MSE ESTIMATION OF MULTICHANNEL SIGNALS WITH MODEL UNCERTAINTIES

Amir Beck, Yonina C. Eldar and Aharon Ben-Tal

Technion–Israel Institute of Technology
Haifa, 32000, Israel
{becka@tx, yonina@ee, abental@ie}.technion.ac.il

ABSTRACT

We consider the problem of multichannel estimation, in which we seek to estimate multiple input vectors that are observed through a set of linear transformations and corrupted by additive noise. The input vectors $\mathbf{x}_k$ are known to satisfy a weighted norm constraint. We discuss both the case where the linear transformations are fixed (certain) and the case where they are only known to reside in some deterministic uncertainty set. We seek the linear estimator that minimizes the worst-case mean-squared error (MSE) across all possible values of the linear transformations and possible values of $\mathbf{x}_k$. We show that for an arbitrary choice of weighting matrix, the minimax MSE estimator can be formulated as a solution to a semidefinite programming problem (SDP). In the case in which the linear transformations are fixed and the norms are unweighed, the minimax MSE multichannel estimator has an explicit closed from solution. Finally, we demonstrate through examples, that the minimax MSE estimator can significantly increase the performance over conventional least-squares based methods.

1. INTRODUCTION

Estimation of multiple signals from multiple outputs is an important problem that appears in a variety of applications, such as blind multichannel estimation [1], speech separation [2] and image restoration [3]. In a multichannel estimation problem, we seek to estimate multiple input vectors $\{\mathbf{x}_k, 0 \le k \le N-1\}$, that are observed through a set of linear transformations $\mathbf{H}_{i,k}$ and corrupted by additive noise. Thus, the ith output vector $\mathbf{y}_i$ is given by the superposition $\mathbf{y}_i = \sum_{k=0}^{N-1} \mathbf{H}_{i,k}\mathbf{x}_k + \mathbf{w}_i$, where $\mathbf{H}_{i,k}$ is the transfer function from the kth input $\mathbf{x}_k$ to the ith output $\mathbf{y}_i$ and $\mathbf{w}_i$ is the ith noise vector.

If the second order statistics of the input vectors and the noise vectors are known, then we can design an estimator to minimize the mean-squared error (MSE). The resulting estimator is the well-known Wiener estimator. However, if the prior statistics of the input vectors are unknown, or if the linear transformations are not known exactly, then the Wiener estimator cannot be implemented.

A straightforward approach to deterministic multichannel estimation in the case of known linear transformations is the least-squares (LS) approach, in which the estimator is designed to minimize the norm of the data error, which is the sum of the norms of the differences between each of the observation vectors $\mathbf{y}_k$ and the corresponding estimated observation vector $\hat{\mathbf{y}}_k$. However, in an estimation context, the objective typically is to minimize the size of the norms of the estimation errors $\hat{\mathbf{x}}_k - \mathbf{x}_k$, where $\hat{\mathbf{x}}_k$ is an estimate of $\mathbf{x}_k$, rather than that of the data error. To develop an estimation method that is based directly on the estimation error, we may seek the estimator that minimizes the MSE, which is equal to the sum of the variance and the squared norm of the bias. However, since the bias generally depends on the unknown parameters $\mathbf{x}_k$, we cannot choose an estimator to directly minimize the MSE.

In this paper we consider the case in which the weighted norm of the unknown vectors $\mathbf{x}_k$ is bounded, *i.e.,* $\|\mathbf{x}_k\|_T = \mathbf{x}_k^*\mathbf{T}\mathbf{x}_k \le L^2$ for some constant L and positive definite weighting matrix $\mathbf{T}$ and we assume that $\mathbf{H}_{i,k}$ are not known exactly but rather given by $\mathbf{H}_{i,k} + \Delta_{i,k}$ where $\mathbf{H}_{i,k}$ is known and $\Delta_{i,k}$ is an unknown perturbation matrix satisfying a norm bound constraint $\|\Delta_{i,k}\| \le \rho_{i,k}$. We then develop a minimax MSE estimator that minimizes the worst case MSE across all possible bounded values of the input vectors $\mathbf{x}_k$ and the transfer matrices $\mathbf{H}_{i,k}$. The multichannel minimax MSE estimator is an extension of the recently proposed linear minimax MSE estimator for the single channel case [4].

