Math. Program., Ser. B 107, 155–187 (2006)

Digital Object Identifier (DOI) 10.1007/s10107-005-0683-3

Amir Beck · Aharon Ben-Tal⋆ · Yonina C. Eldar

Robust Mean-Squared Error Estimation of Multiple Signals in Linear Systems Affected by Model and Noise Uncertainties

Received: August 17, 2004 / Accepted: February 14, 2005
Published online: December 30, 2005 – © Springer-Verlag 2005

Abstract. This paper is a continuation of the work in [11] and [2] on the problem of estimating by a linear estimator, N unobservable input vectors, undergoing the same linear transformation, from noise-corrupted observable output vectors. Whereas in the aforementioned papers, only the matrix representing the linear transformation was assumed uncertain, here we are concerned with the case in which the second order statistics of the noise vectors (*i.e.,* their covariance matrices) are also subjected to uncertainty. We seek a robust mean-squared error estimator immuned against both sources of uncertainty. We show that the optimal robust mean-squared error estimator has a special form represented by an elementary block circulant matrix, and moreover when the uncertainty sets are ellipsoidal-like, the problem of finding the optimal estimator matrix can be reduced to solving an explicit semidefinite programming problem, whose size is independent of N.

Key words. Minimax Mean-Squared Error – Multiple Observations – Robust Estimation – Semidefinite Programming – Block Circulant Matrices – Discrete Fourier Transform

1. Introduction and Summary of Main Results

1.1. Minimax MSE Estimator: Single Signal with Certain Model

A central problem in estimation is to recover a set of unobservable parameters from data corrupted by noise. In many applications, the relation between the parameters vector and the data is given by a linear model

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

where $\mathbf{y} \in \mathbb{R}^n$ is the observed data, $\mathbf{x} \in \mathbb{R}^m$ is the unknown parameter vector, $\mathbf{H}$ is the $n \times m$ model matrix (we always assume that $n \geq m$), and $\mathbf{w} \in \mathbb{R}^n$ is a zero-mean noise vector with positive definite covariance matrix $\mathbf{C} = E(\mathbf{w}\mathbf{w}^T)$. Given the data $\mathbf{y}$, we seek an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ that is close to $\mathbf{x}$ in some sense. This estimation problem arises in a large variety of areas in science and engineering *e.g.,* communication, economics, signal processing, seismology, and control.

A. Beck, A. Ben-Tal: MINERVA Optimization Center, Department of Industrial Engineering, Technion–Istrael Institute of Technology, Haifa 32000, Israel. e-mail: {abental@ie,becka@tx}.technion.ac.il

Y.C. Eldar: Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel. e-mail: yonina@ee.technion.ac.il

Mathematics Subject Classification (2000): 90C22, 90C25

⋆ The research was partially supported by BSF grant #2002038

Owing to lack of statistical information about $\mathbf{x}$, often the parameters are chosen to optimize a criterion based on the observed data $\mathbf{y}$. The celebrated Least Squares (LS) estimator [19, 22, 23, 28], which was first used by Gauss to predict movements of planets [16], seeks the linear estimate $\hat{\mathbf{x}}_{LS}$ of $\mathbf{x}$, which minimizes the norm[1] of the data error *i.e.,*

$$\hat{\mathbf{x}}_{LS} = \operatorname{argmin} \|\mathbf{C}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x})\|^2.$$

In the case where $\mathbf{H}$ has full column rank, $\hat{\mathbf{x}}_{LS}$ is given by

$$\hat{\mathbf{x}}_{LS} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{y}.$$

Notice that the LS estimator has the form

$$\hat{\mathbf{x}}_{LS} = \mathbf{G}\mathbf{y}, \tag{1}$$

where $\mathbf{G}$ is an $m \times n$ matrix. An estimator of the form (1) is called a *linear estimator* and the matrix $\mathbf{G}$ is called the *estimator matrix*. Linear estimators are popular due to two main reasons. First, they are very easy to implement. Second, restricting ourselves to linear estimators often leads to tractable optimization problems. In this paper we shall likewise consider linear estimators.

The LS estimator minimizes the weighted norm of the *data error* $\|\mathbf{C}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x})\|$ and may not provide a good solution in terms of the *estimation error* $\|\mathbf{x} - \hat{\mathbf{x}}\|$. In view of this, Eldar *et al.* [11] suggested to seek an estimator $\hat{\mathbf{x}}$ that minimizes the *mean-squared error* (MSE):

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

and to restrict attention to *linear estimators* of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$. 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 $B(\hat{\mathbf{x}})$:

$$MSE = \underbrace{\mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^T)}_{V(\hat{\mathbf{x}})} + \underbrace{\mathbf{x}^T(\mathbf{I} - \mathbf{G}\mathbf{H})^T(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}}_{\|B(\hat{\mathbf{x}})\|^2}.$$

Since the bias depends on the unknown parameters $\mathbf{x}$, we cannot choose an estimator to directly minimize the MSE. A common approach is to restrict the estimator to be unbiased, *i.e.,* the estimator matrix $\mathbf{G}$ satisfies $\mathbf{G}\mathbf{H} = \mathbf{I}$, and then seek the estimator of this form that minimizes the MSE, which is equal in this case to the variance $\mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^T)$. It is well known that the LS estimator achieves this goal, *i.e.,* $\hat{\mathbf{x}}_{LS} = \mathbf{G}_{LS}y$, where

$$\mathbf{G}_{LS} \stackrel{\triangle}{=} (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1} = \underset{\mathbf{G}:\mathbf{G}\mathbf{H}=\mathbf{I}}{\operatorname{argmin}} \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^T)$$

However, this does not imply that the LS estimator leads to a small variance or a small MSE. In fact, it is well known that for ill-conditioned problems the resulting variance/MSE can be very large [19]. As a result, various modifications of the LS estimator have been proposed.

[1] Throughout the paper we use the following notation: $\|\mathbf{v}\|$ is the Euclidean norm of the vector $\mathbf{v}$ and the norm of matrix $\mathbf{A}$ is the Frobenius norm *i.e.,* $\|\mathbf{A}\| = \sqrt{\mathrm{Tr}(\mathbf{A}^T\mathbf{A})}$.

Among the alternatives are Tikhonov regularization [29] (also known in the statistical literature as the ridge estimator [20]), the shrunken estimator [24], and the covariance shaping LS estimator [15]. In general, these alternatives attempt to reduce the MSE in estimating $\mathbf{x}$ by allowing for a bias, however, they too optimize an objective which does not depend directly on the MSE, but rather depends on the data error $\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}$.

The approach advocated in [11, 12], in order to minimize the MSE, is to use additional apriori information on the signal $\mathbf{x}$, such as an upper bound on its size (norm):

$$\|\mathbf{x}\|_{\mathbf{T}} \leq L,$$

where $\|\mathbf{x}\|_{\mathbf{T}} = (\mathbf{x}^T\mathbf{T}\mathbf{x})^{1/2}$ and $\mathbf{T}$ is a positive definite matrix. This leads to the following optimization problem

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L} E(\|\mathbf{x} - \hat{\mathbf{x}}\|^2). \tag{2}$$

It is shown in [3] that if L is known, then the solution to (2) has a smaller MSE than the LS estimator for *any* $\|\mathbf{x}\|_{\mathbf{T}} \leq L$. Thus, in an MSE sense, this estimator is always preferable to the LS estimator. If no a priori knowledge on the norm bound L is available, then it can be reasonably estimated by say $L = \|\hat{\mathbf{x}}_{LS}\|$. For the case in which $\mathbf{H} = \mathbf{C}_w = \mathbf{I}$, it is proven in [4] that the resulting estimator has a smaller MSE than the LS estimator for *any* $\mathbf{x}$. Simulation results presented in Section 4.1 demonstrate that this performance advantage holds in more general cases as well.

1.2. Minimax MSE Estimator: Single Signal with Uncertain Model

In many applications the model matrix $\mathbf{H}$ is subject to uncertainties. For example, $\mathbf{H}$ may be estimated from noisy data. If the actual data matrix $\mathbf{H}$ deviates from the one assumed, say $\mathbf{H}_0$, then the performance of an estimator designed based on $\mathbf{H}_0$ alone may deteriorate considerably. Indeed, in Section 4.3 an example is given with a small deviation between $\mathbf{H}_0$ and $\mathbf{H}$ which causes the LS estimator to be highly unstable. Various methods have been proposed to account for uncertainties in $\mathbf{H}$. The Total LS method (TLS) [18, 21] finds a pair $(\widehat{\mathbf{H}}, \hat{\mathbf{y}})$, which minimizes the error $\|\widehat{\mathbf{H}} - \mathbf{H}_0\|^2 + \|\hat{\mathbf{y}} - \mathbf{y}\|^2$ subject to the consistency equation $\widehat{\mathbf{y}} \in \mathcal{R}(\widehat{\mathbf{H}})$. The TLS estimator $\hat{\mathbf{x}}_{TLS}$ is then a solution to the system $\widehat{\mathbf{H}}\mathbf{x} = \hat{\mathbf{y}}$. Although the TLS method allows for uncertainties in $\mathbf{H}$, in many cases it results in correction terms that are unnecessarily large. Recently, several methods [8, 17, 26] have been developed to treat the case in which the perturbation to the model matrix $\mathbf{H}$ is bounded. These methods seek the parameters that minimize the *worst-case data error* across all bounded perturbations of $\mathbf{H}$, and possibly bounded perturbations of the data vector. In [9] the authors seek the estimator that minimizes the *best possible data error* over all possible perturbations of $\mathbf{H}$. Here again, the above objectives depend on the data error and not on the estimation error, or the MSE. To address this issue it is suggested in [11] to use the *robust optimization* approach of Ben-Tal and Nemirovsky [5, 6] in combination with the minimax approach (2). This leads to an optimization problem of the form

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}} E(\|\mathbf{x} - \hat{\mathbf{x}}\|^2) \tag{3}$$

where

$$\mathcal{U}_{\mathbf{H}} = \{\mathbf{H}_0 + \Delta_{\mathbf{H}} : \Delta_{\mathbf{H}} \in \mathbb{R}^{n\times m} \|\Delta_{\mathbf{H}}\|, \leq \rho_{\mathbf{H}}\}. \tag{4}$$

An additional major source of uncertainty in signal processing (perhaps more common) is associated with the statistics of the noise vector **w**; Indeed, assuming that its covariance matrix **C** is known exactly is unrealistic. Thus in this paper we seek the minimax MSE estimator that is robust with respect to uncertainty not only in the model matrix **H** but also in the covariance matrix **C**. It is assumed that the only knowledge of **C** is that it resides in an uncertainty set $\mathcal{U}_{\mathbf{C}}$ of the form

$$\mathcal{U}_{\mathbf{C}} = \{\mathbf{C}_0 + \Delta_{\mathbf{C}} : \Delta_{\mathbf{C}} \in \mathbb{R}^{n\times n}, \|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}\}. \tag{5}$$

This leads to the following optimization problem

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L, \mathbf{H}\in\mathcal{U}_{\mathbf{H}}, \mathbf{C}\in\mathcal{U}_{\mathbf{C}}} E(\|\mathbf{x} - \hat{\mathbf{x}}\|^2) \tag{6}$$

We refer to the resulting estimator as the *robust MSE estimator* (RMSE).

In [10] the minimax MSE problem of (6) was considered for the case in which **H** is certain and the uncertainty set on **C** is defined only on the eigenvalues of **C**, but not on the eigenvectors.

A related estimation problem with uncertainties in both the model and the covariance matrices is discussed in [7]. Contrary to our setting, the model in [7] assumes that the input vector **x** is random with a priori gaussian distribution. The aim is to find an estimate of $\bar{x}$ that maximizes a lower bound on the worst case (with respect to the uncertainties) a posteriori probability of the observations. For that purpose a maximum likelihood objective function is used. In [14, 13] RMSE estimators where derived for the case in which **x** is a random vector with uncertain covariance matrix, and both **H** and **C** are also subjected to uncertainty.

In Section 3 we show that the optimal solution of (6) can be found by a semidefinite program (SDP). This result is a generalization of the result derived in [11], where only uncertainty in **H** was assumed. The importance of treating uncertainty in **C** is demonstrated in Section 4.4, where we show an example in which the nominal and actual values of **C** are only slightly different but the LS estimator calculated with the nominal value is very unstable: its MSE is 23.39 while the maximum value of the squared error (over 1000 realizations of the noise vectors) was 258.5, more than 11 times bigger!. In contrast, the robust MSE estimator developed in this paper had MSE equal to 24.34 and its maximum value (over the same 1000 realizations of the noise vector) was only 35.75.

The robust MSE estimator depends on two parameters L and $\rho_{\mathbf{H}}$, which may or may not be known a priori depending on the specific application. When L and $\rho_{\mathbf{H}}$ are unknown, we propose to estimate L by the norm of the TLS estimator, and $\rho_{\mathbf{H}}$ by $\|\mathbf{H}_0 - \widehat{\mathbf{H}}\|$. In Section 4.2, we show an example demonstrating that with these choices the RMSE significantly outperforms the TLS estimator.

