# MINIMAX SUBSPACE SAMPLING IN THE PRESENCE OF NOISE

*Yonina C. Eldar*

Technion–Israel Institute of Technology
Haifa, 32000, Israel
yonina@ee.technion.ac.il

## ABSTRACT

We treat the problem of reconstructing a signal $\boldsymbol{x}$ that lies in a subspace $\mathcal{W}$ from its noisy samples. The samples are modelled as the inner products of $\boldsymbol{x}$ with a set of sampling vectors that span a subspace $\mathcal{S}$, not necessarily equal to $\mathcal{W}$. We consider two approaches to reconstructing $\boldsymbol{x}$ from the noisy samples: a least-squares (LS) method and a minimax mean-squared error (MSE) strategy. We show that if the elements of $\boldsymbol{x}$ are finite, then the minimax MSE approach results in a smaller MSE than the LS approach for all values of $\boldsymbol{x}$. We then generalize the results to the problem of minimizing an inner-product MSE.

## 1. INTRODUCTION

Signal expansions, in which a signal is represented by a set of coefficients, find many applications in signal processing and communications. The expansion coefficients can be regarded as generalized samples of the signal, so that the signal expansion problem can be formulated as a sampling and reconstruction problem.

In a linear signal expansion, the coefficients, or samples, are given by inner products of the signal, denoted by $\boldsymbol{x}$, with sampling vectors $\{\boldsymbol{s}_i, 1 \leq i \leq m\}$. The signal is assumed to lie in arbitrary Hilbert space $\mathcal{H}$, and the sampling vectors span a subspace $\mathcal{S} \subseteq \mathcal{H}$, which we refer to as the sampling space. The problem then is to reconstruct the signal from the samples, using a given set of reconstruction vectors $\{\boldsymbol{w}_i, 1 \leq i \leq m\}$, which span the reconstruction space $\mathcal{W}$. For simplicity, in this paper, we assume that both sets of vectors $\{\boldsymbol{s}_i\}$ and $\{\boldsymbol{w}_i\}$ are linearly independent. We further assume that $\mathcal{H} = \mathbb{C}^n$; the results extend straightforwardly to arbitrary Hilbert spaces.

The sampling framework we consider here, in which we allow for arbitrary sampling and reconstruction spaces, was first considered in [1] for shift-invariant spaces of $L_2$, and later extended in [2, 3, 4, 5] to arbitrary Hilbert spaces. It was shown in [2, 3, 4] that if $\boldsymbol{x} \in \mathcal{W}$, then $\boldsymbol{x}$ can be perfectly reconstructed from the given samples, as long as $\mathcal{H} = \mathcal{W} \oplus \mathcal{S}^\perp$, which, in the finite-dimensional case we consider here, is equivalent to the condition that $\mathcal{W} \cap \mathcal{S}^\perp = \{0\}$. In the more general case in which $\boldsymbol{x} \in \mathcal{H}$, the signal can no longer be reconstructed using only vectors in $\mathcal{W}$. Two possible approaches to reconstruction in this case are consistent reconstruction, first proposed in [1], and minimum squared-norm reconstruction [5].

In [1, 2, 3, 4, 5] it is assumed that the samples are noise-free. Here, we consider the case in which the samples are contaminated by noise, and $\boldsymbol{x}$ lies in $\mathcal{W}$. Our problem is to reconstruct $\boldsymbol{x}$ from the noisy samples such that the reconstructed signal $\hat{\boldsymbol{x}}$ is close to $\boldsymbol{x}$ in some sense.

Sampling in the presence of noise has been investigated for specific classes of signals in [6, 7] (see also references therein). However, the methods proposed are based on a heuristic moving average estimator, that does not have any assertion of optimality, and is not applicable to the general setup we treat here.

The most straightforward approach to reconstructing $\boldsymbol{x}$, is to design $\hat{\boldsymbol{x}}$ to minimize the least-squares (LS) error. This method is developed in Section 2. However, minimizing the LS error does not guarantee that $\hat{\boldsymbol{x}}$ is close to $\boldsymbol{x}$. We may instead consider minimizing the mean-squared error (MSE) between $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}$. Unfortunately, the MSE depends explicitly on $\boldsymbol{x}$ and therefore cannot be minimized directly. Thus, instead, we aim at designing a robust reconstruction whose performance in terms of MSE is good across all possible values of $\boldsymbol{x}$. To this end we rely on a recent minimax estimation framework that has been developed for solving robust estimation problems [8], in which a linear estimator is designed to minimize the worst-case MSE over all bounded norm parameter vectors. Using this framework we develop, in Section 3, a robust reconstruction, and show that the resulting MSE is smaller than the MSE of the LS estimator, for all bounded vectors $\boldsymbol{x}$.

In Section 4 we consider minimizing the expected value of the inner product $|(\hat{\boldsymbol{x}} - \boldsymbol{x})^* \boldsymbol{h}|^2$ for any vector $\boldsymbol{h}$. Here again we consider both a LS approach and a minimax MSE approach, and show that the minimax MSE strategy leads to an estimator whose inner-product MSE (IPMSE) is smaller than that of the LS estimator for all bounded values of $\boldsymbol{x}$.

