CCIT Report #481 May 2004

Minimax Mean-Squared Error Estimation of Multichannel Signals

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

Abstract

We consider the problem of multichannel estimation, in which we seek to estimate N multiple input vectors that are observed through a set of linear transformations and corrupted by additive noise. 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. The input vectors $\mathbf{x}_k$ are known to satisfy a weighted norm constraint. We seek the linear estimator that minimizes the worst-case mean-squared error (MSE) across all possible values of $\mathbf{x}_k$. We show that for an arbitrary choice of weighting, the optimal minimax MSE estimator can be formulated as a solution to a semidefinite programming problem (SDP), which can be solved efficiently. For an *Euclidean norm* bound on $\mathbf{x}_k$, we show that the SDP can be reduced to a simple convex program with $N+1$ variables, or just 3 variables, depending on the specific structure of the underlying model matrix. Moreover, when the linear transformations are fixed, the minimax MSE multichannel estimator reduces to the shrunken estimator of Mayer and Willke, with a specific choice of shrinkage factor, that explicitly takes the prior information into account. Finally, we demonstrate through examples, that the robust minimax MSE estimator can significantly increase the performance over conventional methods *e.g.,* least squares (LS), regularized nonlinear LS and total LS.

*MINERVA Optimization Center, Department of Industrial Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: becka@tx.technion.ac.il.

†Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: yonina@ee.technion.ac.il.

‡MINERVA Optimization Center, Department of Industrial Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: morbt@ie.technion.ac.il.

1 Introduction

Estimation of multiple signals from multiple outputs is an important problem that appears in a variety of applications, such as multiuser communications and space-time coding with antenna arrays, multichannel image restoration, and multiple speaker separation using multiple microphone measurements.

In a multichannel estimation problem, we seek to estimate multiple input vectors $\{\mathbf{x}_k, 0 \leq k \leq N-1\}$, that are observed through a set of linear transformations $\mathbf{H}_{k,i}$ and corrupted by additive noise. Thus, the kth output vector $\mathbf{y}_k$ is given by the superposition $\mathbf{y}_k = \sum_{i=0}^{N-1} \mathbf{H}_{k,i}\mathbf{x}_i$, where $\mathbf{H}_{k,i}$ is the transfer function from the ith input $\mathbf{x}_i$ to the kth output $\mathbf{y}_k$.

There is a vast body of literature that treats the multichannel estimation problem, under the assumption that the input vectors are random with known statistics. 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, or the minimum MSE (MMSE) estimator. However, if the prior statics or the transfer matrices $\mathbf{H}_{i,k}$ are unknown, then the Wiener estimator cannot be implemented.

The blind multichannel estimation problem, in which the transfer matrices are assumed to be unknown, has also received much attention. To enable estimation in this case, it is typically assumed that the input vectors are statistically independent. The simplest example of a blind multichannel signal estimation problem is the scalar case in which each channel $h_{k,i}$ is a scalar, and the channel inputs x_k are independent, identically distributed (iid) random variables. This class of problems is referred to as blind source separation, and has been studied extensively (see *e.g.,* [1, 2, 3, 4]). Blind source separation methods attempt to extract the signals from the observed mixtures by exploiting the fact that the inputs are independent.

In the context of speech, multichannel estimation problems arise when more than one speaker is present and there are several microphones, so that each microphone records the primary speaker together with reflections from other speakers [5, 6, 7]. Traditional noise cancellation schemes usually assume the availability of the interfering signal and ignore the cross-signal interaction [8]. Here again, to deal with multichannel signals, it is typically assumed that the input signals are independent random signals.

Another context in which multichannel estimation plays an important role is in the context of image restoration. Multichannel images arise from measuring a scene using more than one type of sensor. In most applications, the multichannel images are distorted due to motion, out of focus blur, channel cross talk, quantization error and devise noise. Although there is a large amount of work in the literature on restoration of single channel images [9], there has been less work on multichannel restoration. One approach to multichannel image restoration is to assume that the input images are random with known statistics, and then apply an MMSE Wiener estimator to estimate the inputs [10, 11, 12].

A significant amount of prior knowledge is required for MMSE multichannel estimation. Specifically, the statistics of all the inputs and noise vectors must be known. In many practical scenarios, this information may not be available. Although the MMSE estimator minimizes the MSE, its success depends on accurate statistical knowledge of the inputs and noise characteristics. Therefore, in many cases it is more reasonable to assume that the inputs are deterministic but unknown.

A straightforward approach to deterministic multichannel estimation 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 $\hat{\mathbf{y}}_k$. This approach has been used *e.g.,* in the context of image restoration [13, 14]. However, in an estimation context, the objective typically is to minimize the size of the norms of the estimation errors $\hat{\mathbf{x}}_k - \mathbf{x}_k$, 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. 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 (possibly weighted) norm of the unknown vectors $\mathbf{x}_k$ is bounded, and develop robust estimators $\hat{\mathbf{x}}_k$ of $\mathbf{x}_k$ whose performance is reasonably good across all possible values of $\mathbf{x}_k$ in the region of uncertainty. Specifically, we develop a minimax MSE estimator that minimizes the worst-case MSE across all possible bounded values of $\mathbf{x}_k$, *i.e.,* over all values of $\mathbf{x}_k$ such that $\|x_k\|_T = \mathbf{x}_k^* \mathbf{T} \mathbf{x}_k \leq L^2$ for some constant L and weighting matrix $\mathbf{T}$. We then develop minimax MSE estimators which are robust with respect to uncertainty in the model matrices $\mathbf{H}_{i,k}$. To be more specific, 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. The multichannel minimax MSE estimator is an extension of the recently proposed linear minimax MSE estimator for the single channel case [15].

Given a norm constraint on the vectors $\mathbf{x}_k$, one approach to incorporate the norm constraints in the estimation is to use a nonlinear regularized estimator, which is the solution to the optimization problem of minimizing the data error subject to the norm constraints. This regularized nonlinear estimator does not have an explicit expression and finding it involves the usage of an iterative optimization method [16, 17]. This is in contrast to the minimax MSE estimator, which is linear, and in many practical cases has an explicit and simple expression.

In our development, we assume that the norm bound L is known. However, our algorithms can also be implemented when L is not known, by first estimating it from the data. Thus, in practice, no prior information is needed for implementing our proposed estimators. Specifically, in our simulations in Section 6 we estimate L as the norm of the LS estimator. Our experimental results suggests that with this estimated L, the minimax MSE estimator outperforms not only the linear LS estimator but also the nonlinear regularized estimator. Notice that the minimax MSE estimator is linear only if L is known. In the case where L is not known the minimax MSE estimator is not linear but still has an explicit and simple expression.

The mathematical analysis in this paper covers several possible models. The first case is the *direct model* in which we wish to estimate N unknown deterministic parameter vectors $\mathbf{x}_k \in \mathbb{C}^m, 0 \leq k \leq N-1$ from N vector observations $\mathbf{y}_k \in \mathbb{C}^n, 0 \leq k \leq N-1$, where each observation vector $\mathbf{y}_k$ is related to the corresponding parameter vector $\mathbf{x}_k$ through the linear model $\mathbf{y}_k = \mathbf{H}_0\mathbf{x}_k + \mathbf{w}_k,\ 0 \leq k \leq N-1$, where $\mathbf{H}_0$ is an $n \times m$ matrix assumed to have full rank m, and $\mathbf{w}_k,\ 0 \leq k \leq N-1$ are zero-mean random vectors.

If the noise vectors $\mathbf{w}_k$ are uncorrelated, then we may treat this estimation problem as N independent problems, where each problem reduces to the problem considered *e.g.,* in [18, 15], of estimating an unknown vector $\tilde{\mathbf{x}}$ from observations $\tilde{\mathbf{y}} = \widetilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}$ subject to the constraint that $\|\tilde{\mathbf{x}}\|_{\mathbf{T}} \leq L$, where $\tilde{\mathbf{w}}$ is a zero-mean noise vector. If, on the other hand, the noise vectors $\mathbf{w}_k$ are correlated, then we may be able to improve the estimation performance by treating the vectors to be estimated jointly, so that the estimate $\hat{\mathbf{x}}_k$ of $\mathbf{x}_k$ depends on all the observations $\mathbf{y}_l, 0 \leq l \leq N-1$, and not only on $\mathbf{y}_k$. Therefore, in the direct model, we always assume that the noise vectors are correlated.

The second model that our analysis covers is the *symmetric model.* Here, we consider the multichannel problem with $\mathbf{H}_{i,i} = \mathbf{H}_0$ for every i and $\mathbf{H}_{i,k} = \mathbf{H}_1$ for every $i \neq k$. For $N = 2$ (the two channel case), this implies that the within channel transfer function, *i.e.,* the transfer function between each input vector and the corresponding output vector, is identical ($\mathbf{H}_{0,0} = \mathbf{H}_{1,1}$), and the cross channels are also equal ($\mathbf{H}_{0,1} = \mathbf{H}_{1,0}$). The symmetric model is a natural generalization of the symmetric two channel case.

The direct model and the symmetric model are both special cases of the *general model*, in which we assume that the multichannel transfer matrix $\mathbf{H}$ with block matrices $\mathbf{H}_{i,k}$ is a block circulant matrix, so that $\mathbf{H}_{i,k} = \mathbf{H}_{(k-i) mod N}$ where N is the number of channels. The block circulant model has also been used in the context of image restoration [19], and in the context of cyclic convolution filter banks [20]. 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 [21, 22], we can approximate the block-Toeplitz matrix $\mathbf{H}$ by a block circulant matrix. Besides including several cases of practical interest, one of the attributes of the block circulant structure is its analytical tractability. As we show, this structure will allow us to develop the minimax MSE estimator by exploiting properties of block circulant matrices. In particular, the matrix discrete Fourier transform (DFT), *i.e.,* a DFT defined on matrices, will play an important role in our derivations. Therefore, in Section 3, we discuss some properties of circulant matrices and the matrix DFT. As we show, the eigenvalues of a block circulant matrix are exactly the eigenvalues of its discrete Fourier components, and its eigenvectors have a special structure which will be exploited in our analysis.

In Section 4 we first show that if the multichannel transfer matrix $\mathbf{H}$ and the covariance matrix of the concatenated noise vectors $\mathbf{C}$ are block circulant, then with $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, where $\hat{\mathbf{x}}$ is the concatenation of the estimated inputs $\hat{\mathbf{x}}_k$, $\mathbf{G}$ must also be a block circulant matrix. This allows us to formulate the minimax MSE estimator as a solution to a semidefinite programming problem (SDP) [23, 24, 25], which is a tractable convex optimization problem that can be solved efficiently, *e.g.,* using interior point methods [25, 26]. We then develop, in Section 5, a closed form solution to the minimax MSE estimation problem, for the case in which the weighting matrix $\mathbf{T}$ is equal to $\mathbf{I}$. We first show that when $\mathbf{H}$ is certain the optimal estimator is a shrunken estimator proposed by Mayer and Willke [27], with a specific choice of shrinkage factor, that explicitly takes the prior information into account. We then show that, when $\mathbf{H}$ is uncertain, the task of finding the optimal $\mathbf{G}$ reduces to a simple convex optimization problem in $N + 1$ unknowns in the general model or 3 unknowns in the symmetric model. Finally, we demonstrate through examples, in Section 6, that in the case of certain $\mathbf{H}$, the minimax MSE estimator can significantly increase the performance over the conventional multichannel LS approach and over the nonlinear regularized LS approach, and in the case of uncertain $\mathbf{H}$ the minimax MSE estimator can significantly outperform the total LS estimator.

2 Problem Formulation

We denote vectors by boldface lowercase letters and matrices by boldface uppercase letters. The identity matrix of appropriate dimension is denoted by $\mathbf{I}$, $(\cdot)^*$ and $(\cdot)^T$ denote the Hermitian conjugate and the transpose of the corresponding matrices respectively, and $\hat{(\cdot)}$ denotes an estimated vector. 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. 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 $\mathbf{A}$.

Consider the problem of estimating N unknown deterministic parameter vectors $\mathbf{x}_k \in \mathbb{C}^m, 0 \leq k \leq N-1$ from N vector observations $\mathbf{y}_k \in \mathbb{C}^n, 0 \leq k \leq N-1$, where each observation vector $\mathbf{y}_k$ is related to all of the parameter vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ through the linear model

$$\mathbf{y}_k = \sum_{i=0}^{N-1} (\mathbf{H}_{k,i} + \Delta_{k,i})\mathbf{x}_i + \mathbf{w}_k, \quad 0 \leq k \leq N-1. \tag{1}$$

Here $\mathbf{H}_{k,i}$ is the $n \times m$ nominal transfer matrix from the ith input vector $\mathbf{x}_i$ to the kth output vector $\mathbf{y}_k$, $\Delta_{k,i}$ is an $n \times m$ unknown perturbation matrix, and $\mathbf{w}_k$ is the kth noise vector. We assume that $E(\mathbf{w}_k) = 0$ and that

$$E(\mathbf{w}_i\mathbf{w}_j^*) = \mathbf{C}_{i,j}. \tag{2}$$

The matrices $\mathbf{H}_{i,i}$ and $\mathbf{H}_{i,k}, i \neq k$, represent the within channel and cross channel nominal transfer matrices, respectively. We assume that the perturbation matrices satisfy a norm constraint $\|\Delta_{i,k}\| \leq \rho_{i,k}$ and that each vector $\mathbf{x}_k$ is known to satisfy the weighted norm constraint $\|\mathbf{x}_k\|_{\mathbf{T}} \leq L$ for some positive definite matrix $\mathbf{T}$ and scalar $L > 0$, where $\|\mathbf{z}\|^2_{\mathbf{T}} = \mathbf{z}^*\mathbf{T}\mathbf{z}$. It is interesting to note that L and $\rho_{i,k}$ does not have to be given in advance and can be estimated from the LS or total LS estimators (see Section 6 for further details). Thus, practically, our estimator do not require the knowledge of parameters, which are sometimes not known. The input-output relation of (1) is illustrated schematically in Fig. 1 for the special case in which $N = 2$ and $\mathbf{H}$ is exactly known ($\rho_{0,0} = \rho_{0,1} = \rho_{1,0} = \rho_{1,1} = 0$).

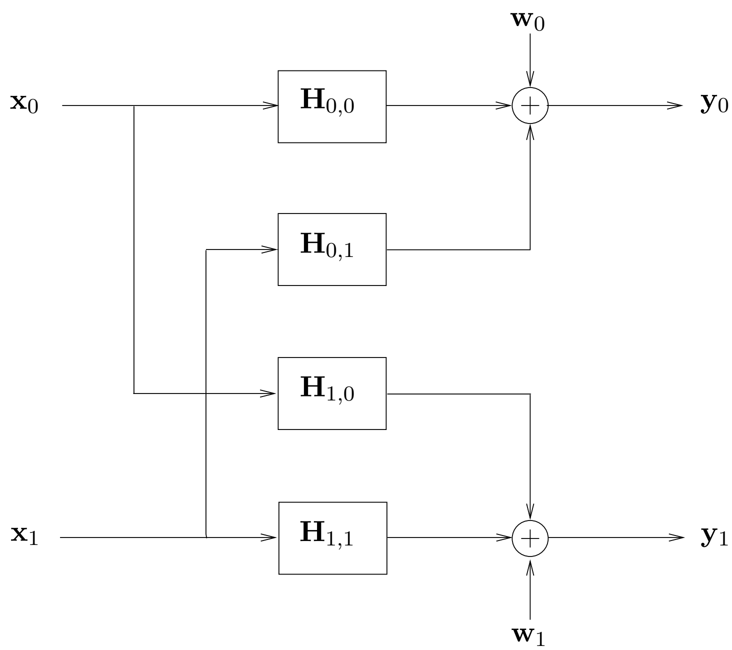


Figure 1: Two channel model.