1.3. Minimax MSE Estimator: Multiple Signals with Uncertain Data

In the main part of the paper, starting from Section 5, we develop robust MSE estimators for a *multiple signals system* with uncertain model matrix $\mathbf{H}$ and uncertain covariance of the noise vectors. Thus we consider the problem of estimating N unknown deterministic parameter vectors $\mathbf{x}_k \in \mathbb{R}^m, 0 \le k \le N-1$ from N observation vectors $\mathbf{y}_k \in \mathbb{R}^n, 0 \le k \le N-1$. Each observation vector $\mathbf{y}_k$ is related to the corresponding parameter vector $\mathbf{x}_k$ through the linear model

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad 0 \le k \le N-1, \tag{7}$$

where $\mathbf{H}$ is an $n \times m$ matrix, and each $\mathbf{w}_k$, $(0 \le k \le N-1)$ is a zero-mean random noise vector with covariance matrix $E(\mathbf{w}_k\mathbf{w}_k^T)$. Each input vector $\mathbf{x}_k$, $(0 \le k \le N-1)$, satisfies a weighted norm constraint $\|\mathbf{x}_k\|_{\mathbf{T}} \le L$. The model is illustrated in Fig 1. Multiple signals models, of which (7) is a special case, where introduced and studied in Beck *et al.* [2], in the case where the covariance matrices are fixed and the uncertainty set associated with $\mathbf{H}$ is an ellipsoidal-like set. Here again, we consider uncertainty in the noise covariance matrices and we seek estimators which are robust against both sources of uncertainty.

In many applications, the noise vectors are correlated. For example, in neurophysiology, in which a multichannel recording array of electrodes sites is used to record spikes from the surrounding neural cell population, the noise vectors are correlated as result of noise emanating from group of cells close to the recording array but not enough for their firing patterns to be detected and sorted by the recording array.

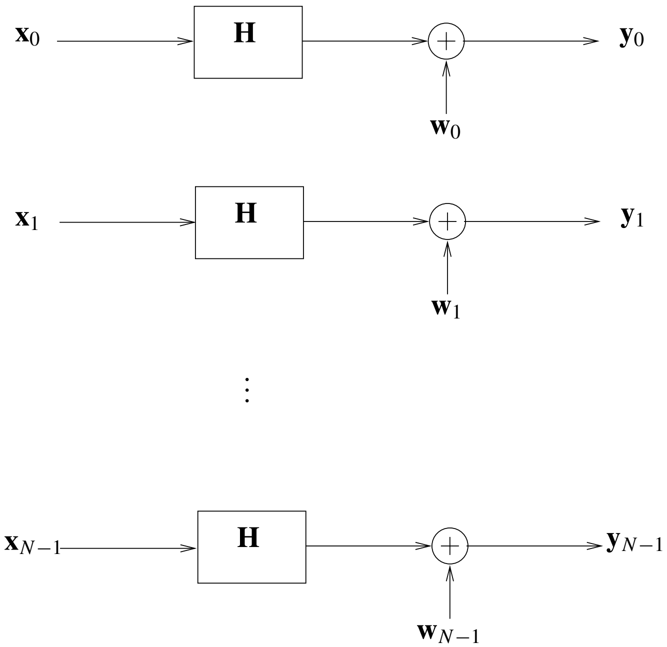

Fig. 1. The multiple signals model

We thus consider the case in which the noise vectors are indeed correlated. In this case, it is better to estimate the vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}$ jointly rather than separately, a fact that will be illustrated by an example in Section 7. To model the correlation between the noise vectors, we note that in many applications the order in which observations occur is immaterial, so that the statistics of $\mathbf{y}_i$ 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,

$$E(\mathbf{w}_i \mathbf{w}_i^T) = \mathbf{C}, \quad \forall 0 \le i \le N-1, \tag{8}$$

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

$$E(\mathbf{w}_i \mathbf{w}_j^T) = \mathbf{B}, \quad \forall 0 \le i \ne j \le N-1, \tag{9}$$

for some matrix $\mathbf{B}$. Moreover, by (9) we have that $E(\mathbf{w}_i \mathbf{w}_j^T) = E(\mathbf{w}_j \mathbf{w}_i^T)$, so that $\mathbf{B}$ is symmetric. Under the above structure, the system (7) can be written as the single signal system

$$\mathbf{y} = \widetilde{\mathbf{H}}\mathbf{x} + \mathbf{w}, \tag{10}$$

where $\widetilde{\mathbf{H}}$ is the $nN \times mN$ matrix defined by:

$$\widetilde{\mathbf{H}} \triangleq \begin{pmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H} \end{pmatrix},$$

$\mathbf{y} = (\mathbf{y}_0^T, \ldots, \mathbf{y}_{N-1}^T)^T$, $\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T$ and $\mathbf{w} = (\mathbf{w}_0^T, \ldots, \mathbf{w}_{N-1}^T)^T$. The covariance matrix of the expanded vector $\mathbf{w}$ is denoted by $\widetilde{\mathbf{C}} \triangleq E(\mathbf{w}\mathbf{w}^T)$ and is given by the following $nN \times nN$ matrix:

$$\widetilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{C} & \cdots & \mathbf{B} \\ \vdots & \vdots & & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{C} \end{pmatrix}. \tag{11}$$

A matrix of the form (11) is called an *elementary block circulant matrix* and is denoted by $\mathcal{M}(\mathbf{C}, \mathbf{B})$.

The multiple signals system (7) can be written as the single signal system (10), whose RMSE estimator is the solution of the problem

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}}^2 \le \tilde{L}^2, \widetilde{\mathbf{H}} \in \mathcal{U}_{\widetilde{\mathbf{H}}}, \widetilde{\mathbf{C}} \in \mathcal{U}_{\widetilde{\mathbf{C}}}} E(\|\mathbf{x} - \hat{\mathbf{x}}\|^2). \tag{12}$$

However, we cannot use the results of Section 3 since now there is a norm constraint on each of the indivudal blocks of $\mathbf{x}$. Furthermore, both $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{C}}$ are *structured*, and so the corresponding uncertainty sets should also have the same structure, *i.e.*, $\mathcal{U}_{\widetilde{\mathbf{H}}}$ should consist of block diagonal matrices and $\mathcal{U}_{\widetilde{\mathbf{C}}}$ should consist of elementary block circulant matrices. Althgouh the matirx $\mathbf{G}$ in (12) is an $mN \times nN$ matrix, we will show in Section 6 that the solution of (12) depends on only two $m \times n$ matrices $\mathbf{G}_0$ and $\mathbf{G}_1$ (regardless of the value of N). This result utilizes the elementary block circulant structure of $\mathbf{C}$ and is independent of any specific structure of the uncertainty sets.

In Section 7 we consider a specific case of the uncertainty sets $\mathcal{U}_{\mathbf{H}}$ and $\mathcal{U}_{\widetilde{\mathbf{C}}}$, under which the above matrices $\mathbf{G}_0$ and $\mathbf{G}_1$ can be computed efficiently by solving an SDP problem. In this case $\mathcal{U}_{\mathbf{H}}$ is the ellipsoidal-like uncertainty set defined in (4), whereas for the uncertainty in the covariance matrix $\widetilde{\mathbf{C}}$, we consider a case where the noise vectors $\mathbf{w}_k$ are composed of a noise component $\mathbf{u}_k$ with known second order statistics and a noise component $\mathbf{v}_k$ with uncertain second order statistics. To be more specific,

$$\mathbf{w}_k = \mathbf{u}_k + \mathbf{v}_k,$$

where $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$ are correlated noise vectors with known covariance matrix $E(\mathbf{u}_k\mathbf{u}_k^T) = \mathbf{C}_0$ and cross correlation covariance matrix $E(\mathbf{u}_i\mathbf{u}_j^T) = \mathbf{B}_0, \quad i \neq j$. The vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}$ are mutually uncorrelated, and are also uncorrelated with $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{N-1}$. Let us denote the uncertain covariance matrix of $\mathbf{v}_k$ by $\Delta_{\mathbf{C}} \stackrel{\triangle}{=} E(\mathbf{v}_k\mathbf{v}_k^T)$, then

$$E(\mathbf{w}_k\mathbf{w}_k^T) = \mathbf{C}_0 + \Delta_{\mathbf{C}}, \quad 0 \leq k \leq N-1$$
$$E(\mathbf{w}_i\mathbf{w}_j^T) = \mathbf{B}_0, \quad i \neq j.$$

We assume that $\Delta_{\mathbf{C}}$ is norm bounded $\|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}$, thus the ellipsoidal-like uncertainty set $\mathcal{U}_{\widetilde{\mathbf{C}}}$ is given by

$$\mathcal{U}_{\widetilde{\mathbf{C}}} = \{\mathcal{M}(\mathbf{C}, \mathbf{B}_0) : \mathbf{C} \in \mathcal{U}_{\mathbf{C}}\},$$

where $\mathcal{U}_{\mathbf{C}}$ is defined in (5). We show that in thef above setting of the uncertainty sets, the optimal solution of (12) can be found by solving an SDP problem whose size is independent of N.

1.4. Summary of the Organization of the Paper

Sections 2 to 4 deal with the single signal problem and sections 5 to 7 with the multiple signals estimation case. In Section 2 we present the problem of the single signal estimation problem and define the notation used throughout the paper. In Section 3 we consider the case of ellipsoidal-like uncertainty sets and show that the robust MSE estimator matrix is the solution to an SDP problem and thus we conclude that finding the robust MSE estimator is a tractable problem. Some examples are discussed in Section 4. The multiple signals estimation problem is presented in Section 5. In Section 6 we prove that under general assumptions on the structure of the uncertainty sets, the RMSE estimator matrix $\mathbf{G}$ can be chosen to be an elementary block circulant matrix. Using this result, we find in Section 7 that the robust MSE estimator in the multiple signals case is the solution of an SDP whose size does not depend on N. Section 7 is concluded with an example which demonstrates the advantage of the RMSE estimator for a multiple signals system as compared to the LS estimator and the RMSE estimator which treats the multiple signals system as N independent single signal systems. Some of the mathematical background needed in the paper is collected in the appendices.

2. Problem Formulation of the Single Signal Estimation Problem

We denote vectors by boldface lowercase letters and matrices by boldface uppercase letters. The identity matrix of appropriate dimension is denoted by $\mathbf{I}$, $(\cdot)^T$ and $\mathcal{R}(\cdot)$ denote the transpose and the range of the corresponding matrix, respectively. For two symmetric matrices $\mathbf{A}$, $\mathbf{B}$ the notation $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. If $\mathbf{A}$ is symmetric then $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$. For a matrix $\mathbf{M}$, $vec(\mathbf{M})$ denotes the vector obtained by stacking the columns of $\mathbf{M}$

In this part we consider the problem of estimating an unknown deterministic parameter vector $\mathbf{x} \in \mathbb{R}^m$ from an observation $\mathbf{y} \in \mathbb{R}^n$, where $\mathbf{y}$ is related to the parameter vector $\mathbf{x}$ through the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}. \tag{13}$$

Here $\mathbf{H}$ is an $n \times m$ $(n \geq m)$ matrix and $\mathbf{w}$ is a zero-mean random noise vector with covariance matrix $\mathbf{C} = E(\mathbf{w}\mathbf{w}^T)$. The matrices $\mathbf{H}$ and $\mathbf{C}$ are "uncertain" and are only known to belong to some uncertainty sets $\mathcal{U}_{\mathbf{H}}$ and $\mathcal{U}_{\mathbf{C}}$ respectively. It is assumed that $\mathbf{x}$ is known to satisfy a weighted norm constraint $\mathbf{x}^T\mathbf{T}\mathbf{x} \leq L$, where $\mathbf{T}$ is a known positive definite matrix.

We estimate $\mathbf{x}$ using a linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m \times n$ matrix $\mathbf{G}$. The matrix $\mathbf{G}$ is chosen to minimize the MSE:

$$\begin{aligned}
MSE &= E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \\
&= E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2) \\
&= E(\|\mathbf{G}(\mathbf{H}\mathbf{x} + \mathbf{w}) - \mathbf{x}\|^2) \\
&= E(\mathbf{w}^T\mathbf{G}^T\mathbf{G}\mathbf{w}) + 2E(\mathbf{w}^T(\mathbf{G}\mathbf{H} - \mathbf{I})\mathbf{x}) + \mathbf{x}^T(\mathbf{I} - \mathbf{G}\mathbf{H})^T(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} \\
&= \mathrm{Tr}(\mathbf{G}\mathbf{C}\mathbf{G}^T) + \mathbf{x}^T(\mathbf{I} - \mathbf{G}\mathbf{H})^T(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x}.
\end{aligned} \tag{14}$$

Since the MSE depends on the unknown parameters $\mathbf{x}$ and on the unknown matrices $\mathbf{H}$ and $\mathbf{C}$, 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 $\mathbf{x}$, $\mathbf{H}$ and $\mathbf{C}$. Thus, we consider the problem

$$\begin{aligned}
&\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}, \mathbf{C} \in \mathcal{U}_{\mathbf{C}}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \\
&= \min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}, \mathbf{C} \in \mathcal{U}_{\mathbf{C}}} \left\{ \mathbf{x}^T(\mathbf{I} - \mathbf{G}\mathbf{H})^T(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^T) \right\}.
\end{aligned} \tag{15}$$

The optimal solution $\mathbf{G}$ to (15) is called the *robust MSE estimator matrix*. and the induced estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ is called the *robust MSE estimator* (RMSE estimator).

In the next section we show that for certain reasonable uncertainty sets, the problem (15) can be reduced to a standard SDP [6, 25, 30], which is the problem of minimizing a linear objective subject to linear matrix inequality constraints, i.e. constraints of the form $\mathbf{A}(\mathbf{x}) \succeq 0$, where $\mathbf{A}$ is a matrix depending affinely on $\mathbf{x}$. The advantage of an SDP formulation is its computational tractability *i.e.,* it can be solved in polynomial time *e.g.,* via interior point methods (see [1, 6, 25, 30]).