When deriving the minimax MSE estimator we assume that the norm bounds L and $\rho_{i,k}$ are known. However, our algorithms can also be implemented when L and $\rho_{i,k}$ are not known, by first estimating them from the data. Thus, in practice, no prior information is needed for implementing the proposed estimators. Our experimental results suggest that with these estimated L and $\rho_{i,k}$, the minimax MSE estimator can significantly increase the performance over conventional methods.

In our development, the multichannel transfer matrix $\mathbf{H}$ with block matrices $\mathbf{H}_{i,k}$ is assumed to be a block circulant matrix, so that $\mathbf{H}_{i,k}$ is equal to $\mathbf{H}_{(k-i)modN}$ where N is the number of channels. The block circulant model has also been used in the context of image restoration [5], and in the context of cyclic convolution filter banks [6]. Moreover, in many practical scenarios it is reasonable to assume that $\mathbf{H}$ is a block Toeplitz matrix so that $\mathbf{H}_{i,k} = \mathbf{H}_{i-k}$. Using the well know convergence properties of Toeplitz matrices [7, 8], we can approximate the block-Toeplitz matrix $\mathbf{H}$ by a block circulant matrix. In Section 3 we show that if $\mathbf{H}$ and the covariance matrix $\mathbf{C}$ of the noise vectors are block circulant, then with $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, where $\hat{\mathbf{x}}$ is the concatenation of the estimated inputs $\hat{\mathbf{x}}_k$, $\mathbf{G}$ can also be chosen as a block circulant matrix. The blocks of $\mathbf{G}$ are a solution to a semidefinite programming problem (SDP), which is a tractable convex optimization problem that can be solved efficiently [9]. We then develop, in Section 4, a closed form solution to the minimax MSE estimation problem for the case where the weighting matrix $\mathbf{T}$ is equal to $\mathbf{I}$ and $\mathbf{H}$ is known. Finally, we demonstrate through examples, in Section 5, that both in the case of certain $\mathbf{H}$ and in the case of unknown $\mathbf{H}$, the minimax MSE estimator can significantly increase the performance over the multichannel LS, nonlinear regularized LS and the Structured Total LS approaches.

Proofs of theorems, which are omitted here for brevity, and a detailed discussion on block circulant matrices and Discrete Fourier

0-7803-8874-7/05/$20.00 ©2005 IEEE

Transforms can be found in [10].

2. PROBLEM FORMULATION AND NOTATION

We denote vectors by boldface lowercase letters and matrices by boldface uppercase letters. The identity matrix is denoted by $\mathbf{I}$, $(\cdot)^*$ and $(\cdot)^T$ denote the Hermitian conjugate and the transpose of the corresponding matrices respectively. For two Hermitian matrices $\mathbf{A}, \mathbf{B}$ the notation $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. The boldface letter $\mathbf{i}$ denotes $\sqrt{-1}$. For an Hermitian matrix $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$. We denote by $\|\mathbf{v}\|$ the Euclidean norm of the vector $\mathbf{v}$ and by $\|\mathbf{A}\| = \sqrt{\mathrm{Tr}(\mathbf{A}^T\mathbf{A})}$ the Frobenius norm of the matrix $\mathbf{A}$.

A matrix $\mathbf{A}$ is *block circulant* (BC) if

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{N-1} \\ \mathbf{A}_{N-1} & \mathbf{A}_0 & \cdots & \mathbf{A}_{N-2} \\ \vdots & & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_0 \end{pmatrix}. \tag{1}$$

The set of all BC matrices is denoted by $\mathcal{B}$. Two interesting special cases of BC matrices are the *elementary BC matrix* in which $\mathbf{A}_1 = \mathbf{A}_2 = \ldots \mathbf{A}_{N-1}$, and the *block diagonal matrix* in which $\mathbf{A}_1 = \mathbf{A}_2 = \ldots \mathbf{A}_{N-1} = \mathbf{0}$. The *Discrete Fourier Transform* (DFT) of a BC matrix given by (1) is a BC matrix with block matrices $\widehat{\mathbf{A}}_0, \ldots, \widehat{\mathbf{A}}_{N-1}$ where

$$\widehat{\mathbf{A}}_j \stackrel{\triangle}{=} \sum_{k=0}^{N-1} \omega^{kj}\mathbf{A}_k, \quad 0 \le j \le N-1, \tag{2}$$

and $\omega = e^{-\frac{2\pi\mathbf{i}}{N}}$.