Using the notation

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{0,0} & \mathbf{H}_{0,1} & \cdots & \mathbf{H}_{0,N-1} \\ \mathbf{H}_{1,0} & \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \mathbf{H}_{N-1,0} & \mathbf{H}_{N-1,1} & \cdots & \mathbf{H}_{N-1,N-1} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_{0,0} & \mathbf{C}_{0,1} & \cdots & \mathbf{C}_{0,N-1} \\ \mathbf{C}_{1,0} & \mathbf{C}_{1,1} & \cdots & \mathbf{C}_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \mathbf{C}_{N-1,0} & \mathbf{C}_{N-1,1} & \cdots & \mathbf{C}_{N-1,N-1} \end{pmatrix}, \quad (3)$$

and

$$\Delta = \begin{pmatrix} \Delta_{0,0} & \Delta_{0,1} & \cdots & \Delta_{0,N-1} \\ \Delta_{1,0} & \Delta_{1,1} & \cdots & \Delta_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \Delta_{N-1,0} & \Delta_{N-1,1} & \cdots & \Delta_{N-1,N-1} \end{pmatrix}, \quad (4)$$

where $\mathbf{H}_{i,k}$ and $\Delta_{i,j}$ are defined by (1) and $\mathbf{C}_{i,j}$ are defined by (2), we can rewrite (1) as

$$\mathbf{y} = (\mathbf{H} + \Delta)\mathbf{x} + \mathbf{w}, \quad (5)$$

where $\mathbf{H}$ is a known $nN \times mN$ matrix, Δ is an unknown $nN \times mN$ perturbation matrix, $\mathbf{x} =$

$(\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T)^T$, $\mathbf{y} = (\mathbf{y}_0^T, \mathbf{y}_1^T, \ldots, \mathbf{y}_{N-1}^T)^T$ and $\mathbf{w} = (\mathbf{w}_0^T, \mathbf{w}_1^T, \ldots, \mathbf{w}_{N-1}^T)^T$ is a zero-mean random vector with covariance given by $\mathbf{C}$. Throughout the paper, we assume that $\mathbf{H}$ has full rank mN.

The most complex model that we consider in our development of the minimax MSE estimator is the *general model*, in which we assume that $\mathbf{H}, \mathbf{C}$ and Δ are block circulant matrices. Specifically, $\mathbf{H}$, $\mathbf{C}$ and Δ are assumed to have the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_{N-1} \\ \mathbf{H}_{N-1} & \mathbf{H}_0 & \cdots & \mathbf{H}_{N-2} \\ \vdots & & & \vdots \\ \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \cdots & \mathbf{C}_{N-1} \\ \mathbf{C}_{N-1} & \mathbf{C}_0 & \cdots & \mathbf{C}_{N-2} \\ \vdots & & & \vdots \\ \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_0, \end{pmatrix} \Delta = \begin{pmatrix} \Delta_0 & \Delta_1 & \cdots & \Delta_{N-1} \\ \Delta_{N-1} & \Delta_0 & \cdots & \Delta_{N-2} \\ \vdots & & & \vdots \\ \Delta_1 & \Delta_2 & \cdots & \Delta_0 \end{pmatrix}, \tag{6}$$

where $\mathbf{H}_0, \ldots, \mathbf{H}_{N-1} \in \mathbb{C}^{n \times m}$ and $\mathbf{C}_0, \ldots, \mathbf{C}_{N-1} \in \mathbb{C}^{n \times n}$. For brevity, throughout the paper we use the notation $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$ to denote the block circulant matrix with first row of matrices $(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$ (see Section 3 for further details). Using this notation, we can write $\mathbf{H}$ of (6) as $\mathbf{H} = \mathcal{C}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1})$ and Δ of (6) as $\Delta = \mathcal{C}(\Delta_0, \Delta_1, \ldots, \Delta_{N-1})$, where $\|\Delta_0\| \leq \rho_0, \|\Delta_1\| \leq \rho_1, \ldots, \|\Delta_{N-1}\| \leq \rho_{N-1}$. We use the following notation for the set of possible values of Δ in the general model

$$\mathcal{U}_G \stackrel{\triangle}{=} \{\Delta = \mathcal{C}(\Delta_0, \Delta_1, \ldots, \Delta_{N-1}) : \|\Delta_0\| \leq \rho_0, \ldots, \|\Delta_{N-1}\| \leq \rho_{N-1}\}. \tag{7}$$

The linear model (5) with block circulant matrices $\mathbf{H}$, Δ and $\mathbf{C}$ will be called the *general model*.

We note that in many practical scenarios it is reasonable to assume that $\mathbf{H}$ is a block Toeplitz matrix so that $\mathbf{H}_{ik} = \mathbf{H}_{i-k}$. Using the well know convergence properties of Toeplitz matrices [21, 22], we can approximate the block-Toeplitz matrix $\mathbf{H}$ by a block circulant matrix of the form (6). Similar ideas have been used to justify the use of the block circulant model in the context of multichannel image restoration [19].

The second model that we consider is the model in which

$$\mathbf{y}_k = (\mathbf{H}_0 + \Delta_0)\mathbf{x}_k + \sum_{i \neq k} (\mathbf{H}_1 + \Delta_1)\mathbf{x}_i + \mathbf{w}_k, \quad 0 \leq k \leq N-1, \tag{8}$$

so that the affect of all of the interfering vectors $\mathbf{x}_i, i \neq k$ on the kth output $\mathbf{y}_k$ is the same (*i.e.,* independent of k). To model the correlation between the noise vectors in this case, we note that in many useful applications the order of the observations $\mathbf{y}_k$ is immaterial, so that the statistics of $\mathbf{y}_k$ and the joint statistics of $\mathbf{y}_i$ and $\mathbf{y}_j$ do not depend on i and j. Equivalently, the statistics of the noise vectors $\mathbf{w}_i$ and the joint statistics of $\mathbf{w}_i$ and $\mathbf{w}_j$ do not depend on i and j. In this case,

$$E(\mathbf{w}_i\mathbf{w}_i^*) = \mathbf{C}_0, \quad 0 \leq i \leq N-1, \tag{9}$$

for some covariance matrix $\mathbf{C}_0 \succeq 0$, and

$$E(\mathbf{w}_i\mathbf{w}_j^*) = \mathbf{C}_1, \quad 0 \leq i \neq j \leq N-1, \tag{10}$$

for some matrix $\mathbf{C}_1$. Since from (10) we have that $E(\mathbf{w}_i\mathbf{w}_j^*) = E(\mathbf{w}_j\mathbf{w}_i^*)$, and it is always true that $E(\mathbf{w}_i\mathbf{w}_j^*) = E^*(\mathbf{w}_j\mathbf{w}_i^*)$, we conclude that under the assumption (10), $E(\mathbf{w}_i\mathbf{w}_j^*) = E^*(\mathbf{w}_i\mathbf{w}_j^*)$, or, $\mathbf{C}_1 = \mathbf{C}_1^*$. Therefore, if the order of the observations is immaterial, as is often the case in practice, then the second order statistics of the noise vectors $\mathbf{w}_i, 0 \leq i \leq N-1$ are given by (9) and (10), where $\mathbf{C}_1$ is an Hermitian matrix. Thus, in this case we have that

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_1 \\ \mathbf{H}_1 & \mathbf{H}_0 & \cdots & \mathbf{H}_1 \\ \vdots & & & \vdots \\ \mathbf{H}_1 & \mathbf{H}_1 & \cdots & \mathbf{H}_0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \cdots & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{C}_0 & \cdots & \mathbf{C}_1 \\ \vdots & & & \vdots \\ \mathbf{C}_1 & \mathbf{C}_1 & \cdots & \mathbf{C}_0 \end{pmatrix}, \Delta = \begin{pmatrix} \Delta_0 & \Delta_1 & \cdots & \Delta_1 \\ \Delta_1 & \Delta_0 & \cdots & \Delta_1 \\ \vdots & & & \vdots \\ \Delta_1 & \Delta_1 & \cdots & \Delta_0 \end{pmatrix}. \tag{11}$$

Throughout the paper we use the notation $\mathbf{H} = \mathcal{M}(\mathbf{H}_0, \mathbf{H}_1)$, $\mathbf{C} = \mathcal{M}(\mathbf{C}_0, \mathbf{C}_1)$ and $\Delta = \mathcal{M}(\Delta_0, \Delta_1)$, where $\|\Delta_0\| \leq \rho_0$ and $\|\Delta_1\| \leq \rho_1$. In this case, model (5) will be called the *symmetric model*. We use the following notation for the set of possible values of Δ in the symmetric model:

$$\mathcal{U}_S \stackrel{\triangle}{=} \{\mathcal{M}(\Delta_0, \Delta_1) : \|\Delta_0\| \leq \rho_0, \|\Delta_1\| \leq \rho_1\}. \tag{12}$$

It is interesting to note that the symmetric model is, in general, *not* a special case of the general model since the symmetric case requires not only that $\mathbf{H}_1 = \mathbf{H}_2 = \ldots = \mathbf{H}_{N-1}$ but also an additional constraint $\Delta_1 = \Delta_2 = \ldots = \Delta_{N-1}$, which implies a different structure of the uncertainty set. However, in the case where $\mathbf{H}$ is known ($\rho_0 = \ldots = \rho_{N-1} = 0$) the symmetric model is indeed a special case of the general model ($\mathbf{H}_1 = \mathbf{H}_2 = \ldots = \mathbf{H}_{N-1}$).

The third model that we consider is the *direct model* in which

$$\mathbf{H}=\begin{pmatrix}\mathbf{H}_0 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_0 & \cdots & \mathbf{0}\\ \vdots & & & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_0\end{pmatrix},\quad \mathbf{C}=\begin{pmatrix}\mathbf{C}_0 & \mathbf{C}_1 & \cdots & \mathbf{C}_1\\ \mathbf{C}_1 & \mathbf{C}_0 & \cdots & \mathbf{C}_1\\ \vdots & & & \vdots\\ \mathbf{C}_1 & \mathbf{C}_1 & \cdots & \mathbf{C}_0,\end{pmatrix},\Delta=\begin{pmatrix}\Delta_0 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \Delta_0 & \cdots & \mathbf{0}\\ \vdots & & & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \Delta_0\end{pmatrix}, \tag{13}$$

i.e., $\mathbf{H}=\mathcal{M}(\mathbf{H}_0,\mathbf{0})$, $\Delta=\mathcal{M}(\Delta_0,\mathbf{0})$ and $\mathbf{C}=\mathcal{M}(\mathbf{C}_0,\mathbf{C}_1)$. In this model,

$$\mathbf{y}_k=(\mathbf{H}_0+\Delta_0)\mathbf{x}_k+\mathbf{w}_k,\quad k=0,1,\ldots,N-1, \tag{14}$$

so that the kth observation is effected only by the kth input vector, as illustrated in Fig. 2. It is

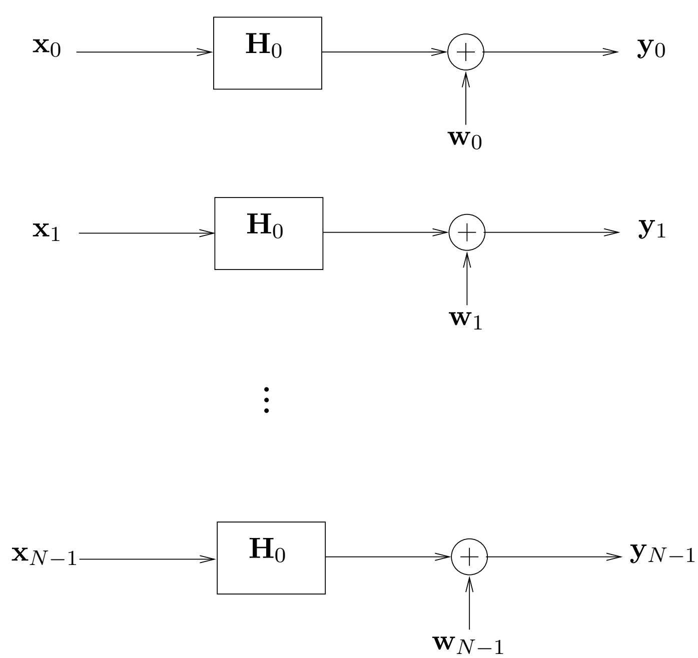


Figure 2: Direct model.

important to note that the direct model is indeed a special case of the symmetric model ($\rho_1=0$ and $\mathbf{H}_1=\mathbf{0}$). The results in the paper are stated only for the general and symmetric models. The results for the direct model can be easily obtained by substituting $\mathbf{H}_1=\mathbf{0}$ and $\rho_1=0$ in the results for the symmetric model. If the noise vectors $\mathbf{w}_k$ are uncorrelated, then under the direct model (14) with known $\mathbf{H}$ we may treat our estimation problem as N independent problems, where each problem reduces to the problem considered *e.g.,* in [18, 15], of estimating an unknown vector $\tilde{\mathbf{x}}$

from an observation $\tilde{\mathbf{y}} = (\mathbf{H}_0 + \Delta)\tilde{\mathbf{x}} + \tilde{\mathbf{w}}$ subject to the constraint that $\|\tilde{\mathbf{x}}\|_{\mathbf{T}} \leq L$, where $\tilde{\mathbf{w}}$ is a zero-mean noise vector. If, on the other hand, the noise vectors $\mathbf{w}_i$ are correlated, then we may be able to improve the estimation performance by treating the vectors to be estimated jointly, so that the estimate $\hat{\mathbf{x}}_j$ of $\mathbf{x}_j$ depends on all the observations $\mathbf{y}_k, 0 \leq k \leq N-1$, and not only on $\mathbf{y}_j$.

Now, consider the general model. 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 $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is given by

$$E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) + \mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))\mathbf{x}. \tag{15}$$

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 $\mathbf{x}_k$ satisfying $\|\mathbf{x}_k\|_{\mathbf{T}} \leq L$ and $\Delta \in \mathcal{U}_G$, where $\mathcal{U}_G$ is defined by (7). Thus, we consider the problem

$$\begin{aligned} &\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \leq L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \leq L, \Delta \in \mathcal{U}_G} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \\ &\min_{\mathbf{G}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \leq L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \leq L, \Delta \in \mathcal{U}_G} \left\{\mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))^*(\mathbf{I} - \mathbf{G}(\mathbf{H} + \Delta))\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*)\right\}. \end{aligned} \tag{16}$$

Problem (16) is reminiscent of the estimation problem considered in [15]. However, whereas in the problem considered in [15] the entire vector $\mathbf{x}$ was norm constrained, in (16) the norm constraint is on sub-vectors of $\mathbf{x}$, which, as we will show, complicates the problem considerably. Furthermore, while in [15] the entire perturbation matrix Δ was norm constrained, in (16) we assume that Δ has a particular structure given by (6), and each of the individual blocks is norm constrained.

The mathematical tools needed to analyze the multichannel model, mainly properties of blocks circulant matrices, are given in Section 3. In Section 4.1, we show that under the assumptions of the general model, $\mathbf{G}$ can be chosen as a block circulant matrix. This enables us, in Section 4.2, to formulate the minimax estimation problem (16) as an SDP problem. From the derived formulation we deduce that in the symmetric model (i.e., $\mathbf{H} = \mathcal{M}(\mathbf{H}_0, \mathbf{H}_1)$, $\Delta = \mathcal{M}(\Delta_0, \Delta_1)$ and $\mathbf{C} = \mathcal{M}(\mathbf{C}_0, \mathbf{C}_1)$), the optimal $\mathbf{G}$ can be chosen as $\mathbf{G} = \mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$. We then find, in Section 5, explicit expressions for the minimax MSE estimator in the case where $\mathbf{T} = \mathbf{I}$. We show that the multichannel minimax MSE estimator with known $\mathbf{H}$ is a shrunken estimator proposed by Mayer and Willke [27], with a specific choice of shrinkage factor. If $\mathbf{H}$ is unknown we show that the computation of the minimax MSE estimator involves the solution of a simple convex optimization problem with $N+1$ unknowns in the general model and with only 3 unknowns in the symmetric model.

