SIAM J. MATRIX ANAL. APPL.
Vol. 29, No. 3, pp. 712–730

MEAN-SQUARED ERROR ESTIMATION FOR LINEAR SYSTEMS WITH BLOCK CIRCULANT UNCERTAINTY*

AMIR BECK†, YONINA C. ELDAR‡, AND AHARON BEN-TAL§

Abstract. We consider the problem of estimating a vector $\mathbf{x}$ in the linear model $\mathbf{A}\mathbf{x} \approx \mathbf{y}$, where $\mathbf{A}$ is a block circulant (BC) matrix with N blocks and $\mathbf{x}$ is assumed to have a weighted norm bound. In the case where both $\mathbf{A}$ and $\mathbf{y}$ are subjected to noise, we propose a minimax mean-squared error (MSE) approach in which we seek the linear estimator that minimizes the worst-case MSE over a BC structured uncertainty region. For an arbitrary choice of weighting, we show that the minimax MSE estimator can be formulated as a solution to a semidefinite programming problem (SDP), which can be solved efficiently. For a Euclidean norm bound on $\mathbf{x}$, the SDP is reduced to a simple convex program with $N + 1$ unknowns. Finally, we demonstrate through an image deblurring example the potential of the minimax MSE approach in comparison with other conventional methods.

Key words. minimax estimation, block circulant structure, semidefinite programming, robust optimization

AMS subject classifications. 90C22, 65F30, 90C90

DOI. 10.1137/050643696

1. Introduction.

Many problems in data fitting and estimation give rise to a system of linear equations $\mathbf{A}\mathbf{x} \approx \mathbf{y}$, where both the matrix $\mathbf{A}$ and the right-hand side $\mathbf{y}$ are contaminated by noise. Given the observation $\mathbf{y}$, we seek an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ that is close in some sense to $\mathbf{x}$. This estimation problem arises in a large variety of areas in science and engineering, e.g., communication, economics, signal processing, seismology, and control.

Several approaches for dealing with uncertainties in the model matrix $\mathbf{A}$ and right-hand side vector $\mathbf{y}$ are known in the literature. In the *total least squares* (TLS) strategy [11, 15], one seeks the minimal norm perturbations $\boldsymbol{\Delta}\mathbf{A}, \boldsymbol{\Delta}\mathbf{y}$ of the nominal model matrix $\mathbf{A}$ and observation vector $\mathbf{y}$ such that the linear system $(\mathbf{A} + \boldsymbol{\Delta}\mathbf{A})\mathbf{x} = \mathbf{y} + \boldsymbol{\Delta}\mathbf{y}$ is consistent. An alternative strategy is the *robust least squares* (RLS) method [10, 22, 6]. Here the underlying assumption is that the perturbation matrix $\boldsymbol{\Delta}\mathbf{A}$ and the perturbation vector $\boldsymbol{\Delta}\mathbf{y}$ belong to some bounded uncertainty set $\mathcal{U}$. The solution (or estimator) is chosen to minimize the worst-case data error (or "residual") over the uncertainty region:

$$\hat{\mathbf{x}}_{\mathrm{RLS}} \in \operatorname*{argmin}_{\mathbf{x}} \max_{(\boldsymbol{\Delta}\mathbf{A}, \boldsymbol{\Delta}\mathbf{y}) \in \mathcal{U}} \|(\mathbf{A} + \boldsymbol{\Delta}\mathbf{A})\mathbf{x} - \mathbf{y} - \boldsymbol{\Delta}\mathbf{y}\|^2. \tag{1.1}$$

Both the RLS and TLS solutions optimize a criterion that is based on the *data*

---

*Received by the editors October 28, 2005; accepted for publication (in revised form) by L. Vandenberghe January 2, 2007; published electronically June 5, 2007.
http://www.siam.org/journals/simax/29-3/64369.html

†MINERVA Optimization Center, Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel (becka@ie.technion.ac.il). The research of this author was partially supported by the Israel Science Foundation under grant 729/04.

‡Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (yonina@ee.technion.ac.il). The research of this author was partially supported by the Israel Science Foundation under grant 536/04 and by the EU 6th framework programme via the NEWCOM network of excellence.

§MINERVA Optimization Center, Department of Industrial Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (abental@ie.technion.ac.il). The research of this author was partially supported by the Israel Science Foundation under grant 729/04.

error ($\|\mathbf{Ax}-\mathbf{y}\|$ or $\|(\mathbf{A}+\mathbf{\Delta A})\mathbf{x}-\mathbf{y}-\mathbf{\Delta y}\|$) and therefore might provide poor solutions in terms of the *estimation error* $\|\mathbf{x}-\hat{\mathbf{x}}\|$. In view of this, the work [8] suggests seeking an estimator $\hat{\mathbf{x}}$ that minimizes the *mean-squared error* (MSE):

$$MSE = E(\|\mathbf{x}-\hat{\mathbf{x}}\|^2),$$

and restricting attention to *linear estimators* of the form $\hat{\mathbf{x}} = \mathbf{Gy}$. The expectation is with respect to the noise vector $\mathbf{\Delta y}$, which is assumed to have a zero mean and a positive definite covariance matrix $\mathbf{C}$. For a linear estimator, the MSE is equal to the sum of the variance $V(\hat{\mathbf{x}})$ and the squared norm of the bias $B(\hat{\mathbf{x}})$:

$$E(\|\mathbf{x}-\hat{\mathbf{x}}\|^2) = \underbrace{\mathrm{Tr}(\mathbf{GCG}^*)}_{V(\hat{\mathbf{x}})} + \underbrace{\mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}}_{\|B(\hat{\mathbf{x}})\|^2}.$$

Since the bias depends on the unknown vector $\mathbf{x}$ and the unknown perturbation matrix $\mathbf{\Delta A}$, we cannot choose an estimator to directly minimize the MSE. The approach advocated in [8, 7], in order to minimize the MSE, is to use additional a priori information on the vector $\mathbf{x}$, such as an upper bound on its weighted norm, $\mathbf{x}^*\mathbf{Tx} \le L^2$, where $\mathbf{T}$ is a positive definite matrix, and minimize the worst-case MSE. This leads to the following optimization problem:

$$\min_{\mathbf{G}} \max_{\mathbf{x}^*\mathbf{Tx}\le L^2,\, \Delta\mathbf{A}\in\mathcal{U}} E(\|\mathbf{x}-\hat{\mathbf{x}}\|^2), \tag{1.2}$$

where $\mathcal{U}$ is an uncertainty set associated with the matrix $\mathbf{A}$. The optimal solution $\mathbf{G}$ of the latter problem is called the *minimax MSE matrix*, and the associated linear estimator $\hat{\mathbf{x}} = \mathbf{Gy}$ is termed the *minimax MSE estimator*. In the case when $\mathcal{U}$ is given by a single norm bound, it was shown in [8] that the optimal $\mathbf{G}$ can be obtained by solving a semidefinite programming (SDP) problem. In practice, if L is unknown, then we can estimate it from the data, for example by using the LS estimator [3].

In this paper we study the minimax MSE estimator when the matrix $\mathbf{A}$ has a *block circulant* (BC) structure:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{N-1} \\ \mathbf{A}_{N-1} & \mathbf{A}_0 & \cdots & \mathbf{A}_{N-2} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_0 \end{pmatrix}, \tag{1.3}$$

where $\mathbf{A}_j \in \mathbb{C}^{n\times m}$, $0 \le j \le N-1$. We use the notation $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})$ for brevity. The BC structure of $\mathbf{A}$ imposes the same structure on the perturbation matrix, i.e., $\mathbf{\Delta A} = \mathcal{C}(\mathbf{\Delta A}_0, \ldots, \mathbf{\Delta A}_{N-1})$ with $\mathbf{\Delta A}_j \in \mathbb{C}^{n\times m}$. We also assume that both the covariance matrix $\mathbf{C}$ and weighting matrix $\mathbf{T}$ are positive definite BC (which includes the case $\mathbf{C} = \sigma^2\mathbf{I}$ and $\mathbf{T} = \mathbf{I}$). Thus, the optimization problem we consider is

$$\min_{\mathbf{G}} \max_{\mathbf{x}^*\mathbf{Tx}\le L^2,\, \Delta\mathbf{A}\in\mathcal{U}_\Delta} \left\{\mathrm{Tr}(\mathbf{GCG}^*) + \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}\right\}, \tag{1.4}$$

where the set $\mathcal{U}_\Delta$, which is the set of possible values of $\mathbf{\Delta A}$, is given by

$$\mathcal{U}_\Delta \triangleq \{\mathbf{\Delta} = \mathcal{C}(\mathbf{\Delta}_0, \ldots, \mathbf{\Delta}_{N-1}) : \|\mathbf{\Delta}_k\| \le \rho_k,\ 0 \le k \le N-1\}. \tag{1.5}$$

Here $\|\mathbf{M}\|$ denotes the Frobenius norm of $\mathbf{M}$.

