independent, zero-mean, Gaussian noise process so that for each z_1 and z_2, $w(z_1,z_2)$ is $\mathcal{N}(0,1)$. By defining the vectors $\mathbf{y}$ and $\mathbf{w}$ with components $y(z_1,z_2)$ and $w(z_1,z_2)$, respectively, and defining a matrix $\mathbf{H}$ with the appropriate elements $H(z_1,z_2)$, the observations $\mathbf{y}$ can be expressed in the form (1).

In Fig. 1 we consider the case in which $m = 5$, $n = 128$, $\sigma = 0.5$, $L = \|\mathbf{x}\|$, and $n\rho = \sqrt{2}$. The values of $k_{\ell,i}, a_\ell$ and ϕ_ℓ are given in Table 1. To estimate the image $x(z_1,z_2)$ from the noisy observations $y(z_1,z_2)$ we consider 4 different estimators: The least-squares (LS) estimator, the MRR estimator of Theorem 3, the MDR estimator of [4], and the minimax MSE estimator (MMX), which is designed to minimize the worst-case MSE over all $\mathbf{x}^*\mathbf{x} \le L^2$, and is given by [3]

$$\hat{\mathbf{x}} = \frac{L^2}{L^2 + \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\right)} (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}^*)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \qquad (12)$$

In Table 2 we report the relative error $\varepsilon = \|\hat{\mathbf{x}} - \mathbf{x}\|/\|\mathbf{x}\|$ corresponding to the 4 estimators. The surprising result is that even though in this example the matrix $\mathbf{H}$ is ill-conditioned, the MRR estimator works pretty well, as can be seen from the results of Table 2, as well as in Fig. 1. Since the error in the LS estimate is so large, we do not show the resulting image. In the images, the "more red" the image, the larger the signal value at that point. As can be seen from the results of Table 2, as well as in Fig. 1, the MRR estimator outperforms the MDR estimator in all of the examples. We observed similar trends in the behavior for different values of the noise variance. In Table 3 we report the relative errors for different values of σ.

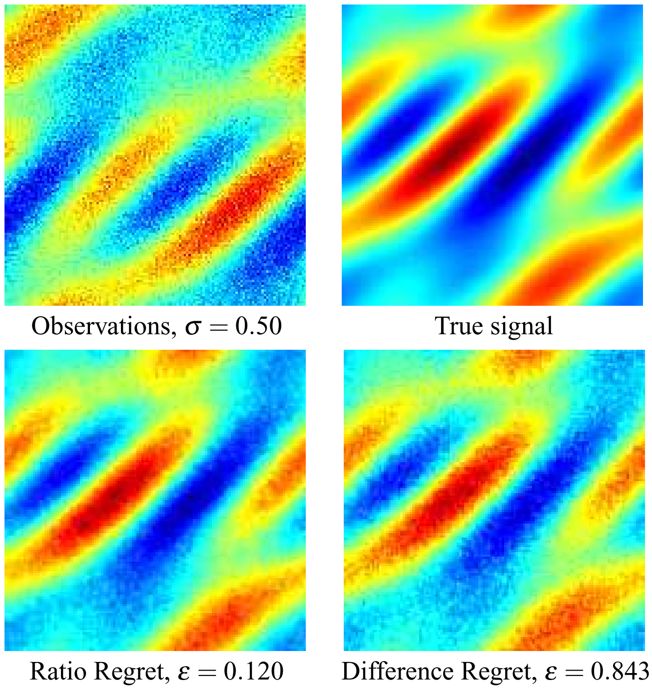

Figure 1: Comparison of the minimax regret estimators.

σ	Estimator	Relative Error
1.5	MMX	1.00
	MDR	0.92
	MRR	0.31
5	MMX	1.00
	MDR	0.96
	MRR	0.69

Table 3: Relative error for different values of σ.

6. ACKNOWLEDGMENT

The author would like to thank Prof. Arakadi Nemirovski for his help with the simulations.

REFERENCES

[1] P. J. Huber, *Robust Statistics*, New York: NY, John Wiley & Sons, Inc., 1981.

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

[3] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation with bounded data uncertainties," *IEEE Trans. Signal Processing*, to appear.

[4] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation with bounded data uncertainties," *IEEE Trans. Signal Processing*, to appear.

[5] Y. C. Eldar, "Minimax MSE-ratio estimation with bounded data uncertainties," submitted to *IEEE Trans. Signal Processing*, 2003.