The next section contains results on block circulant matrices, which are essential for the analysis of the problem (16) in the rest of the paper.

3 Block Circulant Matrices and the Discrete Fourier Transform

The aim of this section is to give a short summary of results on block circulant matrices and the DFT defined on them that will be used later in the paper. Since results on this subject are scattered throughout the literature (e.g. [28, 29, 30, 31]) we find it practical to gather all the important results in this section.

A block circulant matrix is a matrix of the form

$$\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}) \triangleq \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{N-1} \\ \mathbf{A}_{N-1} & \mathbf{A}_0 & \cdots & \mathbf{A}_{N-2} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_0 \end{pmatrix}, \tag{17}$$

where each submatrix $\mathbf{A}_j$ is a $k \times l$ matrix. The dimensions k and l are always clear from the context.

3.1 General Properties

From the definition of block circulant matrices we have the following lemma:

Lemma 3.1 *Let* $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1} \in \mathbb{C}^{k,l}$ *and* $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1} \in \mathbb{C}^{l,m}$. *Then,*

1. $\mathcal{C}^*(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ *is also a block circulant matrix and*

$$\mathcal{C}^*(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}) = \mathcal{C}(\mathbf{A}_0^*, \mathbf{A}_{N-1}^*, \ldots, \mathbf{A}_1^*). \tag{18}$$

2. *The product* $\mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})\mathcal{C}(\mathbf{B}_0, \ldots, \mathbf{B}_{N-1})$ *is a block circulant matrix* $\mathcal{C}(\mathbf{C}_0, \ldots, \mathbf{C}_{N-1})$ *where*

$$\mathbf{C}_j = \sum_{i=0}^{N-1} \mathbf{A}_j \mathbf{B}_{j-i}, \quad 0 \le j \le N-1. \tag{19}$$

Note, that in Lemma 3.1, as well as in the rest of the paper, the indexes are computed modulo N. Thus, for example $\mathbf{B}_N = \mathbf{B}_0$ and $\mathbf{B}_{-1} = \mathbf{B}_{N-1}$.

Remark 1: From equation (18) it follows that a block circulant matrix $\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ where $\mathbf{A}_0, \ldots, \mathbf{A}_{N-1} \in \mathbb{C}^{m\times m}$ is Hermitian if and only if

$$\mathbf{A}_i^* = \mathbf{A}_{N-i}, \ \ 0 \le i \le N-1. \tag{20}$$

Remark 2: Equation (19) is the definition of the *discrete convolution* of $(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ and $(\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1})$ also denoted by:

$$(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1}) = (\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}) * (\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1}). \tag{21}$$

Let $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$. Then the discrete Fourier transform (DFT) of $\mathbf{A}$ is also a block circulant matrix of the same dimensions given by

$$\mathcal{F}(\mathbf{A}) = \mathcal{C}(\widehat{\mathbf{A}}_0, \widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{N-1}), \tag{22}$$

where $\widehat{\mathbf{A}}_j, \ 0 \le j \le N-1$ is defined as:

$$\widehat{\mathbf{A}}_j \stackrel{\triangle}{=} \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i, \ \ 0 \le j \le N-1, \tag{23}$$

and $\omega = e^{-\frac{2\pi i}{N}}$ (here $i = \sqrt{-1}$). In the sequel, we will also use the notation[1]

$$\mathcal{F}_j(\mathbf{A}) \stackrel{\triangle}{=} \widehat{\mathbf{A}}_j = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i, \ \ 0 \le j \le N-1. \tag{24}$$

The matrices $\mathcal{F}_j(\mathbf{A})$ are called the *discrete Fourier components*. The inverse DFT (IDFT), denoted by $\mathcal{F}^{-1}$, is defined by $\mathcal{F}^{-1}(\mathbf{A}) = (\widetilde{\mathbf{A}}_0, \widetilde{\mathbf{A}}_1, \ldots, \widetilde{\mathbf{A}}_{N-1})$, where

$$\widetilde{\mathbf{A}}_j = \frac{1}{N} \sum_{i=0}^{N-1} \omega^{-ij} \mathbf{A}_i, \ \ 0 \le j \le N-1. \tag{25}$$

We also use the notation $\mathcal{F}_j^{-1}(\mathbf{A}) \stackrel{\triangle}{=} \widetilde{\mathbf{A}}_j$. It is not difficult to see that for every $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})$ we have:

[1] We use two notations for the DFT components; depending on the context one notation is better suited than the other.

$$\mathcal{F}^{-1}(\mathcal{F}(\mathbf{A})) = \mathbf{A}, \quad \mathcal{F}(\mathcal{F}^{-1}(\mathbf{A})) = \mathbf{A}. \tag{26}$$

The following properties of $\mathcal{F}_j$ are generalizations of well known properties of the DFT:

Lemma 3.2 *Suppose that* $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$, $\mathbf{B} = \mathcal{C}(\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1})$ *and* $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$, *where* $\mathbf{A}_j, \mathbf{C}_j \in \mathbb{C}^{k\times l}$ *and* $\mathbf{B}_j \in \mathbb{C}^{l\times m}$, $0 \le j \le N-1$. *Then for every* $0 \le j \le N-1$ *the following holds:*

1. $(\mathcal{F}_j(\mathbf{A}))^* = \mathcal{F}_j(\mathbf{A}^*)$.

2. $\mathcal{F}_j(\mathbf{I}_{mN}) = \mathbf{I}_m$.

3. $\mathcal{F}_j(\mathbf{A}+\mathbf{C}) = \mathcal{F}_j(\mathbf{A}) + \mathcal{F}_j(\mathbf{C})$.

4. $\mathcal{F}_j(\mathbf{A}\mathbf{B}) = \mathcal{F}_j(\mathbf{A})\mathcal{F}_j(\mathbf{B})$.

5. *If* $k = l$ *and* $\mathbf{A}$ *is invertible then* $\mathcal{F}_j(\mathbf{A}^{-1}) = (\mathcal{F}_j(\mathbf{A}))^{-1}$.

Indeed, given a length-M sequence $x[n]$ with DFT $X[k]$, property 1 in Lemma 3.2 is analogous to the scalar DFT property $X^*[k] = \mathcal{F}\left(x^*[-n \bmod M]\right)$. Property 5 is analogous to the convolution property (taking into account that matrix multiplication represents convolution).

An important special case of block circulant matrices are matrices of the form:

$$\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1) \stackrel{\triangle}{=} \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_1) = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 & \cdots & \mathbf{A}_1 \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_1 & \cdots & \mathbf{A}_0 \end{pmatrix}. \tag{27}$$

In this case there are only two different DFT components:

$$\begin{aligned} \mathcal{F}_0(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) &= \mathbf{A}_0 + (N-1)\mathbf{A}_1, && (28) \\ \mathcal{F}_j(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) &= \mathbf{A}_0 - \mathbf{A}_1, \;\; 1 \le j \le N-1. && (29) \end{aligned}$$

It is also easy to see that there are only two different inverse DFT components:

$$\begin{aligned} \mathcal{F}_0^{-1}(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) &= \frac{1}{N}(\mathbf{A}_0 + (N-1)\mathbf{A}_1), && (30) \\ \mathcal{F}_j^{-1}(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) &= \frac{1}{N}(\mathbf{A}_0 - \mathbf{A}_1), \;\; 1 \le j \le N-1. && (31) \end{aligned}$$

3.2 Eigenvalues of Hermitian Block Circulant Matrices

In this section, we consider the eigenvalues and eigenvectors of an Hermitian block circulant matrix. We use the following notation: Let $\mathbf{A} \in \mathbb{C}^{k\times k}$ be an Hermitian matrix. Then a matrix $\mathbf{U} \in \mathbb{C}^{k\times k}$ is called an *eigenvector matrix* of $\mathbf{A}$ if its columns are independent eigenvectors of $\mathbf{A}$. Theorem 3.1 below shows that the eigenvalues of a block circulant matrix are exactly the eigenvalues of its discrete Fourier components. This theorem will be important in the analysis later in the paper.

Theorem 3.1 *Let $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1} \in \mathbb{C}^{k\times k}$ be matrices such that $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ is an Hermitian matrix. For each $0 \leq j \leq N-1$ let $\mathbf{U}_j$ be an eigenvector matrix of $\mathcal{F}_j(\mathbf{A}) = \sum_{i=0}^{N-1} \omega^{ij}\mathbf{A}_i$, where $\omega = e^{-\frac{2\pi i}{N}}$, and let $\lambda_{j,0}, \lambda_{j,1}, \ldots, \lambda_{j,k-1}$ be the eigenvalues of $\mathcal{F}_j(\mathbf{A})$. Then:*

1. *An eigenvector matrix of $\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ is the matrix*

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_0 & \mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_{N-1} \\ \mathbf{U}_0 & \omega\mathbf{U}_1 & \omega^2\mathbf{U}_2 & \cdots & \omega^{N-1}\mathbf{U}_{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{U}_0 & \omega^{N-1}\mathbf{U}_1 & \omega^{2(N-1)}\mathbf{U}_2 & \cdots & \omega^{(N-1)(N-1)}\mathbf{U}_{N-1} \end{pmatrix}. \tag{32}$$

2. *The eigenvalues of $\mathbf{A}$ are the $N \cdot k$ eigenvalues $\lambda_{j,i}$ $, 0 \leq i \leq k-1, 0 \leq j \leq N-1$.*

Proof: see Appendix A. □

4 Minimax MSE Multichannel Estimator

We now use the properties of block circulant matrices and the DFT discussed in the previous section in order to find the $\mathbf{G}$ which is the optimal solution to (16). Section 4.1 establishes the fact that $\mathbf{G}$ can always be chosen as a block circulant matrix. In Section 4.2 we use this structure of $\mathbf{G}$ to find an SDP formulation of the estimation problem (16), 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 $\mathbf{x}$ [23, 24, 25]. The advantage in this formulation is that it readily lends itself to efficient computational methods. Indeed, by exploiting the many well known algorithms for solving SDPs [24, 23], *e.g.,* interior point methods[2] [25, 26],

[2]Interior point methods are iterative algorithms that terminate once a pre-specified accuracy has been reached. A worst case analysis of interior point methods shows that the effort required to solve an SDP to a given accuracy grows no faster than a polynomial of the problem size. In practice, the algorithms behave much better than predicted by the worst case analysis, and in fact in many cases the number of iterations is almost constant in the size of the problem.

the optimal estimator can be computed efficiently in polynomial time. Furthermore, SDP-based algorithms are guaranteed to converge to the global optimum.

4.1 The Structure of G

In Theorem 4.1 below, we show that in the general model (5) and (6) the minimax MSE multichannel matrix $\mathbf{G}$ that is the solution to (16) can be chosen as $\mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ for some $\mathbf{G}_0, \ldots, \mathbf{G}_{N-1} \in \mathbb{C}^{m\times n}$. In this case $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})\mathbf{y}$, so that

$$\hat{\mathbf{x}}_k = \sum_{i=0}^{N-1} \mathbf{G}_{i+k}\mathbf{y}_i, \quad 0 \le k \le N-1. \tag{33}$$

Note, that (33) 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 4.1 *Let* $\mathbf{x} = \left(\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T\right)^T$ *denote the unknown deterministic parameters in the model* $\mathbf{y} = (\mathbf{H} + \Delta)\mathbf{x} + \mathbf{w}$*, where* $\mathbf{H} = \mathcal{C}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1})$ *and* $\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1}$ *are known* $n \times m$ *matrices,* $\Delta \in \mathcal{U}_G = \{\Delta = \mathcal{C}(\Delta_0, \ldots, \Delta_{N-1}) : \|\Delta_0\| \le \rho_0, \ldots, \|\Delta_{N-1}\| \le \rho_{N-1}\}$ *and* $\mathbf{w}$ *is nonzero-mean random vector with covariance* $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$*. Then there exists an optimal solution* $\mathbf{G}$ *to*

$$\begin{aligned} &\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \\ &\min_{\mathbf{G}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} \left\{\mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\Delta))\mathbf{x} + Tr(\mathbf{G}\mathbf{C}\mathbf{G}^*)\right\}, \end{aligned}$$

which is equal to $\mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ *for some* $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1} \in \mathbb{C}^{m\times n}$*.*

Proof: Before we begin the proof, we introduce some notation. The set of all permutations of $\{0, 1, \ldots, N-1\}$ is denoted by S_N. For every permutation $\sigma \in S_N$ and a positive integer l, we associate an $lN \times lN$ matrix $\mathbf{P}_{\sigma,l}$ comprised of $N \times N$ blocks of size $l \times l$. The (i, j) block of $\mathbf{P}_{\sigma,l}$ is defined as:

$$(\mathbf{P}_{\sigma,l})_{i,j} = \delta_{j,\sigma(i)}\mathbf{I}_l, \tag{34}$$

where

$$\delta_{i,j} = \begin{cases} 0, & i \ne j, \\ 1, & i = j \end{cases} \tag{35}$$

is the kronecker delta. For example, if $N=4$ and $\sigma(0)=1, \sigma(1)=0, \sigma(2)=2$ and $\sigma(3)=3$, then,

$$\mathbf{P}_{\sigma,4} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_4 \end{pmatrix}, \tag{36}$$

where $\mathbf{I}_4$ is the identity matrix of size 4×4. We will be interested particularly in a special class of permutations:

$$\mathcal{A} = \{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}, \tag{37}$$

where $\sigma_k(i) = (i+k) \bmod N$. For example, if N=4, then

$$\mathbf{P}_{\sigma_0,4} = \begin{pmatrix} \mathbf{I}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_4 \end{pmatrix}, \quad \mathbf{P}_{\sigma_2,4} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_4 & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{38}$$

Permutation matrices $\mathbf{P}_{\sigma,l}$ satisfy some interesting properties that will be useful later on in the proof:

1. For every $\sigma \in S_N$ and positive integer l, $\mathbf{P}_{\sigma,l}\mathbf{P}^*_{\sigma,l} = \mathbf{P}^*_{\sigma,l}\mathbf{P}_{\sigma,l} = \mathbf{I}$.
2. For every $\sigma \in S_N$ and for every block vector $\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T$, $\mathbf{P}\mathbf{x} = \mathbf{y}$ where

$$\mathbf{y}_k = \mathbf{x}_{\sigma(k)}, \quad 0 \leq k \leq N-1. \tag{39}$$

3. For every block circulant matrix $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots \mathbf{A}_{N-1})$, where $\mathbf{A}_k \in \mathbb{C}^{m,n}$ and every permutation σ in the class $\mathcal{A}$, we have that $\mathbf{P}_{\sigma,m}\mathbf{A}\mathbf{P}^*_{\sigma,n} = \mathbf{A}$, or equivalently, $\mathbf{P}_{\sigma,m}\mathbf{A} = \mathbf{A}\mathbf{P}_{\sigma,n}$.