The BC model has previously been used in a variety of signal processing problems, including image restoration [17], cyclic convolution filter banks [20], texture synthesis and recognition [24], and detection techniques for CDMA systems [26]. Moreover, in many practical scenarios $\mathbf{A}$ is a block Toeplitz matrix which can be approximated by a BC matrix [12, 9]. We refer the reader to the example in section 5 that describes a usage of this Toeplitz/circulant approximation in an image deblurring context. The BC structure also includes the multiple observation model in which the matrix $\mathbf{A}$ is a block diagonal matrix with the same diagonal matrix (corresponding to $\mathbf{A}_1 = \mathbf{A}_2 = \cdots = \mathbf{A}_{N-1} = \mathbf{0}$ in (1.3)). The minimax MSE estimator for the multiple observation model was studied in [2].

Besides including several cases of practical interest, one of the attributes of the BC structure is its analytical tractability. In fact the minimax MSE problem (1.4) is intractable for most choices of uncertainty sets $\mathcal{U}_\Delta$. However, in the BC model we are able to exploit properties of BC matrices (in particular, the matrix discrete Fourier transform (DFT)) that will enable us to develop a computationally tractable scheme for computing the minimax MSE estimator.

The BC model has been investigated in the context of structured TLS problems in [1], where it was shown that by using the matrix DFT, the problem can be decomposed into several unstructured TLS problems.

The paper is organized as follows. We begin by reviewing in section 2 some properties of BC matrices and the matrix DFT. In section 3 we first show that under the BC model, the optimal minimax MSE estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is such that $\mathbf{G}$ is a BC matrix. This allows us to formulate the minimax MSE estimator as a solution to an SDP, which is a tractable (i.e., polynomial solvable) convex optimization problem that can be solved, e.g., using interior point methods [21, 25, 4]. In section 4 we treat the case where the weighting matrix $\mathbf{T}$ is the identity matrix $\mathbf{I}$. When the matrix $\mathbf{A}$ is *known*, we derive an explicit formula for the minimax MSE estimator. When $\mathbf{A}$ is *uncertain* but the noise vector consists of independent and identically distributed random variables ($\mathbf{C} = \sigma^2\mathbf{I}$), we show that the task of computing the minimax MSE estimator reduces to solving a simple convex program in $N+1$ variables. Finally, we demonstrate through an image deblurring example, in section 5, the potential of the minimax MSE approach in comparison with other conventional strategies.

Notation. 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 a 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 $\|\mathbf{A}\| = \sqrt{\mathrm{Tr}(\mathbf{A}^*\mathbf{A})}$ the Frobenius norm of the matrix $\mathbf{A}$. For a given matrix $\mathbf{M}$, $\mathbf{m} = \mathrm{vec}(\mathbf{M})$ denotes the vector obtained by stacking the columns of $\mathbf{M}$.

2. BC matrices and the DFT. The aim of this short section is to give a brief review of results on BC matrices and the DFT defined on them that will be used later in the paper. These results can also be found in [1, 2], and they are presented here for completeness.

We begin by noting that the result of multiplication, addition, and conjugation of BC matrices is also a BC matrix. Let $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$; then the DFT of $\mathbf{A}$ is also a BC matrix of the same dimensions given by

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

where $\mathbf{F}_j(\mathbf{A})$ are defined as

$$\mathbf{F}_j(\mathbf{A}) \triangleq \sum_{k=0}^{N-1} \omega^{kj}\mathbf{A}_k, \quad 0 \le j \le N-1,$$

with $\omega = e^{-\frac{2\pi i}{N}}$ (here $i = \sqrt{-1}$). The matrices $\mathbf{F}_j(\mathbf{A})$ are called the *discrete Fourier components*. The inverse DFT, denoted by $\mathbf{F}^{-1}$, is defined by $\mathbf{F}^{-1}(\mathbf{A}) = (\mathbf{F}_0^{-1}(\mathbf{A}), \mathbf{F}_1^{-1}(\mathbf{A}), \dots, \mathbf{F}_{N-1}^{-1}(\mathbf{A}))$, where

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

Note that $\mathbf{F}^{-1}$ is indeed an inverse of $\mathbf{F}$ in the sense that for every BC matrix $\mathbf{A}$

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

The following properties of $\mathbf{F}_j$ are generalizations of well-known properties of the DFT.

LEMMA 2.1. *Suppose that* $\mathbf{A}$, $\mathbf{B}$, *and* $\mathbf{C}$ *are BC matrices. Then for every* $0 \le j \le N-1$ *the following hold:*

1. $(\mathbf{F}_j(\mathbf{A}))^* = \mathbf{F}_j(\mathbf{A}^*)$.
2. $\mathbf{F}_j(\mathbf{I}_{mN}) = \mathbf{I}_m$.
3. $\mathbf{F}_j(\mathbf{A}+\mathbf{C}) = \mathbf{F}_j(\mathbf{A}) + \mathbf{F}_j(\mathbf{C})$.
4. $\mathbf{F}_j(\mathbf{AB}) = \mathbf{F}_j(\mathbf{A})\mathbf{F}_j(\mathbf{B})$.
5. *If* $\mathbf{A}$ *is square and invertible, then* $\mathbf{F}_j(\mathbf{A}^{-1}) = (\mathbf{F}_j(\mathbf{A}))^{-1}$.

Theorem 2.1 shows that the eigenvalues of a Hermitian BC matrix are exactly the eigenvalues of its discrete Fourier components. Theorem 2.1 below is an extension of a well-known result on circulant matrices to the case of Hermitian block circulant matrices; for a proof, see, e.g., [2, Theorem A.1].

THEOREM 2.1. *Let* $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1} \in \mathbb{C}^{k\times k}$ *be matrices such that* $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1})$ *is a Hermitian matrix. For each* $0 \le j \le N-1$, *let* $\lambda_{j,0}, \lambda_{j,1}, \dots, \lambda_{j,k-1}$ *be the eigenvalues of* $\mathbf{F}_j(\mathbf{A})$. *Then the eigenvalues of* $\mathbf{A}$ *are the* $N \cdot k$ *eigenvalues* $\lambda_{j,i}$, $0 \le i \le k-1$, $0 \le j \le N-1$.

3. Minimax MSE estimator for BC systems. We now use the properties of BC matrices and the DFT discussed in the previous section in order to find a $\mathbf{G}$ which is a solution to (1.4). Section 3.1 establishes the fact that $\mathbf{G}$ can always be chosen as a BC matrix. In section 3.2 we use this structure of $\mathbf{G}$ to find an SDP formulation of the estimation problem (1.4), where an SDP is the problem of minimizing a linear objective subject to linear matrix inequality (LMI) constraints, i.e., constraints of the form $\mathcal{B}(\mathbf{x}) \succeq 0$, where the matrix $\mathcal{B}$ depends linearly on $\mathbf{x}$. 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, e.g., interior point methods [21, 25, 23], the optimal estimator can be computed efficiently in polynomial time. Furthermore, SDP-based algorithms are guaranteed to converge to the global optimum.