## 2. LEAST-SQUARES RECONSTRUCTION

Suppose we are given noisy samples of a signal $\boldsymbol{x} \in \mathbb{C}^n$. The noise-free samples are modelled as $\{c_i = \boldsymbol{s}_i^* \boldsymbol{x}, 1 \leq i \leq m\}$, where $\{\boldsymbol{s}_i, 1 \leq i \leq m\}$ are the sampling vectors and are assumed to be linearly independent, and $(\cdot)^*$ is the Hermitian conjugate. Denoting by $\boldsymbol{S}$ the matrix of columns $\boldsymbol{s}_i$, the noisy samples are

$$\boldsymbol{y} = \boldsymbol{S}^* \boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n}$ is a zero-mean noise vector with covariance matrix $\boldsymbol{C}$. We construct an approximation $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$ using a given set of linearly independent reconstruction vectors $\{\boldsymbol{w}_i, 1 \leq i \leq m\}$, that are the columns of the matrix $\boldsymbol{W}$. Thus, the reconstruction $\hat{\boldsymbol{x}}$ has the form

$$\hat{\boldsymbol{x}} = \sum_{i=1}^{m} d_i \boldsymbol{w}_i = \boldsymbol{W}\boldsymbol{d}, \tag{2}$$

for some coefficients $\{d_i\}$ that are a linear transformation of the samples $\{c_i\}$, so that $\boldsymbol{d} = \boldsymbol{G}\boldsymbol{c}$ for some $m \times m$ matrix $\boldsymbol{G}$. The sampling and reconstruction scheme is illustrated in Fig. 1.

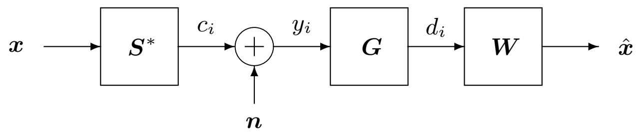


**Fig. 1**. General sampling and reconstruction scheme.

Denoting by $\mathcal{W}$ the reconstruction space spanned by the reconstruction vectors $\{\boldsymbol{w}_i\}$, it follows from (2) that $\hat{\boldsymbol{x}} \in \mathcal{W}$. Thus, if $\boldsymbol{x}$ lies out of $\mathcal{W}$, then it cannot be perfectly reconstructed from the samples $c_i$, even in the noise-free case $\boldsymbol{n} = 0$. If, on the other hand, $\boldsymbol{x} \in \mathcal{W}$, then it was shown in [2, 3] that $\boldsymbol{x}$ can be perfectly reconstructed from the samples $c_i$ as long as $\mathcal{W} \cap \mathcal{S}^{\perp} = \{0\}$, where $\mathcal{S}^{\perp}$ is the orthogonal complement of $\mathcal{S}$ in $\mathcal{H}$. In this case perfect reconstruction can be obtained by choosing

$$\boldsymbol{G} = (\boldsymbol{S}^*\boldsymbol{W})^{-1}, \tag{3}$$

which results in

$$\hat{\boldsymbol{x}} = \boldsymbol{W}(\boldsymbol{S}^*\boldsymbol{W})^{-1}\boldsymbol{S}^*\boldsymbol{x} = \boldsymbol{E}_{\mathcal{W}\mathcal{S}^{\perp}}\boldsymbol{x}, \tag{4}$$

where $\boldsymbol{E}_{\mathcal{W}\mathcal{S}^{\perp}}$ is the *oblique projection* onto $\mathcal{W}$ along $\mathcal{S}^{\perp}$, and is the unique operator satisfying that $\boldsymbol{E}_{\mathcal{W}\mathcal{S}^{\perp}}\boldsymbol{w} = \boldsymbol{w}$ for any $\boldsymbol{w} \in \mathcal{W}$, and $\boldsymbol{E}_{\mathcal{W}\mathcal{S}^{\perp}}\boldsymbol{s} = 0$ for any $\boldsymbol{s} \in \mathcal{S}^{\perp}$. Clearly, if $\boldsymbol{x}$ is in $\mathcal{W}$, then $\hat{\boldsymbol{x}}$ given by (4) is equal to $\boldsymbol{x}$. We note that, as shown in [2], if $\mathcal{W} \cap \mathcal{S}^{\perp} = \{0\}$, then $\boldsymbol{S}^*\boldsymbol{W}$ is invertible.

When the samples are corrupted by noise, $\boldsymbol{G}$ given by (3) no longer guarantees perfect reconstruction of $\boldsymbol{x} \in \mathcal{W}$. Thus, our problem is to design $\boldsymbol{G}$ such that $\hat{\boldsymbol{x}}$ is close to $\boldsymbol{x}$, for any $\boldsymbol{x} \in \mathcal{W}$.