We are now ready to prove that there is an optimal $\mathbf{G}$ which is block circulant. First, we show that if $\mathbf{G}$ is an optimal solution to (16) then so is $\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}$ for every permutation $\sigma \in \mathcal{A}$. To this end we prove that $\Gamma(\mathbf{G}) = \Gamma(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})$ where

$$\Gamma(\mathbf{G}) = \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \leq L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \leq L, \Delta \in \mathcal{U}_G} \{\mathbf{x}^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I} - \mathbf{G}(\mathbf{H}+\Delta))\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*)\} \tag{40}$$

is the objective function in the minimization problem (16). Indeed, (the number of the property used is indicated):

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) &\overset{1}{=} \mathrm{Tr}(\mathbf{P}^*_{\sigma,m}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{C}\mathbf{G}^*) = \mathrm{Tr}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{C}\mathbf{G}^*\mathbf{P}^*_{\sigma,m}) \\
&\overset{3}{=} \mathrm{Tr}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}\mathbf{C}\mathbf{P}_{\sigma,n}\mathbf{G}^*\mathbf{P}^*_{\sigma,m}) = \mathrm{Tr}((\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})\mathbf{C}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})^*),
\end{aligned}
$$

and,

$$
\begin{aligned}
&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L,\Delta\in\mathcal{U}_G} \{\mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))\mathbf{x}\} \overset{2}{=} \\
&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L,\Delta\in\mathcal{U}_G} \{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))\mathbf{P}_{\sigma,m}\mathbf{x}\} \overset{1}{=} \\
&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L,\Delta\in\mathcal{U}_G} \{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*\mathbf{P}^*_{\sigma,m}\mathbf{P}_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))\mathbf{P}_{\sigma,m}\mathbf{x}\} \overset{1}{=} \\
&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L,\Delta\in\mathcal{U}_G} \{\mathbf{x}^*(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}(\mathbf{H}+\Delta)\mathbf{P}^*_{\sigma,m})^*(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}(\mathbf{H}+\Delta)\mathbf{P}^*_{\sigma,m})\mathbf{x}\} \overset{3}{=} \\
&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L,\Delta\in\mathcal{U}_G} \{\mathbf{x}^*(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})(\mathbf{H}+\Delta))^*(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})(\mathbf{H}+\Delta))\mathbf{x}\},
\end{aligned}
$$

where in the last equality we used the fact that $\mathbf{H}+\Delta$ is a block circulant matrix. Since (16) is a convex optimization problem, if $\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}$ is an optimal solution to (16) for all $\sigma \in \mathcal{A}$, then so is the convex combination $\frac{1}{N}\sum_{\sigma\in\mathcal{A}}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}$. We can immediately show that $\frac{1}{N}\sum_{\sigma\in\mathcal{A}}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n} = \mathcal{C}(\mathbf{G}_0,\mathbf{G}_1,\ldots,\mathbf{G}_{N-1})$ for some matrices $\mathbf{G}_0,\mathbf{G}_1,\ldots,\mathbf{G}_{N-1}\in\mathbb{C}^{m\times n}$. Specifically, if

$$
\mathbf{G} = \begin{pmatrix} \mathbf{G}_{00} & \mathbf{G}_{01} & \cdots & \mathbf{G}_{0,N-1} \\ \mathbf{G}_{10} & \mathbf{G}_{11} & \cdots & \mathbf{G}_{1,N-1} \\ \vdots & \vdots & & \vdots \\ \mathbf{G}_{N-1,0} & \mathbf{G}_{N-1,1} & \cdots & \mathbf{G}_{N-1,N-1} \end{pmatrix}, \tag{41}
$$

then $\mathbf{G}_k = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{G}_{i,i+k}$, $0\le k\le N-1$ completing the proof of the theorem. $\square$

4.2 SDP Formulation of the Estimation Problem

We now use Theorem 4.1 to develop an SDP formulation of (16). We first consider the inner maximization problem

$$\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))\mathbf{x}. \tag{42}$$

As a result of Theorem 4.1 we have that $\mathbf{G}=\mathcal{C}(\mathbf{G}_0,\mathbf{G}_1,\ldots,\mathbf{G}_{N-1})$. Since $\mathbf{I}$ and $\mathbf{H}+\Delta$ are also block circulant matrices, Lemma 3.1 implies that $(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))$ is a block circulant matrix. Therefore, there exists $\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1}\in\mathbb{C}^{m\times m}$ such that

$$(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))=\mathcal{C}(\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1}), \tag{43}$$

and $\mathcal{C}(\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1})$ is an Hermitian matrix. Using (43), (42) can be expressed as:

$$\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L} \mathbf{x}^*\mathcal{C}(\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1})\mathbf{x}. \tag{44}$$

The following lemma is the key result which enables us to solve (44).

Lemma 4.1 *Let $\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1}\in\mathbb{C}^{m\times m}$ be matrices such that $\mathcal{C}(\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1})$ is an Hermitian matrix. Let $\mathbf{T}$ be a positive definite matrix and let $L>0$ be a constant. Then,*

$$\begin{aligned}&\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L} \mathbf{x}^*\mathcal{C}(\mathbf{A}_0,\ldots,\mathbf{A}_{N-1})\mathbf{x}=\\ &NL^2\max_{0\le j\le N-1}\left\{\lambda max\left(\mathbf{T}^{-1/2}\left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{A}_i\right)\mathbf{T}^{-1/2}\right)\right\},\end{aligned} \tag{45}$$

where $w=e^{-\frac{2\pi i}{N}}$. Furthermore,

$$\max_{\|\mathbf{x}_0\|_{\mathbf{T}}\le L,\ldots,\|\mathbf{x}_{N-1}\|_{\mathbf{T}}\le L} \mathbf{x}^*\mathcal{C}(\mathbf{A}_0,\ldots,\mathbf{A}_{N-1})\mathbf{x}=\max_{\|\mathbf{x}\|^2_{\mathbf{T}}\le NL^2}\mathbf{x}^*\mathcal{C}(\mathbf{A}_0,\ldots,\mathbf{A}_{N-1})\mathbf{x}.$$

Proof: See Appendix B. □

The expression $\sum_{i=0}^{N-1}\omega^{ij}\mathbf{A}_i$ which appears in (45) is exactly $\mathcal{F}_j((\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))$. By the properties listed in Lemma 3.2 we can deduce that for every $0\le j\le N-1$,

$$
\begin{aligned}
\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i &= \mathcal{F}_j((\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))) \\
&= \mathcal{F}_j\left((\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*\right)\mathcal{F}_j\left(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta)\right) \\
&= \left(\mathcal{F}_j(\mathbf{I})-\mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\mathbf{H}+\Delta)\right)^*\left(\mathcal{F}_j(\mathbf{I})-\mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\mathbf{H}+\Delta)\right) \\
&= \left(\mathbf{I}-\mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\mathbf{H}+\Delta)\right)^*\left(\mathbf{I}-\mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\mathbf{H}+\Delta)\right) \\
&= \left(\mathbf{I}-\mathcal{F}_j(\mathbf{G})(\mathcal{F}_j(\mathbf{H})+\mathcal{F}_j(\Delta))\right)^*\left(\mathbf{I}-\mathcal{F}_j(\mathbf{G})(\mathcal{F}_j(\mathbf{H})+\mathcal{F}_j(\Delta))\right). \qquad (46)
\end{aligned}
$$

Substituting (46) into (45) we conclude that

$$
\max_{\|\mathbf{x}_0\|_{\mathbf{T}} \leq L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \leq L, \Delta \in \mathcal{U}_G} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))^*(\mathbf{I}-\mathbf{G}(\mathbf{H}+\Delta))\mathbf{x} = NL^2 \max_{\Delta \in \mathcal{U}_G} \max_{0 \leq j \leq N-1} \alpha_j(\Delta) \quad (47)
$$

where

$$
\alpha_j(\Delta) = \lambda_{\max}\left(\mathbf{T}^{-1/2}(\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j))^*(\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j)\mathbf{T}^{-1/2}\right), \quad 0 \leq j \leq N-1, \qquad (48)
$$

and $\widehat{\mathbf{G}}_j, \widehat{\mathbf{H}}_j$ and $\widehat{\Delta}_j$ are defined by

$$
\begin{aligned}
\widehat{\mathbf{G}}_j &= \mathcal{F}_j(\mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})) = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{G}_i, \quad 0 \leq j \leq N-1, \\
\widehat{\mathbf{H}}_j &= \mathcal{F}_j(\mathcal{C}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1})) = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{H}_i, \quad 0 \leq j \leq N-1, \\
\widehat{\Delta}_j &= \mathcal{F}_j(\mathcal{C}(\Delta_0, \Delta_1, \ldots, \Delta_{N-1})) = \sum_{i=0}^{N-1} \omega^{ij} \Delta_i, \quad 0 \leq j \leq N-1. \qquad (49)
\end{aligned}
$$

We can express (47) as the solution to the problem

$$
\min_{\tau} NL^2 \tau \qquad (50)
$$

subject to

$$\mathbf{T}^{-1/2}(\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j))^*(\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j))\mathbf{T}^{-1/2} \preceq \tau\mathbf{I}, \quad \forall \Delta \in \mathcal{U}_G, \quad 0 \le j \le N-1. \tag{51}$$

We now rely on the following important result:

Lemma 4.2 (Schur's complement [23]) *Let*

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{C} \end{pmatrix}$$

be a Hermitian matrix with $\mathbf{C} \succ 0$*. Then* $\mathbf{M} \succeq 0$ *if and only if* $\Delta_{\mathbf{C}} \succeq 0$*, where* $\Delta_{\mathbf{C}}$ *is the Schur complement of* $\mathbf{C}$ *in* $\mathbf{M}$ *and is given by*

$$\Delta_{\mathbf{C}} = \mathbf{A} - \mathbf{B}^*\mathbf{C}^{-1}\mathbf{B}.$$

□

Using Lemma 4.2 we can rewrite the constraint (51) as

$$\begin{pmatrix} \tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j))^* \\ (\mathbf{I}-\widehat{\mathbf{G}}_j(\widehat{\mathbf{H}}_j+\widehat{\Delta}_j))\mathbf{T}^{-1/2} & \mathbf{I} \end{pmatrix} \succeq 0, \quad \forall \Delta \in \mathcal{U}_G, \quad 0 \le j \le N-1, \tag{52}$$

which is equivalent to

$$\mathbf{R}_j \succeq \mathbf{P}_j^*\widehat{\Delta}_j\mathbf{Q}_j + \mathbf{Q}_j^*\widehat{\Delta}_j^*\mathbf{P}_j, \quad \forall \Delta \in \mathcal{U}_G, \quad 0 \le j \le N-1, \tag{53}$$

where

$$\mathbf{R}_j = \begin{pmatrix} \tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\widehat{\mathbf{G}}_j\widehat{\mathbf{H}}_j)^* \\ (\mathbf{I}-\widehat{\mathbf{G}}_j\widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} \end{pmatrix}, \quad \mathbf{P}_j = \begin{pmatrix} \mathbf{0} & \widehat{\mathbf{G}}_j^* \end{pmatrix}, \quad \mathbf{Q}_j = \begin{pmatrix} \mathbf{T}^{-1/2} & \mathbf{0} \end{pmatrix}. \tag{54}$$

We now exploit the following lemma, the proof of which is provided in Appendix C:

Lemma 4.3 *Given matrices* $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ *with* $\mathbf{R} = \mathbf{R}^*$*,*

$$\mathbf{R} \succeq \mathbf{P}^*\widehat{\mathbf{X}}_j\mathbf{Q} + \mathbf{Q}^*\widehat{\mathbf{X}}_j^*\mathbf{P}, \quad \forall \mathbf{X} \in \mathcal{U}_G$$

if and only if there exists a $\lambda \geq 0$ *such that*

$$\begin{pmatrix} \mathbf{R} - \lambda \mathbf{Q}^*\mathbf{Q} & -\rho \mathbf{P}^* \\ -\rho \mathbf{P} & \lambda \mathbf{I} \end{pmatrix} \succeq 0,$$

where $\rho = \sum_{j=0}^{N-1} \rho_j$.

From Lemma 4.3 it follows that (53) is satisfied if and only if there exists $\lambda_j \geq 0, 0 \leq j \leq N-1$ such that

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \widehat{\mathbf{G}}_j \widehat{\mathbf{H}}_j)^* & \mathbf{0} \\ (\mathbf{I} - \widehat{\mathbf{G}}_j \widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho \widehat{\mathbf{G}}_j \\ \mathbf{0} & -\rho \widehat{\mathbf{G}}_j^* & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \leq j \leq N-1, \tag{55}$$

with $\rho = \sum_{j=0}^{N-1} \rho_j$. Thus, it follows that the problem (16) reduces to

$$\min_{\tau, \lambda_0, \ldots, \lambda_{N-1}, \mathbf{G}} \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) + NL^2\tau \tag{56}$$

subject to (55).

Since $\mathbf{C}$ and $\mathbf{G}$ are both block circulant matrices, by Lemma 3.1 the product $\mathbf{G}\mathbf{C}\mathbf{G}^*$ is also a block circulant matrix. Let

$$\mathbf{G}\mathbf{C}\mathbf{G}^* = \mathcal{C}(\mathbf{S}_0, \mathbf{S}_2, \ldots, \mathbf{S}_{N-1}), \tag{57}$$

for some $\mathbf{S}_0, \mathbf{S}_2, \ldots, \mathbf{S}_{N-1} \in \mathbb{C}^{m \times m}$. Then, $\mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) = N\mathrm{Tr}(\mathbf{S}_0)$. But,

$$N\mathbf{S}_0 = N\mathcal{F}_0^{-1}(\mathcal{F}(\mathbf{G}\mathbf{C}\mathbf{G}^*)) = \sum_{j=0}^{N-1} \mathcal{F}_j(\mathbf{G}\mathbf{C}\mathbf{G}^*) = \sum_{j=0}^{N-1} \mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\mathbf{C})\mathcal{F}_j(\mathbf{G}^*) = \sum_{j=0}^{N-1} \widehat{\mathbf{G}}_j \widehat{\mathbf{C}}_j \widehat{\mathbf{G}}_j^*. \tag{58}$$

Making the change of variables

$$\mathbf{A}_j = \widehat{\mathbf{G}}_j, \ \ 0 \leq j \leq N-1 \tag{59}$$

we arrive at the following formulation of the minimax problem (16):

$$\min_{\tau,\lambda_0,\ldots,\lambda_{N-1},\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1}} NL^2\tau + \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{A}_j\widehat{\mathbf{C}}_j\mathbf{A}_j^*) \tag{60}$$

subject to

$$\begin{pmatrix} \tau\mathbf{I} - \lambda_j\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j)^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho\mathbf{A}_j \\ \mathbf{0} & -\rho\mathbf{A}_j^* & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1, \tag{61}$$

which is equivalent to

$$\min_{\tau,t_0,\ldots,t_{N-1},\mathbf{A}_0,\ldots,\mathbf{A}_{N-1},\lambda_0,\ldots,\lambda_{N-1}} \sum_{j=0}^{N-1} t_j + NL^2\tau \tag{62}$$

subject to the LMI (61) and

$$\mathrm{Tr}(\mathbf{A}_j\widehat{\mathbf{C}}_j\mathbf{A}_j^*) \le t_j, \quad 0 \le j \le N-1. \tag{63}$$

Using Lemma 4.2, (63) can be expressed as the LMI

$$\begin{pmatrix} t_j & \mathbf{a}_j^* \\ \mathbf{a}_j & \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1 \tag{64}$$

$\mathbf{a}_j = vec(\mathbf{A}_j\widehat{\mathbf{C}}_j^{1/2}),\ 0 \le j \le N-1$, and $\mathbf{m} = vec(\mathbf{M})$ denotes the vector obtained by stacking the columns of $\mathbf{M}$. Thus, our problem reduces to that of (62) subject to (64) and (61), which is an SDP.

We summarize the our results in the following two theorems. In Theorem 4.2 we present the SDP formulation for the general model and in Theorem 4.3 we present the SDP formulation of the symmetric model.