3.1. The structure of G. Before proceeding, we introduce some notation. The set of all permutations of $\{0, 1, \dots, N-1\}$ 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 (k,j) block of $\mathbf{P}_{\sigma,l}$ is defined as

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

where

$$\delta_{k,j} = \begin{cases} 0, & k \neq j, \\ 1, & k = j \end{cases}$$

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

$$\mathbf{P}_{\sigma,3} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{pmatrix},$$

where $\mathbf{I}_3$ is the identity matrix of size 3×3. We will be interested particularly in a special class of permutations,

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

where $\sigma_k(j) = (j + k) \bmod N$. For example, if $N = 3$, then

$$\mathbf{P}_{\sigma_0,3} = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{pmatrix}, \quad \mathbf{P}_{\sigma_2,3} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \end{pmatrix}.$$

Permutation matrices $\mathbf{P}_{\sigma,l}$ satisfy some interesting properties that will be useful later on in the proof of Theorem 3.1.

Property A. For every $\sigma \in S_N$ and positive integer l, $\mathbf{P}_{\sigma,l}\mathbf{P}^*_{\sigma,l} = \mathbf{P}^*_{\sigma,l}\mathbf{P}_{\sigma,l} = \mathbf{I}$.

Property B. For every BC 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, $\mathbf{P}_{\sigma,m}\mathbf{A} = \mathbf{A}\mathbf{P}_{\sigma,n}$.

The main result of this section is presented in Theorem 3.1, where we show that the solution of (1.4) is a BC matrix, i.e., $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ for some $\mathbf{G}_0, \ldots, \mathbf{G}_{N-1} \in \mathbb{C}^{m \times n}$.

THEOREM 3.1. *Let* $\mathbf{x}$ *denote the unknown vector in the model* $\mathbf{y} = (\mathbf{A} + \mathbf{\Delta A})\mathbf{x} + \mathbf{\Delta y}$, *where* $\mathbf{A}$ *is a BC matrix,* $\mathbf{\Delta A}$ *is an unknown perturbation matrix satisfying* $\mathbf{\Delta A} \in \mathcal{U}_\Delta$ *with* $\mathcal{U}_\Delta$ *given by* (1.5), *and* $\mathbf{\Delta y}$ *is zero-mean random vector with a positive definite BC covariance matrix* $\mathbf{C}$. *Let* $\mathbf{T}$ *be a positive definite BC matrix. Then the problem*

$$\min_{\mathbf{G}} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2,\, \Delta\mathbf{A} \in \mathcal{U}_\Delta} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$$

has a unique solution $\mathbf{G}$, *which is a BC matrix.*

Proof. We first rewrite problem (1.2) as

$$\min_{\mathbf{G} \in \mathbb{C}^{m \times n}} \Gamma(\mathbf{G}), \tag{3.1}$$

where

$$\Gamma(\mathbf{G}) = \max_{\mathbf{x}^*\mathbf{T}\mathbf{x} \leq L^2,\, \Delta\mathbf{A} \in \mathcal{U}_\Delta} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2).$$

The function Γ can be decomposed as follows (see (1.4)):

$$\Gamma(\mathbf{G}) = \theta_1(\mathbf{G}) + \theta_2(\mathbf{G}),$$

where

$$\begin{aligned}\theta_1(\mathbf{G}) &= \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*),\\ \theta_2(\mathbf{G}) &= \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta} \varphi(\mathbf{G},\mathbf{x},\mathbf{\Delta A}),\end{aligned}$$

and $\varphi(\mathbf{G},\mathbf{x},\mathbf{\Delta A}) \overset{\triangle}{=} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}$. It is easy to see that the positive definiteness of $\mathbf{C}$ implies strict convexity of θ_1. Moreover, since φ is a convex function with respect to $\mathbf{G}$, it follows that θ_2, being a maximum of convex functions, is also a convex function. Thus, $\Gamma = \theta_1 + \theta_2$ is a strictly convex function, and hence it has a unique optimal solution.

Using Properties A and B, we have

$$\begin{aligned}\theta_1(\mathbf{G}) = \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) &\overset{A}{=} \mathrm{Tr}(\mathbf{P}^*_{\sigma,m}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{C}\mathbf{G}^*)\\ &= \mathrm{Tr}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{C}\mathbf{G}^*\mathbf{P}^*_{\sigma,m})\\ &\overset{B}{=} \mathrm{Tr}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}\mathbf{C}\mathbf{P}_{\sigma,n}\mathbf{G}^*\mathbf{P}^*_{\sigma,m})\\ &= \mathrm{Tr}((\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})\mathbf{C}(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})^*)\\ &= \theta_1(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})\end{aligned}$$

and

$$\begin{aligned}\theta_2(\mathbf{G}) &= \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}\}\\ &= \max_{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}\mathbf{T}\mathbf{P}_{\sigma,m}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{P}_{\sigma,m}\mathbf{x}\}\\ &\overset{B}{=} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{P}_{\sigma,m}\mathbf{x}\}\\ &\overset{A}{=} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*\mathbf{P}_{\sigma,m}\mathbf{P}^*_{\sigma,m}(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{P}_{\sigma,m}\mathbf{x}\}\\ &\overset{A}{=} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}(\mathbf{A}+\mathbf{\Delta A})\mathbf{P}^*_{\sigma,m})^*(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}(\mathbf{A}+\mathbf{\Delta A})\mathbf{P}^*_{\sigma,m})\mathbf{x}\}\\ &\overset{B}{=} \max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\Delta\mathbf{A}\in\mathcal{U}_\Delta}\{\mathbf{x}^*(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}\}\\ &= \theta_2(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}).\end{aligned}$$

Therefore, $\Gamma(\mathbf{G}) = \Gamma(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n})$. We conclude that if $\mathbf{G}$ is an optimal solution of (3.1), then so is $\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}$ for all $\sigma \in \mathcal{A}$. Hence, by the convexity of Γ it follows that the convex combination $\frac{1}{N}\sum_{\sigma\in\mathcal{A}}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^*_{\sigma,n}$ is also an optimal solution. However, it can be easily verified 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

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

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

3.2. SDP formulation of the estimation problem. We now use Theorem 3.1 to develop an SDP formulation of (1.4). We first consider the inner maximization problem

$$\max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}. \tag{3.2}$$

As a result of Theorem 3.1, we can assume that $\mathbf{G}$ is a BC matrix. Since $\mathbf{I}$, $\mathbf{T}$, and $\mathbf{A}+\mathbf{\Delta A}$ are also BC matrices, it follows that $\mathbf{H}\equiv \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{T}^{-1/2}$ is a BC matrix. By the properties listed in Lemma 2.1, we can deduce that for every $0\le j\le N-1$

$$\begin{aligned}\mathbf{F}_j(\mathbf{H}) &= \mathbf{F}_j(\mathbf{T})^{-1/2}\mathbf{F}_j((\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A})))\mathbf{F}_j(\mathbf{T})^{-1/2}\\ &= \mathbf{S}_j\left(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A}))\right)^*\left(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A}))\right)\mathbf{S}_j,\end{aligned} \tag{3.3}$$

where $\mathbf{S}_j=\mathbf{F}_j(\mathbf{T})^{-1/2}$ and $\mathbf{E}_j=\mathbf{F}_j(\mathbf{G})$. Therefore, by Theorem 2.1, we have

$$\begin{aligned}&\max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le L^2,\,\mathbf{\Delta A}\in\mathcal{U}_\Delta} \mathbf{x}^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))^*(\mathbf{I}-\mathbf{G}(\mathbf{A}+\mathbf{\Delta A}))\mathbf{x}\\ &\qquad = NL^2\max_{\mathbf{\Delta A}\in\mathcal{U}_\Delta}\max_{0\le j\le N-1}\alpha_j(\mathbf{\Delta A}),\end{aligned} \tag{3.4}$$

where $\alpha_j(\mathbf{\Delta A})$ is given by

$$\lambda_{\max}\left(\mathbf{S}_j(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))^*(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))\mathbf{S}_j\right). \tag{3.5}$$