3. SDP Formulation of the Estimation Problem with Ellipsoidal-like Uncertainty Sets

In this section we consider the following choice of uncertainty sets of **C** and **H**

$$\mathcal{U}_{\mathbf{C}} = \{\mathbf{C}_0 + \Delta_{\mathbf{C}} : \Delta_{\mathbf{C}} \in \mathbb{R}^{n\times n}, \|\Delta_{\mathbf{C}}\| \le \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}\}, \quad (16)$$

$$\mathcal{U}_{\mathbf{H}} = \{\mathbf{H}_0 + \Delta_{\mathbf{H}} : \Delta_{\mathbf{H}} \in \mathbb{R}^{n\times m} \|\Delta_{\mathbf{H}}\| \le \rho_{\mathbf{H}}\}, \quad (17)$$

where $\mathbf{C}_0$ is an $n \times n$ positive definite matrix, $\mathbf{H}_0$ is an $n \times m$ $(n \ge m)$ matrix assumed to have full rank m and $\rho_{\mathbf{C}}$ and $\rho_{\mathbf{H}}$ are nonnegative constants. $\mathbf{H}_0$ and $\mathbf{C}_0$ represent the nominal values of **H** and **C** respectively while $\Delta_{\mathbf{H}}$ and $\Delta_{\mathbf{C}}$ represent unknown perturbation matrices, which are norm bounded. Notice that $\mathcal{U}_{\mathbf{C}}$ is not an ellipsoidal set due to the additional constraints that force all matrices in $\mathcal{U}_{\mathbf{C}}$ to be positive semidefinite. The main result of this section is Theorem 1, which states that the problem of finding the optimal matrix **G** solving the problem (15), with uncertainty sets (16) and (17), can be formulated as an SDP problem. Before proving the main result, some technical lemmas are required.

Lemma 1. *Let* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *be a symmetric PSD matrix. Then,*

$$\max_{\mathbf{X}\in\mathbb{R}^{n\times n}, \|\mathbf{X}\|\le\delta, \mathbf{X}=\mathbf{X}^T, \mathbf{C}_0+\mathbf{X}\succeq\mathbf{0}} \{Tr(\mathbf{XQ})\} = \delta\sqrt{Tr(\mathbf{Q}^T\mathbf{Q})} \quad (18)$$

The maximum is attained at $\mathbf{X} = \frac{\delta}{\sqrt{Tr(\mathbf{Q}^T\mathbf{Q})}}\mathbf{Q}$.

Proof. For all matrices **X** with $\|\mathbf{X}\| \le \delta$ we have

$$\mathrm{Tr}(\mathbf{XQ}) \le \|\mathbf{X}\|\|\mathbf{Q}\| \le \delta\sqrt{\mathrm{Tr}(\mathbf{Q}^T\mathbf{Q})},$$

where the first inequality follows from the Cauchy-Schwartz inequality. Now, the matrix

$$\bar{\mathbf{X}} = \frac{\delta}{\sqrt{\mathrm{Tr}(\mathbf{Q}^T\mathbf{Q})}}\mathbf{Q}^T \quad (19)$$

is symmetric, PSD (by the assumption on **Q**) and $\|\bar{\mathbf{X}}\| = \delta$, thus $\bar{\mathbf{X}}$ is a feasible solution of problem (18) with a corresponding objective function value equal to the upper bound $\delta\|\mathbf{Q}\|$. Thus, $\bar{\mathbf{X}}$ is optimal. □

We will also need the following result

Lemma 2 ([11]). *Given matrices* $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ *with* $\mathbf{R} = \mathbf{R}^T$*,*

$$\mathbf{R} \succeq \mathbf{P}^T\mathbf{ZQ} + \mathbf{Q}^T\mathbf{Z}^T\mathbf{P}, \quad \forall\mathbf{Z} : \|\mathbf{Z}\| \le \delta$$

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

$$\begin{pmatrix} \mathbf{R} - \lambda\mathbf{Q}^T\mathbf{Q} & -\delta\mathbf{P}^T \\ -\delta\mathbf{P} & \lambda\mathbf{I} \end{pmatrix} \succeq 0.$$

With the help of the above lemmas we are able to prove our main result.

Theorem 1. *Let* $\mathbf{x} \in \mathbb{R}^m$ *denote the unknown parameters in the model* $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$*, where* $\mathbf{H} \in \mathcal{U}_{\mathbf{H}}$ *is an* $n \times m$ *matrix,* $\mathbf{x}$ *is known to satisfy the weighted norm constraint* $\|\mathbf{x}\|_{\mathbf{T}} \leq L$*, and* $\mathbf{w} \in \mathbb{R}^n$ *is a zero-mean random noise vector with covariance* $\mathbf{C} \in \mathcal{U}_{\mathbf{C}}$ *where*

$$\mathcal{U}_{\mathbf{C}} = \{\mathbf{C}_0 + \Delta_{\mathbf{C}} : \|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}\},$$
$$\mathcal{U}_{\mathbf{H}} = \{\mathbf{H}_0 + \Delta_{\mathbf{H}} : \|\Delta_{\mathbf{H}}\| \leq \rho_{\mathbf{H}}\}.$$

Then the robust MSE estimator matrix, i.e., *the matrix* $\mathbf{G}$ *that solves*

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}, \mathbf{C} \in \mathcal{U}_{\mathbf{C}}} \left\{\mathbf{x}^T(\mathbf{I} - \mathbf{G}\mathbf{H})^T(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{x} + Tr(\mathbf{G}\mathbf{C}\mathbf{G}^T)\right\} \tag{20}$$

is the optimal solution of the following SDP problem in the variables $\tau, \lambda, t_1, t_2 \in \mathbb{R}$, $\mathbf{X} \in \mathbb{R}^{n \times n}$ *and* $\mathbf{G} \in \mathbb{R}^{m \times n}$*:*

$$\min_{\tau, \lambda, t_1, t_2, \mathbf{X}, \mathbf{G}} \{L^2\tau + t_1 + \rho_{\mathbf{C}} t_2\}$$

subject to

$$\begin{pmatrix} \tau\mathbf{I} - \lambda\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H}_0)^T & \mathbf{0} \\ (\mathbf{I} - \mathbf{G}\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{G} \\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{G}^T & \lambda\mathbf{I} \end{pmatrix} \succeq \mathbf{0}$$
$$\begin{pmatrix} t_1 & \mathbf{g}^T \\ \mathbf{g} & \mathbf{I} \end{pmatrix} \succeq \mathbf{0}$$
$$\begin{pmatrix} \mathbf{X} & \mathbf{G}^T \\ \mathbf{G} & \mathbf{I} \end{pmatrix} \succeq \mathbf{0}$$
$$\begin{pmatrix} t_2 & \mathbf{x}^T \\ \mathbf{x} & t_2\mathbf{I} \end{pmatrix} \succeq \mathbf{0},$$

where $\mathbf{x} = vec(\mathbf{X})$ *and* $\mathbf{g} = vec(\mathbf{G}\mathbf{C}_0^{1/2})$.

Proof. Maximizing the inner objective function of (15) with respect to $\mathbf{x}$, problem (20) reduces to the following equivalent formulation:

$$\min_{\mathbf{G}} \max_{\|\Delta_{\mathbf{H}}\| \leq \rho_{\mathbf{H}}, \|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}} \Upsilon(\mathbf{G}, \Delta_{\mathbf{H}}, \Delta_{\mathbf{C}}), \tag{21}$$

where

$$\Upsilon(\mathbf{G}, \Delta_{\mathbf{H}}, \Delta_{\mathbf{C}}) = tr(\mathbf{G}(\mathbf{C} + \Delta_{\mathbf{C}})\mathbf{G}^T) + L^2\lambda_{\max}\left(\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}(\mathbf{H}_0 + \Delta_{\mathbf{H}}))^T(\mathbf{I} - \mathbf{G}(\mathbf{H}_0 + \Delta_{\mathbf{H}}))\mathbf{T}^{-1/2}\right). \tag{22}$$

Since $\Upsilon(\mathbf{G}, \Delta_{\mathbf{H}}, \Delta_{\mathbf{C}})$ is separable with respect to $\Delta_{\mathbf{H}}$ and $\Delta_{\mathbf{C}}$, it follows that the optimal value of the inner maximization problem in (21) is the sum of the optimal values of two independent maximization problems

$$\max_{\|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}} \mathrm{Tr}(\mathbf{G}(\mathbf{C}_0 + \Delta_{\mathbf{C}})\mathbf{G}^T) \tag{23}$$

and

$$\max_{\|\Delta_{\mathbf{H}}\|\leq\rho_{\mathbf{H}}} L^2\lambda_{\max}(\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))^T(\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))\mathbf{T}^{-1/2}). \tag{24}$$

Let us consider first (23). Since

$$\mathrm{Tr}(\mathbf{G}(\mathbf{C}_0+\Delta_{\mathbf{C}})\mathbf{G}^T)=\mathrm{Tr}(\mathbf{G}\mathbf{C}_0\mathbf{G}^T)+\mathrm{Tr}(\mathbf{G}\Delta_{\mathbf{C}}\mathbf{G}^T)=\mathrm{Tr}(\mathbf{G}\mathbf{C}_0\mathbf{G}^T)+\mathrm{Tr}(\Delta_{\mathbf{C}}\mathbf{G}^T\mathbf{G}),$$

we are left with the optimization problem

$$\max_{\|\Delta_{\mathbf{C}}\|\leq\rho_{\mathbf{C}},\Delta_{\mathbf{C}}=\Delta_{\mathbf{C}}^T,\mathbf{C}_0+\Delta_{\mathbf{C}}\succeq\mathbf{0}} \mathrm{Tr}(\Delta_{\mathbf{C}}\mathbf{G}^T\mathbf{G}). \tag{25}$$

Invoking Lemma 1 with $\mathbf{Q}=\mathbf{G}^T\mathbf{G}$, we obtain that the solution of (23) is $\mathrm{Tr}(\mathbf{G}\mathbf{C}_0\mathbf{G}^T)+\rho_{\mathbf{C}}\sqrt{\mathrm{Tr}(\mathbf{G}^T\mathbf{G}\mathbf{G}^T\mathbf{G})}$.

Now, let us consider (24). We can express (24) as the solution to the problem

$$\min_{\tau} L^2\tau \tag{26}$$

subject to

$$\mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))^T(\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))\mathbf{T}^{-1/2}\preceq\tau\mathbf{I},\quad \forall\Delta_{\mathbf{H}}:\|\Delta_{\mathbf{H}}\|\leq\rho_{\mathbf{H}}. \tag{27}$$

By Schur's complement (see Lemma 7 in the appendix), the constraint (27) is equivalent to

$$\begin{pmatrix}\tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))^T\\ (\mathbf{I}-\mathbf{G}(\mathbf{H}_0+\Delta_{\mathbf{H}}))\mathbf{T}^{-1/2} & \mathbf{I}\end{pmatrix}\succeq 0,\quad \forall\Delta_{\mathbf{H}}:\|\Delta_{\mathbf{H}}\|\leq\rho_{\mathbf{H}}. \tag{28}$$

Now, (28) can be rewritten equivalently as

$$\mathbf{R}\succeq\mathbf{P}^T\Delta_{\mathbf{H}}\mathbf{Q}+\mathbf{Q}^T\Delta_{\mathbf{H}}^T\mathbf{P},\quad \forall\Delta_{\mathbf{H}}:\|\Delta_{\mathbf{H}}\|\leq\rho_{\mathbf{H}}, \tag{29}$$

where

$$\begin{aligned}\mathbf{R}&=\begin{pmatrix}\tau\mathbf{I} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H}_0)^T\\ (\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I}\end{pmatrix}\\ \mathbf{P}&=\begin{pmatrix}\mathbf{0} & \mathbf{G}^T\end{pmatrix}\\ \mathbf{Q}&=\begin{pmatrix}\mathbf{T}^{-1/2} & \mathbf{0}\end{pmatrix}.\end{aligned}$$

We now use Lemma 2, which states that (29) is satisfied if and only if there exists a $\lambda\geq 0$ such that

$$\begin{pmatrix}\tau\mathbf{I}-\lambda\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{G}\mathbf{H}_0)^T & \mathbf{0}\\ (\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{G}\\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{G}^T & \lambda\mathbf{I}\end{pmatrix}\succeq 0. \tag{30}$$

To summarize, problem (21) reduces to

$$\min_{\tau,\lambda,\mathbf{G}} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}_0\mathbf{G}^T) + \rho_{\mathbf{C}}\sqrt{\mathrm{Tr}(\mathbf{G}\mathbf{G}^T\mathbf{G}\mathbf{G}^T)} + L^2\tau \right\} \tag{31}$$

subject to (30). Obviously, (31) subject to (30) is equivalent to:

$$\min_{\tau,\lambda,t_1,t_2,\mathbf{G}} \{t_1 + \rho_{\mathbf{C}} t_2 + L^2\tau\} \tag{32}$$

subject to

$$\begin{pmatrix} \tau\mathbf{I} - \lambda\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{G}\mathbf{H}_0)^T & \mathbf{0} \\ (\mathbf{I} - \mathbf{G}\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{G} \\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{G}^T & \lambda\mathbf{I} \end{pmatrix} \succeq 0 \tag{33}$$

$$\mathrm{Tr}(\mathbf{G}\mathbf{C}_0\mathbf{G}^T) \leq t_1 \tag{34}$$

$$\sqrt{\mathrm{Tr}(\mathbf{G}\mathbf{G}^T\mathbf{G}\mathbf{G}^T)} \leq t_2. \tag{35}$$

Denoting $\mathbf{g} = vec(\mathbf{G}\mathbf{C}_0^{1/2})$, we obtain that constraint (34) is equivalent to $\mathbf{g}^T\mathbf{g} \leq t_1$, which is the same as

$$\begin{pmatrix} t_1 & \mathbf{g}^T \\ \mathbf{g} & \mathbf{I} \end{pmatrix} \succeq 0$$

Finally, constraint (35) can be expressed as a set of two constraints:

$$\mathbf{G}^T\mathbf{G} \preceq \mathbf{X}, \quad \sqrt{\mathrm{Tr}(\mathbf{X}^T\mathbf{X})} \leq t_2,$$

which, by Schur's complement (Lemma 7 in the appendix), are the same as

$$\begin{pmatrix} \mathbf{X} & \mathbf{G}^T \\ \mathbf{G} & \mathbf{I} \end{pmatrix} \succeq \mathbf{0}, \quad \begin{pmatrix} t_2 & \mathbf{x}^T \\ \mathbf{x} & t_2\mathbf{I} \end{pmatrix} \succeq \mathbf{0}$$

and the proof is completed. □

4. Examples for the Single Signal Case

In this section we illustrate the results of Section 3 through some examples. The first example shows that in the case where **H** and **C** are known, the RMSE estimator is better than the common LS estimator even if we do not have a priori knowledge on L. In the second example, we consider the case where **C** is known and **H** is uncertain and we compare the RMSE estimator to the TLS estimator. We show that the RMSE estimator outperforms the TLS estimator in the case where L and $\rho_{\mathbf{H}}$ are not known but rather are estimated from the TLS result. The third example will illustrate the robustness of the RMSE estimator with respect to **H** and the last example will illustrate the robustness of the RMSE estimator with respect to **C** . The SDP programs needed to compute the RMSE estimator (see Theorem 1) where solved by SeDuMi (see [27]).