Theorem 4.2 (SDP Formulation for the General Model) *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* $\mathbf{H} = \mathcal{C}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1})$ *and* $\mathbf{H}_0, \ldots, \mathbf{H}_{N-1}$ *are known* $n \times m$ *matrices,* Δ *is an unknown matrix satisfying* $\Delta \in \mathcal{U}_G = \{\Delta = \mathcal{C}(\Delta_0, \ldots, \Delta_{N-1}) : \|\Delta_0\| \le \rho_0, \ldots, \|\Delta_{N-1}\| \le \rho_{N-1}\}$ *and* $\mathbf{w}$ *is zero-mean* *random vector with covariance* $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$*. Then there exists a solution to*

$$\begin{array}{l} \displaystyle\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \\ \displaystyle\min_{\mathbf{G}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} \left\{ \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} + Tr(\mathbf{G}\mathbf{C}\mathbf{G}^*) \right\}, \end{array}$$

given by $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ *where*

$$\mathbf{G}_j = \frac{1}{N} \sum_{i=0}^{N-1} \omega^{-ij} \mathbf{A}_i, \quad 0 \le j \le N-1.$$

Here $\omega = e^{-\frac{2\pi i}{N}}$*, and* $\mathbf{A}_j,\ \ 0 \le j \le N-1$ *are the solutions to the SDP*

$$\min_{\tau, \lambda_0, \ldots, \lambda_{N-1}, t_0, \ldots, t_{N-1}, \mathbf{A}_0, \ldots, \mathbf{A}_{N-1}} NL^2\tau + \sum_{j=0}^{N-1} t_j$$

subject to

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

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

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

$$\begin{aligned} \widehat{\mathbf{H}}_j &= \sum_{i=0}^{N-1} \omega^{ij} \mathbf{H}_i, \quad 0 \le j \le N-1, \\ \widehat{\mathbf{C}}_j &= \sum_{i=0}^{N-1} \omega^{ij} \mathbf{C}_i, \quad 0 \le j \le N-1. \end{aligned}$$

Using similar ideas, we can derive the minimax MSE estimator for the symmetric model:

Theorem 4.3 (SDP Formulation for the Symmetric Model) *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* $\mathbf{H} = \mathcal{M}(\mathbf{H}_0, \mathbf{H}_1)$ *and* $\mathbf{H}_0, \mathbf{H}_1$ *are known* $n \times m$ *matrices,* Δ *is an unknown matrix satisfying* $\Delta \in \mathcal{U}_S = \{\Delta = \mathcal{M}(\Delta_0, \Delta_1) : \|\Delta_0\| \le \rho_0, \|\Delta_1\| \le \rho_1\}$ *and* $\mathbf{w}$ *is zero-mean random vector with covariance* $\mathbf{C} =$

$\mathcal{M}(\mathbf{C}_0, \mathbf{C}_1)$. *Then there exists a solution to*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) =$$
$$\min_{\mathbf{G}} \max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L, \Delta \in \mathcal{U}_G} \left\{ \mathbf{x}^*(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} + Tr(\mathbf{G}\mathbf{C}\mathbf{G}^*) \right\},$$

given by $\mathbf{G} = \mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$ *where*

$$\mathbf{G}_0 = \frac{1}{N}(\mathbf{A}_0 + (N-1)\mathbf{A}_1), \quad \mathbf{G}_1 = \frac{1}{N}(\mathbf{A}_0 - \mathbf{A}_1).$$

Here $\omega = e^{-\frac{2\pi i}{N}}$, *and* $\mathbf{A}_0, \mathbf{A}_1$ *are the solutions to the SDP*

$$\min_{\tau, \lambda_0, \lambda_1, t_0, t_1, \mathbf{A}_0, \mathbf{A}_1} NL^2\tau + t_0 + (N-1)t_1$$

subject to

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

where $\mathbf{a}_j = vec(\mathbf{A}_j \widehat{\mathbf{C}}_j^{1/2})$, $\widehat{\rho}_0 = \rho_0 + (N-1)\rho_{N-1}$, $\widehat{\rho}_1 = \rho_0 + \rho_1$ *and*

$$\begin{aligned} \widehat{\mathbf{H}}_0 &= \mathbf{H}_0 + (N-1)\mathbf{H}_1, \quad \widehat{\mathbf{H}}_1 = \mathbf{H}_0 - \mathbf{H}_1, \\ \widehat{\mathbf{C}}_0 &= \mathbf{C}_0 + (N-1)\mathbf{C}_1, \quad \widehat{\mathbf{C}}_1 = \mathbf{C}_0 - \mathbf{C}_1. \end{aligned}$$

Proof: The proof is almost identical to the proof of Theorem 4.2 and only the two major differences between the proofs will be given here. The first difference is that in the symmetric case, $\mathbf{G}$ is of the form $\mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$ and thus has only two different DFT components $\mathbf{A}_0 = \mathbf{G}_0 + (N-1)\mathbf{G}_1$ and $\mathbf{A}_1 = \mathbf{G}_1 - \mathbf{G}_1$. In the same manner, $\mathbf{H}$, Δ and $\mathbf{C}$ also have only two different DFT components. As a result, the inner maximization with respect to $\mathbf{x}$ (42) reduces to

$$\max\{\alpha_0(\Delta), \alpha_1(\Delta)\} + \mathrm{Tr}(\mathbf{A}_0 \widehat{\mathbf{C}}_0 \mathbf{A}_0^T) + (N-1)\mathrm{Tr}(\mathbf{A}_1 \widehat{\mathbf{C}}_1 \mathbf{A}_1^T) \tag{65}$$

where $\alpha_0(\Delta)$ and $\alpha_1(\Delta)$ are defined as in (48). The second main difference is that we need to use a slightly different version Lemma 4.3. The new version of Lemma 4.3 states that for each $j = 0, 1$

$$\mathbf{R} \succeq \mathbf{P}^* \widehat{\mathbf{X}}_j \mathbf{Q} + \mathbf{Q}^* \widehat{\mathbf{X}}_j^* \mathbf{P}, \forall \mathbf{X} \in \mathcal{U}_S \tag{66}$$

if and only if there exists a $\lambda \geq 0$ such

$$\begin{pmatrix} \mathbf{R} - \lambda \mathbf{Q}^* \mathbf{Q} & -\hat{\rho}_j \mathbf{P}^* \\ -\hat{\rho}_j \mathbf{P} & \lambda \mathbf{I} \end{pmatrix} \succeq 0,$$

where $\hat{\rho}_0 = \rho_0 + (N-1)\rho_1$ and $\hat{\rho}_1 = \rho_0 + \rho_1$.$\square$

Remark: In the case where the block matrices $\mathbf{H}_0, \ldots, \mathbf{H}_{N-1}$ are known exactly, *i.e.,* in the case where

$$\rho_0 = \rho_1 = \ldots = \rho_{N-1} = 0 \tag{67}$$

the LMI (61) reduces to

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \leq j \leq N-1. \tag{68}$$

Since $\lambda_j \geq 0$ we have that (68) is the same as

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)^* \\ (\mathbf{I} - \mathbf{A}_j \widehat{\mathbf{H}}_j)\mathbf{T}^{-1/2} & \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \leq j \leq N-1. \tag{69}$$

Thus, we can formulate a simplified SDP for the minimax MSE estimator in the case where $\mathbf{H}$ is known exactly.

In the next section we develop explicit expressions for the minimax MSE estimator in the case $\mathbf{T} = \mathbf{I}$ for both the known and unknown $\mathbf{H}$ scenarios.

5 Minimax MSE Estimator for $\mathbf{T} = \mathbf{I}$

In this section we discuss a special case of minimax MSE estimator problem where $\mathbf{T} = \mathbf{I}$. For the case in which $\mathbf{H}$ is certain we find an explicit expression for the optimal minimax MSE estimator. In the case of uncertain $\mathbf{H}$ show that the SDP problem of Theorem 4.2 can be reduced to a simple convex optimization problem in $N+1$ unknowns. Moreover, for the symmetric model case, the SDP of Theorem 4.3 is reduced to a simple convex optimization problem in 3 unknowns regardless of the value of N.

5.1 Minimax MSE Estimator for $\mathbf{T} = \mathbf{I}$ with Known $\mathbf{H}$

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 $\mathbf{H}$ 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 [15] where it was shown that the minimax MSE estimator for the case $\mathbf{T} = \mathbf{I}$ is given by

$$\hat{\mathbf{x}} = \alpha(\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}^{-1}y, \tag{70}$$

with $\alpha = \frac{NL^2}{NL^2+B}$ where

$$B = \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}\right). \tag{71}$$

The estimator of (70) is a shrunken estimator proposed by Mayer and Willke [27], which is simply a scaled version of the LS estimator with an optimal choice of shrinkage factor.

Note that the dominant computation in (70) and (71) is the inversion of the $mN \times mN$ matrix $\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}$, which requires $O(m^3N^3)$ operations. This number is prohibitively large even for medium size problems. On the other hand, the calculation stemming from Theorem 5.1 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 $O(m^3N)$ operations. For example, if $N = 100$ then our computation is 10000 cheaper than the direct computation.

Theorem 5.1 *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* $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{w}$*, where* $\mathbf{H} = \mathcal{C}(\mathbf{H}_0, \ldots, \mathbf{H}_{N-1})$ *and* $\mathbf{H}_j,\ 0 \le j \le N-1$ *are known* $n \times m$ *matrices, and* $\mathbf{w}$ *is a zero-mean random vector with covariance* $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$*. 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*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\|\le L,\dots,\|\mathbf{x}_{N-1}\|\le L} E\left(\|\hat{\mathbf{x}}-\mathbf{x}\|^2\right)$$

is given by

$$\hat{\mathbf{x}}_k = \sum_{i=0}^{N-1} \mathbf{G}_{i+k}\mathbf{y}_i \tag{72}$$

where

$$\mathbf{G}_j = \frac{1}{N}\sum_{i=0}^{N-1} \omega^{-ij}\mathbf{A}_i, \quad 0 \le j \le N-1,$$

and

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

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*

$$\widehat{\mathbf{H}}_j = \sum_{i=0}^{N-1}\omega^{ij}\mathbf{H}_i, \quad \widehat{\mathbf{C}}_j = \sum_{i=0}^{N-1}\omega^{ij}\mathbf{C}_i, \quad 0 \le j \le N-1.$$

Proof: We know from [15] that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ where

$$\mathbf{G} = \frac{NL^2}{NL^2+B}(\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}^{-1},$$

and $B = \mathrm{Tr}((\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1})$, which is equal to $\sum_{i=1}^{mN}\frac{1}{\lambda_i}$, where $\lambda_1, \lambda_2, \dots, \lambda_{mN}$ are the eigenvalues of $\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}$. From Theorem 3.1 it follows that $B = \sum_{j=0}^{N-1}\mathrm{Tr}\left((\widehat{\mathbf{H}}_j^*\widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{H}}_j)^{-1}\right)$. By Theorem 4.1, $\mathbf{G}$ is a block circulant matrix and thus is equal to $\mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_{N-1})$ for some $\mathbf{G}_0, \mathbf{G}_1, \dots, \mathbf{G}_{N-1} \in \mathbb{C}^{m\times n}$. Using the properties listed in Lemma 3.2, we can calculate the jth DFT component $\mathbf{A}_j = \mathbf{G}_j$,

$$\mathbf{A}_j = \mathcal{F}_j\left(\frac{NL^2}{NL^2+B}(\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}^{-1}\right) = \frac{NL^2}{NL^2+B}(\widehat{\mathbf{H}}_j^*\widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{H}}_j)^{-1}\widehat{\mathbf{H}}_j^*\widehat{\mathbf{C}}_j^{-1}.$$

Applying the inverse DFT we obtain the desired expression for $\mathbf{G}_j$:

$$\mathbf{G}_j = \mathcal{F}_j^{-1}(\mathcal{C}(\widehat{\mathbf{G}}_0, \widehat{\mathbf{G}}_1, \ldots, \widehat{\mathbf{G}}_{N-1})) = \frac{1}{N}\sum_{i=0}^{N-1} \omega^{-ij}\mathbf{A}_i\,,$$

and the result follows. $\Box$

As can be expected intuitively, when $L \to \infty$, $\hat{\mathbf{x}}$ of (72) reduces to the LS estimator. Indeed, when the norm of $\mathbf{x}$ can be made arbitrarily large, the MSE will also be arbitrarily large unless the bias is equal to zero. Therefore, in this limit, the worst-case estimation error is minimized by choosing an estimator with zero bias that minimizes the variance, which leads to the LS estimator.

5.2 Minimax Estimator for $\mathbf{T} = \mathbf{I}, \mathbf{C} = \boldsymbol{\sigma}^2\mathbf{I}$ and with Unknown H

We now show that in the case where $\mathbf{T} = \mathbf{I}$ and $\mathbf{C} = \sigma^2\mathbf{I}$, the minimax MSE estimator reduces to a simple convex optimization problem in $N+1$ unknowns in the general model and in 3 unknowns in the symmetric models (regardless of the value of N), solved very efficiently. Specifically, we have the following theorem.

Theorem 5.2 (Minimax MSE Estimator for the General Model with $\mathbf{T} = \mathbf{I}$ and $\mathbf{C} = \sigma^2\mathbf{I}$) *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 $\mathbf{H} = \mathcal{C}(\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N-1})$ and $\mathbf{H}_0, \ldots, \mathbf{H}_{N-1}$ are known $n \times m$ matrices, Δ is an unknown matrix satisfying $\Delta \in \mathcal{U}_G = \{\Delta = \mathcal{C}(\Delta_0, \ldots, \Delta_{N-1}) : \|\Delta_0\| \leq \rho_0, \ldots, \|\Delta_{N-1}\| \leq \rho_{N-1}\}$ and $\mathbf{w}$ is zero-mean random vector with covariance $\mathbf{C} = \sigma^2\mathbf{I}$. For every $0 \leq j \leq N-1$, Let $\widehat{\mathbf{H}}_j = \mathbf{U}_j\Sigma_j\mathbf{V}_j^*$ be the singular value decomposition of $\widehat{\mathbf{H}}_j = \sum_{i=0}^{N-1}\omega^{ij}\mathbf{H}_i$, where Σ_j is an $n \times m$ diagonal matrix with diagonal elements $\sigma_{j,i} > 0, \quad 1 \leq i \leq m$, and $\mathbf{U}_j$ and $\mathbf{V}_j$ are unitary matrices. Then there exists a solution to*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\| \leq L, \ldots, \|\mathbf{x}_{N-1}\| \leq L, \Delta \in \mathcal{U}_G} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$$

given by $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ where

$$\mathbf{G}_j = \frac{1}{N}\sum_{i=0}^{N-1}\omega^{-ij}\mathbf{A}_i, \quad 0 \leq j \leq N-1.$$

Here $\omega = e^{-\frac{2\pi i}{N}}$, and

$$\mathbf{A}_j = \mathbf{V}_j\mathbf{Z}_j\mathbf{V}_j^*(\widehat{\mathbf{H}}_j^*\widehat{\mathbf{H}}_j)^{-1/2}\widehat{\mathbf{H}}_j^*, \quad 0 \leq j \leq N-1$$

where $\mathbf{Z}_j$ *is an* $m \times m$ *diagonal matrix with diagonal elements* $z_{j,i} = f_{j,i}(\tau, \lambda_j)$, *with*

$$f_{j,i}(\tau, \lambda_j) = \frac{\sigma_{j,i}\lambda_j - \sqrt{\lambda_j(\tau - \lambda_j)\left(\sigma_{j,i}^2\lambda_j - \rho^2(1 + \lambda_j - \tau)\right)}}{(\tau - \lambda_j)\rho^2 + \sigma_{j,i}^2\lambda_j}, \quad 1 \le i \le m, 0 \le j \le N-1,$$

$\rho = \sum_{j=0}^{N-1} \rho_j$ *and* $\lambda_0, \ldots, \lambda_{N-1}$ *and* τ *are the solution to the convex optimization problem*

$$\min_{\tau, \lambda_0, \ldots, \lambda_{N-1}} \left\{ \sigma^2 \sum_{j=0}^{N-1} \sum_{i=1}^{m} f_{j,i}^2(\tau, \lambda_j) + NL^2\tau \right\}$$

subject to

$$\begin{aligned} \lambda_j \sigma_{j,i}^2 &\ge \rho^2(1 + \lambda_j - \tau), \qquad 1 \le i \le m, 0 \le j \le N-1; \\ \lambda_j &\ge 0, \qquad 0 \le j \le N-1; \\ \tau &\ge \lambda_j, \qquad 0 \le j \le N-1. \end{aligned}$$

Proof: From Theorem 4.1, the optimal estimator $\mathbf{G}$ is equal to $\mathcal{C}(\mathbf{G}_0, \ldots, \mathbf{G}_{N-1})$ where $\mathbf{G}_j = \frac{1}{N}\sum_{i=0}^{N-1} \omega^{-ij}\mathbf{A}_i$ and $(\mathbf{A}_j)_{j=0}^{N-1}$ is the solution to

$$\min_{\tau, \mathbf{A}_0, \ldots, \mathbf{A}_{N-1}, \lambda_0, \ldots, \lambda_{N-1}} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\widehat{\mathbf{A}}_j \widehat{\mathbf{A}}_j^*) + NL^2\tau \right\}, \tag{73}$$

subject to

$$\mathbf{M}_j \stackrel{\triangle}{=} \begin{pmatrix} (\tau - \lambda_j)\mathbf{I} & (\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j)^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j) & \mathbf{I} & -\rho\mathbf{A}_j \\ \mathbf{0} & -\rho\mathbf{A}_j^* & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1. \tag{74}$$

The proof of the proposition is comprised of three parts. First, we show that the optimal solution $(\mathbf{A}_j)_{j=0}^{N-1}$ to (73) and (74) is of the form

$$\mathbf{A}_j = \mathbf{V}_j\mathbf{Z}_j\mathbf{V}_j^* \left(\widehat{\mathbf{H}}_j^*\widehat{\mathbf{H}}_j\right)^{-1/2} \widehat{\mathbf{H}}_j^*, \ \ 0 \le j \le N-1, \tag{75}$$

for some $m \times m$ matrices $\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{N-1}$. We then show that $\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{N-1}$ can be chosen as diagonal matrices. Finally, we find the diagonal elements of $\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{N-1}$.