The most straightforward approach is to design $\boldsymbol{G}$ to minimize the (weighted) LS data error, which is given by

$$\begin{aligned} \epsilon_{\text{LS}} &= (\boldsymbol{S}^*\hat{\boldsymbol{x}} - \boldsymbol{y})^* \boldsymbol{C}^{-1} (\boldsymbol{S}^*\hat{\boldsymbol{x}} - \boldsymbol{y}) \\ &= (\boldsymbol{S}^*\boldsymbol{W}\boldsymbol{G}\boldsymbol{y} - \boldsymbol{y})^* \boldsymbol{C}^{-1} (\boldsymbol{S}^*\boldsymbol{W}\boldsymbol{G}\boldsymbol{y} - \boldsymbol{y}). \end{aligned} \tag{5}$$

Clearly, $\epsilon_{\text{LS}}$ is minimized with $\boldsymbol{G}_{\text{LS}} = (\boldsymbol{S}^*\boldsymbol{W})^{-1}$. The LS estimator is then

$$\hat{\boldsymbol{x}}_{\text{LS}} = \boldsymbol{W}\boldsymbol{G}_{\text{LS}}\boldsymbol{y} = \boldsymbol{W}(\boldsymbol{S}^*\boldsymbol{W})^{-1}\boldsymbol{y}. \tag{6}$$

Since $\boldsymbol{y} = \boldsymbol{S}^*\boldsymbol{x} + \boldsymbol{n}$,

$$\hat{\boldsymbol{x}}_{\text{LS}} = \boldsymbol{E}_{\mathcal{W}\mathcal{S}^{\perp}}\boldsymbol{x} + \boldsymbol{W}(\boldsymbol{S}^*\boldsymbol{W})^{-1}\boldsymbol{n} = \boldsymbol{x} + \boldsymbol{W}(\boldsymbol{S}^*\boldsymbol{W})^{-1}\boldsymbol{n}. \tag{7}$$

Although the LS approach is a very popular estimation method, in some cases it can result in a large MSE, which is a measure of the average estimation error. In the next section we show that if the elements of $\boldsymbol{x}$ are finite, then we can design an estimator with smaller MSE than that of the LS estimator, for all values of $\boldsymbol{x}$.

## 3. MINIMAX MSE RECONSTRUCTION

In practice, the elements of $\boldsymbol{x}$ are typically finite, so that the squared-norm $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^*\boldsymbol{x}$ can be bounded by a constant $U^2$. We now show that if $\|\boldsymbol{x}\| \leq U$, then there exists an estimator with MSE that is always smaller than that of the LS estimator.

To develop an estimator with small MSE we may seek the transformation $\boldsymbol{G}$ that minimizes the MSE between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. However, computing the MSE $E\left\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right\}$ shows that

$$\begin{aligned} E\left\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right\} = \boldsymbol{x}^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{x} \\ + \text{Tr}(\boldsymbol{W}\boldsymbol{G}\boldsymbol{C}\boldsymbol{G}^*\boldsymbol{W}^*), \end{aligned} \tag{8}$$

which depends explicitly on $\boldsymbol{x}$ and therefore cannot be minimized.

Following the minimax MSE approach of [8], we consider instead minimizing the worst-case MSE over all bounded vectors $\boldsymbol{x}$, which leads to the problem

$$\begin{aligned} \min_{\boldsymbol{G}} \{ &\text{Tr}(\boldsymbol{W}\boldsymbol{G}\boldsymbol{C}\boldsymbol{G}^*\boldsymbol{W}^*) + \\ &\max_{\|\boldsymbol{x}\| \leq U, \boldsymbol{x} \in \mathcal{W}} \boldsymbol{x}^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{x}\}. \end{aligned} \tag{9}$$

To develop a solution to (9), we first treat the inner maximization:

$$\begin{aligned} &\max_{\|\boldsymbol{x}\| \leq U, \boldsymbol{x} \in \mathcal{W}} \boldsymbol{x}^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{x} = \\ &= \max_{\|\boldsymbol{x}\| \leq U} \boldsymbol{x}^*\boldsymbol{P}_{\mathcal{W}}(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{P}_{\mathcal{W}}\boldsymbol{x} \\ &= U^2\lambda_{\max}\left(\boldsymbol{P}_{\mathcal{W}}(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{P}_{\mathcal{W}}\right), \end{aligned} \tag{10}$$

where $\boldsymbol{P}_{\mathcal{W}}$ is the orthogonal projection onto $\mathcal{W}$, and $\lambda_{\max}(\boldsymbol{A})$ denotes the largest eigenvalue of $\boldsymbol{A}$. Thus, problem (9) becomes

$$\begin{aligned} \min_{\boldsymbol{G}} \{ &\text{Tr}(\boldsymbol{W}\boldsymbol{G}\boldsymbol{C}\boldsymbol{G}^*\boldsymbol{W}^*) + \\ &+ U^2\lambda_{\max}\left(\boldsymbol{P}_{\mathcal{W}}(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)^*(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*)\boldsymbol{P}_{\mathcal{W}}\right)\}. \end{aligned} \tag{11}$$