We can therefore express (3.4) as the solution to the problem

$$\min_{\tau} NL^2\tau \tag{3.6}$$

subject to

$$\mathbf{S}_j(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))^*(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))\mathbf{S}_j\preceq\tau\mathbf{I} \tag{3.7}$$

for every $\mathbf{\Delta A}\in\mathcal{U}_\Delta$. Invoking Schur's lemma [4], we can rewrite the constraint (3.7) as

$$\begin{pmatrix}\tau\mathbf{I} & \mathbf{S}_j(\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))^*\\ (\mathbf{I}-\mathbf{E}_j(\mathbf{F}_j(\mathbf{A})+\mathbf{F}_j(\mathbf{\Delta A})))\mathbf{S}_j & \mathbf{I}\end{pmatrix}\succeq 0,$$

which can be further written as

$$\mathbf{R}_j\succeq\mathbf{P}_j^*\mathbf{F}_j(\mathbf{\Delta A})\mathbf{Q}_j+\mathbf{Q}_j^*\mathbf{F}_j(\mathbf{\Delta A})^*\mathbf{P}_j\quad\forall\mathbf{\Delta A}\in\mathcal{U}_\Delta, \tag{3.8}$$

where

$$\begin{aligned}\mathbf{R}_j &= \begin{pmatrix}\tau\mathbf{I} & \mathbf{S}_j(\mathbf{I}-\mathbf{E}_j\mathbf{F}_j(\mathbf{A}))^*\\ (\mathbf{I}-\mathbf{E}_j\mathbf{F}_j(\mathbf{A}))\mathbf{S}_j & \mathbf{I}\end{pmatrix},\\ \mathbf{P}_j &= \begin{pmatrix}\mathbf{0} & \mathbf{E}_j^*\end{pmatrix},\quad \mathbf{Q}_j=\begin{pmatrix}\mathbf{S}_j & \mathbf{0}\end{pmatrix}.\end{aligned}$$

We now exploit the following lemma, the proof of which is very similar to the proof of Lemma 2 in [8] and thus is omitted here.

LEMMA 3.1. *Given matrices* $\mathbf{P},\mathbf{Q},\mathbf{R}$ *with* $\mathbf{R}=\mathbf{R}^*$ *and the set* $\mathcal{U}_\Delta$ *in* (1.5)*, the statement*

$$\mathbf{R}\succeq\mathbf{P}^*\mathbf{F}_j(\mathbf{X})\mathbf{Q}+\mathbf{Q}^*\mathbf{F}_j(\mathbf{X})^*\mathbf{P}\ \textit{for every}\ \mathbf{X}\in\mathcal{U}_\Delta\ \textit{and}\ 0\le j\le N-1$$

holds 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 3.1, it follows that (3.8) is satisfied if and only if there exists $\lambda_j \geq 0$, $0 \leq j \leq N-1$, such that

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{F}_j(\mathbf{T})^{-1} & \mathbf{S}_j(\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A}))^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A}))\mathbf{S}_j & \mathbf{I} & -\rho \mathbf{E}_j \\ \mathbf{0} & -\rho \mathbf{E}_j^* & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0 \tag{3.9}$$

with $\rho = \sum_{j=0}^{N-1} \rho_j$. Summarizing the above derivations, we see that problem (1.4) reduces to

$$\min_{\tau, \lambda_j, \mathbf{G}} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) + NL^2 \tau \right\}$$

subject to (3.9).

Since $\mathbf{C}$ and $\mathbf{G}$ are both BC matrices, the product $\mathbf{G}\mathbf{C}\mathbf{G}^*$ is also a BC matrix. Let $\mathbf{G}\mathbf{C}\mathbf{G}^* = \mathcal{C}(\mathbf{S}_0, \mathbf{S}_2, \ldots, \mathbf{S}_{N-1})$ for some $\mathbf{S}_0, \mathbf{S}_2, \ldots, \mathbf{S}_{N-1} \in \mathbb{C}^{m \times m}$. Then $\mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^*) = N \,\mathrm{Tr}(\mathbf{S}_0)$. However,

$$N\mathbf{S}_0 = N\mathbf{F}_0^{-1}(\mathbf{F}(\mathbf{G}\mathbf{C}\mathbf{G}^*)) = \sum_{j=0}^{N-1} \mathbf{F}_j(\mathbf{G}\mathbf{C}\mathbf{G}^*) = \sum_{j=0}^{N-1} \mathbf{E}_j \mathbf{F}_j(\mathbf{C}) \mathbf{E}_j^*. \tag{3.10}$$

We thus arrive at the following formulation of problem (1.4):

$$\min_{\tau, \lambda_j, \mathbf{E}_j} \left\{ NL^2 \tau + \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{E}_j \mathbf{F}_j(\mathbf{C}) \mathbf{E}_j^*) \right\} \tag{3.11}$$

subject to

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{F}_j(\mathbf{T})^{-1} & \mathbf{S}_j(\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A}))^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A}))\mathbf{S}_j & \mathbf{I} & -\rho \mathbf{E}_j \\ \mathbf{0} & -\rho \mathbf{E}_j^* & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \leq j \leq N-1, \tag{3.12}$$

which is equivalent to

$$\min_{\tau, t_j, \mathbf{E}_j, \lambda_j} \left\{ \sum_{j=0}^{N-1} t_j + NL^2 \tau \right\} \tag{3.13}$$

subject to the LMI (3.12) and

$$\mathrm{Tr}(\mathbf{E}_j \mathbf{F}_j(\mathbf{C}) \mathbf{E}_j^*) \leq t_j, \quad 0 \leq j \leq N-1, \tag{3.14}$$

which can clearly be expressed as an LMI (see (3.15)). Thus, our problem reduces finally to an SDP.

We summarize our results in Theorem 3.2, where we present the SDP formulation for the circulant model.

THEOREM 3.2 (SDP formulation). *Consider the setting of Theorem* 3.1. *Then the unique solution to* (1.4) *is the BC matrix* $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \ldots, \mathbf{G}_{N-1})$*, where*

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

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

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

subject to

$$(3.15) \qquad \begin{pmatrix} t_j & \mathbf{e}_j^* \\ \mathbf{e}_j & \mathbf{I} \end{pmatrix} \succeq 0, \quad 0 \le j \le N-1,$$

$$\begin{pmatrix} \tau \mathbf{I} - \lambda_j \mathbf{F}_j(\mathbf{T})^{-1} & \mathbf{S}_j (\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A}))^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{E}_j \mathbf{F}_j(\mathbf{A})) \mathbf{S}_j & \mathbf{I} & -\rho \mathbf{E}_j \\ \mathbf{0} & -\rho \mathbf{E}_j^* & \lambda_j \mathbf{I} \end{pmatrix} \succeq 0,$$

where $\mathbf{e}_j = \text{vec}(\mathbf{E}_j \mathbf{F}_j(\mathbf{C})^{1/2})$, $\mathbf{S}_j = \mathbf{F}_j(\mathbf{T})^{-1/2}$, *and* $\rho = \sum_{j=0}^{N-1} \rho_j$.

4. Minimax MSE estimator for $\mathbf{T} = \mathbf{I}$. In this section we discuss a special case of the minimax MSE estimator problem where $\mathbf{T} = \mathbf{I}$. When $\mathbf{A}$ is certain, we find an explicit expression for the optimal minimax MSE estimator. In the case of uncertain $\mathbf{A}$, we show that the SDP problem of Theorem 3.2 can be reduced to a simple convex optimization problem in $N + 1$ unknowns.

4.1. Minimax MSE estimator for $\mathbf{T} = \mathbf{I}$ with known $\mathbf{A}$. In the case of known $\mathbf{A}$, we return to the problem of a single system $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{\Delta y}$ with $\mathbf{x}^* \mathbf{T} \mathbf{x} \le L^2$. This problem was discussed in [8], where it was shown that the minimax MSE estimator for the case $\mathbf{T} = \mathbf{I}$ is given by $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ with

$$(4.1) \qquad \mathbf{G} = \alpha (\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{C}^{-1},$$

where $\alpha = \frac{L^2}{L^2 + B}$ and

$$(4.2) \qquad B = \text{Tr}\left((\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A})^{-1} \right).$$

The estimator of (4.1) is a *shrunken estimator* proposed by Mayer and Willke [19], which is simply a scaled version of the LS estimator with an optimal choice of shrinkage factor.