Consider the problem of estimating N unknown deterministic parameter vectors $\mathbf{x}_k \in \mathbb{C}^m, 0 \le k \le N-1$ from N vector observations $\mathbf{y}_k \in \mathbb{C}^n, 0 \le k \le N-1$, where

$$\mathbf{y} = (\mathbf{H} + \Delta)\mathbf{x} + \mathbf{w}. \tag{3}$$

Here $\mathbf{y} = (\mathbf{y}_0^T, \ldots, \mathbf{y}_{N-1}^T)^T$, $\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T$, $\mathbf{H}$ is a known $nN \times mN$ matrix, Δ is an unknown $nN \times mN$ perturbation matrix and $\mathbf{w} = (\mathbf{w}_0^T, \ldots, \mathbf{w}_{N-1}^T)^T$ is a zero-mean random vector with covariance $\mathbf{C}$. We assume that $\mathbf{H}, \mathbf{C}$ and Δ are BC matrices. An interesting special case of (3) is the multiple observation problem, in which $\mathbf{H}$ and Δ are both block diagonal. The problem then reduces to that of estimating N input vectors $\mathbf{x}_k$ from N observations $\mathbf{y}_k$, given by

$$\mathbf{y}_k = (\mathbf{H}_0 + \Delta_0)\mathbf{x}_k + \mathbf{w}_k, \quad 0 \le k \le N-1. \tag{4}$$

Another interesting case is when $\mathbf{H}$ and Δ are elementary BC matrices. In this case the observations are given by

$$\mathbf{y}_k = (\mathbf{H}_0 + \Delta_0)\mathbf{x}_k + \sum_{i \ne k}(\mathbf{H}_1 + \Delta_1)\mathbf{x}_i + \mathbf{w}_k, \quad 0 \le k \le N-1, \tag{5}$$

which implies that the within channel transfer function, *i.e.,* the transfer function between each input vector and the corresponding output vector, is identical ($\mathbf{H}_{i,i} = \mathbf{H}_0$), and the cross channels are also equal ($\mathbf{H}_{i,j} = \mathbf{H}_1, i \ne j$).

The set of all possible values of Δ is

$$\mathcal{U}_\Delta \stackrel{\triangle}{=} \{\Delta \in \mathcal{B} : \|\Delta_k\| \le \rho_k, 0 \le k \le N-1\} \tag{6}$$

and the set of all possible values of $\mathbf{x}$ is

$$\mathcal{U}_\mathbf{x} \stackrel{\triangle}{=} \{\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T : \|\mathbf{x}_k\|_\mathbf{T} \le L, 0 \le k \le N-1\}. \tag{7}$$

We estimate $\mathbf{x}$ using a linear estimator so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $mN \times nN$ matrix $\mathbf{G}$. The MSE of the estimator is

$$E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))\mathbf{x}. \tag{8}$$

Since the MSE depends on the unknown perturbation matrix Δ and on the unknown parameters $\mathbf{x}$, in general we cannot construct an estimator to directly minimize the MSE. Instead, we seek the linear estimator that minimizes the worst-case MSE across all possible values of $\mathbf{x}$ and Δ satisfying $\mathbf{x} \in \mathcal{U}_\mathbf{x}$ and $\Delta \in \mathcal{U}_\Delta$. Thus, we consider the problem

$$\min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{\mathbf{x} \in \mathcal{U}_\mathbf{x}, \Delta \in \mathcal{U}_\Delta} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \tag{9}$$

where $E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ is given in (8). Problem (9) is reminiscent of the estimation problem considered in [4]. However, whereas in the problem considered in [4] the entire vector $\mathbf{x}$ was norm constrained, in (9) the norm constraint is on sub-vectors of $\mathbf{x}$, which complicates the problem considerably. Furthermore, while in [4] the entire perturbation matrix Δ was norm constrained, in (9) we assume that Δ is a BC matrix and each of the individual blocks of Δ is norm constrained.