Let $\boldsymbol{W}$ have an SVD $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{Z}\Sigma\boldsymbol{V}^*$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices of size $n$ and $m$ respectively, $\Sigma$ is a size $m$ diagonal matrix with positive diagonal elements, and $\boldsymbol{Z}$ is the $n \times m$ matrix defined by $\boldsymbol{Z} = [\boldsymbol{I}_m \ \ 0]^*$. Then $\boldsymbol{P}_{\mathcal{W}} = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{U}^*$, and

$$\begin{aligned} &\lambda_{\max}\left((\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}})^*(\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}})\right) = \\ &= \lambda_{\max}\left((\boldsymbol{I} - \Sigma\boldsymbol{V}^*\boldsymbol{G}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z})^*(\boldsymbol{I} - \Sigma\boldsymbol{V}^*\boldsymbol{G}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z})\right), \end{aligned} \tag{12}$$

where we used the facts that $\boldsymbol{Z}^*\boldsymbol{Z} = \boldsymbol{I}$ and for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\lambda_{\max}(\boldsymbol{A}\boldsymbol{B}) = \lambda_{\max}(\boldsymbol{B}\boldsymbol{A})$. Defining $\boldsymbol{H} = \boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z}$, and $\boldsymbol{M} = \Sigma\boldsymbol{V}^*\boldsymbol{G}$, our problem becomes

$$\min_{\boldsymbol{M}} \left\{\text{Tr}(\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^*) + U^2\lambda_{\max}\left((\boldsymbol{I} - \boldsymbol{M}\boldsymbol{H})^*(\boldsymbol{I} - \boldsymbol{M}\boldsymbol{H})\right)\right\}. \tag{13}$$

Once we find $\boldsymbol{M}$, the optimal value of $\boldsymbol{G}$ is given by

$$\boldsymbol{G} = \boldsymbol{V}\Sigma^{-1}\boldsymbol{M}. \tag{14}$$

The problem of (13) is equal to the minimax MSE problem of [8], the solution of which is incorporated in the following theorem:

**Theorem 1.** *[8] Let $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{H}$ is an $n \times m$ matrix with rank $m$, and $\boldsymbol{n}$ is a zero-mean random vector with covariance $\boldsymbol{C}$, and let $\gamma = Tr((\boldsymbol{H}^*\boldsymbol{C}^{-1}\boldsymbol{H})^{-1})$. Then the solution to*

$$\begin{aligned} &\min_{\hat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y}} \max_{\|\boldsymbol{x}\| \leq U} E\left\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right\} = \\ &= \min_{\boldsymbol{M}} \left\{\lambda_{\max}\left((\boldsymbol{I} - \boldsymbol{M}\boldsymbol{H})^*(\boldsymbol{I} - \boldsymbol{M}\boldsymbol{H})\right) + Tr\left(\boldsymbol{M}^*\boldsymbol{C}\boldsymbol{M}\right)\right\} \end{aligned}$$

*is $\boldsymbol{M} = [U^2/(U^2 + \gamma)](\boldsymbol{H}^*\boldsymbol{C}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{C}^{-1}$.*

Since the columns of $\boldsymbol{U}\boldsymbol{Z}$ span $\mathcal{W}$, it follows from [2] that $\boldsymbol{H}$ is invertible. Therefore, from Theorem 1 the optimal $\boldsymbol{M}$ is

$$\boldsymbol{M} = \alpha(\boldsymbol{H}^*\boldsymbol{C}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^*\boldsymbol{C}^{-1} = \alpha\boldsymbol{H}^{-1}, \tag{15}$$

where

$$\alpha = \frac{U^2}{U^2 + \text{Tr}\left((\boldsymbol{H}\boldsymbol{H}^*)^{-1}\boldsymbol{C}\right)} = \frac{U^2}{U^2 + \text{Tr}\left((\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S})^{-1}\boldsymbol{C}\right)}. \tag{16}$$

The minimax MSE matrix $\boldsymbol{G}_{\mathrm{MX}}$ is then given from (14) as

$$\boldsymbol{G}_{\mathrm{MX}}=\alpha\left(\boldsymbol{S}^{*}\boldsymbol{U}\boldsymbol{Z}\Sigma\boldsymbol{V}^{*}\right)^{-1}=\alpha\left(\boldsymbol{S}^{*}\boldsymbol{W}\right)^{-1}, \qquad (17)$$

and the minimax MSE reconstruction, denoted $\hat{\boldsymbol{x}}_{\mathrm{MX}}$, is

$$\hat{\boldsymbol{x}}_{\mathrm{MX}}=\boldsymbol{W}\boldsymbol{G}_{\mathrm{MX}}\boldsymbol{y}=\alpha\boldsymbol{W}\left(\boldsymbol{S}^{*}\boldsymbol{W}\right)^{-1}\boldsymbol{y}. \qquad (18)$$