4.1. *Known* **H** *and* **C**

In this subsection we present a typical example demonstrating that, even when **H** and **C** are certain, the estimation error associated with minimizing the worst case MSE (minimax MSE estimator) is consistently smaller than the corresponding error associated with the LS estimator $\hat{\mathbf{x}}_{LS}$. No a priori knowledge is assumed on L, the upper bound on the signal norm. Instead, L is estimated from the LS solution itself, *i.e.,* $L = \|\hat{\mathbf{x}}_{LS}\|$, as discussed in [4]. Let

$$\mathbf{H}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.6 & 1 & 0 & 0 \\ 0.3 & 0.6 & 1 & 0 \\ 0.2 & 0.3 & 0.6 & 1 \\ 0 & 0.2 & 0.3, & 0.6 \\ 0 & 0 & 0.2 & 0.3 \end{pmatrix}, \quad \mathbf{C}_0 = \sigma^2 \mathbf{I}, \quad \mathbf{T} = \mathbf{I}. \tag{36}$$

$\mathbf{H}_0$ of the above structure represents typically a convolution with an LTI filter. The signal $\mathbf{x}$ has a norm 5 and was randomly generated from the uniform distribution of all vectors of norm 5. We compare two estimators:

1. The Least Squares estimator (LS), which is given by

$$\hat{\mathbf{x}}_{LS} = (\mathbf{H}_0^T \mathbf{C}_0^{-1} \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{C}_0^{-1} \mathbf{y}.$$

2. The minimax MSE estimator (MMSE) with known **H** and **C**. This estimator is of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, where **G** is the solution to the SDP problem in Theorem 1 with $\rho_{\mathbf{H}} = \rho_{\mathbf{C}} = 0$, and in the case $\mathbf{T} = \mathbf{I}$ reduces to $\alpha \hat{\mathbf{x}}_{LS}$ where $\alpha = L^2/(L^2 + \mathrm{Tr}((\mathbf{H}_0^* \mathbf{C}_0^{-1} \mathbf{H}_0)^{-1}))$. We estimate L as the norm of the LS estimator *i.e.,* $\widehat{L} = \|\hat{\mathbf{x}}_{LS}\|$.

For each of the estimation methods, we calculated the MSE. In Fig. 2 we plot the MSE averaged over 1000 noise realizations as a function of σ. The noise vectors where generated from a multivariate normal distribution with zero mean and covariance matrix $\mathbf{C}_0$.

From Fig. 2 it is clear the the MMSE estimator is better than the LS estimator and more so as the variance grows.

4.2. *Comparison of RMSE and TLS with Known* **C** *and Uncertain* **H**

In this subsection we consider the case in which **C** is known ($\rho_{\mathbf{C}} = 0$) and **H** is uncertain. $\mathbf{H}_0$, $\mathbf{C}_0$ and **T** are defined as in (36). The signal $\mathbf{x}$ has a norm 5 and was randomly generated from the uniform distribution of all vectors of norm 5. We compare two estimators:

1. The Total Least Squares estimator (TLS). The TLS problem seeks to minimize $\|(\mathbf{H}_0; \mathbf{y}) - (\hat{\mathbf{H}}; \hat{\mathbf{y}})\|$ subject to $\hat{\mathbf{y}} \in \mathcal{R}(\hat{\mathbf{H}})$. Once a minimizing pair $(\hat{\mathbf{H}}; \hat{\mathbf{y}})$ is found, the estimator $\hat{\mathbf{x}}_{TLS}$ is defined to be the solution to $\hat{\mathbf{H}}\mathbf{x} = \hat{\mathbf{y}}$ (see *e.g.,* [21] and references within).
2. The RMSE estimator with $\rho_{\mathbf{C}} = 0$, $\rho_{\mathbf{H}} = \|\hat{\mathbf{H}} - \mathbf{H}_0\|$ and $L = \|\hat{\mathbf{x}}_{TLS}\|$, which is the solution of the SDP problem in Theorem 1.

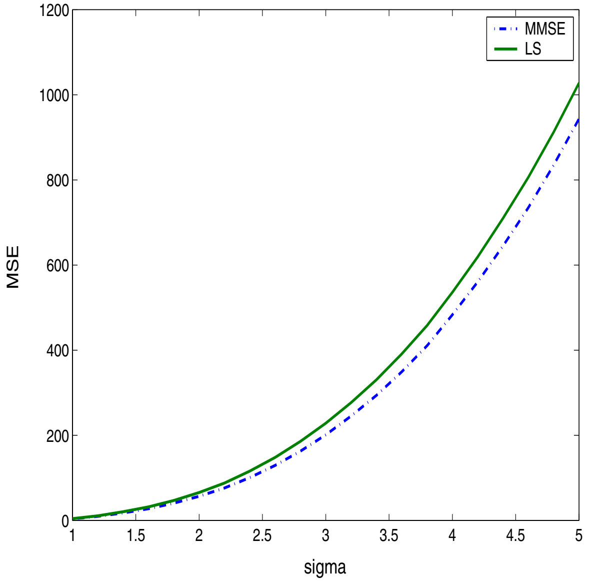

Fig. 2. MSE as a function of σ for the LS and MMSE estimators with known $\mathbf{H}$ and $\mathbf{C}$ and unknown L

$\mathbf{H}$ was chosen to be $\mathbf{H} = \mathbf{H}_0 + \Delta_{\mathbf{H}}$, where $\Delta_{\mathbf{H}}$ was randomly generated from the uniform distribution of all 6×4 matrices of norm 0.5. $\Delta_{\mathbf{H}}$ and $\mathbf{x}$ remain fixed in all noise realizations. For each of the estimation methods, we calculated the MSE. In Fig. 3 we plot the MSE averaged over 400 noise realizations as a function of σ. The noise vectors where generated from a multivariate normal distribution with expectation zero and covariance matrix $\mathbf{C}_0$.

The results show that the RMSE estimator significantly outperforms the TLS estimator. It is interesting to note that for large values of σ the TLS solution becomes very unstable. For example, for $\sigma = 2$ the MSE of the RMSE estimator was 19.6 and the MSE of the TLS solution was 254.7(!).

4.3. Robustness of the RMSE Estimator with respect to $\mathbf{H}$

In this subsection we demonstrate the robustness of RMSE with respect to $\mathbf{H}$. In our example

$$\mathbf{H}_0 = \begin{pmatrix} a & 0 & 0 & 0 \\ 0.4 & a & 0 & 0 \\ 0.6 & 0.4 & a & 0 \\ 1 & 0.6 & 0.4 & a \\ 0 & 1 & 0.6, & 0.4 \\ 0 & 0 & 1 & 0.6 \end{pmatrix}, \quad \mathbf{C}_0 = 0.5^2\mathbf{I}, \quad \mathbf{T} = \mathbf{I}, \quad L = 10, \quad \rho_{\mathbf{H}} = 0.02. \tag{37}$$

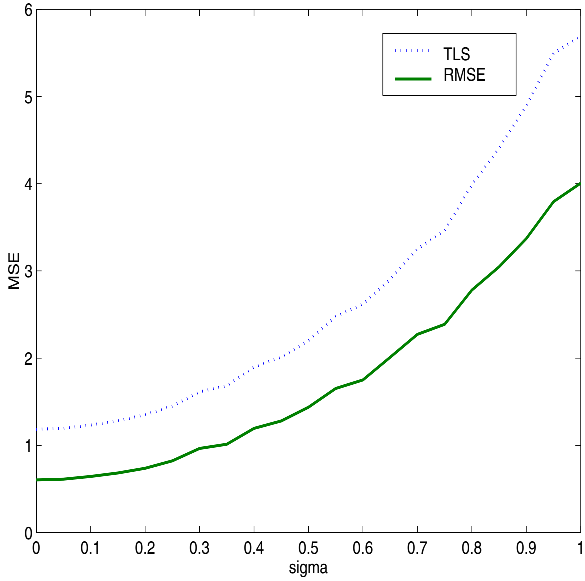

Fig. 3. MSE as a function of σ for the TLS and RMSE with unknown $\mathbf{H}$

The signal $\mathbf{x}$ has a norm 10 and was randomly generated from the uniform distribution of all vectors of norm 10. Suppose that the real value of $\mathbf{H}$ is

$$\mathbf{H} = \begin{pmatrix} 0.95a & 0 & 0 & 0 \\ 0.4 & 0.95a & 0 & 0 \\ 0.6 & 0.4 & 0.95a & 0 \\ 1 & 0.6 & 0.4 & 0.95a \\ 0 & 1 & 0.6, & 0.4 \\ 0 & 0 & 1 & 0.6 \end{pmatrix}, \tag{38}$$

We calculated the MSE for the LS estimator and for the RMSE estimator by averaging over 1000 realizations of the noise vector. The results are summarized in the table below.

a	$\sigma_4(\mathbf{H}_0)$	*MSE*		*sd*		*maximum*	
		LS	RMSE	LS	RMSE	LS	RMSE
0.02	0.104	21.93	34.71	29.75	7.93	273.7	91.9
0.06	0.073	46.69	58.27	65.55	6.55	603.6	107.5
0.10	0.051	102.86	81.14	133.09	3.16	845.8	97.5
0.14	0.045	132.0	85.35	189.93	2.45	1468.5	105.6

$\sigma_4(\mathbf{H}_0)$ is the smallest singular value of $\mathbf{H}_0$, which is equal to $\sqrt{\lambda_{\min}(\mathbf{H}_0^T\mathbf{H}_0)}$. The parameter a controls the ill-posedness of the problem: as it grows the problem becomes more ill-posed. *sd* is the standard deviation of the squared error and *maximum* is the maximum (over the 1000 realizations of the noise) of the squared error. Observing *sd* and *maximum* in the table, the results clearly demonstrate the severe instability of the

LS estimator in face of uncertainty in $\mathbf{H}$, and on the other hand the robustness of the RMSE estimator. Notice also that for the more severe ill-posedness cases ($a \geq 0.10$), the RMSE estimator has even a smaller MSE compared to the LS estimator.

4.4. *Robustness of the RMSE Estimator with respect to* C

In this subsection we demonstrate the robustness of the RMSE estimator with respect to the covariance matrix $\mathbf{C}$. In our example $\mathbf{H}_0$ is defined as in (36), $\mathbf{T} = \mathbf{I}$, $\rho_{\mathbf{H}} = 0$, $\rho_{\mathbf{C}} = 0.7$ and $L = 6$. The signal $\mathbf{x}$ has a norm 6 and was randomly generated from the uniform distribution of all vectors of norm 6. The nominal value of the covariance matrix $\mathbf{C}_0$ and the actual value of the covariance matrix $\mathbf{C}$ are given by

$$\mathbf{C}_0 = \begin{pmatrix} 1+a & 1 & 1 & 1 & 1 & 1 \\ 1 & 1+a & 1 & 1 & 1 & 1 \\ 1 & 1 & 1+a & 1 & 1 & 1 \\ 1 & 1 & 1 & 1+a & 1 & 1 \\ 1 & 1 & 1 & 1 & 1+a & 1 \\ 1 & 1 & 1 & 1 & 1 & 1+a \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1+0.5a & 1 & 1 & 1 & 1 & 1 \\ 1 & 1+0.5a & 1 & 1 & 1 & 1 \\ 1 & 1 & 1+0.5a & 1 & 1 & 1 \\ 1 & 1 & 1 & 1+0.5a & 1 & 1 \\ 1 & 1 & 1 & 1 & 1+0.5a & 1 \\ 1 & 1 & 1 & 1 & 1 & 1+0.5a \end{pmatrix}.$$

We calculated the MSE for the LS estimator and for the RMSE estimator by averaging over 1000 realizations of the noise vector. The results are summarized in the table below

a	*MSE*		*sd*		*maximum*	
	LS	RMSE	LS	RMSE	LS	RMSE
0.1	7.86	21.59	10.49	1.37	81.86	28.82
0.2	17.03	23.22	23.77	6.4	218.98	37.28
0.3	23.39	24.34	31.32	6.4	258.5	35.75
0.4	33.93	25.5	46.1	8.24	437.98	44.97
0.5	42.73	26.36	58.98	8.42	444.53	44.66

The same phenomena encountered in the case where $\mathbf{H}$ was uncertain are present in the case where $\mathbf{C}$ is uncertain. The RMSE becomes increasingly better than the LS estimator as a grows. The robustness of the RMSE estimator and the non robustness of the LS estimator is evident from the table.