Note that the dominant computation in (4.1) and (4.2) is the inversion of the $mN \times mN$ matrix $\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A}$, which requires $O(m^3 N^3)$ operations. This number is prohibitively large even for medium size problems. On the other hand, the calculation stemming from Theorem 4.1, which exploits the BC structure, requires the inversion of N DFT components, each an $m \times m$ matrix resulting in a total of only $O(m^3 N)$ operations. For example, if $N = 100$, then our computation is 10000 cheaper than the direct computation.

THEOREM 4.1. *Let* $\mathbf{x}$ *denote the vector of unknown parameters in the model* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{\Delta y}$, *where* $\mathbf{A}$ *is a known BC matrix and* $\mathbf{\Delta y}$ *is a zero-mean random vector with a positive definite BC covariance matrix* $\mathbf{C}$. *Then the solution to the problem*

$\min_{\mathbf{G}} \max_{\|\mathbf{x}\|^2 \le L^2} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ *is given by the BC matrix* $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$, *where*

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

Here

$$\mathbf{E}_j = \frac{L^2}{L^2 + B} \left(\mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1} \mathbf{F}_j(\mathbf{A}) \right)^{-1} \mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1}, \quad 0 \le j \le N-1,$$

and $B = \sum_{j=0}^{N-1} \operatorname{Tr}\left((\mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1} \mathbf{F}_j(\mathbf{A}))^{-1} \right)$.

Proof. First, we note that B of (4.2) is equal to $\sum_{i=1}^{mN} \frac{1}{\lambda_i}$, where $\lambda_1, \lambda_2, \ldots, \lambda_{mN}$ are the eigenvalues of $\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A}$. From Theorem 2.1, it follows that

$$B = \sum_{j=0}^{N-1} \operatorname{Tr}\left((\mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1} \mathbf{F}_j(\mathbf{A}))^{-1} \right).$$

By Theorem 3.1, $\mathbf{G}$ is a BC matrix and thus is equal to $\mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$ for some $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1} \in \mathbb{C}^{m \times n}$. Using the properties listed in Lemma 2.1, we can calculate the jth DFT component of $\mathbf{G}$ (denoted by $\mathbf{E}_j$):

$$\begin{aligned} \mathbf{E}_j &= \mathbf{F}_j \left(\frac{L^2}{L^2 + B} (\mathbf{A}^* \mathbf{C}^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{C}^{-1} \right) \\ &= \frac{L^2}{L^2 + B} (\mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1} \mathbf{F}_j(\mathbf{A}))^{-1} \mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{C})^{-1}. \end{aligned}$$

Applying the inverse DFT, we obtain the desired expression for $\mathbf{G}_j$, and the result follows. ∎

As can be expected intuitively, when $L \to \infty$, the minimax MSE estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ of Theorem 4.1 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 solution.

4.2. Minimax estimator for $\mathbf{T} = \mathbf{I}$, $\mathbf{C} = \sigma^2\mathbf{I}$, and unknown model matrix. 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.

THEOREM 4.2. *Consider the setting of Theorem* 3.1 *with* $\mathbf{C} = \sigma^2 \mathbf{I}$. *For every* $0 \le j \le N-1$, *let* $\mathbf{F}_j(\mathbf{A}) = \mathbf{U}_j \Sigma_j \mathbf{V}_j^*$ *be the singular value decomposition of* $\mathbf{F}_j(\mathbf{A})$ *(the* j*th DFT component of* $\mathbf{A}$*), where* Σ_j *is an* $n \times m$ *diagonal matrix with diagonal elements* $\sigma_{j,k} > 0$, $1 \le k \le m$, *and* $\mathbf{U}_j$ *and* $\mathbf{V}_j$ *are unitary matrices. Then the unique solution to* $\min_{\mathbf{G}} \max_{\|\mathbf{x}\|^2 \le L^2, \Delta\mathbf{A} \in \mathcal{U}_\Delta} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ *is given by* $\mathbf{G} = \mathcal{C}(\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N-1})$, *where* $\mathbf{G}_j = \mathbf{F}_j^{-1}(\mathbf{E})$, $\mathbf{E} = \mathcal{C}(\mathbf{E}_0, \ldots, \mathbf{E}_{N-1})$, *with*

$$\mathbf{E}_j = \mathbf{V}_j \mathbf{Z}_j \mathbf{V}_j^* (\mathbf{F}_j(\mathbf{A})^* \mathbf{F}_j(\mathbf{A}))^{-1/2} \mathbf{F}_j(\mathbf{A})^*, \quad 0 \le j \le N-1,$$

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

$$f_{j,k}(\tau, \lambda_j) = \frac{\sigma_{j,k}\lambda_j - \sqrt{\lambda_j(\tau - \lambda_j)\left(\sigma_{j,k}^2 \lambda_j - \rho^2(1 + \lambda_j - \tau)\right)}}{(\tau - \lambda_j)\rho^2 + \sigma_{j,k}^2 \lambda_j},$$

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

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

subject to

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

Proof. From Theorem 3.2, the optimal estimator $\mathbf{G}$ is equal to $\mathcal{C}(\mathbf{G}_0, \dots, \mathbf{G}_{N-1})$, where $\mathbf{G}_j = \frac{1}{N} \sum_{k=0}^{N-1} \omega^{-kj} \mathbf{E}_k$ and $(\mathbf{E}_j)_{j=0}^{N-1}$ is the solution to

$$\min_{\tau, \mathbf{E}_j, \lambda_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{F}_j(\mathbf{E})\mathbf{F}_j(\mathbf{E})^*) + NL^2\tau \right\}, \tag{4.3}$$

subject to

$$\mathbf{M}_j \triangleq \begin{pmatrix} (\tau - \lambda_j)\mathbf{I} & (\mathbf{I} - \mathbf{E}_j\mathbf{F}_j(\mathbf{A}))^* & \mathbf{0} \\ (\mathbf{I} - \mathbf{E}_j\mathbf{F}_j(\mathbf{A})) & \mathbf{I} & -\rho\mathbf{E}_j \\ \mathbf{0} & -\rho\mathbf{E}_j^* & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0. \tag{4.4}$$

The proof of the theorem is comprised of three parts. First, we show that the optimal solution $(\mathbf{E}_j)_{j=0}^{N-1}$ to (4.3) and (4.4) is of the form

$$\mathbf{E}_j = \mathbf{V}_j\mathbf{Z}_j\mathbf{V}_j^* \left(\mathbf{F}_j(\mathbf{A})^*\mathbf{F}_j(\mathbf{A})\right)^{-1/2} \mathbf{F}_j(\mathbf{A})^*, \quad 0 \leq j \leq N-1, \tag{4.5}$$

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

We begin by showing that the optimal $(\mathbf{E}_j)_{j=0}^{N-1}$ has the form (4.5). The constraint (4.4) is equivalent to $\mathbf{Q}_j\mathbf{M}_j\mathbf{Q}_j^* \succeq 0$ for any invertible $\mathbf{Q}_j$. Choosing

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

(4.4) becomes

$$\begin{pmatrix} (\tau - \lambda_j)\mathbf{I} & \mathbf{V}_j^*(\mathbf{I} - \mathbf{E}_j\mathbf{F}_j(\mathbf{A}))^*\mathbf{V}_j & \mathbf{0} \\ \mathbf{V}_j^*(\mathbf{I} - \mathbf{E}_j\mathbf{F}_j(\mathbf{A}))\mathbf{V}_j & \mathbf{I} & -\rho\mathbf{V}_j^*\mathbf{E}_j\mathbf{U}_j \\ \mathbf{0} & -\rho\mathbf{U}_j^*\mathbf{E}_j^*\mathbf{V} & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0. \tag{4.6}$$