Comparing (18) with (6), it follows that $\hat{\boldsymbol{x}}_{\mathrm{MX}}=\alpha\hat{\boldsymbol{x}}_{\mathrm{LS}}$. Thus,

$$\hat{\boldsymbol{x}}_{\mathrm{MX}}=\alpha\boldsymbol{x}+\alpha\boldsymbol{W}\left(\boldsymbol{S}^{*}\boldsymbol{W}\right)^{-1}\boldsymbol{n}. \qquad (19)$$

### 3.1. MSE Performance

We now compare the MSE $E_{\mathrm{LS}}=E\left\{\|\boldsymbol{x}-\hat{\boldsymbol{x}}_{\mathrm{LS}}\|^2\right\}$ of the LS estimator with the MSE $E_{\mathrm{MX}}=E\left\{\|\boldsymbol{x}-\hat{\boldsymbol{x}}_{\mathrm{MX}}\|^2\right\}$ of the minimax MSE estimator. From (7),

$$E_{\mathrm{LS}}=E\left\{\|\boldsymbol{W}(\boldsymbol{S}^{*}\boldsymbol{W})^{-1}\boldsymbol{n}\|^2\right\}=\mathrm{Tr}\left((\boldsymbol{S}^{*}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S})^{-1}\boldsymbol{C}\right)\stackrel{\triangle}{=}\gamma, \qquad (20)$$

where we used the fact that

$$\begin{aligned}(\boldsymbol{W}^{*}\boldsymbol{S})^{-1}\boldsymbol{W}^{*}\boldsymbol{W}(\boldsymbol{S}^{*}\boldsymbol{W})^{-1}&=\\ &=\left(\boldsymbol{S}^{*}\boldsymbol{W}(\boldsymbol{W}^{*}\boldsymbol{W})^{-1}\boldsymbol{W}^{*}\boldsymbol{S}\right)^{-1}\\ &=\left(\boldsymbol{S}^{*}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\right)^{-1}. \qquad (21)\end{aligned}$$

Using (19), the MSE $E_{\mathrm{MX}}$ is

$$E_{\mathrm{MX}}=(1-\alpha)^2\|\boldsymbol{x}\|^2+\alpha^2\gamma. \qquad (22)$$

Since $\|\boldsymbol{x}\|\leq U$, we have that

$$E_{\mathrm{MX}}\leq(1-\alpha)^2U^2+\alpha^2\gamma=\frac{U^2\gamma}{\gamma+U^2}\leq\gamma=E_{\mathrm{LS}}, \qquad (23)$$

where we used the fact that from (16) and (20), $\alpha=U^2/(U^2+\gamma)$. Thus, $E_{\mathrm{MX}}\leq E_{\mathrm{LS}}$ for all $\boldsymbol{x}\in\mathcal{W}$ such that $\|\boldsymbol{x}\|\leq U$.

### 3.2. Tikhonov Reconstruction

The LS approach does not account for the fact that $\|\boldsymbol{x}\|\leq U$. To take advantage of this information, we may seek the reconstruction $\hat{\boldsymbol{x}}=\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}$ which minimizes the LS error subject to $\|\hat{\boldsymbol{x}}\|\leq U$. Thus, $\hat{\boldsymbol{x}}$ can be determined by minimizing the Lagrangian

$$\left(\boldsymbol{S}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{y}\right)^{*}\boldsymbol{C}^{-1}\left(\boldsymbol{S}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{y}\right)+\lambda\boldsymbol{y}^{*}\boldsymbol{G}^{*}\boldsymbol{W}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}, \qquad (24)$$

where from the Karush-Kuhn-Tucker conditions $\lambda\geq 0$ and

$$\lambda(\boldsymbol{y}^{*}\boldsymbol{G}^{*}\boldsymbol{W}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-U^2)=0. \qquad (25)$$

Differentiating (24) with respect to $\boldsymbol{G}$ and equating to 0,

$$\boldsymbol{W}^{*}\boldsymbol{S}\boldsymbol{C}^{-1}\left(\boldsymbol{S}^{*}\boldsymbol{W}\boldsymbol{G}-\boldsymbol{I}\right)\boldsymbol{y}\boldsymbol{y}^{*}+\lambda\boldsymbol{W}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}\boldsymbol{y}^{*}=0, \qquad (26)$$

which is satisfied for all $\boldsymbol{y}$ if

$$\boldsymbol{G}=\left(\boldsymbol{W}^{*}\left(\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^{*}+\lambda\boldsymbol{I}\right)\boldsymbol{W}\right)^{-1}\boldsymbol{W}^{*}\boldsymbol{S}\boldsymbol{C}^{-1}. \qquad (27)$$

Using some algebraic manipulations, $\boldsymbol{G}$ of (27) can be written as

$$\boldsymbol{G}_{\mathrm{TIK}}=(\boldsymbol{W}^{*}\boldsymbol{W})^{-1}\boldsymbol{W}^{*}\boldsymbol{S}\left(\lambda\boldsymbol{C}+\boldsymbol{S}^{*}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\right)^{-1}. \qquad (28)$$