5. The Multiple Signals Estimation Problem

In this part we consider the problem of estimating N unknown deterministic parameter vectors $\mathbf{x}_k \in \mathbb{R}^m$, $0 \leq k \leq N-1$ from N vector observations $\mathbf{y}_k \in \mathbb{R}^n$, $0 \leq k \leq N-1$, where each observation vector $\mathbf{y}_k$ is related to the corresponding parameter vector $\mathbf{x}_k$ through the linear model

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k, \quad 0 \le k \le N-1. \tag{39}$$

Here $\mathbf{H}$ is an $n \times m$ matrix, and $\mathbf{w}_k,\ 0 \le k \le N-1$ are zero-mean random noise vectors with covariance matrix $\mathbf{C} = E(\mathbf{w}_k\mathbf{w}_k^T)$ and cross correlation $\mathbf{B} = E(\mathbf{w}_k\mathbf{w}_l^T)$. The system (39) can be written as

$$\mathbf{y} = \widetilde{\mathbf{H}}\mathbf{x} + \mathbf{w}, \tag{40}$$

where $\widetilde{\mathbf{H}}$ is the $nN \times mN$ matrix defined by:

$$\widetilde{\mathbf{H}} \triangleq \begin{pmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H} \end{pmatrix},$$

$\mathbf{y} = (\mathbf{y}_0^T, \ldots, \mathbf{y}_{N-1}^T)^T$, $\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T$ and $\mathbf{w} = (\mathbf{w}_0^T, \ldots, \mathbf{w}_{N-1}^T)^T$. The covariance matrix of the expended vector $\mathbf{w}$ is denoted by $\widetilde{\mathbf{C}} \triangleq E(\mathbf{w}\mathbf{w}^T)$ and is given by the following $nN \times nN$ matrix:

$$\widetilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{C} & \cdots & \mathbf{B} \\ \vdots & \vdots & & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{C} \end{pmatrix}. \tag{41}$$

As in the single signal case, we assume that the model matrix and the noise statistics are not known exactly. Specifically, we assume that $\mathbf{H}$ and $\widetilde{\mathbf{C}}$ belong to some uncertainty sets $\mathcal{U}_{\mathbf{H}}$ and $\mathcal{U}_{\widetilde{\mathbf{C}}}$ respectively. As was mentioned in the introduction, we impose the following natural assumption on $\mathcal{U}_{\widetilde{\mathbf{C}}}$:

Assumption A. $\mathcal{U}_{\widetilde{\mathbf{C}}}$ contains only PSD elementary block circulant matrices.

Consider the case where $\mathbf{H}$ and $\mathbf{C}$ are known. Clearly, if the noise vectors were uncorrelated (*i.e.,* $\mathbf{B}$ is the zero matrix $\mathbf{0}$) then we can treat our estimation problem as N independent single signal problems, where each problem reduces to the problem, considered in [11], of estimating an unknown vector $\tilde{\mathbf{x}}$ from observations $\tilde{\mathbf{y}} = \mathbf{H}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}$ subject to the constraint that $\|\tilde{\mathbf{x}}\|_{\mathbf{T}} \le L$, where $\tilde{\mathbf{w}}$ is a zero-mean noise vector. If, on the other hand, the noise vectors are correlated, then we may be able to improve the estimation performance by treating the vectors to be estimated jointly, since the estimator $\hat{\mathbf{x}}_k$ of $\mathbf{x}_k$ may depend on all the observations $\mathbf{y}_l$, $0 \le l \le N-1$, and not just on $\mathbf{y}_k$.

We estimate $\mathbf{x}$ using a linear estimator

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}, \tag{42}$$

where $\mathbf{G}$ is an $mN \times nN$ matrix. Thus, the optimization problem we aim to solve is

$$\begin{aligned}&\min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L, \mathbf{H}\in\mathcal{U}_{\mathbf{H}}, \widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)\\ &= \min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L, \mathbf{H}\in\mathcal{U}_{\mathbf{H}}, \widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \left\{\mathbf{x}^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} + \mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T)\right\}. \quad (43)\end{aligned}$$

One of the main results, proven in Section 6, is that an optimal matrix $\mathbf{G}$, can be chosen to have a very unique structure. Specifically, we find that $\mathbf{G}$ can be chosen to be a block matrix of the form

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \cdots & \mathbf{G}_1 \\ \mathbf{G}_1 & \mathbf{G}_0 & \cdots & \mathbf{G}_1 \\ \vdots & \vdots & & \vdots \\ \mathbf{G}_1 & \mathbf{G}_1 & \cdots & \mathbf{G}_0 \end{pmatrix}. \quad (44)$$

Matrices of the structure (44) are called *elementary block circulant* matrices and are a special case of block circulant matrices (see Appendix A for further details) . Substituting (44) in (42) we have that the optimal linear estimator has the following structure:

$$\hat{\mathbf{x}}_k = \mathbf{G}_0\mathbf{y}_k + \sum_{i\neq k} \mathbf{G}_1\mathbf{y}_i, \quad 0 \le k \le N-1. \quad (45)$$

The latter formula implies the intuitively appealing result that all the vectors $y_l,\ l \neq k$ have the same effect on the estimator of $\mathbf{x}_k$. The structure (44) suggests a considerable computational simplification of the solution to (43) since we need to find only *two* $n \times m$ matrices $\mathbf{G}_0, \mathbf{G}_1$ rather then a huge $mN \times nN$ matrix $\mathbf{G}$.

6. The Structure of the Optimal Linear Estimator Matrix G: General Uncertainty Sets

In this section $\mathcal{U}_{\mathbf{H}}$ is an arbitrary uncertainty set and $\mathcal{U}_{\widetilde{\mathbf{C}}}$ is assumed to satisfy assumption A but otherwise no specific structure of $\mathcal{U}_{\mathbf{H}}$ and $\mathcal{U}_{\widetilde{\mathbf{C}}}$ is assumed. In Theorem 2 below, we show that the optimal linear estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ solving problem (43) can be chosen with $\mathbf{G}$ being an elementary block circulant matrix.

Theorem 2. *Let* $\mathbf{x} = \left(\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T\right)^T$ *denote the unknown parameters in the model* $\mathbf{y} = \widetilde{\mathbf{H}}\mathbf{x} + \mathbf{w}$, *where* $\widetilde{\mathbf{H}} = \mathcal{M}(\mathbf{H}, \mathbf{0})$ *and* $\mathbf{H}$ *is an* $n \times m$ *matrix that belongs to an uncertainty set* $\mathcal{U}_{\mathbf{H}}$, $\mathbf{w}$ *is zero-mean random vector with an* $nN \times nN$ *covariance matrix* $\widetilde{\mathbf{C}}$ *that belongs to an uncertainty set* $\mathcal{U}_{\widetilde{\mathbf{C}}}$ *satisfying assumption A. If there exists an optimal solution to*

$$\begin{aligned}&\min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L, \mathbf{H}\in\mathcal{U}_{\mathbf{H}}, \widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)\\ &= \min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L, \mathbf{H}\in\mathcal{U}_{\mathbf{H}}, \widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \left\{\mathbf{x}^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} + Tr(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T)\right\},\end{aligned}$$

then there exists an optimal solution $\mathbf{G}$ *which is equal to* $\mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$ *for some* $\mathbf{G}_0, \mathbf{G}_1 \in \mathbb{R}^{m\times n}$.

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:

$$(\mathbf{P}_{\sigma,l})_{i,j} = \delta_{j,\sigma(i)}\mathbf{I}_l, \tag{46}$$

where

$$\delta_{i,j} = \begin{cases} 0, & i \neq j, \\ 1, & i = j \end{cases} \tag{47}$$

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

$$\mathbf{P}_{\sigma,4} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_4 \end{pmatrix}, \tag{48}$$

where $\mathbf{I}_4$ is the identity matrix of size 4×4. 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, $\mathbf{P}_{\sigma,l}\mathbf{P}^T_{\sigma,l} = \mathbf{P}^T_{\sigma,l}\mathbf{P}_{\sigma,l} = \mathbf{I}$.
2. For every $\sigma \in S_N$ and for every block vector $\mathbf{x} = (\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T)^T$, $\mathbf{Px} = \mathbf{y}$ where

$$\mathbf{y}_k = \mathbf{x}_{\sigma(k)}, \quad 0 \leq k \leq N-1. \tag{49}$$

3. For every elementary block circulant matrix $\mathbf{A} = \mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)$, where $\mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{m,n}$ and every permutation $\sigma \in S_N$, we have that $\mathbf{P}_{\sigma,m}\mathbf{A}\mathbf{P}^T_{\sigma,n} = \mathbf{A}$, 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 elementary block circulant. Let $\mathbf{G}$ be an optimal solution to (43), we claim that so is $\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^T_{\sigma,n}$ for every permutation $\sigma \in S_N$. To this end we prove that $\Gamma(\mathbf{G}) = \Gamma(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}^T_{\sigma,n})$ where

$$\Gamma(\mathbf{G}) = \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}, \widetilde{\mathbf{C}} \in \mathcal{U}_{\widetilde{\mathbf{C}}}} \left\{ \mathbf{x}^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} + \mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T) \right\} \tag{50}$$

is the objective function in the minimization problem (43). Indeed, (the number of the property used is indicated):

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

and,

$$\begin{aligned}
&\max_{\|\mathbf{x}_i\|_{\mathbf{T}}\le L,\widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \{\mathbf{x}^T(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})\mathbf{x}\}\\
&\overset{2}{=} \max_{\|\mathbf{x}_i\|_{\mathbf{T}}\le L,\mathbf{H}\in\mathcal{U}_{\mathbf{H}},\widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \{\mathbf{x}^T\mathbf{P}_{\sigma,m}^T(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})\mathbf{P}_{\sigma,m}\mathbf{x}\}\\
&\overset{1}{=} \max_{\|\mathbf{x}_i\|_{\mathbf{T}}\le L,\mathbf{H}\in\mathcal{U}_{\mathbf{H}},\widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \{\mathbf{x}^T\mathbf{P}_{\sigma,m}^T(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})^T\mathbf{P}_{\sigma,m}^T\mathbf{P}_{\sigma,m}(\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})\mathbf{P}_{\sigma,m}\mathbf{x}\}\\
&\overset{1}{=} \max_{\|\mathbf{x}_i\|_{\mathbf{T}}\le L,\mathbf{H}\in\mathcal{U}_{\mathbf{H}},\widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \{\mathbf{x}^T(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}\widetilde{\mathbf{H}}\mathbf{P}_{\sigma,m}^T)^T(\mathbf{I}-\mathbf{P}_{\sigma,m}\mathbf{G}\widetilde{\mathbf{H}}\mathbf{P}_{\sigma,m}^T)\mathbf{x}\}\\
&\overset{3}{=} \max_{\|\mathbf{x}_i\|_{\mathbf{T}}\le L,\mathbf{H}\in\mathcal{U}_{\mathbf{H}},\widetilde{\mathbf{C}}\in\mathcal{U}_{\widetilde{\mathbf{C}}}} \{\mathbf{x}^T(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T)\widetilde{\mathbf{H}})^T(\mathbf{I}-(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T)\widetilde{\mathbf{H}})\mathbf{x}\},
\end{aligned}$$

where in the last equality we used the fact that $\widetilde{\mathbf{H}}$ is a block diagonal matrix and as a result also a elementary block circulant matrix. Since (43) is a convex optimization problem, if $\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T$ is an optimal solution to (43) for all $\sigma \in S_N$, then so is the convex combination $\frac{1}{N!}\sum_{\sigma\in S_N}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T$. We will now show that $\widetilde{\mathbf{G}}\overset{\triangle}{=}\frac{1}{N!}\sum_{\sigma\in S_N}\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T = \mathcal{M}(\mathbf{G}_0,\mathbf{G}_1)$ for some matrices $\mathbf{G}_0,\mathbf{G}_1\in\mathbb{R}^{m\times n}$. Specifically, we will show that if

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{00} & \mathbf{G}_{01} & \cdots & \mathbf{G}_{0,N-1}\\ \mathbf{G}_{10} & \mathbf{G}_{11} & \cdots & \mathbf{G}_{1,N-1}\\ \vdots & \vdots & & \vdots\\ \mathbf{G}_{N-1,0} & \mathbf{G}_{N-1,1} & \cdots & \mathbf{G}_{N-1,N-1}\end{pmatrix}, \tag{51}$$

then

$$\mathbf{G}_0 = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{G}_{i,i},$$
$$\mathbf{G}_1 = \frac{1}{N(N-1)}\sum_{i\neq j}\mathbf{G}_{i,j}.$$

Indeed, for every $\sigma\in S_N$ we have

$$(\mathbf{P}_{\sigma,m}\mathbf{G}\mathbf{P}_{\sigma,n}^T)_{i,j} = \mathbf{G}_{\sigma^{-1}(i),\sigma^{-1}(j)}$$

As a result,

$$\widetilde{\mathbf{G}}_{i,j} = \frac{1}{N!}\sum_{\sigma\in S_N}\mathbf{G}_{\sigma^{-1}(i),\sigma^{-1}(j)}. \tag{52}$$

Suppose that $i\neq j$. For every $k\neq l$, the number of times the expression $\mathbf{G}_{k,l}$ occurs in (52) is equal to the number of permutation $\sigma\in S_N$ that satisfy $\sigma^{-1}(i)=k$ and $\sigma^{-1}(j)=l$ which is (N-2)!. Thus,

$$\widetilde{\mathbf{G}}_{i,j} = \frac{1}{N!}\sum_{k\neq l}(N-2)!\mathbf{G}_{k,l} = \frac{1}{N(N-1)}\sum_{k\neq l}\mathbf{G}_{k,l}$$

Thus, $\widetilde{\mathbf{G}}_{i,j}$ is the same matrix for every $i \neq j$ which is precisely the matrix $\mathbf{G}_1$ above. A similar argument proves that $\widetilde{\mathbf{G}}_{i,i} = \mathbf{G}_0$, completing the proof of the theorem. □

7. SDP Formulation of the Multiple Signals Estimation Problem: Ellipsoidal-like Uncertainty Sets

In Section 6 we have shown that under Assumption A, but otherwise regardless of the specific choice of the uncertainty sets $\mathcal{U}_{\mathbf{H}}$ and $\mathcal{U}_{\widetilde{\mathbf{C}}}$, the RMSE estimator matrix $\mathbf{G}$ can be chosen to be a elementary block circulant matrix. In this section we deal with the case where $\mathcal{U}_{\mathbf{H}}$ is an ellipsoidal-like set (see (17)). As for $\mathcal{U}_{\widetilde{\mathbf{C}}}$, we deal with the case where no uncertainty is associated with the cross correlation matrix $E(\mathbf{w}_i\mathbf{w}_j^T), i \neq j$. Thus, the uncertainty set $\mathcal{U}_{\widetilde{\mathbf{C}}}$ is of the form[2]:

$$\mathcal{U}_{\widetilde{\mathbf{C}}} = \{\mathcal{M}(\mathbf{C}, \mathbf{B}_0) : \mathbf{C} \in \mathcal{U}_{\mathbf{C}}\}$$

where $\mathcal{U}_{\mathbf{C}}$ is defined as in (16). We also assume that the nominal value of the covariance matrix $\widetilde{\mathbf{C}}$ is a positive definite matrix. *i.e.,*

Assumption B. $\mathcal{M}(\mathbf{C}_0, \mathbf{B}_0)$ is a positive definite matrix.