Making the change of variables

$$\mathbf{B}_j \triangleq \mathbf{V}_j^*\mathbf{E}_j\mathbf{U}_j, \tag{4.7}$$

so that

$$\mathbf{E}_j = \mathbf{V}_j\mathbf{B}_j\mathbf{U}_j^*, \tag{4.8}$$

the problem of (4.3) and (4.6) can be expressed as

$$\min_{\tau,\lambda_j,\mathbf{B}_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{B}_j^*\mathbf{B}_j) + NL^2\tau \right\} \tag{4.9}$$

subject to

$$\begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{B}_j\Sigma_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{B}_j\Sigma_j) & \mathbf{I} & -\rho\mathbf{B}_j \\ \mathbf{0} & -\rho\mathbf{B}_j^* & \lambda_j \end{pmatrix} \succeq 0. \tag{4.10}$$

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 $\sigma_{j,k}$, $1 \le k \le m$, for every $0 \le j \le N-1$. Then we can express the constraint (4.10) as

$$\mathbf{L}(\mathbf{B}_j) \triangleq \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j)^* & \mathbf{0} & \mathbf{0} \\ (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{Z}_j & -\rho\mathbf{W}_j \\ \mathbf{0} & -\rho\mathbf{Z}_j^* & \lambda_j\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\rho\mathbf{W}_j^* & \mathbf{0} & \lambda_j\mathbf{I} \end{pmatrix} \succeq 0. \tag{4.11}$$

Clearly, if (4.11) is satisfied, then

$$\mathbf{K}(\mathbf{Z}_j) \triangleq \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{Z}_j\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{Z}_j \\ \mathbf{0} & -\rho\mathbf{Z}_j^* & \lambda_j \end{pmatrix} \succeq 0. \tag{4.12}$$

Now let $\mathbf{B}_j = (\mathbf{Z}_j\ \mathbf{W}_j)$ be any matrix satisfying (4.11), and define $\widetilde{\mathbf{B}}_j = (\mathbf{Z}_j\ \mathbf{0})$. Then

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

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

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

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 (4.9) and (4.10) reduces to

$$\min_{\tau,\mathbf{Z}_j,\lambda_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \mathrm{Tr}(\mathbf{Z}_j^*\mathbf{Z}_j) + NL^2\tau \right\}, \tag{4.13}$$

subject to (4.12). Once we find the optimal $(\mathbf{Z}_j)_{j=0}^{N-1}$, the optimal $(\mathbf{E}_j)_{j=0}^{N-1}$ can be found from (4.8) as

$$\mathbf{E}_j = \mathbf{V}_j\mathbf{Z}_j(\mathbf{I}\ \mathbf{0})\mathbf{U}_j^* = \mathbf{V}_j\mathbf{Z}_j\mathbf{V}_j^*(\mathbf{F}_j(\mathbf{A})^*\mathbf{F}_j(\mathbf{A}))^{-1/2}\mathbf{F}_j(\mathbf{A})^*,$$

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 (4.12), then for every

$0 \le j \le N-1$

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

$$= \begin{pmatrix} (\tau-\lambda_j)\mathbf{I} & (\mathbf{I}-\mathbf{J}\mathbf{Z}_j\mathbf{J}\widetilde{\Sigma}_j)^* & \mathbf{0} \\ (\mathbf{I}-\mathbf{J}\mathbf{Z}_j\mathbf{J}\widetilde{\Sigma}_j) & \mathbf{I} & -\rho\mathbf{J}\mathbf{Z}_j\mathbf{J} \\ \mathbf{0} & -\rho\mathbf{J}\mathbf{Z}_j^*\mathbf{J} & \lambda_j \end{pmatrix} \succeq 0, \tag{4.14}$$

where $\mathbf{J}$ is any diagonal matrix with diagonal elements ±1, $\widetilde{\mathbf{J}} = \mathrm{diag}(\mathbf{J},\mathbf{J},\mathbf{J})$, and we have used the fact that diagonal matrices commute and that $\mathbf{J}^*\mathbf{J} = \mathbf{J}^2 = \mathbf{I}$. It follows from (4.14) 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 [18], 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 $0 \le j \le N-1$.

Denote the diagonal elements of $\mathbf{Z}_j$ by $z_{j,k}$, $1 \le k \le m$, and let $\mathrm{diag}(\alpha_1,\ldots,\alpha_m)$ denote the $m\times m$ diagonal matrix with diagonal elements α_j. By permuting the rows and the columns of the matrix $\mathbf{K}(\mathbf{Z}_j)$, it can be seen that the constraint $\mathbf{K}(\mathbf{Z}_j) \succeq \mathbf{0}$ can be written as

$$\begin{pmatrix} \tau-\lambda_j & 1-\sigma_{j,k}z_{j,k} & 0 \\ 1-\sigma_{j,k}z_{j,k} & 1 & -\rho z_{j,k} \\ 0 & -\rho z_{j,k} & \lambda_j \end{pmatrix}, \quad 1 \le k \le m. \tag{4.15}$$

Thus, the problem of (4.13) and (4.12) becomes

$$\min_{\tau, z_{j,k}, \lambda_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \sum_{i=1}^{m} z_{j,k}^2 + NL^2\tau \right\} \tag{4.16}$$

subject to

$$\begin{pmatrix} \tau-\lambda_j & 1-\sigma_{j,k}z_{j,k} & 0 \\ 1-\sigma_{j,k}z_{j,k} & 1 & -\rho z_{j,k} \\ 0 & -\rho z_{j,k} & \lambda_j \end{pmatrix} \succeq 0 \tag{4.17}$$

for every $1 \le k \le m$, $0 \le j \le N-1$. We now show that the problem of (4.16) subject to (4.17) can be further simplified. First, we note that to satisfy (4.17) we must have that

$$\tau \ge \max_{0\le j\le N-1} \lambda_j.$$

Suppose first that $\tau > \max_{0\le j\le N-1}\lambda_j$. In this case, by Schur's lemma, (4.17) is equivalent to

$$\begin{pmatrix} 1 & -\rho z_{j,k} \\ -\rho z_{j,k} & \lambda_j \end{pmatrix} - \frac{1}{\tau-\lambda_j}\begin{pmatrix} 1-\sigma_{j,k}z_{j,k} \\ 0 \end{pmatrix}\begin{pmatrix} 1-\sigma_{j,k}z_{j,k} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1-\frac{(1-\sigma_{j,k}z_{j,k})^2}{\tau-\lambda} & -\rho z_{j,k} \\ -\rho z_{j,k} & \lambda_j \end{pmatrix} \succeq 0. \tag{4.18}$$

Now a 2×2 matrix is positive semidefinite if and only if the diagonal elements and the determinant are nonnegative. Therefore, (4.18) is equivalent to the conditions

$$\lambda_j \geq 0, \tag{4.19}$$

$$\tau - \lambda_j \geq (1 - \sigma_{j,k} z_{j,k})^2, \tag{4.20}$$

$$\lambda_j \left(1 - \frac{(1 - \sigma_{j,k} z_{j,k})^2}{\tau - \lambda_j} \right) - \rho^2 z_{j,k}^2 \geq 0. \tag{4.21}$$

Clearly, (4.21) and (4.19) together imply (4.20). Furthermore, we can express (4.21) as

$$z_{j,k}^2 \left((\lambda_j - \tau)\rho^2 - \sigma_{j,k}^2 \lambda_j \right) + 2 z_{j,k} \sigma_{j,k} \lambda_j + \lambda_j (\tau - \lambda_j - 1) \geq 0. \tag{4.22}$$

Since the coefficient multiplying $z_{j,k}^2$ in (4.22) is negative, it follows that there exists a $z_{j,k}$ satisfying (4.22) if and only if the discriminant is nonnegative, i.e, if and only if

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

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

$$\lambda_j \sigma_{j,k}^2 \geq \rho^2 (1 + \lambda_j - \tau). \tag{4.23}$$

If (4.23) is satisfied, then the set of $z_{j,k}$'s satisfying (4.22) are

$$z_{j,k}^- \leq z_{j,k} \leq z_{j,k}^+,$$

where $z_{j,k}^- \leq z_{j,k}^+$ are the roots of the quadratic function in (4.22). Since we would like to choose $z_{j,k}$ to minimize (4.16), it follows that the optimal $z_{j,k}$ is

$$\begin{aligned} z_{j,k} &= f_{j,k}(\tau, \lambda_j) \\ &= \frac{\sigma_{j,k}\lambda_j - \sqrt{\lambda_j(\tau - \lambda_j)\left(\sigma_{j,k}^2 \lambda_j - \rho^2(1 + \lambda_j - \tau)\right)}}{(\tau - \lambda_j)\rho^2 + \sigma_{j,k}^2 \lambda_j}. \end{aligned} \tag{4.24}$$

Thus, if $\tau > \max_{0 \leq j \leq N-1} \lambda_j$, then the optimal value of $z_{j,k}$ is given by (4.24), where, in addition, conditions (4.23) and (4.19) must be satisfied.