If $\lambda=0$, then $\boldsymbol{G}_{\mathrm{TIK}}=(\boldsymbol{S}^{*}\boldsymbol{W})^{-1}=\boldsymbol{G}_{\mathrm{LS}}$. This solution is valid only if $\|\boldsymbol{W}\boldsymbol{G}_{\mathrm{LS}}\boldsymbol{y}\|\leq U$. Otherwise, $\lambda>0$, and to satisfy (25) $\lambda$ must be chosen such that $\|\boldsymbol{W}\boldsymbol{G}_{\mathrm{TIK}}\boldsymbol{y}\|=U$. To show that such a $\lambda$ always exists, define $\mathcal{G}(\lambda)=\boldsymbol{y}^{*}\boldsymbol{G}_{\mathrm{TIK}}^{*}\boldsymbol{W}^{*}\boldsymbol{W}\boldsymbol{G}_{\mathrm{TIK}}\boldsymbol{y}-U^2$, so that $\lambda$ is a positive root of $\mathcal{G}(\lambda)$. Clearly, $\mathcal{G}(\lambda)$ is monotonically decreasing in $\lambda$. In addition, $\mathcal{G}(0)>0$, and $\mathcal{G}(\lambda)\rightarrow -U^2$ as $\lambda\rightarrow\infty$, therefore $\mathcal{G}(\lambda)$ has exactly one positive root.

We see that in general the Tikhonov reconstruction is *nonlinear*, and does not have an explicit solution; the parameter $\lambda$ does not have a closed form, but is rather determined as the solution of a data-dependent, nonlinear equation.

## 4. MINIMAX INNER-PRODUCT MSE

Suppose now that instead of seeking to minimize the total MSE $E\left\{\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|^2\right\}=\sum_{i=1}^{n}E\left\{|\hat{\boldsymbol{x}}_i-\boldsymbol{x}_i|^2\right\}$, we wish to minimize the MSE of the $i$th component

$$E\left\{|\hat{\boldsymbol{x}}_i-\boldsymbol{x}_i|^2\right\}=E\left\{|(\hat{\boldsymbol{x}}-\boldsymbol{x})^{*}\boldsymbol{e}_i|^2\right\}, \qquad (29)$$

where $\boldsymbol{e}_i$ is the $i$th indicator vector, with components all equal zero besides the $i$th component which is equal to 1. More generally, we can consider the IPMSE

$$E_{\boldsymbol{h}}=E\left\{|(\hat{\boldsymbol{x}}-\boldsymbol{x})^{*}\boldsymbol{h}|^2\right\}=E\left\{|(\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{x})^{*}\boldsymbol{h}|^2\right\}, \qquad (30)$$

for any nonzero vector $\boldsymbol{h}\in\mathbb{C}^n$.

One approach is to design a LS estimator, which minimizes

$$\epsilon_{\mathrm{LS}}(\boldsymbol{h}) \;=\; \|\left(\boldsymbol{S}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{y}\right)^{*}\boldsymbol{C}^{-1/2}\boldsymbol{h}\|^2. \qquad (31)$$

Clearly the error of (31) is minimized with $\boldsymbol{G}=(\boldsymbol{S}^{*}\boldsymbol{W})^{-1}$, which results in $\hat{\boldsymbol{x}}=\hat{\boldsymbol{x}}_{\mathrm{LS}}$ of (6).

We now show, that as in the case of minimizing the total MSE, we can develop an estimator with smaller IPMSE than the LS estimator for all bounded-norm vectors $\boldsymbol{x}$ and all choices of $\boldsymbol{h}$ by minimizing a worst-case IPMSE.

Specifically, we now derive the estimator that minimizes the worst-case MSE $E_{\boldsymbol{h}}$ over all bounded norm vectors $\|\boldsymbol{x}\|\leq U$ in $\mathcal{W}$. Thus, we seek the matrix $\boldsymbol{G}$ that is the solution to the problem

$$\min_{\hat{\boldsymbol{x}}=\boldsymbol{G}\boldsymbol{y}}\max_{\boldsymbol{x}\in\mathcal{W},\|\boldsymbol{x}\|\leq U}E\left\{|(\hat{\boldsymbol{x}}-\boldsymbol{x})^{*}\boldsymbol{h}|^2\right\}. \qquad (32)$$

To develop a solution to (32), we note that

$$\begin{aligned}&E\left\{|(\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{x})^{*}\boldsymbol{h}|^2\right\}=\\ &=\;\boldsymbol{h}^{*}E\left\{(\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{x})(\boldsymbol{W}\boldsymbol{G}\boldsymbol{y}-\boldsymbol{x})^{*}\right\}\boldsymbol{h}\\ &=\;|\boldsymbol{x}^{*}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)^{*}\boldsymbol{h}|^2+\boldsymbol{h}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{C}\boldsymbol{G}^{*}\boldsymbol{W}^{*}\boldsymbol{h}. \qquad (33)\end{aligned}$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned}&\max_{\boldsymbol{x}\in\mathcal{W},\|\boldsymbol{x}\|\leq U}|\boldsymbol{x}^{*}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)^{*}\boldsymbol{h}|^2=\\ &=\;\max_{\|\boldsymbol{x}\|\leq U}|\boldsymbol{x}^{*}\boldsymbol{P}_{\mathcal{W}}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)^{*}\boldsymbol{h}|^2\\ &=\;U^2\boldsymbol{h}^{*}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)\boldsymbol{P}_{\mathcal{W}}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)^{*}\boldsymbol{h}. \qquad (34)\end{aligned}$$