3. MINIMAX MSE ESTIMATOR

We now show that the estimation problem (9) can be formulated as an SDP, where an SDP is the problem of minimizing a linear objective subject to linear matrix inequality (LMI) constraints, which are matrix constraints of the form $\mathbf{A}(\mathbf{x}) \succeq 0$, where the matrix $\mathbf{A}$ depends linearly on $\mathbf{x}$ [9]. The advantage in this formulation is that it readily lends itself to efficient polynomial time methods [9]. In Theorem 1 below we present the SDP formulation. A byproduct of this theorem is that the minimax MSE estimator matrix $\mathbf{G}$ can be chosen as a BC matrix. In this case the relation $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ can be written as

$$\hat{\mathbf{x}}_k = \sum_{i=0}^{N-1} \mathbf{G}_{i+k}\mathbf{y}_i, \quad 0 \le k \le N-1, \tag{10}$$

where the indices are calculated modulo N. Note, that (10) implies the intuitive result that the vector $\mathbf{y}_l$ has the same effect on the estimator of $\mathbf{x}_{l+j}$ as $\mathbf{y}_k$ on the estimator of $\mathbf{x}_{k+j}$, for every l, k, j.

Theorem 1 *Let $\mathbf{x} = \left(\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T\right)^T$ denote the vector of unknown parameters in the model $\mathbf{y} = (\mathbf{H} + \Delta)\mathbf{x} + \mathbf{w}$, where $\mathbf{H}$ is a BC matrix, Δ is an unknown matrix satisfying $\Delta \in \mathcal{U}_\Delta$ and $\mathbf{w}$ is zero-mean random vector with a BC covariance matrix $\mathbf{C}$. Then there exists a solution to $\min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{\mathbf{x} \in \mathcal{U}_\mathbf{x}, \Delta \in \mathcal{U}_\Delta} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ given by a BC matrix $\mathbf{G}$ where*

$$\mathbf{G}_j = \frac{1}{N}\sum_{k=0}^{N-1} \omega^{-kj}\mathbf{A}_k, \quad 0 \le j \le N-1.$$

Here $\omega = e^{-\frac{2\pi\mathbf{i}}{N}}$, and $\mathbf{A}_0, \ldots, \mathbf{A}_{N-1}$ are the solutions to the SDP

$$\min_{\tau, \lambda_j, t_j, \mathbf{A}_j} \left\{ NL^2\tau + \sum_{j=0}^{N-1} t_j \right\}$$

subject to

$$\begin{pmatrix} t_j & \mathbf{a}_j^* \\ \mathbf{a}_j & \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1,$$

$$\begin{pmatrix} \tau\mathbf{I} - \lambda_j \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho \mathbf{A}_j \\ \mathbf{0} & -\rho \mathbf{A}_j^* & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0,$$

where $\mathbf{a}_j = vec(\mathbf{A}_j \widehat{\mathbf{C}}_j^{1/2})$, $\rho = \sum_{j=0}^{N-1} \rho_j$ *and*

$$\widehat{\mathbf{H}}_j = \sum_{k=0}^{N-1} \omega^{kj} \mathbf{H}_k, \widehat{\mathbf{C}}_j = \sum_{k=0}^{N-1} \omega^{kj} \mathbf{C}_k, \;\; 0 \le j \le N-1.$$

4. MINIMAX MSE ESTIMATOR FOR $\mathbf{T} = \mathbf{I}$ AND KNOWN $\mathbf{H}$

We now consider a special case of the minimax MSE problem in which $\mathbf{T} = \mathbf{I}$ and $\mathbf{H}$ is known exactly. The following lemma is a key result which enables us to simplify (9) in this case.

Lemma 4.1 *Let* $\mathbf{A}$ *be a BC matrix and let* $L > 0$. *Then,*

$$\max_{\|\mathbf{x}_0\| \le L, \ldots, \|\mathbf{x}_{N-1}\| \le L} \mathbf{x}^* \mathbf{A} \mathbf{x} = NL^2 \max_{0 \le j \le N-1} \left\{ \lambda_{max}(\widehat{\mathbf{A}}_j) \right\}.$$

Furthermore, $\max_{\|\mathbf{x}_j\| \le L} \mathbf{x}^* \mathbf{A} \mathbf{x} = \max_{\|\mathbf{x}\|^2 \le NL^2} \mathbf{x}^* \mathbf{A} \mathbf{x}$.