Next, suppose that $\tau = \lambda_j$ for some j. In this case, to ensure that (4.17) is satisfied, we must have that

$$z_{j,i} = \frac{1}{\sigma_{j,k}}, \tag{4.25}$$

$$\lambda_j \geq \frac{\rho^2}{\sigma_{j,k}^2}. \tag{4.26}$$

We can immediately verify that (4.25) and (4.26) are special cases of (4.24) and (4.23) with $\tau = \lambda_j$. We therefore conclude that the optimal value of $z_{j,k}$ is given by (4.24) subject to (4.23) and (4.19). Substituting the optimal value of $z_{j,k}$ into (4.16), our problem becomes

$$\min_{\tau, \lambda_j} \left\{ \sigma^2 \sum_{j=0}^{N-1} \sum_{k=1}^{m} f_{j,k}^2(\tau, \lambda_j) + N L^2 \tau \right\} \tag{4.27}$$

subject to

$$\lambda_j \sigma_{j,k}^2 \geq \rho^2(1+\lambda_j-\tau), \quad 1 \leq k \leq m, \quad 0 \leq j \leq N-1,$$
$$\lambda_j \geq 0, \quad 0 \leq j \leq N-1,$$
$$\tau \geq \lambda_j, \quad 0 \leq j \leq N-1. \tag{4.28}$$

Since the problem of (4.16) subject to (4.17) is convex, and the reduced problem (4.27) subject to (4.28) is obtained by minimizing over some of the variables in (4.16), the reduced problem is also convex, completing the proof of the theorem. □

Remark 4.1. The line of analysis employed in Theorem 4.2 can also be carried out when $\mathbf{T} = (\mathbf{A}^*\mathbf{A})^\alpha$ for some real number α. The resulting optimization problem is very similar to the one derived in Theorem 4.2. In some applications such as the image deblurring examples described in [8], choosing a negative α provides better results than the Euclidean weighting (i.e., $\alpha = 0$).

5. An image deblurring example. To illustrate the effectiveness of the minimax MSE approach, we consider an image deblurring example from the "Regularization Tools" [14].

We consider the square system

$$\mathbf{A}_{\text{true}}\mathbf{x}_{\text{true}} = \mathbf{y}_{\text{true}},$$

where $\mathbf{x}_{\text{true}} \in \mathbb{R}^{1024}$ is obtained by stacking the columns of the 32×32 image and $\mathbf{A}_{\text{true}}$ is a 1024×1024 matrix that represents an atmospheric turbulence blur originating from [13] and implemented in the function blur(n,3,0.5) from the "Regularization Tools" [14] (3 is the half bandwidth and 0.5 is the standard deviation associated with the corresponding point spread function). The image corresponding to $\mathbf{x}_{\text{true}}$ is shown at the top of Figure 1. The matrix $\mathbf{A}_{\text{true}}$ is a block Toeplitz matrix with half bandwidth 3. We note that, in fact, any matrix representing a two-dimensional convolution has a block Toeplitz structure [16].

The observed matrix $\mathbf{A}$ was generated by the function blur(n,3,0.7), and so essentially the uncertainty in the model matrix is due to lack of knowledge of the standard deviation. The observed vector was generated by adding white noise $\mathbf{y} = \mathbf{y}_{\text{true}} + \sigma\mathbf{e}$, where each component of $\mathbf{e} \in \mathbb{R}^{1024}$ was generated from a standard normal distribution.

In our experiment the standard deviation σ was chosen to be 0.1, which results with the noisy image shown in Figure 1 (Observation). We considered several estimation methods:

- *Least Squares.* The LS estimator is given by $\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{y}$. As can be seen in Figure 1, the resulting image is of a poor quality.
- *Structured TLS.* The structured TLS (STLS) solution $\hat{\mathbf{x}}_{\text{STLS}}$ to the problem is the $\mathbf{x}$-part of the solution to the optimization problem

  $$\min_{\Delta\mathbf{A},\Delta\mathbf{y},\mathbf{x}} \{\|\mathbf{\Delta A}\|^2 + \|\mathbf{\Delta y}\|^2 : (\mathbf{A}+\mathbf{\Delta A})\mathbf{x} = \mathbf{y}+\mathbf{\Delta y},\ \mathbf{\Delta A} \text{ is BC}\}.$$

  The STLS problem with BC structure can be solved by decomposing the problem into several unstructured TLS problems (for details see [1]). As can be seen from Figure 1, the STLS method generates an even worse image than $\hat{\mathbf{x}}_{\text{LS}}$. This poor performance of the STLS solution stems from the fact that the unstructured TLS solution is a deregularization [15] of the LS solution and as such is rather unstable. The STLS solution for BC systems is constructed

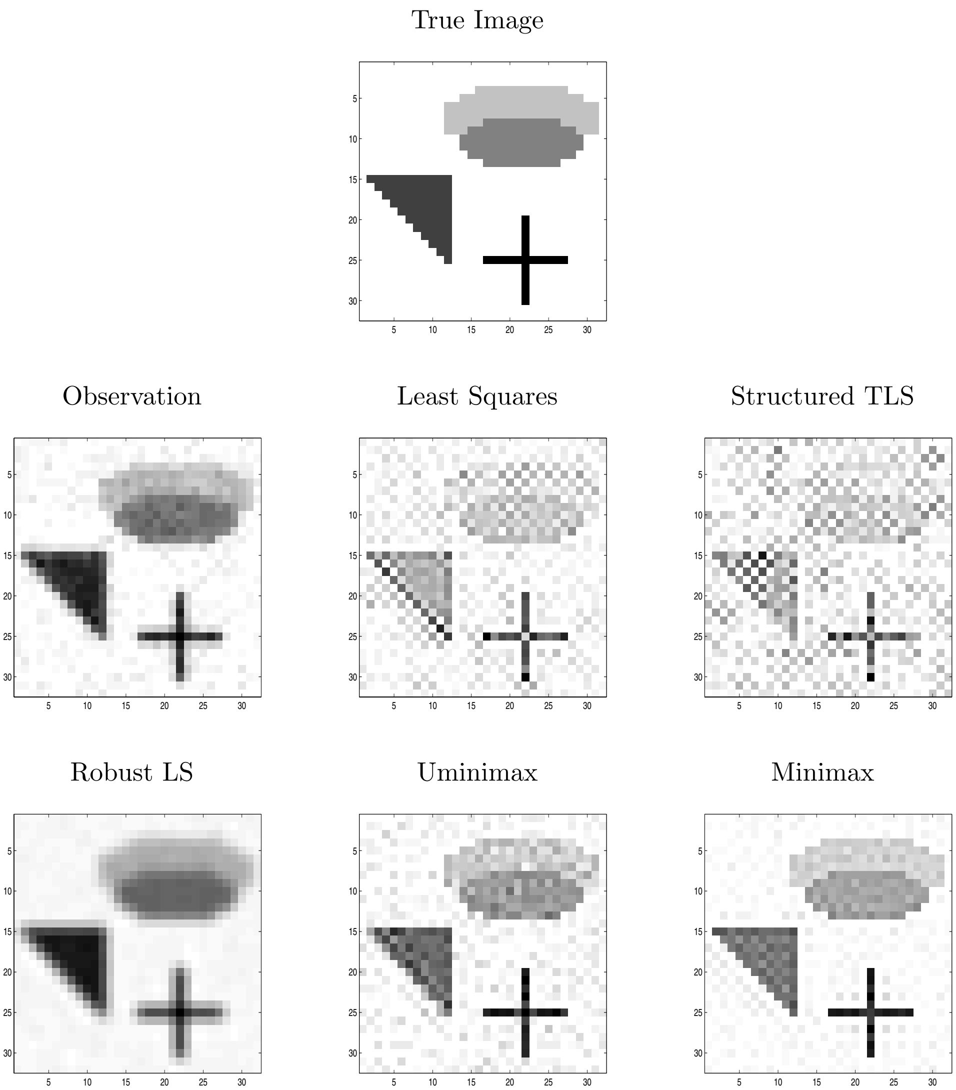

FIG. 1. *Comparison between different estimators.*

from several solutions of unstructured TLS problems and is therefore unstable as well.