We begin by showing that the optimal $(\mathbf{A}_j)_{j=0}^{N-1}$ has the form (75). The constraint (74) is equivalent to $\mathbf{Q}_j\mathbf{M}_j\mathbf{Q}_j^* \succeq 0$ for any invertible $\mathbf{Q}_j$. Choosing

$$\mathbf{Q}_j = \begin{pmatrix} \mathbf{V}_j^* & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_j^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_j^* \end{pmatrix}, \quad 0 \le j \le N-1, \tag{76}$$

(74) becomes

$$\begin{pmatrix} (\tau - \lambda_j)\mathbf{I} & \mathbf{V}_j^*(\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j)^*\mathbf{V}_j & \mathbf{0} \\ \mathbf{V}_j^*(\mathbf{I} - \mathbf{A}_j\widehat{\mathbf{H}}_j)\mathbf{V}_j & \mathbf{I} & -\rho\mathbf{V}_j^*\mathbf{A}_j\mathbf{U}_j \\ \mathbf{0} & -\rho\mathbf{U}_j^*\mathbf{A}_j^*\mathbf{V} & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1. \tag{77}$$

Make the following change of variables

$$\mathbf{B}_j \stackrel{\triangle}{=} \mathbf{V}_j^*\mathbf{A}_j\mathbf{U}_j, \tag{78}$$

so that

$$\mathbf{A}_j = \mathbf{V}_j\mathbf{B}_j\mathbf{U}_j^*, \tag{79}$$

the problem of (73) and (77) can be expressed as

$$\min_{\tau,\lambda_0,\ldots,\lambda_{N-1},\mathbf{B}_0,\ldots,\mathbf{B}_{N-1}} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{B}_j^*\mathbf{B}_j) + NL^2\tau \right\}, \tag{80}$$

subject to

$$\begin{pmatrix} (\tau - \lambda_j)\mathbf{I} & (\mathbf{I} - \mathbf{B}_j\Sigma_j)^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{B}_j\Sigma_j) & \mathbf{I} & -\rho\mathbf{B}_j \\ \mathbf{0} & -\rho\mathbf{B}_j^* & \lambda_j \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1. \tag{81}$$

Let $\mathbf{B}_j = (\mathbf{Z}_j \ \mathbf{W}_j)$ where $\mathbf{Z}_j$ is the $m \times m$ matrix consisting of the first m columns of $\mathbf{B}_j$ and let $\widetilde{\Sigma}_j$ denote the $m \times m$ matrix with diagonal elements $\sigma_{j,i}, 1 \le i \le m$ for every $0 \le j \le N-1$.

Then we can express the constraint (81) as

$$\mathbf{L}(\mathbf{B}_j) \triangleq \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j)^* & \mathbf{0} & \mathbf{0} \\ (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j & \mathbf{I} & -\rho\mathbf{Z}_j & -\rho\mathbf{W}_j \\ \mathbf{0} & -\rho\mathbf{Z}_j^* & \lambda_j\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\rho\mathbf{W}_j^* & \mathbf{0} & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1. \tag{82}$$

Clearly, if (82) is satisfied, then

$$\mathbf{K}(\mathbf{Z}_j) \triangleq \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{Z}_j \\ \mathbf{0} & -\rho\mathbf{Z}_j^* & \lambda_j \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1. \tag{83}$$

Now, let $\mathbf{B}_j = (\mathbf{Z}_j\ \mathbf{W}_j)$ be any matrix satisfying (82), and define $\widetilde{\mathbf{B}}_j = [\mathbf{Z}_j\ \mathbf{0}]$. Then,

$$\mathbf{L}(\widetilde{\mathbf{B}}_j) = \begin{pmatrix} \mathbf{K}(\mathbf{Z}_j) & \mathbf{0} \\ \mathbf{0} & \lambda_j \end{pmatrix} \succeq \mathbf{0}, \tag{84}$$

since $\mathbf{K}(\mathbf{Z}_j) \succeq \mathbf{0}$. In addition,

$$\mathrm{Tr}(\widetilde{\mathbf{B}}_j^*\widetilde{\mathbf{B}}_j) = \mathrm{Tr}(\mathbf{Z}_j^*\mathbf{Z}_j) \le \mathrm{Tr}(\mathbf{Z}_j^*\mathbf{Z}_j) + \mathrm{Tr}(\mathbf{W}_j^*\mathbf{W}_j) = \mathrm{Tr}(\mathbf{B}_j^*\mathbf{B}_j). \tag{85}$$

Therefore, the optimal value of $\mathbf{B}_j$ satisfies $\mathbf{W}_j = \mathbf{0}$ for every $0 \le j \le N-1$, so that the problem of (80) and (81) reduces to

$$\min_{\tau,\mathbf{Z}_0,\ldots,\mathbf{Z}_{N-1},\lambda_0,\ldots,\lambda_{N-1}} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{Z}_j^*\mathbf{Z}_j) + NL^2\tau \right\}, \tag{86}$$

subject to (83). Once we find the optimal $(\mathbf{Z}_j)_{j=0}^{N-1}$, the optimal $(\mathbf{A}_j)_{j=0}^{N-1}$ can be found from (79) as

$$\mathbf{A}_j = \mathbf{V}_j\mathbf{Z}_j[\mathbf{I}\ \mathbf{0}]\mathbf{U}_j^* = \mathbf{V}_j\mathbf{Z}_j\mathbf{V}_j^*(\widehat{\mathbf{H}}_j^*\widehat{\mathbf{H}}_j)^{-1/2}\widehat{\mathbf{H}}_j^*, \quad 0 \le j \le N-1, \tag{87}$$

thus completing the first part of the proof.

We now show that the optimal values of $(\mathbf{Z}_j)_{j=0}^{N-1}$ can be chosen as diagonal matrices. To this end, we first note that if $(\mathbf{Z}_j)_{j=0}^{N-1}$ satisfies (83), then for every $0 \leq j \leq N-1$

$$\begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J} \end{pmatrix} \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{Z}_j \\ \mathbf{0} & -\rho\mathbf{Z}_j^* & \lambda_j \end{pmatrix} \begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J} \end{pmatrix} =$$
$$= \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{J}\mathbf{Z}_j\mathbf{J}\widetilde{\Sigma}_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{J}\mathbf{Z}_j\mathbf{J}\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{J}\mathbf{Z}_j\mathbf{J} \\ \mathbf{0} & -\rho\mathbf{J}\mathbf{Z}_j^*\mathbf{J} & \lambda_j \end{pmatrix} \succeq 0. \tag{88}$$

Here $\mathbf{J}$ is any diagonal matrix with diagonal elements ± 1, and we used the fact that diagonal matrices commute and that $\mathbf{J}^*\mathbf{J} = \mathbf{J}^2 = \mathbf{I}$. It follows from (88) that $\mathbf{K}(\widetilde{\mathbf{Z}}_j) \succeq \mathbf{0}$ for any $\mathbf{J}$, where $\widetilde{\mathbf{Z}}_j = \mathbf{J}\mathbf{Z}_j\mathbf{J}$. In addition, we have that $\mathrm{Tr}(\widetilde{\mathbf{Z}}_j^*\widetilde{\mathbf{Z}}_j) = \mathrm{Tr}(\mathbf{Z}_j^*\mathbf{Z}_j)$. Therefore, if $(\mathbf{Z}_j)_{j=0}^{N-1}$ is an optimal solution, then so is $(\mathbf{J}\mathbf{Z}_j\mathbf{J})_{j=0}^{N-1}$. Since our problem is convex, the set of optimal solutions is also convex [32], which implies that $(\mathbf{Z}'_j)_{j=0}^{N-1} = ((1/2^m)\sum_{\mathbf{J}} \mathbf{J}\mathbf{Z}_j\mathbf{J})_{j=0}^{N-1}$ is also a solution, where the summation is over all 2^m diagonal matrices $\mathbf{J}$ with diagonal elements ± 1. It is easy to see that $\mathbf{Z}'_j$ is a diagonal matrix. Therefore, we have shown that there exists an optimal diagonal solution $\mathbf{Z}_j$ for every $0 \leq j \leq N-1$.

Denote the diagonal elements of $\mathbf{Z}_j$ by $z_{j,i}$, $1 \leq i \leq m$, and let $\mathrm{diag}(\alpha_1, \dots \alpha_m)$ denote the $m \times m$ diagonal matrix with diagonal elements α_i. Then the constraint $\mathbf{K}(\mathbf{Z}_j) \succeq \mathbf{0}$ can be written as

$$\begin{pmatrix} \mathrm{diag}(\tau-\lambda_j, \dots, \tau-\lambda_j) & \mathrm{diag}(1-\sigma_{j,1}z_{j,1}, \dots, 1-\sigma_{j,m}z_{j,m}) & \mathbf{0} \\ \mathrm{diag}(1-\sigma_{j,1}z_{j,1}, \dots, 1-\sigma_{j,m}z_{j,m}) & \mathbf{I} & -\rho\,\mathrm{diag}(z_{j,1}, \dots, z_{j,m}) \\ \mathbf{0} & -\rho\,\mathrm{diag}(z_{j,1}, \dots, z_{j,m}) & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0. \tag{89}$$

By permuting the rows and the columns of the matrix in (89), we can transform it into a block diagonal matrix, where the ith block is

$$\begin{pmatrix} \tau-\lambda_j & 1-\sigma_{j,i}z_{j,i} & 0 \\ 1-\sigma_{j,i}z_{j,i} & 1 & -\rho z_{j,i} \\ 0 & -\rho z_{j,i} & \lambda_j \end{pmatrix}, \tag{90}$$

so that (89) is satisfied if and only if each of the matrices (90) is positive semidefinite. Thus, the problem of (86) and (83) become

$$\min_{\tau, z_{j,i}, \lambda_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \sum_{i=1}^{m} z_{j,i}^2 + NL^2\tau \right\} \tag{91}$$

subject to

$$\begin{pmatrix} \tau - \lambda_j & 1 - \sigma_{j,i} z_{j,i} & 0 \\ 1 - \sigma_{j,i} z_{j,i} & 1 & -\rho z_{j,i} \\ 0 & -\rho z_{j,i} & \lambda_j \end{pmatrix} \succeq 0, \quad 1 \leq i \leq m, 0 \leq j \leq N-1. \tag{92}$$

We now show that the problem of (91) subject to (92) can be further simplified. First we note that to satisfy (92) we must have that

$$\tau \geq \max_{0 \leq j \leq N-1} \lambda_j. \tag{93}$$

Suppose first that $\tau > \max_{0 \leq j \leq N-1} \lambda_j$. In this case, using Lemma 4.2, (92) is equivalent to

$$\begin{aligned} &\begin{pmatrix} 1 & -\rho z_{j,i} \\ -\rho z_{j,i} & \lambda_j \end{pmatrix} - \tfrac{1}{\tau - \lambda_j} \begin{pmatrix} 1 - \sigma_{j,i} z_{j,i} \\ 0 \end{pmatrix} \begin{pmatrix} 1 - \sigma_{j,i} z_{j,i} & 0 \end{pmatrix} = \\ &\qquad \begin{pmatrix} 1 - \frac{(1 - \sigma_{j,i} z_{j,i})^2}{\tau - \lambda} & -\rho z_{j,i} \\ -\rho z_{j,i} & \lambda_j \end{pmatrix} \succeq 0. \end{aligned} \tag{94}$$

Now, a 2×2 matrix is positive semidefinite if and only if the diagonal elements and the determinant are nonnegative. Therefore, (94) is equivalent to the conditions

$$\lambda_j \geq 0; \tag{95}$$

$$\tau - \lambda_j \geq (1 - \sigma_{j,i} z_{j,i})^2; \tag{96}$$

$$\lambda_j \left(1 - \tfrac{(1 - \sigma_{j,i} z_{j,i})^2}{\tau - \lambda_j} \right) - \rho^2 z_{j,i}^2 \geq 0. \tag{97}$$

Clearly, (97) and (95) together imply (96). Furthermore, we can express (97) as

$$z_{j,i}^2 \left((\lambda_j - \tau)\rho^2 - \sigma_{j,i}^2 \lambda_j \right) + 2 z_{j,i} \sigma_{j,i} \lambda_j + \lambda_j (\tau - \lambda_j - 1) \geq 0. \tag{98}$$

Since the coefficient multiplying $z_{j,i}^2$ in (98) is negative, it follows that there exists a $z_{j,i}$ satisfying (98) if and only if the discriminant is nonnegative, *i.e.,* if and only if

$$\sigma_{j,i}^2\lambda_j + \left((\tau-\lambda_j)\rho^2 + \sigma_{j,i}^2\lambda_j\right)(\tau-\lambda_j-1) \geq 0, \tag{99}$$

which, using the fact that $\tau - \lambda_j > 0$ for every $0 \leq j \leq N-1$, is equivalent to

$$\lambda_j\sigma_{j,i}^2 \geq \rho^2(1+\lambda_j-\tau). \tag{100}$$

If (100) is satisfied, then the set of $z_{j,i}$'s satisfying (98) are

$$z_{j,i}^- \leq z_{j,i} \leq z_{j,i}^+, \tag{101}$$

where $z_{j,i}^- \leq z_{j,i}^+$ are the roots of the quadratic function in (98). Since we would like to choose $z_{j,i}$ to minimize (91), it follows that the optimal $z_{j,i}$ is

$$z_{j,i} = f_{j,i}(\tau,\lambda_j) = \frac{\sigma_{j,i}\lambda_j - \sqrt{\lambda_j(\tau-\lambda_j)\left(\sigma_{j,i}^2\lambda_j - \rho^2(1+\lambda_j-\tau)\right)}}{(\tau-\lambda_j)\rho^2 + \sigma_{j,i}^2\lambda_j}. \tag{102}$$

Thus, if $\tau > \max_{0\leq j\leq N-1}\lambda_j$, then the optimal value of $z_{j,i}$ is given by (102), where in addition conditions (100) and (95) must be satisfied.