Using Lemma 4.1 we can replace the set of constraints $\|\mathbf{x}_0\| \le L, \ldots, \|\mathbf{x}_{N-1}\| \le L$ with the single constraint $\|\mathbf{x}\|^2 \le NL^2$. Thus, in the case of known $\mathbf{H}$ (but not in the case of unknown $\mathbf{H}$), we return to the problem of a single system $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}\|^2 \le NL^2$. This problem was discussed in [4] where it was shown that the minimax MSE estimator for the case $\mathbf{T} = \mathbf{I}$ is given by

$$\hat{\mathbf{x}} = \alpha(\mathbf{H}^* \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^* \mathbf{C}^{-1} y, \tag{11}$$

with $\alpha = \frac{NL^2}{NL^2 + B}$ and $B = \mathrm{Tr}\left((\mathbf{H}^* \mathbf{C}^{-1} \mathbf{H})^{-1}\right)$. The estimator of (11) is a shrunken estimator proposed by Mayer and Willke [11], which is simply a scaled version of the LS estimator with an optimal choice of shrinkage factor.

The dominant computation in (11) is the inversion of the $mN \times mN$ matrix $\mathbf{H}^* \mathbf{C}^{-1} \mathbf{H}$, which requires $O(m^3 N^3)$ operations. This number is prohibitively large even for medium size problems. On the other hand, the calculation stemming from Theorem 2 below, which exploits the block circulant structure, requires the inversion of N DFT components, each an $m \times m$ matrix resulting in a total of only $O(m^3 N)$ operations. For example, if $N = 100$ then our computation is 10000 cheaper than the direct computation.

Theorem 2 *Let* $\mathbf{x} = \left(\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T\right)^T$ *denote the vector of unknown parameters in the model* $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, *where* $\mathbf{H}$ *is a known BC matrix and* $\mathbf{w}$ *is a zero-mean random vector with a BC covariance matrix* $\mathbf{C}$. *Suppose that* $\mathbf{T} = \mathbf{I}$. *For every* $0 \le j \le N-1$, *let* $\widehat{\mathbf{H}}_j^* \widehat{\mathbf{C}}_j^{-1} \widehat{\mathbf{H}}_j = \mathbf{V}_j \Sigma_j \mathbf{V}_j^*$ *where* Σ_j *is a diagonal matrix with diagonal elements* $\sigma_{j,1}, \sigma_{j,2}, \ldots, \sigma_{j,m} > 0$ *and* $\mathbf{V}_j$ *is a unitary matrix. Then the solution to the problem* $\min_{\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}} \max_{\mathbf{x} \in \mathcal{U}_\mathbf{x}} E\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right)$ *is given by*

$$\hat{\mathbf{x}}_k = \sum_{j=0}^{N-1} \mathbf{G}_{j+k} \mathbf{y}_j \tag{12}$$

where

$$\mathbf{G}_j = \frac{1}{N} \sum_{k=0}^{N-1} \omega^{-kj} \mathbf{A}_k, \quad 0 \le j \le N-1,$$

and

$$\mathbf{A}_j = \frac{NL^2}{NL^2 + B} \left(\widehat{\mathbf{H}}_j^* \widehat{\mathbf{C}}_j^{-1} \widehat{\mathbf{H}}_j\right)^{-1} \widehat{\mathbf{H}}_j^* \widehat{\mathbf{C}}_j^{-1}, 0 \le j \le N-1.$$

Here $B = \sum_{j=0}^{N-1} \sum_{i=1}^{m} \frac{1}{\sigma_{j,i}} = \sum_{j=0}^{N-1} Tr\left((\widehat{\mathbf{H}}_j^* \widehat{\mathbf{C}}_j^{-1} \widehat{\mathbf{H}}_j)^{-1}\right)$ *and*

$$\widehat{\mathbf{H}}_j = \sum_{k=0}^{N-1} \omega^{kj} \mathbf{H}_k, \quad \widehat{\mathbf{C}}_j = \sum_{k=0}^{N-1} \omega^{kj} \mathbf{C}_k, \;\; 0 \le j \le N-1.$$

5. EXAMPLES

In this section we present two examples of the minimax MSE estimator. In the first example, $\mathbf{H}$ is known, while in the second example, $\mathbf{H}$ is subject to uncertainty.