- *Robust LS.* We also considered the RLS method defined in (1.1), where the uncertainty set $\mathcal{U}$ is given by a simple norm constraint $\mathcal{U} = \{(\mathbf{\Delta A}, \mathbf{\Delta y}) : \|(\mathbf{\Delta A}, \mathbf{\Delta y})\| \le \rho_{\mathrm{R}}\}$ and ρ_{R} is chosen as $1.1 \cdot \|(\mathbf{A} - \mathbf{A}_{\mathrm{true}}, \mathbf{y} - \mathbf{y}_{\mathrm{true}})\|$. The resulting figure is quite blurred. The reason for not using a complicated set such as $\mathcal{U}_\Delta$ (given in (1.5)) to describe the uncertainty in $\mathbf{A}$ is that problem (1.1) appears to be intractable in this case, since the uncertainty set involves *several* norm constraints. Another alternative would be to use the structured RLS problem [10] and to relax the multiple norm constraints in $\mathcal{U}_\Delta$ into a single norm constraint. However, the generated SDP needed to be solved in our example here is too large to handle with standard software.

- *Uminimax*. Unstructured minimax is the minimax estimator for the unstructured case (see [8]). This estimator minimizes the worst-case MSE across all values of $\mathbf{x}$ satisfying $\mathbf{x}^*\mathbf{T}\mathbf{x} \le L^2$ and perturbation matrices $\mathbf{\Delta A}$ satisfying $\|\mathbf{\Delta A}\| \le \rho_B$. Note, however, that it ignores the special structure of $\mathbf{\Delta A}$. We have chosen the parameters L, ρ_B to be 10 percent larger than their true values (for example, L was chosen to be $1.1 \cdot \mathbf{x}_{\text{true}}^*\mathbf{T}\mathbf{x}_{\text{true}}$). $\mathbf{T}$ was chosen to be $(\mathbf{A}^*\mathbf{A})^{-1}$. This choice of $\mathbf{T}$ reflects the fact that components corresponding to small singular values of $\mathbf{A}^*\mathbf{A}$ should receive a smaller weight than components corresponding to large singular values. The resulting image for this method is of good quality.
- *Minimax*. Finally, we compared the above-mentioned methods with the minimax MSE estimator for BC systems developed in this paper. In implementing the Minimax estimator, we have used a BC approximation of the block Toeplitz matrix $\mathbf{A}$ as follows:

$$\begin{pmatrix}
\mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0
\end{pmatrix}$$

$$\Downarrow$$

$$\begin{pmatrix}
\mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{A}_{-2} & \mathbf{A}_{-1} \\
\mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{A}_{-2} \\
\mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 \\
\mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{-1} & \mathbf{A}_0 & \mathbf{A}_1 \\
\mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_{-2} & \mathbf{A}_{-1} & \mathbf{A}_0
\end{pmatrix}.$$

  The approximation is made by adding three block matrices to the northeast and southwest corners of $\mathbf{A}$. As in the Uminimax estimator, all parameters are chosen to be 10 percent larger than their true value. It can be seen that Minimax gives even a better result than Uminimax.

We note that the Minimax estimate was not calculated by solving the SDP formulation of Theorem 3.2, since its size was too big for standard software such as SeDuMi [23]. Instead we applied a gradient projection algorithm with armijo-type line search [5] on the convex optimization formulation of Theorem 4.2. In this algorithm the dominant computational effort is the calculation of the orthogonal projection onto the polyhedral feasible set, which amounts to solving a quadratic minimization problem in 1025 variables. Since the linear system describing the feasible set is extremely sparse, the CPU time required to calculate a single projection (using SeDuMi) was a small fraction of a second. The resulting image was obtained after 10 iterations in an overall CPU time of 0.8 seconds (on a Pentium 4, 1.8 Ghz). The stopping criterion was chosen to be $|f_k - f_{k-1}| < \varepsilon$, where $\varepsilon = 10^{-3}$ and f_j denotes the objective function value at the jth iteration. We noticed that the quality of the image does not improve if we choose a smaller value of ε.

As can be seen from this example, the structured minimax MSE estimator gives better results than the LS, STLS, RLS, and Uminimax estimators.

REFERENCES

[1] A. Beck and A. Ben-Tal, *A global solution for the structured total least squares problem with block circulant matrices*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 238–255.

[2] A. Beck, A. Ben-Tal, and Y. C. Eldar, *Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties*, Math. Program., 107 (2006), pp. 155–187.

[3] Z. Ben-Haim and Y. C. Eldar, *Blind Minimax Estimation*, CCIT report 550, Electrical Engineering Department, Technion—Israel Institute of Technology, Haifa, Israel, 2005, IEEE Trans. Inform. Theory, submitted.

[4] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MPS/SIAM Ser. Optim. 2, SIAM, Philadelphia, 2001.

[5] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.

[6] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Parameter estimation in the presence of bounded data uncertainties*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 235–252.

[7] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, *Linear minimax regret estimation of deterministic parameters with bounded data uncertainties*, IEEE Trans. Signal Process., 52 (2004), pp. 2177–2188.

[8] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, *Robust mean-squared error estimation in the presence of model uncertainties*, IEEE Trans. Signal Process., 53 (2005), pp. 168–181.

[9] H. Gazzah, P. A. Regalia, and J. Delmas, *Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification*, IEEE Trans. Inform. Theory, 47 (2001), pp. 1243–1251.

[10] L. El Ghaoui and H. Lebret, *Robust solutions to least-squares problems with uncertain data*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1035–1064.

[11] G. H. Golub and C. F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893.

[12] R. Gray, *Toeplitz and Circulant Matrices: A Review*, Tech. report 6504-1, Information System Laboratory, Stanford University, Stanford, CA, 1977.

[13] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3 (1993), pp. 253–315.

[14] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[15] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers Appl. Math. 9, SIAM, Philadelphia, 1991.

[16] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice–Hall, Englewood Cliffs, NJ, 1989.

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

[18] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, New York, 1968.

[19] L. S. Mayer and T. A. Willke, *On biased estimation in linear models*, Technometrics, 15 (1973), pp. 497–508.

[20] H. Murakami, *Discrete wavelet transform based on cyclic convolution*, IEEE Trans. Signal Process., 52 (2004), pp. 165–174.

[21] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[22] A. H. Sayed and S. Chandrasekaran, *Parameter estimation with multiple sources and levels of uncertainties*, IEEE Trans. Signal Process., 48 (2000), pp. 680–692.

[23] J. F. Sturm, *Using SeDuMi* 1.02*, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11/12 (1999), pp. 625–653.

[24] M. Sznaier, O. Camps, and M. C. Mazzaro, *Finite horizon model reduction of a class of neutrally stable systems with applications to texture synthesis and recognition*, in Proceedings of the 43rd IEEE Conference on Decision and Control, 2004.
[25] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.
[26] M. Vollmer, M. Haardt, and J. Gotze, *Comparative study of joint-detection techniques for TD-CDMA based mobile radio systems*, IEEE Select. Areas Comm., 19 (2001), pp. 1461–1475.