We will show that in this setting of the uncertainty governing $\mathbf{H}$ and $\widetilde{\mathbf{C}}$, the problem of computing the solution $\mathbf{G}$ to problem (43), can be formulated as an SDP problem.

Notice that in general, the inner maximization problem in (43)

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \leq L} \left\{ \mathbf{x}^T (\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T (\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} \right\}, \tag{53}$$

is an NP-hard problem. Indeed, even in the simplest instance of problem (53)

$$\max\{\mathbf{x}^T \mathbf{R}\mathbf{x} : |\mathbf{x}_i| \leq 1, i = 0, \ldots, N-1\}$$

is already NP-Hard. However, by exploiting the special structure of $\mathbf{G}$ and $\widetilde{\mathbf{H}}$, we will be able to solve (53) efficiently.

Since $\mathbf{G}$ and $\widetilde{\mathbf{H}}$ are both elementary block circulant matrices, it follows immediately that $(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T (\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})$ is also an elementary block circulant matrix. Therefore, there exists $\mathbf{S}_0, \mathbf{S}_1 \in \mathbb{R}^{m \times m}$ such that

$$(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T (\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}}) = \mathcal{M}(\mathbf{S}_0, \mathbf{S}_1), \tag{54}$$

and $\mathcal{M}(\mathbf{S}_0, \mathbf{S}_1)$ is a symmetric matrix. Notice that if we define $\mathbf{S}_2, \mathbf{S}_3, \ldots, \mathbf{S}_{N-1}$ to be equal to $\mathbf{S}_1$ then we have

$$\mathcal{M}(\mathbf{S}_0, \mathbf{S}_1) = \mathcal{C}(\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{N-1}). \tag{55}$$

Using (54) and (55), (53) can be expressed as:

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \leq L} \mathbf{x}^T \mathcal{C}(\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{N-1})\mathbf{x}. \tag{56}$$

The following lemma is the key result which enables us to solve the maximization problem (53).

[2] This structure of $\mathcal{U}_{\widetilde{\mathbf{C}}}$ was motivated in the introduction.

Lemma 3. *Let $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{N-1} \in \mathbb{R}^{m \times m}$ be matrices such that $\mathcal{C}(\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{N-1})$ is a symmetric matrix. Let $\mathbf{T}$ be a positive definite matrix and let $L > 0$ be a constant. Then,*

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L} \mathbf{x}^T \mathcal{C}(\mathbf{S}_0, \ldots, \mathbf{S}_{N-1})\mathbf{x} = \max_{\sum_{i=0}^{N-1} \|\mathbf{x}_i\|_{\mathbf{T}}^2 \le NL^2} \mathbf{x}^T \mathcal{C}(\mathbf{S}_0, \ldots, \mathbf{S}_{N-1})\mathbf{x}.$$

Furthermore,

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L} \mathbf{x}^T \mathcal{C}(\mathbf{S}_0, \ldots, \mathbf{S}_{N-1})\mathbf{x} = NL^2 \max_{0 \le j \le N-1} \left\{ \lambda_{\max} \left(\mathbf{T}^{-1/2} \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{S}_i \right) \mathbf{T}^{-1/2} \right) \right\}, \tag{57}$$

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

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

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L} \mathbf{x}^T \mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})\mathbf{x} = \max_{\|\mathbf{z}_i\| \le L} \mathbf{z}^T \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z} \tag{58}$$

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

$$\underbrace{\max_{\|\mathbf{z}_i\| \le L} \mathbf{z}^T \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z}}_{(P_1)} \le \underbrace{\max_{\|\mathbf{z}\|^2 \le NL^2} \mathbf{z}^T \mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})\mathbf{z}}_{(P_2)}. \tag{59}$$

The value of the solution of (P_2) is $NL^2\lambda_{\max}\mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})$ 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, from Remark 2 following Theorem 4, we infer that for every eigenvalue we can find a corresponding eigenvector $\mathbf{z}$ of square norm NL^2 that satisfies $\|\mathbf{z}_0\| = \|\mathbf{z}_1\| = \cdots = \|\mathbf{z}_{N-1}\| = L$. From this it follows that $val(P_1) = val(P_2)$ and that the optimal value of (P_1) is equal to

$$NL^2\lambda_{\max}(\mathcal{C}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_{N-1})) \overset{\text{Theorem 4}}{=} NL^2 \max_{0 \le j \le N-1} \left\{ \lambda_{\max} \left(\sum_{i=0}^{N-1} \omega^{ij} \tilde{\mathbf{A}}_i \right) \right\}$$
$$= NL^2 \max_{0 \le j \le N-1} \left\{ \lambda_{\max} \left(\mathbf{T}^{-1/2} \left(\sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i \right) \mathbf{T}^{-1/2} \right) \right\}, \tag{60}$$

completing the proof. □

The next Theorem is the main result of this section. It states that the problem of finding the RMSE estimator matrix can be formulated as an SDP problem, which moreover, is of a size independent of the number of input vectors N.

Theorem 3. *Let* $\mathbf{x} = (\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{N-1}^T)^T$ *denote the vector of unknown parameters in the model* $\mathbf{y} = \widetilde{\mathbf{H}}\mathbf{x} + \mathbf{w}$*, where* $\widetilde{\mathbf{H}} = \mathcal{M}(\mathbf{H}, \mathbf{0})$ *with* $\mathbf{H}$ *being an* $n \times m$ *matrix, and* $\mathbf{w}$ *a zero-mean random vector with covariance* $\widetilde{\mathbf{C}} = \mathcal{M}(\mathbf{C}, \mathbf{B}_0)$*. Consider the uncertainty sets*

$$\mathcal{U}_{\mathbf{H}} = \{\mathbf{H}_0 + \Delta_{\mathbf{H}} : \|\Delta_{\mathbf{H}}\| \leq \rho_{\mathbf{H}}\},$$
$$\mathcal{U}_{\mathbf{C}} = \{\mathbf{C}_0 + \Delta_{\mathbf{C}} : \|\Delta_{\mathbf{C}}\| \leq \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}\},$$

and assume that $\mathcal{M}(\mathbf{C}_0, \mathbf{B}_0)$ *is a PSD matrix. Then the optimal matrix* $\mathbf{G}$ *that solves*

$$\min_{\mathbf{G}} \max_{\|\mathbf{x}_i\|_{\mathbf{T}} \leq L, \mathbf{H} \in \mathcal{U}_{\mathbf{H}}, \mathbf{C} \in \mathcal{U}_{\mathbf{C}}} \left\{\mathbf{x}^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} + Tr(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T)\right\} \tag{61}$$

is equal to $\mathcal{M}(\mathbf{G}_0, \mathbf{G}_1)$*, with* $\mathbf{G}_0 = \frac{1}{N}(\mathbf{A}_0 + (N-1)\mathbf{A}_1)$ *and* $\mathbf{G}_1 = \frac{1}{N}(\mathbf{A}_0 - \mathbf{A}_1)$*, where* $\mathbf{A}_0$ *and* $\mathbf{A}_1$ *are the optimal solutions of the following SDP problem in the variables* $\tau, \lambda_0, \lambda_1, t_0, t_1, t_2 \in \mathbb{R}, \mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{m \times n}$ *and* $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{R}^{n \times n}$*:*

$$\min_{\mathbf{A}_0, \mathbf{A}_1, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \tau, \lambda_0, \lambda_1, t_0, t_1, t_2} \{NL^2\tau + t_0 + (N-1)t_1 + \rho_{\mathbf{C}} t_2\}$$

subject to

$$\begin{pmatrix} \tau\mathbf{I} - \lambda_i\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)^T & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{A}_i \\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{A}_i^T & \lambda_i\mathbf{I} \end{pmatrix} \succeq 0, \quad i = 0, 1$$
$$\begin{pmatrix} t_i & \mathbf{a}_i^T \\ \mathbf{a}_i & \mathbf{I} \end{pmatrix} \succeq 0, \quad i = 0, 1,$$
$$\begin{pmatrix} \mathbf{Y} & \mathbf{A}_0^T \\ \mathbf{A}_0 & \mathbf{I} \end{pmatrix} \succeq \mathbf{0},$$
$$\begin{pmatrix} \mathbf{Z} & \mathbf{A}_1^T \\ \mathbf{A}_1 & \mathbf{I} \end{pmatrix} \succeq \mathbf{0},$$
$$\mathbf{Y} + (N-1)\mathbf{Z} \preceq \mathbf{X},$$
$$\begin{pmatrix} t_2 & \mathbf{x}^T \\ \mathbf{x} & t_2\mathbf{I} \end{pmatrix} \succeq \mathbf{0},$$

where $\mathbf{a}_0 = vec(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)^{1/2})$*,* $\mathbf{a}_1 = vec(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)^{1/2})$ *and* $\mathbf{x} = vec(\mathbf{X})$*.*

Proof. First, note that the expression $\sum_{i=0}^{N-1} \omega^{ij}\mathbf{S}_i$, which appears in (57), is exactly $\mathcal{F}_j((\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}}))$. By the properties listed in Lemma 5, we can deduce that for every $0 \leq j \leq N-1$,

$$\begin{aligned} \sum_{i=0}^{N-1} \omega^{ij}\mathbf{S}_i &= \mathcal{F}_j((\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})) \\ &= \mathcal{F}_j\left((\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T\right)\mathcal{F}_j\left(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}}\right) \\ &= \left(\mathcal{F}_j(\mathbf{I}) - \mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\widetilde{\mathbf{H}})\right)^T\left(\mathcal{F}_j(\mathbf{I}) - \mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\widetilde{\mathbf{H}})\right) \\ &= \left(\mathbf{I} - \mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\widetilde{\mathbf{H}})\right)^T\left(\mathbf{I} - \mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\widetilde{\mathbf{H}})\right). \end{aligned} \tag{62}$$

Since $\widetilde{\mathbf{H}} = \mathcal{M}(\mathbf{H}_0 + \Delta_{\mathbf{H}}, \mathbf{0})$ we have that

$$\mathcal{F}_j(\widetilde{\mathbf{H}}) = \mathbf{H}_0 + \Delta_{\mathbf{H}} \tag{63}$$

for every $0 \le j \le N-1$. Also, by (84) and (85) the following holds

$$\mathcal{F}_j(\mathbf{G}) = \begin{cases} \mathbf{G}_0 + (N-1)\mathbf{G}_1 & j = 0 \\ \mathbf{G}_0 - \mathbf{G}_1 & 1 \le j \le N-1 \end{cases} \tag{64}$$

Thus, substituting (62), (63) and (64) into (57) we conclude that

$$\max_{\|\mathbf{x}_i\|_{\mathbf{T}} \le L} \mathbf{x}^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^T(\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})\mathbf{x} = NL^2 \max\{\alpha(\Delta_{\mathbf{H}}), \beta(\Delta_{\mathbf{H}})\}, \tag{65}$$

where

$$\alpha(\Delta_{\mathbf{H}}) = \lambda_{\max}\left(\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_0(\mathbf{H}_0 + \Delta_{\mathbf{H}}))^T(\mathbf{I} - \mathbf{A}_0(\mathbf{H}_0 + \Delta_{\mathbf{H}})\mathbf{T}^{-1/2}\right),$$
$$\beta(\Delta_{\mathbf{H}}) = \lambda_{\max}\left(\mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_1(\mathbf{H}_0 + \Delta_{\mathbf{H}}))^T(\mathbf{I} - \mathbf{A}_1(\mathbf{H}_0 + \Delta_{\mathbf{H}})\mathbf{T}^{-1/2}\right),$$

and

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

The inner maximization problem of (61) is the sum of two independent maximization problems:

$$\max_{\|\Delta_{\mathbf{H}}\| \le \rho_{\mathbf{H}}} NL^2 \max\{\alpha(\Delta_{\mathbf{H}}), \beta(\Delta_{\mathbf{H}})\} \tag{66}$$

and

$$\max_{\|\Delta_{\mathbf{C}}\| \le \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}} \mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T). \tag{67}$$

First, the maximization problem (66) can be expressed as

$$\begin{aligned} &\min\ NL^2\tau \\ &\text{s.t.} \\ &\quad \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_i(\mathbf{H}_0 + \Delta_{\mathbf{H}}))^T(\mathbf{I} - \mathbf{A}_i(\mathbf{H}_0 + \Delta_{\mathbf{H}}))\mathbf{T}^{-1/2} \preceq \tau\mathbf{I} \quad \forall \Delta_{\mathbf{H}} : \|\Delta_{\mathbf{H}}\| \le \rho_{\mathbf{H}} \\ &\quad i = 0, 1. \end{aligned}$$