Next, suppose that $\tau = \lambda_j$ for some j. In this case, to ensure that (92) is satisfied we must have that

$$z_{j,i} = \tfrac{1}{\sigma_{j,i}}; \tag{103}$$

$$\lambda_j \geq \tfrac{\rho^2}{\sigma_{j,i}^2}. \tag{104}$$

We can immediately verify that (103) and (104) are special cases of (102) and (100) with $\tau = \lambda_j$. We therefore conclude that the optimal value of $z_{j,i}$ is given by (102), subject to (100) and (95). Substituting the optimal value of $z_{j,i}$ into (91), our problem becomes

$$\min_{\tau,\lambda_0,\ldots,\lambda_{N-1}} \left\{\sigma^2\sum_{j=0}^{N-1}\sum_{i=1}^{m} f_{j,i}^2(\tau,\lambda_j) + NL^2\tau\right\}, \tag{105}$$

subject to

$$\begin{aligned}
\lambda_j \sigma_{j,i}^2 &\geq \rho^2(1+\lambda_j-\tau), \qquad 1 \leq i \leq m, 0 \leq j \leq N-1, \\
\lambda_j &\geq 0, \qquad 0 \leq j \leq N-1, \\
\tau &\geq \lambda_j, \qquad 0 \leq j \leq N-1.
\end{aligned} \tag{106}$$

Since the problem of (91) subject to (92) is convex, and the reduced problem (105) subject to (106) is obtained by minimizing over one of the variables in (91), the reduced problem is also convex, completing the proof of the theorem. □

We have shown in Theorem 5.2 that in the general model, the computation of the minimax MSE estimator with unknown $\mathbf{H}$ is reduced to solving a convex optimization problem with $N+1$ unknowns. Theorem 5.3 below states that in the symmetric model, the computation is reduced to solving a convex optimization problem with only 3 unknowns, regardless of the value of N. The analysis in the symmetric model is very similar to the analysis in the general model, thus, Theorem 5.3 is presented without a proof.

Theorem 5.3 (Minimax MSE Estimator for Symmetric Model with $\mathbf{T}=\mathbf{I}$ and $\mathbf{C}=\sigma^2\mathbf{I}$) *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 $\mathbf{H} = \mathcal{M}(\mathbf{H}_0, \mathbf{H}_1)$ and $\mathbf{H}_0, \mathbf{H}_1$ are known $n \times m$ matrices, Δ is an unknown matrix satisfying $\Delta \in \mathcal{U}_S = \{\Delta = \mathcal{M}(\Delta_0, \Delta_1) : \|\Delta_0\| \leq \rho_0, \|\Delta_1\| \leq \rho_1\}$ and $\mathbf{w}$ is zero-mean random vector with covariance $\mathbf{C} = \sigma^2\mathbf{I}$. Let $\widehat{\mathbf{H}}_0 = \mathbf{H}_0 + (N-1)\mathbf{H}_1$ and $\widehat{\mathbf{H}}_1 = \mathbf{H}_0 - \mathbf{H}_1$ and let $\widehat{\mathbf{H}}_j = \mathbf{U}_j \Sigma_j \mathbf{V}_j^*$ be the singular value decomposition of $\widehat{\mathbf{H}}_j$ for $j = 0, 1$. Σ_j is an $n \times m$ diagonal matrix with diagonal elements $\sigma_{j,i} > 0, \quad 1 \leq i \leq m$. Then there exists a solution to*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\mathbf{x}_0\| \leq L, \ldots, \|\mathbf{x}_{N-1}\| \leq L, \Delta \in \mathcal{U}_S} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)$$

given by $\mathbf{G} = \mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$ where

$$\mathbf{G}_0 = \frac{1}{N}(\mathbf{A}_0 + (N-1)\mathbf{A}_1), \quad \mathbf{G}_1 = \frac{1}{N}(\mathbf{A}_0 - \mathbf{A}_1).$$

$\mathbf{A}_j$ are of the form

$$\mathbf{A}_j = \mathbf{V}_j \mathbf{Z}_j \mathbf{V}_j^* (\widehat{\mathbf{H}}_j^* \widehat{\mathbf{H}}_j)^{-1/2} \widehat{\mathbf{H}}_j^*, \quad j = 0, 1,$$

where $\mathbf{Z}_j$ *is an* $m \times m$ *diagonal matrix with diagonal elements* $z_{j,i} = f_{j,i}(\tau, \lambda_j)$, *where*

$$f_{j,i}(\tau, \lambda_j) = \frac{\sigma_{j,i}\lambda_j - \sqrt{\lambda_j(\tau - \lambda_j)\left(\sigma_{j,i}^2\lambda_j - \hat{\rho}_j^2(1 + \lambda_j - \tau)\right)}}{(\tau - \lambda_j)\hat{\rho}_j^2 + \sigma_{j,i}^2\lambda_j}, \quad 1 \le i \le m, j = 0, 1,$$

$\hat{\rho}_0 = \rho_0 + (N-1)\rho_1$, $\hat{\rho}_1 = \rho_0 + \rho_1$ *and* λ_0, λ_1 *and* τ *are the solution to the convex optimization problem*

$$\min_{\tau,\lambda_0,\lambda_1} \left\{ \sigma^2 \sum_{i=1}^m f_{0,i}^2(\tau, \lambda_j) + (N-1)\sigma^2 \sum_{i=1}^m f_{1,i}^2(\tau, \lambda_j) + NL^2 \right\}$$

subject to

$$\begin{aligned} \lambda_j \sigma_{j,i}^2 &\ge \rho_j^2(1 + \lambda_j - \tau), \qquad 1 \le i \le m, j = 0, 1; \\ \lambda_j &\ge 0; \qquad j = 0, 1 \\ \tau &\ge \lambda_j, \qquad j = 0, 1. \end{aligned}$$

6 Examples

We now consider several examples illustrating the minimax MSE estimates with $\mathbf{T} = \mathbf{I}$ both for the case of known $\mathbf{H}$, and for the case of unknown $\mathbf{H}$.

6.1 Known H

To demonstrate the multichannel minimax MSE estimator of Theorem 5.1, we give an example of the symmetric model. We consider the two channel case ($N = 2$) in which

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

with

$$
\mathbf{H}_0 = \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0.2 & 0.1 & 0 & 0 \\ 0.4 & 0.2 & 0.1 & 0 \\ 1 & 0.4 & 0.2 & 0.1 \\ 0 & 1 & 0.4, & 0.2 \\ 0 & 0 & 1 & 0.4 \end{pmatrix}, \quad \mathbf{H}_1 = \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0.4 & 0.1 & 0 & 0 \\ 1 & 0.4 & 0.1 & 0 \\ 0 & 1 & 0.4 & 0.1 \\ 0 & 0 & 1 & 0.4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{107}
$$

so that $\mathbf{H}_0$ and $\mathbf{H}_1$ represent convolution with LTI filters. 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 three estimation methods:

1. Least Squares (LS).

2. Minimax MSE with L estimated as the norm of the LS estimator.

3. Regularized LS estimator (RLS). This is a regularization of the LS estimator, that takes the norm constraints into account and is given by:

$$
\hat{\mathbf{x}} = \operatorname*{argmin}_{\|\mathbf{x}_0\| \leq L, \|\mathbf{x}_1\| \leq L} \|\mathbf{C}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x})\|^2. \tag{108}
$$

Like in the Minimax MSE estimator, L is approximated as the norm of the LS estimator. Notice that (108) does not define the classical Tikhonov regularization estimator since we have two norm constraints, one for each subvector. This nonlinear estimator does not have an explicit expression. In order to calculate it we have implemented a gradient projection algorithm (see *e.g.,* [33]). Specifically, the iterations are defined by:

Initial step: Take an arbitrary $\mathbf{x}^0$.

General step: For every $k = 0, 1, 2, \ldots$ define:

$$
\mathbf{z}^{k+1} = \mathbf{x}^k - t_k \left(\mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}\mathbf{x}^k - \mathbf{H}^*\mathbf{C}^{-1}y\right), \tag{109}
$$

for $i = 0, 1$,

$$
\mathbf{x}_i^{k+1} = \begin{cases} \mathbf{z}_i^{k+1}, & \text{if} \|\mathbf{z}_i^{k+1}\| \leq L, \\ \frac{\mathbf{z}_i^{k+1}}{\|\mathbf{z}_i^{k+1}\|} L, & \text{else.} \end{cases} \tag{110}
$$

In Fig. 3 we plot the MSE averaged over 400 noise realizations as a function of $10log\left(\frac{\sigma^2}{\|\mathbf{x}_0\|^2}\right)$ using each of the methods above (for the LS we have an analytic expression for the MSE, so the 400 realizations of the noise are relevant only for RLS and the minimax MSE).

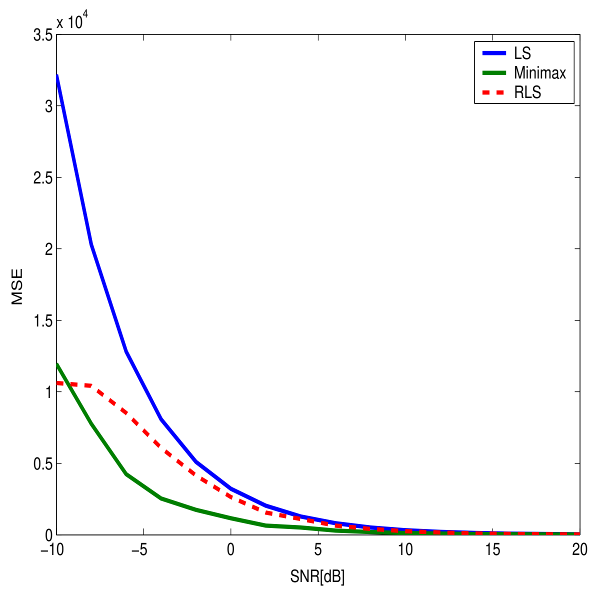
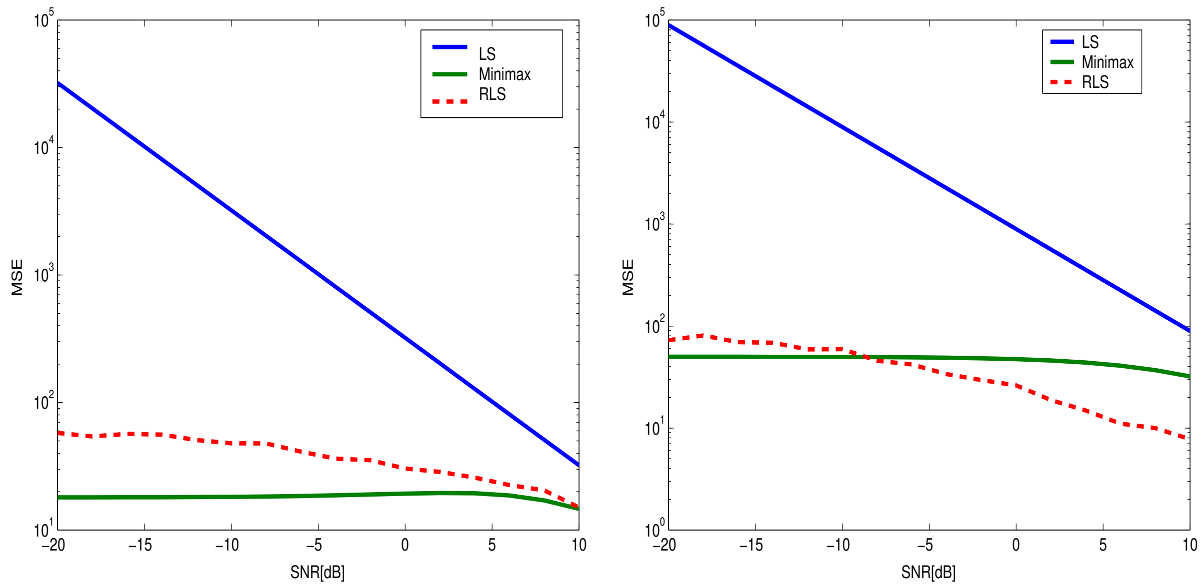


Figure 3: MSE as a function of SNR for the LS, RLS and minimax MSE estimators for unknown L.

It is clear that the minimax MSE estimator is the best of the three estimators for $SNR > -9dB$, the LS is the worst of the three and RLS is somewhere in between these two estimators. This is somewhat a surprising result since the RLS estimator is a highly nonlinear estimator even for a known L, while the minimax MSE is only a linear estimator for a known L. Note, that even for an unknown L the minimax MSE estimator is much easier to compute, since it has an explicit expression, while the RLS estimator is computed via an iterative method, whether or not L is known. An empirical reason for this phenomena can be shown in the following experiment. Suppose that we don't estimate L and we know theoretically that $L = 5$. Now, in Fig. 4 we can see our three estimators with $L = 5$. Fig. 4 (a) is the case of a signal with $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 3$ and Fig. 4 (b) is the case where the norm is tight *i.e.,* $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 5$. It is obvious from the plots that if the norm bound L is not tight (plot (b)) then the RLS estimator is worse than minimax MSE. However, if L is tight (as in the plot (b) then the RLS estimator for $SNR > -5dB$ is better than the minimax MSE. Thus, the RLS estimator is very sensitive to the choice of L and requires tight bounds on the norms (a very difficult requirement). As a result, if we don't have a tight bound for the norms then RLS will be inferior to the minimax MSE.



(a) (b)

Figure 4: MSE as a function of SNR for the LS, RLS and minimax MSE estimators for $L = 5$, (a) $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 3$, (b) $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = 5$

6.2 Unknown H

To demonstrate the multichannel minimax MSE estimator of Theorem 4.3 for uncertain transfer functions, we use the same $\mathbf{H}_0$, $\mathbf{H}_1$ as in (107). We consider the two channel case ($N = 2$) in which

$$\begin{aligned} \mathbf{y}_0 &= (\mathbf{H}_0 + \Delta_0)\mathbf{x}_0 + (\mathbf{H}_1 + \Delta_1)\mathbf{x}_1 + \mathbf{w}_0\,; \\ \mathbf{y}_1 &= (\mathbf{H}_1 + \Delta_1)\mathbf{x}_0 + (\mathbf{H}_0 + \Delta_0)\mathbf{x}_1 + \mathbf{w}_1\,. \end{aligned} \tag{111}$$

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 considered two estimation methods:

1. Structured total LS (STLS). This is a modified TLS method where here 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 of the form $\mathcal{M}(\mathbf{H}_0, \mathbf{H}_1)$. The implementation of the STLS requires only the calculation of the singular value decompositions of two $n \times (m+1)$ matrices.

2. Minimax MSE with L, ρ_0 and ρ_1 estimated from the STLS estimator in the following manner: The STLS outputs a matrix $\widehat{\mathbf{H}} = \mathcal{M}(\mathbf{H}_0 + \Delta_0, \mathbf{H}_1 + \Delta_1)$ a vector $\hat{\mathbf{y}}$ such that $\hat{\mathbf{y}} \in Im(\widehat{\mathbf{H}})$ and an estimator $\hat{\mathbf{x}}_{STLS}$ that is a solution to the system $\widehat{\mathbf{H}}\mathbf{x} = \hat{\mathbf{y}}$. We estimate then L as $\|\hat{\mathbf{x}}_{STLS}\|$, and ρ_0 and ρ_1 are estimated as $\|\Delta_0\|$ and $\|\Delta_1\|$ respectively.

In Fig. 5 we plot the MSE averaged over 400 noise realizations as a function of $10log\left(\frac{\sigma^2}{\|\mathbf{x}_0\|^2}\right)$ using each of the methods above.