Substituting (34) into (33), our problem becomes

$$\min_{\boldsymbol{G}}\left\{U^2|\boldsymbol{P}_{\mathcal{W}}\left(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{G}\boldsymbol{S}^{*}\right)^{*}\boldsymbol{h}|^2+\boldsymbol{h}^{*}\boldsymbol{W}\boldsymbol{G}\boldsymbol{C}\boldsymbol{G}^{*}\boldsymbol{W}^{*}\boldsymbol{h}\right\}. \qquad (35)$$

Since the objective in (35) is convex, the optimal $\boldsymbol{G}$ can be found by setting the derivative to zero, resulting in

$$\boldsymbol{W}\boldsymbol{h}\boldsymbol{h}^{*}\boldsymbol{W}\boldsymbol{G}\left(\boldsymbol{S}^{*}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}+(1/U^2)\boldsymbol{C}\right)=\boldsymbol{W}\boldsymbol{h}\boldsymbol{h}^{*}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}, \qquad (36)$$

which is satisfied for any choice of $\boldsymbol{h}$ if

$$\boldsymbol{G}_{\mathrm{IPMX}} = (\boldsymbol{W}^*\boldsymbol{W})^{-1}\boldsymbol{W}^*\boldsymbol{S}\left(\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S} + (1/U^2)\boldsymbol{C}\right)^{-1}. \quad (37)$$

The minimax IPMSE estimator is then given by

$$\hat{\boldsymbol{x}}_{\mathrm{IPMX}} = \boldsymbol{W}\boldsymbol{G}\boldsymbol{y} = \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\left(\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S} + (1/U^2)\boldsymbol{C}\right)^{-1}\boldsymbol{y}. \quad (38)$$

Comparing (37) with (28), it follows that the minimax IPMSE estimator has the same from as the Tikhonov estimator, where the nonlinear parameter $\lambda$ in the Tikhonov estimator is replaced by the constant $1/U^2$.

### 4.1. Inner-Product MSE Performance

We now show that the IPMSE using the estimator of (38) is smaller than that of the LS estimator, for all choices of $\boldsymbol{h}$ and $\|\boldsymbol{x}\| \leq U$.

The IPMSE when using the LS estimator is

$$E_{\mathrm{LS}}(\boldsymbol{h}) = \boldsymbol{h}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}(\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S})^{-1}\boldsymbol{C}(\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S})^{-1}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{h}. \quad (39)$$

Since $\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S} \in \mathcal{W}$, we have that $\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S} = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{X}$ for some $m \times m$ invertible matrix $\boldsymbol{X}$, so that

$$E_{\mathrm{LS}}(\boldsymbol{h}) = \boldsymbol{h}^*\boldsymbol{U}\boldsymbol{Z}\left(\boldsymbol{Z}^*\boldsymbol{U}^*\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^*\boldsymbol{U}^*\boldsymbol{h}. \quad (40)$$

The IPMSE when using the minimax MSE estimator is

$$\begin{aligned} E_{\mathrm{MX}}(\boldsymbol{h}) = &|\boldsymbol{x}^*(\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}})\boldsymbol{h}|^2 + \\ &+\boldsymbol{h}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{h}, \end{aligned} \quad (41)$$

where we used the fact that $\boldsymbol{x} = \boldsymbol{P}_{\mathcal{W}}\boldsymbol{x}$, and defined

$$\boldsymbol{B} = (\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S} + (1/U^2)\boldsymbol{C})^{-1}. \quad (42)$$

Now, for any $\|\boldsymbol{x}\| \leq U$, we have that

$$\begin{aligned} &|\boldsymbol{x}^*(\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}})\boldsymbol{h}|^2 \leq \\ &\leq \quad U^2\boldsymbol{h}^*(\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}})^2\boldsymbol{h} \\ &= \quad U^2\boldsymbol{h}^*\left(\boldsymbol{P}_{\mathcal{W}} - 2\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}} + \boldsymbol{T}\right)\boldsymbol{h}, \end{aligned} \quad (43)$$

where $\boldsymbol{T} = \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}$. Noting that

$$U^2\boldsymbol{T} + \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}} = U^2\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}, \quad (44)$$

and substituting (44) into (43), we have that

$$\begin{aligned} E_{\mathrm{MX}}(\boldsymbol{h}) &\leq U^2\boldsymbol{h}^*\left(\boldsymbol{P}_{\mathcal{W}} - \boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\right)\boldsymbol{h} \\ &= U^2\boldsymbol{h}^*\boldsymbol{P}_{\mathcal{W}}\left(\boldsymbol{I} + U^2\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\right)^{-1}\boldsymbol{P}_{\mathcal{W}}\boldsymbol{h}, \end{aligned} \quad (45)$$

where we used the Matrix Inversion Lemma [9].