As in the proof of Theorem 1, the constraints of the above minimization problem can be expressed as

$$\begin{pmatrix} \tau\mathbf{I} - \lambda_i\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)^T & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{A}_i \\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{A}_i^T & \lambda_i\mathbf{I} \end{pmatrix} \succeq 0, \quad i = 0, 1.$$

Now, the objective function of (67) can also be simplified. Since $\mathbf{G}$ and $\widetilde{\mathbf{C}}$ are elementary block circulant then so is $\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T$. That is, there exists $\mathbf{R}_0$ and $\mathbf{R}_1$ such that $\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T = \mathcal{M}(\mathbf{R}_0, \mathbf{R}_1)$. Thus, $\mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T) = N\mathrm{Tr}(\mathbf{R}_0)$. Furthermore,

$$\begin{aligned}
\mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T) &= N\mathrm{Tr}(\mathbf{R}_0) \\
&\overset{(81)}{=} N\mathrm{Tr}(\mathcal{F}_0^{-1}(\mathcal{F}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T))) \\
&= \sum_{j=0}^{N-1} \mathrm{Tr}(\mathcal{F}_j(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T)) \\
&= \sum_{j=0}^{N-1} \mathrm{Tr}(\mathcal{F}_j(\mathbf{G})\mathcal{F}_j(\widetilde{\mathbf{C}})\mathcal{F}_j(\mathbf{G})^T) \\
&\overset{\mathcal{F}_j(\widehat{\mathbf{C}})=\mathbf{C}}{=} \sum_{j=0}^{N-1} \mathrm{Tr}(\widehat{\mathbf{G}}_j\widehat{\mathbf{C}}_j\widehat{\mathbf{G}}_j^T) \\
&= \mathrm{Tr}(\mathbf{A}_0(\mathbf{C} + (N-1)\mathbf{B}_0)\mathbf{A}_0^T) + (N-1)\mathrm{Tr}(\mathbf{A}_1(\mathbf{C} - \mathbf{B}_0)\mathbf{A}_1^T)
\end{aligned}$$

Substituting $\mathbf{C} = \mathbf{C}_0 + \Delta_{\mathbf{C}}$ we obtain

$$\begin{aligned}
\mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T) = {}& \mathrm{Tr}(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)\mathbf{A}_0^T) + (N-1)\mathrm{Tr}(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)\mathbf{A}_1^T) \\
&+ \mathrm{Tr}(\Delta_{\mathbf{C}}(\mathbf{A}_0^T\mathbf{A}_0 + (N-1)\mathbf{A}_1^T\mathbf{A}_1)).
\end{aligned}$$

From Lemma 1 we have:

$$\begin{aligned}
\max_{\|\Delta_{\mathbf{C}}\| \le \rho_{\mathbf{C}}, \Delta_{\mathbf{C}} = \Delta_{\mathbf{C}}^T, \mathbf{C}_0 + \Delta_{\mathbf{C}} \succeq \mathbf{0}} \mathrm{Tr}(\mathbf{G}\widetilde{\mathbf{C}}\mathbf{G}^T) = {}& \mathrm{Tr}(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)\mathbf{A}_0^T) \\
&+ (N-1)\mathrm{Tr}(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)\mathbf{A}_1^T) \\
&+ \sqrt{\mathrm{Tr}\left((\mathbf{A}_0^T\mathbf{A}_0 + (N-1)\mathbf{A}_1^T\mathbf{A}_1)^2\right)}
\end{aligned}$$

Thus, (43) can be expressed as

$$\min_{\mathbf{A}_0, \mathbf{A}_1, \tau, \lambda_0, \lambda_1, t_0, t_1, t_2} \{NL^2\tau + t_0 + (N-1)t_1 + \rho_{\mathbf{C}} t_2\}$$

subject to

$$\begin{aligned}
\begin{pmatrix} \tau\mathbf{I} - \lambda_i\mathbf{T}^{-1} & \mathbf{T}^{-1/2}(\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)^T & \mathbf{0} \\ (\mathbf{I} - \mathbf{A}_i\mathbf{H}_0)\mathbf{T}^{-1/2} & \mathbf{I} & -\rho_{\mathbf{H}}\mathbf{A}_i \\ \mathbf{0} & -\rho_{\mathbf{H}}\mathbf{A}_i^T & \lambda_i\mathbf{I} \end{pmatrix} &\succeq 0, \quad i = 0, 1 \\
\mathrm{Tr}(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)\mathbf{A}_0^T) &\le t_0 \\
\mathrm{Tr}(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)\mathbf{A}_1^T) &\le t_1 \\
\sqrt{\mathrm{Tr}\left((\mathbf{A}_0^T\mathbf{A}_0 + (N-1)\mathbf{A}_1^T\mathbf{A}_1)^2\right)} &\le t_2.
\end{aligned}$$

By Schur's complement (Lemma 7 in the appendix), the constraints:

$$\mathrm{Tr}(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)\mathbf{A}_0^T) \le t_0, \quad \mathrm{Tr}(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)\mathbf{A}_1^T) \le t_1$$

are equivalent to

$$\begin{pmatrix} t_i & \mathbf{a}_i^T \\ \mathbf{a}_i & \mathbf{I} \end{pmatrix} \succeq 0, \quad i = 0, 1,$$

where $\mathbf{a}_0 = vec(\mathbf{A}_0(\mathbf{C}_0 + (N-1)\mathbf{B}_0)^{1/2})$ and $\mathbf{a}_1 = vec(\mathbf{A}_1(\mathbf{C}_0 - \mathbf{B}_0)^{1/2})$. Finally, the constraint

$$\sqrt{\operatorname{Tr}\left((\mathbf{A}_0^T\mathbf{A}_0 + (N-1)\mathbf{A}_1^T\mathbf{A}_1)^2\right)} \leq t_2$$

is equivalent to the following set of constraints:

$$\mathbf{A}_0^T\mathbf{A}_0 \preceq \mathbf{Y}, \tag{68}$$

$$\mathbf{A}_1^T\mathbf{A}_1 \preceq \mathbf{Z}, \tag{69}$$

$$\mathbf{Y} + (N-1)\mathbf{Z} \preceq \mathbf{X}$$

$$\sqrt{\operatorname{Tr}(\mathbf{X}^T\mathbf{X})} \leq t_2 \tag{70}$$

Constraints (68), (69) and (70) can be expressed as

$$\begin{pmatrix} \mathbf{Y} & \mathbf{A}_0^T \\ \mathbf{A}_0 & \mathbf{I} \end{pmatrix} \succeq \mathbf{0}, \quad i = 0, 1,$$

$$\begin{pmatrix} \mathbf{Z} & \mathbf{A}_1^T \\ \mathbf{A}_1 & \mathbf{I} \end{pmatrix} \succeq \mathbf{0}, \quad i = 0, 1,$$

$$\begin{pmatrix} t_2 & \mathbf{x}^T \\ \mathbf{x} & t_2\mathbf{I} \end{pmatrix} \succeq \mathbf{0},$$

where $\mathbf{x} = vec(\mathbf{X})$. The proof of the theorem is completed. □

Example: We now demonstrate the RMSE estimator in the multiple signals scenario. $\mathbf{H}_0$ and $\mathbf{H}$ are defined as in (37) and (38) respectively. We consider the case where $N = 20, L = 5, \rho_{\mathbf{H}} = 0.06$ and $\rho_{\mathbf{C}} = 0.3$. The nominal noise covariance matrix is given by

$$\widetilde{\mathbf{C}} = \mathcal{M}(\mathbf{C}_0, \mathbf{B}_0),$$

where

$$\mathbf{C}_0 = \begin{pmatrix} 2&1&1&1&1&1 \\ 1&2&1&1&1&1 \\ 1&1&2&1&1&1 \\ 1&1&1&2&1&1 \\ 1&1&1&1&2&1 \\ 1&1&1&1&1&2 \end{pmatrix}, \quad \mathbf{B}_0 = t\mathbf{I}_6$$

The "real" covariance matrix is given by $\mathcal{M}(\mathbf{C}, \mathbf{B}_0)$ where $\mathbf{C} = \mathbf{C}_0 - 0.1\mathbf{I}_6$. t is a parameter that quantifies the amount of correlation between the noise vectors. In our numerical experiments we considered the values $t = 0.1$ (weak correlation) and $t = 0.8$ (strong correlation). $\mathbf{x} = (\mathbf{x}_0^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_{19}^T)$ with subvectors $\mathbf{x}_i$ satisfying $\|\mathbf{x}_0\| = \|\mathbf{x}_1\| = \cdots = \|\mathbf{x}_{19}\| = 5$. L is defined to be 5. We considered three estimation methods:

1. JOI - The RMSE estimator that jointly estimates $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{19}$. This estimator is of the form $\hat{\mathbf{x}} = \mathbf{G}_J \mathbf{y}$, where $\mathbf{G}_J$ is the solution to the SDP problem in Theorem 3.
2. SIN - The RMSE estimator that estimates $\mathbf{x}_i$ for every $0 \leq i \leq 19$ separately. This estimator is of the form $\hat{\mathbf{x}_i} = \mathbf{G}_S \mathbf{y}_i$, where $\mathbf{G}_S$ is the solution to the SDP problem in Theorem 1.
3. LS - The LS estimator corresponding to the system (40):

$$\mathbf{x}_{LS} = (\widetilde{\mathbf{H}}\mathbf{C}^{-1}\widetilde{\mathbf{H}})^{-1}\widetilde{\mathbf{H}}^T\mathbf{C}^{-1}\mathbf{y}.$$

We calculated the MSE for the LS, SIN and JOI estimators by averaging over 1000 realizations of the noise vector. The results are summarized in the table below

a	t	$\sigma_4(\mathbf{H}_0)$	*MSE*			*sd*			*maximum*		
			LS	SIN	JOI	LS	SIN	JOI	LS	SIN	JOI
0.8	0.1	0.61	117.6	101.6	101.2	28.6	23.3	21.8	249.1	207.5	204.3
0.4	0.1	0.25	418	257.5	252.7	143.8	50.6	41.6	1197.2	536.1	488.6
0.2	0.1	0.08	3361	484.5	450.9	1300	10	7.7	9088.4	515.2	510.2
0.8	0.8	0.61	114.3	98.1	85.6	96.89	74.68	68.42	1160.4	858	745.3
0.4	0.8	0.25	421.8	259.1	225.3	483.8	152.9	118.4	4152	1416	1181.3
0.2	0.8	0.08	2994	483.5	449.5	4112	21.1	16.3	36266	596	566

The same phenomena encountered in the single signal case is present here. Both SIN and JOI become increasingly better than the LS estimator as a **grows** and the problem becomes more ill-posed. The robustness of the RMSE estimator and the non robustness of the LS estimator are evident from the table.

Moreover, as was expected, for $t = 0.1$ (weak correlation) JOI is only slightly better than SIN and in the case $t = 0.8$, the advantage of JOI over SIN is more significant.

Appendix

A. Block Circulant Matrices and the Discrete Fourier Transform

Results on the eigen-structure of block circulant matrices proved to be essential for the study of multiple signals systems, see [2], and they are likewise essential in our study. For the sake of completeness we give in this appendix a short summary of those results that are needed to prove the main results in sections 6 and 7.

A block circulant matrix is a matrix of the form

$$\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}) \stackrel{\triangle}{=} \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_{N-1} \\ \mathbf{A}_{N-1} & \mathbf{A}_0 & \cdots & \mathbf{A}_{N-2} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_0 \end{pmatrix}, \tag{71}$$

where each submatrix $\mathbf{A}_j$ is a $k \times l$ matrix. The dimensions k and l will be clear from the context and are not therefore part of the notation.

A.1. General Properties

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

Lemma 4. *Let* $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1} \in \mathbb{R}^{k\times l}$ *and* $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1} \in \mathbb{R}^{l\times m}$ *Then,*

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

$$\mathcal{C}^T(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1}) = \mathcal{C}(\mathbf{A}_0^T, \mathbf{A}_{N-1}^T, \ldots, \mathbf{A}_1^T). \tag{72}$$

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*

$$\mathbf{C}_j = \sum_{i=0}^{N-1} \mathbf{A}_j \mathbf{B}_{j-i}, \quad 0 \le j \le N-1. \tag{73}$$

Note, that in Lemma 4, as well as throughout 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 (72) it follows that a block circulant matrix $\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ where $\mathbf{A}_0, \ldots, \mathbf{A}_{N-1} \in \mathbb{R}^{m\times m}$ is symmetric if and only if

$$\mathbf{A}_i^T = \mathbf{A}_{N-i}, \quad \forall 0 \le i \le N-1. \tag{74}$$

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

$$\mathcal{F}(\mathbf{A}) = \mathcal{C}(\widehat{\mathbf{A}}_0, \widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{N-1}), \tag{75}$$

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

$$\widehat{\mathbf{A}}_j \stackrel{\triangle}{=} \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i, \quad 0 \le j \le N-1, \tag{76}$$

and $\omega = e^{-\frac{2\pi \mathbf{i}}{N}}$ (here $\mathbf{i} = \sqrt{-1}$). In the sequel, we will also use the notation[3]

$$\mathcal{F}_j(\mathbf{A}) \stackrel{\triangle}{=} \widehat{\mathbf{A}}_j = \sum_{i=0}^{N-1} \omega^{ij} \mathbf{A}_i, \quad 0 \le j \le N-1. \tag{77}$$

$\mathcal{F}_j(\mathbf{A})$ are called the *discrete fourier components*. The inverse DFT (IDFT), denoted by $\mathcal{F}^{-1}$, is defined by:

$$\mathcal{F}^{-1}(\mathbf{A}) = (\widetilde{\mathbf{A}}_0, \widetilde{\mathbf{A}}_1, \ldots, \widetilde{\mathbf{A}}_{N-1}), \tag{78}$$

where

$$\widetilde{\mathbf{A}}_j = \frac{1}{N} \sum_{i=0}^{N-1} \omega^{-ij} \mathbf{A}_i, \quad 0 \le j \le N-1. \tag{79}$$

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

We also use the notation

$$\mathcal{F}_j^{-1}(\mathbf{A})\stackrel{\triangle}{=}\widetilde{\mathbf{A}}_j. \tag{80}$$

It is not difficult to see that for every $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1})$ we have:

$$\mathcal{F}^{-1}(\mathcal{F}(\mathbf{A})) = \mathbf{A}, \tag{81}$$

$$\mathcal{F}(\mathcal{F}^{-1}(\mathbf{A})) = \mathbf{A}. \tag{82}$$

The following properties of $\mathcal{F}_j$ are generalizations to the block circulant case of well known properties of the DFT for circulant matrices:

Lemma 5. *Let* $\mathbf{A}, \mathbf{B}$ *and* $\mathbf{C}$ *be block circulant matrices:* $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ *,* $\mathbf{B} = \mathcal{C}(\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1})$ *and* $\mathbf{C} = \mathcal{C}(\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N-1})$ *where* $\mathbf{A}_j, \mathbf{C}_j \in \mathbb{R}^{k\times l}, \mathbf{B}_j \in \mathbb{R}^{l\times m},\ 0 \le j \le N-1$. *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. $\mathcal{F}_j(\mathbf{AB}) = \mathcal{F}_j(\mathbf{A})\mathcal{F}_j(\mathbf{B})$.