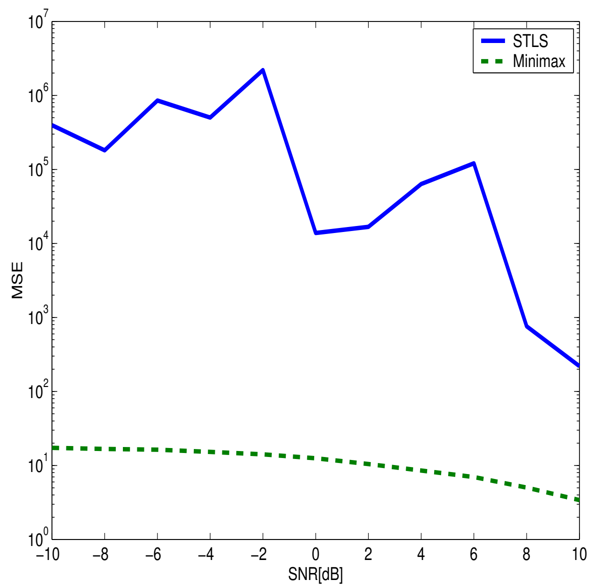


Figure 5: MSE as a function of SNR for the STLS and minimax MSE estimators for unknown L, ρ_0 and ρ_1.

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.

A Proof of Theorem 3.1

Proof of Theorem 3.1: First, let us establish that $\mathbf{U}_j$ exists for every $0 \le j \le N-1$. In order to show this, we prove that the matrix $\widehat{\mathbf{A}}_j = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i$ is Hermitian. This then implies that for every $0 \le j \le N-1$, $\widehat{\mathbf{A}}_j$ has k independent eigenvectors with real eigenvalues. Now,

$$\left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right)^* = \sum_{i=0}^{N-1} \omega^{N-ij} \mathbf{A}_i^* \overset{(20)}{=} \sum_{i=0}^{N-1} \omega^{(N-i)j} \mathbf{A}_{N-i} = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i, \tag{112}$$

so that $\widehat{\mathbf{A}}_j$ is Hermitian and as a result has an eigenvector matrix denoted by $\mathbf{U}_j$. From the definition of an eigenvector matrix we have that $\widehat{\mathbf{A}}_j \mathbf{U}_j = \mathbf{U}_j \mathbf{D}$ where $\mathbf{D} = Diag(\lambda_{j,0}, \lambda_{j,1}, \ldots, \lambda_{j,k-1})$. Now,

$$\mathbf{A} \begin{pmatrix} \mathbf{U}_j \\ \omega^j \mathbf{U}_j \\ \vdots \\ \omega^{(N-1)j} \mathbf{U}_j \end{pmatrix} = \begin{pmatrix} \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right) \mathbf{U}_j \\ \omega^j \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right) \mathbf{U}_j \\ \vdots \\ \omega^{(N-1)j} \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right) \mathbf{U}_j \end{pmatrix} = \begin{pmatrix} \mathbf{U}_j \mathbf{D} \\ \omega^j \mathbf{U}_j \mathbf{D} \\ \vdots \\ \omega^{(N-1)j} \mathbf{U}_j \mathbf{D} \end{pmatrix}, \tag{113}$$

which implies that the columns of $\left(\mathbf{U}_j^T \;\; \omega^j \mathbf{U}_j^T \;\; \cdots \;\; \omega^{(N-1)j} \mathbf{U}_j^T \right)^T$ are k eigenvectors of $\mathbf{A}$ with eigenvalues $\lambda_{j,0}, \lambda_{j,1}, \ldots, \lambda_{j,k-1}$.

The only fact left to prove is that the matrix (32) is invertible. Assume that

$$\begin{pmatrix} \mathbf{U}_0 & \mathbf{U}_1 & \mathbf{U}_2 & \cdot & \mathbf{U}_{N-1} \\ \mathbf{U}_0 & \omega \mathbf{U}_1 & \omega^2 \mathbf{U}_2 & \cdot & \omega^{N-1} \mathbf{U}_{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{U}_0 & \omega^{N-1} \mathbf{U}_1 & \omega^{2(N-1)} \mathbf{U}_2 & \cdots & \omega^{(N-1)(N-1)} \mathbf{U}_{N-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{114}$$

Denoting $\mathbf{x}_j = \mathbf{U}_j \alpha_j$ for $0 \le j \le N-1$, (114) is equivalent to the set of equations

$$\begin{array}{ccccccc} \mathbf{x}_0 & + & \mathbf{x}_1 & + \ldots + & \mathbf{x}_{N-1} & = & 0 \\ \mathbf{x}_0 & + & \omega \mathbf{x}_1 & + \ldots + & \omega^{N-1} \mathbf{x}_{N-1} & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ \mathbf{x}_0 & + & \omega^{N-1} \mathbf{x}_1 & + \ldots + & \omega^{(N-1)(N-1)} \mathbf{x}_{N-1} & = & 0 \end{array} \tag{115}$$

Since the Fourier matrix is invertible we have $\mathbf{x}_j = 0$ for every $0 \le j \le N-1$ and so $\mathbf{U} \alpha_j = \mathbf{x}_j = 0$ which implies that $\alpha_j = 0$ for every $0 \le j \le N-1$, completing the proof. □

B Proof of Lemma 4.1

Proof of Lemma 4.1: By making the change of variables $\mathbf{z}_k = \mathbf{T}^{1/2}\mathbf{x}_k$ we have:

$$\max_{\|\mathbf{x}_0\|_{\mathbf{T}} \le L, \ldots, \|\mathbf{x}_{N-1}\|_{\mathbf{T}} \le L} \mathbf{x}^* \mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})\mathbf{x} = \max_{\|\mathbf{z}_0\| \le L, \ldots, \|\mathbf{z}_{N-1}\| \le L} \mathbf{z}^* \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z} \tag{116}$$

where $\tilde{\mathbf{A}}_i = \mathbf{T}^{-1/2}\mathbf{A}_j\mathbf{T}^{-1/2} \quad 0 \le i \le N-1$. If we relax the constraint set of our maximization problem, then we obtain the following simple relation:

$$\underbrace{\max_{\|\mathbf{z}_0\| \le L, \ldots, \|\mathbf{z}_{N-1}\| \le L} \mathbf{z}^* \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z}}_{(P_1)} \le \underbrace{\max_{\|\mathbf{z}\|^2 \le NL^2} \mathbf{z}^* \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z}}_{(P_2)}. \tag{117}$$

The value of the solution of (P_2) is $L^2 N \lambda_{\max} \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})$ and it is attained at an eigenvector of $\mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})$ with square norm of NL^2. But, by the structure of the eigenvector matrix (32) we see that for every eigenvalue we can find an eigenvector $\mathbf{z}$ of square norm NL^2 that satisfies $\|\mathbf{z}_0\| = \|\mathbf{z}_1\| = \ldots = \|\mathbf{z}_{N-1}\| = L$. From this it follows that $val(P_1) = val(P_2)$ and that the optimal value of (P_1) is equal to:

$$\begin{aligned} NL^2 \lambda_{\max}(\mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})) &= NL^2 \max_{0 \le j \le N-1} \left\{ \lambda_{\max} \left(\sum_{i=0}^{N-1} \omega^{ij} \tilde{\mathbf{A}}_i \right) \right\} \\ = \quad NL^2 \max_{0 \le j \le N-1} &\left\{ \lambda_{\max} \left(\mathbf{T}^{-1/2} \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right) \mathbf{T}^{-1/2} \right) \right\}, \end{aligned} \tag{118}$$

completing the proof. □

C Proof of Lemma 4.3

To prove the proposition, we first note that

$$\mathbf{R} \succeq \mathbf{P}^* \widehat{\mathbf{Z}}_j \mathbf{Q} + \mathbf{Q}^* \widehat{\mathbf{Z}}_j^* \mathbf{P}, \quad \forall \mathbf{Z} \in \mathcal{U}_G \tag{119}$$

if and only if for every $\mathbf{x}$,

$$\begin{aligned}
\mathbf{x}^*\mathbf{R}\mathbf{x} &\geq \max_{\mathbf{Z}\in\mathcal{U}_G} \left\{\mathbf{x}^*\mathbf{P}^*\widehat{\mathbf{Z}}_j\mathbf{Q}\mathbf{x} + \mathbf{x}^*\mathbf{Q}^*\widehat{\mathbf{Z}}_j^*\mathbf{P}\mathbf{x}\right\} \\
&= \max_{\mathbf{Z}\in\mathcal{U}_G} \left\{\mathbf{x}^*\mathbf{P}^*\left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{Z}_i\right)\mathbf{Q}\mathbf{x} + \mathbf{x}^*\mathbf{Q}^*\left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{Z}_i\right)^*\mathbf{P}\mathbf{x}\right\} \\
&= \max_{\mathbf{Z}\in\mathcal{U}_G} \left\{\sum_{i=0}^{N-1}\left(\mathbf{x}^*\mathbf{P}^*\omega^{ij}\mathbf{Z}_i\mathbf{Q}\mathbf{x} + \mathbf{x}^*\mathbf{Q}^*\omega^{-ij}\mathbf{Z}_i^*\mathbf{P}\mathbf{x}\right)\right\} \\
&= \sum_{i=0}^{N-1}\max_{\|\mathbf{Z}_i\|\leq\rho_i}\left\{\mathbf{x}^*\mathbf{P}^*\omega^{ij}\mathbf{Z}_i\mathbf{Q}\mathbf{x} + \mathbf{x}^*\mathbf{Q}^*\omega^{-ij}\mathbf{Z}_i^*\mathbf{P}\mathbf{x}\right\} \\
&= \sum_{i=0}^{N-1} 2\rho_i\|\mathbf{P}\mathbf{x}\|\,\|\mathbf{Q}\mathbf{x}\| \\
&= 2\rho\|\mathbf{P}\mathbf{x}\|\,\|\mathbf{Q}\mathbf{x}\|. \qquad (120)
\end{aligned}$$

Using the Cauchy-Schwarz inequality, we can express (120) as

$$\mathbf{x}^*\mathbf{R}\mathbf{x} - 2\rho\mathbf{y}^*\mathbf{P}\mathbf{x} \geq 0, \quad \forall \mathbf{x}, \mathbf{y} : \|\mathbf{y}\| \leq \|\mathbf{Q}\mathbf{x}\|. \qquad (121)$$

We now rely on the following lemma [34, p. 23]:

Lemma C.1 *[$\mathcal{S}$-procedure] Let $P(\mathbf{z}) = \mathbf{z}^*\mathbf{R}\mathbf{z} + 2\mathbf{u}^*\mathbf{z} + v$ and $\mathcal{Q}(\mathbf{z}) = \mathbf{z}^*\mathbf{B}\mathbf{z} + 2\mathbf{x}^*\mathbf{z} + y$ be two quadratic functions of $\mathbf{z}$, where $\mathbf{R}$ and $\mathbf{B}$ are symmetric and there exists a $\mathbf{z}_0$ satisfying $P(\mathbf{z}_0) > 0$. Then the implication*

$$P(\mathbf{z}) \geq 0 \Rightarrow Q(\mathbf{z}) \geq 0$$

holds true if and only if there exists an $\alpha \geq 0$ such that

$$\begin{pmatrix} \mathbf{B} - \alpha\mathbf{R} & \mathbf{x} - \alpha\mathbf{u} \\ \mathbf{x}^* - \alpha\mathbf{u}^* & y - \alpha v \end{pmatrix} \succeq 0.$$

Since $\|\mathbf{y}\| \leq \|\mathbf{Q}\mathbf{x}\|$ is equivalent to $\mathbf{x}^*\mathbf{Q}^*\mathbf{Q}\mathbf{x} - \mathbf{y}^*\mathbf{y} \geq 0$, we can use Lemma C.1 to conclude that (121) is satisfied if and only if there exists a $\lambda \geq 0$ such that

$$\begin{pmatrix} \mathbf{R} - \lambda\mathbf{Q}^*\mathbf{Q} & -\rho\mathbf{P}^* \\ -\rho\mathbf{P} & \lambda\mathbf{I} \end{pmatrix} \succeq 0, \qquad (122)$$

completing the proof. $\square$

References

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

[2] J.-F. Cardoso. Source separation using higher-order moments. In *Proc. ICASSP*, volume 4, pages 2109–2112, Glasgow, U.K., 1989.

[3] P. Comon. Independent component analysis: A new concept? *Signal Process.*, 36:287–314, Apr. 1994.

[4] L. Tong, Y. Inouye, and R. Liu. Waveform preserving blind estimation of multiple independent sources. *IEEE Trans. Signal Processing*, 41:2461–2470, July 1993.

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

[6] E. Weinstein, M. Feder, and A. V. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. Speech Audio Processing*, 1:405–413, Oct. 1993.

[7] D. Yellin and E. Weinstein. Criteria for multi-channel signal separation. *IEEE Trans. Speech Audio Processing*, 1:405–413, Oct. 1993.

[8] B. Widrow *et. al.* Adaptive noise cancelling: Principles and application. *Proc. IEEE*, pages 1692–1716, 1975.

[9] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

[10] B. R. Hunt and O. Kubler. Karhunen-Loeve multispectral image restoration, part I: Theory. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP–32(3):592–600, June 1984.

[11] N. P. Galatsanos and R. T. Chin. Restoration of color images by multichannel Kalman filtering. *IEEE Trans. on Signal Processing*, 39(10):2237–2252, 1991.

[12] N. P. Galatsanos and R. T. Chin. Digital restoration of multichannel images. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP–37(3):415–421, Mar. 1989.

[13] N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery. Least squares restoration of multichannel images. *IEEE Trans. Acoust., Speech, Signal Processing*, 39(10):2222–2236, 1991.

[14] Moon Gi Kang and Aggelos K. Katsaggelos. Simultaneous multichannel image restoration and estimation of the regularization parameters. *IEEE Trans. on Image Processing*, 6(5), May 1997.

[15] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski. Robust mean-squared error estimation in the presence of model uncertainties. *IEEE Trans. Signal Processing*, to appear.

[16] M. G. Kang and A. K. Katsaggelos. General choice of the regularization functional in regularized image restoration. *IEEE Trans. Image Processing*, 4(5):594–602, May 1995.

[17] M. G. Kang and A. K. Katsaggelos. Simultaneous multichannel image restoration and estimation of the regularization parameters. *IEEE Trans. Image Processing*, 6(5):774–778, May 1997.

[18] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski. Linear minimax regret estimation with bounded data uncertainties. Technical Report CCIT 437, Technion—Israel Institute of Technology, Haifa, Israel, July 2003. *IEEE Trans. Signal Processing*, to appear.

[19] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos. A general framework for frequency domain multi-channel signal processing. *IEEE Trans. Image Processing*, 2(3):417–420, July 1993.

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

[21] R. Gray. Toeplitz and circulant matrices: A review. Technical Report 6504-1, Inform. Sys. Lab., Stanford Univ., Stanford, CA, Apr. 1977.

[22] 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*, 47:1243–1251, Mar. 2001.

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

[24] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):40–95, Mar. 1996.

[25] Y. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PE: SIAM, 1994.

[26] F. Alizadeh. *Combinatorial Optimization With Interior Point Methods and Semi-Definite Matrices*. PhD thesis, University of Minnesota, Minneapolis, MN, Oct. 1991.

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

[28] G. Beylkin. Discrete radon transform. *IEEE Trans. on Acoustic, Speech, and Signal Processing*, ASSP-35(2), 1987.

[29] R. Vescovo. Inversion of block-circulant matrices and circular array approach. *IEEE Trans. on Propagat.*, 45(10), 1997.

[30] A. S. Willsky. *Digital Signal Processing and Control and Estimation Theory*. MIT Press, Cambridge, Massachusetts, and London, England, 1979.

[31] P. J. Davis. *Circulant Matrices*. John Wiley & Sons, Inc, New York, 1979.

[32] D. G. Luenberger. *Optimization by Vector Space Methods*. New York, NY: John Willey & Sons, 1968.

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

[34] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.