Comparing (45) with (40) and recalling that $\boldsymbol{P}_{\mathcal{W}} = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{U}^*$, it follows that $E_{\mathrm{MX}}(\boldsymbol{h}) \leq E_{\mathrm{LS}}(\boldsymbol{h})$ for all $\boldsymbol{h}$ if

$$\begin{aligned} &U^2\boldsymbol{Z}^*\left(\boldsymbol{I} + U^2\boldsymbol{U}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{U}\right)^{-1}\boldsymbol{Z} \\ &\qquad \preceq \left(\boldsymbol{Z}^*\boldsymbol{U}^*\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z}\right)^{-1}, \end{aligned} \quad (46)$$

where $\boldsymbol{A} \preceq \boldsymbol{B}$ means that $\boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite. Now,

$$\begin{aligned} &\boldsymbol{I} + U^2\boldsymbol{U}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{U} = \\ &= \quad \boldsymbol{I} + U^2\boldsymbol{Z}\boldsymbol{Z}^*\boldsymbol{U}^*\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z}\boldsymbol{Z}^* \\ &= \quad \begin{bmatrix} \boldsymbol{I} + U^2\boldsymbol{K} & 0 \\ 0 & \boldsymbol{I} \end{bmatrix}, \end{aligned} \quad (47)$$

where $\boldsymbol{K} = \boldsymbol{Z}^*\boldsymbol{U}^*\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{U}\boldsymbol{Z}$. Therefore,

$$\boldsymbol{Z}^*\left(\boldsymbol{I} + U^2\boldsymbol{U}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{S}^*\boldsymbol{P}_{\mathcal{W}}\boldsymbol{U}\right)^{-1}\boldsymbol{Z} = \left(\boldsymbol{I} + U^2\boldsymbol{K}\right)^{-1}, \quad (48)$$

and (46) becomes

$$U^2\left(\boldsymbol{I} + U^2\boldsymbol{K}\right)^{-1} \preceq \boldsymbol{K}^{-1}. \quad (49)$$

Since $U^2\left(\boldsymbol{I} + U^2\boldsymbol{K}\right)^{-1} = \left((1/U^2)\boldsymbol{I} + \boldsymbol{K}\right)^{-1}$, it follows immediately that (49) is satisfied and $E_{\mathrm{MX}}(\boldsymbol{h}) \leq E_{\mathrm{LS}}(\boldsymbol{h})$ for all $\boldsymbol{h}$.

## 5. CONCLUSION

We considered reconstructing a signal $\boldsymbol{x} \in \mathcal{W}$ from its samples $\boldsymbol{s}_i^*\boldsymbol{x}$ that are contaminated by noise, where the vectors $\{\boldsymbol{s}_i\}$ span a subspace $\mathcal{S}$ such that $\mathcal{W} \cap \mathcal{S}^\perp = \{0\}$. We first derived the estimator resulting from the popular LS approach. We then showed that if $\boldsymbol{x}$ is finite, then we can obtain an estimator whose MSE is smaller than the MSE of the LS estimator for all values of $\boldsymbol{x} \in \mathcal{W}$, by minimizing the worst-case MSE. Similarly, we showed that we can obtain an estimator whose IPMSE is always smaller than that of the LS estimator by minimizing the worst-case IPMSE.

## 6. ACKNOWLEDGMENT

The author wishes to thank Tsvi Dvorkind for fruitful discussions.

## 7. REFERENCES

[1] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2915–2925, Nov. 1994.

[2] Y. C. Eldar, "Sampling and reconstruction in arbitrary spaces and oblique dual frame vectors," *J. Fourier Analys. Appl.*, vol. 1, no. 9, pp. 77–96, Jan. 2003.

[3] Y. C. Eldar, "Sampling without input constraints: Consistent reconstruction in arbitrary spaces," in *Sampling, Wavelets and Tomography*, A. I. Zayed and J. J. Benedetto, Eds., pp. 33–60. Boston, MA: Birkhauser, 2004.

[4] T. Werther and Y. C. Eldar, "General framework for consistent sampling in Hilbert spaces," to appear in *International Journal of Wavelets, Multiresolution and Information Processing*.

[5] Y. C. Eldar and T. Dvorkind, "A minimum squared-error framework for sampling and reconstruction in arbitrary spaces," submitted to *IEEE Trans. Signal Processing*, July 2004.

[6] A. Krzyzak, E. Rafajłowicz, and M. Pawlak, "Moving average restoration of bandlimited signals from noisy observations," *IEEE Trans. Signal Processing*, vol. 45, pp. 2967–2976, Dec. 1997.

[7] M. Pawlak, E. Rafajłowicz, and A. Krzyzak, "Postfiltering versus prefiltering for signal recovery from noisy samples," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3195–3212, Dec. 2003.

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

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1985.