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

An important special case of block circulant matrices are *elementary block circulant* matrices, which are matrices of the form:

$$\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)\stackrel{\triangle}{=}\mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_1) = \begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \cdots & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 & \cdots & \mathbf{A}_1 \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_1 & \cdots & \mathbf{A}_0 \end{pmatrix}. \tag{83}$$

In this case there are only two DFT components:

$$\mathcal{F}_0(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) = \mathbf{A}_0 + (N-1)\mathbf{A}_1, \tag{84}$$

$$\mathcal{F}_j(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) = \mathbf{A}_0 - \mathbf{A}_1, \quad 1 \le j \le N-1. \tag{85}$$

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

$$\mathcal{F}_0^{-1}(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) = \frac{1}{N}(\mathbf{A}_0 + (N-1)\mathbf{A}_1), \tag{86}$$

$$\mathcal{F}_j^{-1}(\mathcal{M}(\mathbf{A}_0, \mathbf{A}_1)) = \frac{1}{N}(\mathbf{A}_0 - \mathbf{A}_1), \quad 1 \le j \le N-1. \tag{87}$$

A.2. Eigenvalues of Symmetric Block Circulant Matrices

In this subsection, we consider the eigenvalues and eigenvectors of a symmetric block circulant matrix. We use the following notation: Let $\mathbf{A} \in \mathbb{R}^{k\times k}$ be a symmetric matrix, a matrix $\mathbf{U} \in \mathbb{R}^{k\times k}$ is called an *eigenvector matrix* of $\mathbf{A}$ if its columns are linearly independent eigenvectors of $\mathbf{A}$. Theorem 4 below shows that the eigenvalues of a block circulant matrix are exactly the eigenvalues of its discrete fourier components.

Theorem 4. *Let $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1} \in \mathbb{R}^{k\times k}$ be matrices such that $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{N-1})$ is a symmetric matrix (see (74)). For each $0 \le j \le N-1$ 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 $\omega = e^{-\frac{2\pi\mathbf{i}}{N}}$, and let $\lambda_{j,0}, \lambda_{j,1}, \dots, \lambda_{j,k-1}$ be the eigenvalues of $\mathcal{F}_j(\mathbf{A})$. Then:*

1. *An eigenvector matrix of $\mathbf{A}$ is the matrix*

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_0 & \mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_{N-1} \\ \mathbf{U}_0 & \omega\mathbf{U}_1 & \omega^2\mathbf{U}_2 & \cdots & \omega^{N-1}\mathbf{U}_{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{U}_0 & \omega^{N-1}\mathbf{U}_1 & \omega^{2(N-1)}\mathbf{U}_2 & \cdots & \omega^{(N-1)(N-1)}\mathbf{U}_{N-1} \end{pmatrix}. \tag{88}$$

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

Proof. 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,

$$\left(\sum_{i=0}^{N-1} \omega^{ij}\mathbf{A}_i\right)^* = \sum_{i=0}^{N-1} \omega^{N-ij}\mathbf{A}_i^* \overset{(74)}{=} \sum_{i=0}^{N-1} \omega^{(N-i)j}\mathbf{A}_{N-i} = \sum_{i=0}^{N-1} \omega^{ij}\mathbf{A}_i, \tag{89}$$

so that $\widehat{\mathbf{A}}_j$ is Hermitian and as a result has an eigenvector matrix, which we denote 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 $\mathbf{D} = Diag(\lambda_{j,0}, \lambda_{j,1}, \dots, \lambda_{j,k-1})$. Now,

$$\mathbf{A}\begin{pmatrix} \mathbf{U}_j \\ \omega^j\mathbf{U}_j \\ \vdots \\ \omega^{(N-1)j}\mathbf{U}_j \end{pmatrix} = \begin{pmatrix} \left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{A}_i\right)\mathbf{U}_j \\ \omega^j\left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{A}_i\right)\mathbf{U}_j \\ \vdots \\ \omega^{(N-1)j}\left(\sum_{i=0}^{N-1}\omega^{ij}\mathbf{A}_i\right)\mathbf{U}_j \end{pmatrix} = \begin{pmatrix} \mathbf{U}_j\mathbf{D} \\ \omega^j\mathbf{U}_j\mathbf{D} \\ \vdots \\ \omega^{(N-1)j}\mathbf{U}_j\mathbf{D} \end{pmatrix}, \tag{90}$$

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 $\lambda_{j,0}, \lambda_{j,1}, \dots, \lambda_{j,k-1}$.

The only fact left to prove is that the matrix (88) is invertible. Assume that

$$\begin{pmatrix} \mathbf{U}_0 & \mathbf{U}_1 & \mathbf{U}_2 & \cdot & \mathbf{U}_{N-1} \\ \mathbf{U}_0 & \omega\mathbf{U}_1 & \omega^2\mathbf{U}_2 & \cdot & \omega^{N-1}\mathbf{U}_{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{U}_0 & \omega^{N-1}\mathbf{U}_1 & \omega^{2(N-1)}\mathbf{U}_2 & \cdots & \omega^{(N-1)(N-1)}\mathbf{U}_{N-1} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{91}$$

where $\alpha_i \in \mathbb{R}^k$. Denoting $\mathbf{x}_j = \mathbf{U}_j \alpha_j$ for $0 \le j \le N-1$, (91) is equivalent to the set of equations

$$\begin{array}{llll} \mathbf{x}_0 + & \mathbf{x}_1 & + \cdots + & \mathbf{x}_{N-1} & = 0 \\ \mathbf{x}_0 + & \omega \mathbf{x}_1 & + \cdots + & \omega^{N-1}\mathbf{x}_{N-1} & = 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{x}_0 + & \omega^{N-1}\mathbf{x}_1 & + \cdots + & \omega^{(N-1)(N-1)}\mathbf{x}_{N-1} & = 0 \end{array} \tag{92}$$

Since the Fourier matrix

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & & \omega^{(N-1)(N-1)} \end{pmatrix}$$

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 $0 \le j \le N-1$. The proof is completed. □

Remark 2. A direct result of the structure (88) of the eigenvector matrix **U** is the fact that for each eigenvalue, there exists a corresponding eigenvector (a column of the matrix **U**) whose subvectors all have the same norm.

A.3. Inversion of a Block Circulant Matrix

A direct consequence of Theorem 4 is that a block circulant matrix is invertible if and only if all its discrete fourier components are invertible. This is evident from the fact that a matrix is invertible if and only if it does not have a zero eigenvalue. In this case, we can find an explicit expression for the inverse of the block circulant matrix using the properties of the DFT listed in Lemma 5.

Lemma 6. *Let* $\mathbf{A} = \mathcal{C}(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1})$ *where* $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N-1} \in \mathbb{R}^{k \times k}$. *Then* $\mathbf{A}$ *is invertible if and only if* $\mathcal{F}_j(\mathbf{A})$ *are invertible for every* $0 \le j \le N-1$. *In that case,* $\mathbf{B} = \mathbf{A}^{-1}$ *is also a block circulant matrix* $\mathbf{B} = \mathcal{C}(\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_{N-1})$ *where*

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

B. Known Results

Lemma 7 (Schur's complement). *Let*

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

be a symmetric 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*

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

Lemma 8 (S-lemma). *Let $P(\mathbf{z}) = \mathbf{z}^T\mathbf{A}\mathbf{z} + 2\mathbf{u}^T\mathbf{z} + v$ and $Q(\mathbf{z}) = \mathbf{z}^T\mathbf{B}\mathbf{z} + 2\mathbf{x}^T\mathbf{z} + y$ be two quadratic functions of $\mathbf{z}$, where $\mathbf{A}$ 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

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

References

1. Alizadeh, F.: Optimization over the positive-definite cone: interior point methods and combinatorial applications. In: Pardalos, P. (ed.), Advances in optimization and parallel computing. North-Holland, the Netherlands, 1992
2. Beck, A., Eldar, Y.C., Ben-Tal, A.: Minimax mean-squared error estimation of multichannel signals in the presence of bounded data uncertainties. submitted to *SIAM J. Matrix Anal. Appl.*
3. Ben-Haim, Z., Eldar, Y.C.: Maximum set estimators with bounded estimation error. submitted to *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3172–3182, Aug. 2005
4. Ben-Haim, Z., Eldar, Y.C.: Minimax estimators dominating the least-squares estimator. submitted to *Int. Conf. Acoust., Speech, Signal Processing (ICASSP-2005)*, pp. 53–56, Mar. 2005
5. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23** (4), 769–805 (1998)
6. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization. MPS-SIAM Series on Optimization, 2001
7. Calafiore, G., El Ghaoui, L.: Robust maximum likelihood estimation in the linear model. Automatica **37**, 573–580 (2001)
8. Chandrasekaran, S., Golub, G.H., Gu, M., Sayed, A.H.: Parameter estimation in the presence of bounded data uncertainties. SIAM J. Matrix Anal. Appl. **19** (1), 235–252 (1998)
9. Chandrasekaran, S., Gu, M., Sayed, A.H., Schubert, K.E.: The degenerate bounded error-in-variables model. SIAM J. Matrix Anal. Appl. **23** (1), 138–166 (2001)
10. Eldar, Y.C.: Minimax MSE estimation of deterministic parameters with noise covariance uncertainties. submitted to *IEEE Trans. Signal Processing*, to appear
11. Eldar, Y.C., Ben-Tal, A., Nemirovski, A.: Robust mean-squared error estimation in the presence of model uncertainties. IEEE Trans. Signal Processing, vol. 53, pp. 168–181, Jan. 2005
12. Eldar, Y.C., Ben-Tal, A., Nemirovski, A.: Linear minimax regret estimation of deterministic parameters with bounded data uncertainties. IEEE Trans. Signal Processing **52**, 2177–2188 (2004)
13. Eldar, Y.C., Merhav, N.: Robust linear estimation under a minimax mse–ratio criterion. Technical Report CCIT 449, Technion–Israel Institute of Technology, Haifa, Israel, Oct. 2003. submitted to *IEEE Trans. Signal Processing*, vol. 52, pp. 2177–2188, Aug. 2004
14. Eldar, Y.C., Merhav, N.: A competitive minimax approach to robust estimation in linear models. IEEE Trans. Signal Processing **52**, 1931–1946 (2004)
15. Eldar, Y.C., Oppenheim, A.V.: Covariance shaping least-squares estimation. IEEE Trans. Signal Processing **51**, 686–697 (2003)
16. Gauss, K.G.: Theory of motion of heavenly bodies. Dover, New York, NY, 1963
17. El Ghaoui, L., Lebret, H.: Robust solution to least-squares problems with uncertain data. SIAM J. Matrix Anal. Appl. **18** (4), 1035–1064 (1997)
18. Golub, G.H., Van Loan, C.F.: An analysis of the total least-squares problem. SIAM J. Numer. Anal. **17** (6), 883–893 (1980)
19. Golub, G.H., Van Loan, C.F.: Matrix computations. Baltimore MD: Johns Hopkins Univ. Press, 3rd edn., 1996
20. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics **12**, 55–67 (1970)
21. Van Huffel, S., Vandewalle, J.: The total least-squares problem: computational aspects and analysis. Volume 9 of *Frontier in Applied Mathematics*. Philadelphia, PA: SIAM, 1991
22. Kailath, T.: Lectures On linear least-squares estimation. Springer, Wein, New York, 1976
23. Kay, S.M.: Fundamentals of statistical signal processing: estimation theory. Prentice Hall, Inc., Upper Saddle River, NJ, 1993

24. Mayer, L.S., Willke, T.A.: On biased estimation in linear models. Technometrics, **15**, 497–508 (1973)
25. Nesterov, Y., Nemirovski, A.: Interior-point polynomial algorithms in convex programming. SIAM, Philadelphia, PE, 1994
26. Sayed, A.H., Nascimento, V., Chandrasekaran, S.: Estimation and control with bounded data uncertainties. Linear Alg. Appl. **284** (1), 259–306 (1998)
27. Strum, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optimization methods and software, **11–12**, 625–653 (1999)
28. Therrien, C.W.: Discrete random signals and statistical signal processing. Prentice Hall, Inc., Englewood Cliffs, NJ, 1992
29. Tikhonov, A.N., Arsenin, V.Y.: Solution of Ill-posed problems. V.H. Winston, Washington, DC, 1977
30. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38** (1), 40–95 (1996)