5.1. Known H

The first example illustrates the minimax MSE estimator of Theorem 2 for known $\mathbf{H}$ and $\mathbf{T} = \mathbf{I}$. We consider the two channel case ($N = 2$) in which

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{H}_0 \mathbf{x}_0 + \mathbf{H}_1 \mathbf{x}_1 + \mathbf{w}_0; \\ \mathbf{y}_1 &= \mathbf{H}_1 \mathbf{x}_0 + \mathbf{H}_0 \mathbf{x}_1 + \mathbf{w}_1, \end{aligned}$$

where $\mathbf{H}_0$ and $\mathbf{H}_1$ represent convolution with LTI filters with impulse responses h_0 and h_1 respectively, with $h_0 = (1\ 0.4\ 0.2\ 0.1)^T$ and $h_1 = (1\ 0.4\ 0.1)^T$. The noise covariance matrix is given by $\mathbf{C} = \sigma^2 \mathbf{I}$ for some σ^2. To evaluate the performance of the minimax MSE estimator, we generate a random vector $\mathbf{x}$ with subvectors $\mathbf{x}_0$ and $\mathbf{x}_1$ such that $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 3$.

We consider three estimation methods: Least Squares (LS), Minimax MSE and Regularized LS (RLS), which is given by

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\|\mathbf{x}_0\| \le L, \|\mathbf{x}_1\| \le L} \|\mathbf{C}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x})\|^2. \tag{13}$$

Thus, the RLS is a modified version of the LS method which considers the norm constraints on $\mathbf{x}_i$. Both in the Minimax MSE estimator and in the RLS estimator, L is approximated as the norm of the LS estimator. The RLS estimator does not have an explicit expression. In order to calculate it we have implemented a gradient projection algorithm (see *e.g.,* [12]), defined by:
Initial step: Take an arbitrary $\mathbf{x}^0$.
General step: For every $k = 0, 1, 2, \ldots$ define:

$$\mathbf{z}^{k+1} = \mathbf{x}^k - t_k \left(\mathbf{H}^* \mathbf{C}^{-1} \mathbf{H} \mathbf{x}^k - \mathbf{H}^* \mathbf{C}^{-1} y\right), \tag{14}$$

$$\text{for } i = 0, 1, \quad \mathbf{x}_i^{k+1} = \begin{cases} \mathbf{z}_i^{k+1}, & \text{if } \|\mathbf{z}_i^{k+1}\| \le L, \\ \frac{\mathbf{z}_i^{k+1}}{\|\mathbf{z}_i^{k+1}\|} L, & \text{else.} \end{cases} \tag{15}$$

In Fig. 1 we plot the MSE averaged over 400 noise realizations as a function of the SNR defined by $10 \log\left(\|\mathbf{x}_0\|^2 / \sigma^2\right)$, using each of the methods above (for the LS we have an analytic

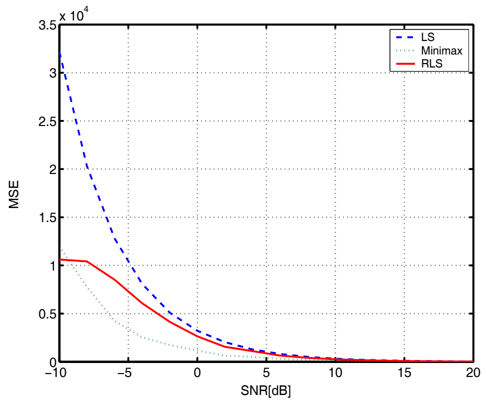

Fig. 1. MSE as a function of SNR for the LS, RLS and minimax MSE estimators for unknown L.

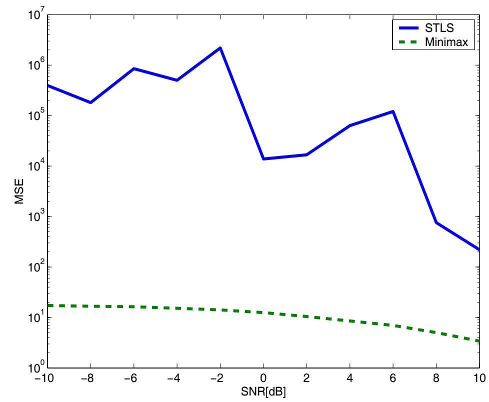

Fig. 2. MSE as a function of SNR for the STLS and minimax MSE estimators for unknown L, ρ_0 and ρ_1.

expression for the MSE, so the 400 realizations of the noise are relevant only for the RLS and the minimax MSE estimators). It is clear that the minimax MSE estimator is the best of the three estimators for SNR>-9 dB, the LS is the worst of the three and RLS is somewhere in between these two estimators. Note, that the RLS is iterative, and therefore computationally more demanding then the minimax MSE estimator.

5.2. Unknown H

To demonstrate the multichannel minimax MSE estimator of Theorem 1 for uncertain transfer functions, we use the same $\mathbf{H}_0$, $\mathbf{H}_1$ as in the previous example. We consider the two channel case ($N = 2$) in which

$$\begin{aligned} \mathbf{y}_0 &= (\mathbf{H}_0 + \Delta_0)\mathbf{x}_0 + (\mathbf{H}_1 + \Delta_1)\mathbf{x}_1 + \mathbf{w}_0\,; \\ \mathbf{y}_1 &= (\mathbf{H}_1 + \Delta_1)\mathbf{x}_0 + (\mathbf{H}_0 + \Delta_0)\mathbf{x}_1 + \mathbf{w}_1\,. \end{aligned} \quad (16)$$

The noise covariance matrix is given by $\mathbf{C} = \sigma^2\mathbf{I}$ for some σ^2. To evaluate the performance of the minimax MSE estimator, we generate a random vector $\mathbf{x}$ with subvectors $\mathbf{x}_0$ and $\mathbf{x}_1$ such that $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 3$ and random perturbation matrices Δ_0 and Δ_1 with norm 0.02. We consider two estimation methods: Structured Total Least Squares (STLS) (see [13]), which is a modified Total LS method where we seek a pair $(\widehat{\mathbf{H}}, \mathbf{y})$ that minimizes the error $\|\widehat{\mathbf{H}} - \mathbf{H}\|^2 + \|\hat{\mathbf{y}} - \mathbf{y}\|^2$ subject to the consistency equation $\hat{\mathbf{y}} \in Im(\widehat{\mathbf{H}})$ and the constraint that $\mathbf{H}$ is a BC matrix, and the Minimax MSE estimator with L, ρ_0 and ρ_1 estimated from the STLS estimator.

In Fig. 2 we plot the MSE averaged over 400 noise realizations as a function of the SNR using each of the methods above. It is clear that the minimax estimator outperforms the STLS estimator even though the norm bounds are unknown. Moreover, the STLS exhibits an unstable behavior in the sense that the estimation error has a *huge* variance. For example, for the same value of σ, we got that the square of the estimation error was 141 in one realization of the noise and approximately 10^6 in another realization.

6. REFERENCES

[1] P. Comon, "Separation of stochastic processes," *Proc. Workshop Higher-Order Spectral Analysis*, pp. 174–179, July 1989.

[2] S. Shamsunder and G. Giannakis, "Multichannel blind signal separation and reconstruction," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 515–527, Nov. 1997.

[3] B. R. Hunt and O. Kubler, "Karhunen-Loeve multispectral image restoration, part I: Theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–32, no. 3, pp. 592–600, June 1984.

[4] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168–181, Jan. 2005.

[5] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, "A general framework for frequency domain multi-channel signal processing," *IEEE Trans. Image Processing*, vol. 2, no. 3, pp. 417–420, July 1993.

[6] H. Murakami, "Discrete wavelet transform based on cyclic convolution," *IEEE Trans. Signal Processing*, pp. 165–174, Jan. 2004.

[7] R. Gray, "Toeplitz and circulant matrices: A review," Tech. Rep. 6504-1, Inform. Sys. Lab., Stanford Univ., Stanford, CA, Apr. 1977.

[8] P. A. Regalia H. Gazzah and J. Delmas, "Asymptotic eigenvalue distribution of block toeplitz matrices and application to blind simo channel identification," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1243–1251, Mar. 2001.

[9] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, 2001.

[10] A. Beck, Y. C. Eldar, and A. Ben-Tal, "Minimax mean-squared error estimation of multichannel signals," submitted to *IEEE Trans. Inform. Theory*.

[11] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, Aug. 1973.

[12] D. P. Bertsekas, *Nonlinear Programming*, Belmont MA: Athena Scientific, second edition, 1999.

[13] A. Beck and A. Ben-Tal, "A global solution for the structured total least squares problem with block circulant matrices," submitted to *SIAM J. Matrix Anal. Appl.*