Optimal Parameter Estimation from Shift-Invariant Subspaces

Richard J. Vaccaro
Department of Electrical and Computer Engineering
The University of Rhode Island
Kingston, RI USA

Introduction

Shift-invariant subspaces appear in a variety of signal processing problems including:

Problem 1 Modeling scalar data as the sum of exponentially damped sinusoidal components. Estimate amplitudes, phases, frequencies, and damping factors from Hankel matrices formed from the data. The data could be real or complex valued.

Problem 2 Multichannel signal modeling. Several signals are observed, each with the same frequencies and damping factors, but with independent amplitudes and phases.

Summary of Talk

- We consider a standard method for estimating parameters that reduces to the need to solve a system of noisy linear equations $\tilde{\mathbf{A}}\tilde{\mathbf{X}} \approx \tilde{\mathbf{B}}$ in which the perturbations in $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ have a special structure.

- When $\mathbf{X}$ is square and nonsingular, the equation $\mathbf{AX} = \mathbf{B}$ says that col$(\mathbf{A}) =$ col$(\mathbf{B})$. We first estimate this underlying subspace. Parameters are then calculated from the subspace.

- The result is a two step (non-iterative) method, called OSE (optimal subspace estimation), for estimating parameters. Error variances reach the Cramer-Rao lower bound.

Data Model for Single Channel Data

Consider the continuous-time signal model

$$y(t) = \sum_{i=1}^{q} A_i e^{-\delta_i t} \cos(2\pi f_i + \phi_i)$$

The sampled, discrete-time signal is

$$y[k] = y(t)|_{t=(k-1)T}, \;\; k = 1, \cdots, N.$$

The sampled signal can be modeled as the impulse response of a discrete-time state-space system.

State-Space Approach to Signal Modeling

The sampled signal can be modeled as the impulse response of a discrete-time state-space system. This representation is as follows

$$y[k] = \mathbf{h}\mathbf{F}^{k-1}\mathbf{g}, \ k = 1, \cdots, N$$

where $\mathbf{h}$ is a $1 \times p$ vector, $\mathbf{F}$ is a $p \times p$ matrix, and $\mathbf{g}$ is a $p \times 1$ vector. The value of p is twice the number of sinusoidal components in the data ($p = 2q$).

Recall: $\text{eig}(\mathbf{F}) = \{e^{-\delta_i + j2\pi f_i T}\}, \{e^{-\delta_i - j2\pi f_i T}\}, \ i = 1, \cdots, q.$

Data Model and Hankel Matrix Factorization

The formula $y[k] = \mathbf{h}\mathbf{F}^{k-1}\mathbf{g}$ implies:

$$\begin{bmatrix} y[1] & y[2] & \cdots & y[N-m+1] \\ y[2] & y[3] & \cdots & y[N-m+2] \\ \vdots & \vdots & \vdots & \vdots \\ y[m] & y[m+1] & \cdots & y[N] \end{bmatrix} = \begin{bmatrix} \mathbf{h} \\ \mathbf{hF} \\ \vdots \\ \mathbf{hF}^{m-1} \end{bmatrix} [\mathbf{g} \quad \mathbf{Fg} \quad \cdots \quad \mathbf{F}^{N-}$$

or

$$\mathbf{H} = \mathbf{OC}$$

Realization Formulas

A model $(\mathbf{F}, \mathbf{g}, \mathbf{h})$ is obtained from a factorization of $\mathbf{H}$ as follows:

Let

$$\begin{aligned} \overline{\mathbf{O}} &= \text{ rows 1 to } m-1 \text{ of } \mathbf{O}, \\ \underline{\mathbf{O}} &= \text{ rows 2 to } m \text{ of } \mathbf{O}. \end{aligned}$$

The model is obtained as follows

$$\begin{aligned} \mathbf{h} &= \text{ first row of } \mathbf{O}, \\ \mathbf{g} &= \text{ first column of } \mathbf{C}, \\ \overline{\mathbf{O}}\mathbf{F} &= \underline{\mathbf{O}} \text{ which may be solved to yield } \mathbf{F} = \overline{\mathbf{O}}^L \underline{\mathbf{O}} \end{aligned}$$

where $\overline{\mathbf{O}}^L$ is the left inverse of $\overline{\mathbf{O}}$.

Obtaining a Model from Data

Given a data sequence $y[k]$ form the Hankel matrix $\mathbf{H}$ and factor it by computing the singular value decomposition (SVD) of $\mathbf{H}$ (throw away the zero singular values):

$$\mathbf{H} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T = (\mathbf{U}_1)(\mathbf{\Sigma}_1 \mathbf{V}_1^T) \stackrel{\text{def}}{=} \mathbf{O}\mathbf{C}.$$

The SVD provides a factorization of $\mathbf{H}$, and a state-space model may be obtained from this factorization using the realization formulas.

Obtaining State-Space Models from Noisy Data

Given a noisy data vector

$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{n}$$

form a Hankel matrix

$$\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{N}.$$

The SVD of $\tilde{\mathbf{H}}$ is

$$\tilde{\mathbf{H}} = [\,\tilde{\mathbf{U}}_1 \quad \tilde{\mathbf{U}}_2\,] \begin{bmatrix} \tilde{\mathbf{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_1^T \\ \tilde{\mathbf{V}}_2^T \end{bmatrix}$$

Obtaining State-Space Models from Noisy Data

The rank of $\tilde{\mathbf{H}}$ is full, although it can be approximated by a matrix of rank p. Truncate the SVD to get such an approximation

$$\tilde{\mathbf{H}} \approx \tilde{\mathbf{U}}_1 \tilde{\mathbf{\Sigma}}_1 \tilde{\mathbf{V}}_1^T.$$

Note that the approximation is not a Hankel matrix, and if we try to use the realization formulas we get

$$\overline{\tilde{\mathbf{U}}}_1 \boxed{\mathbf{F}} \approx \underline{\tilde{\mathbf{U}}}_1$$

Solving Overdetermined Equations: Least Squares

In order to solve the overdetermined system of equations

$$\mathbf{A}\mathbf{F} \approx \mathbf{B}$$

all columns of $\mathbf{A}$ and $\mathbf{B}$ must lie in a p-dimensional subspace. The LS approach chooses this subspace to be the column-space of $\mathbf{A}$ and projects the columns of $\mathbf{B}$ into this subspace. The LS solution is

$$\mathbf{F}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B}$$

The LS approach is optimal only when the columns of $\mathbf{A}$ are noise-free and the perturbations on the elements of $\mathbf{B}$ are i.i.d. random variables.

The Total Least Squares Approach

The TLS approach chooses the p-dimensional subspace by minimizing the sum of squared distances from every element of $\mathbf{A}$ and $\mathbf{B}$ to the subspace. This is the optimal subspace only when the perturbations on the elements of $[\,\mathbf{A} \quad \mathbf{B}\,]$ are i.i.d. random variables.

$$[\,\mathbf{A} \quad \mathbf{B}\,] = [\,\bar{\mathbf{U}}_1 \quad \bar{\mathbf{U}}_2\,] \begin{bmatrix} \bar{\mathbf{\Sigma}}_1 & 0 \\ 0 & \bar{\mathbf{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}_{11} & \bar{\mathbf{V}}_{12} \\ \bar{\mathbf{V}}_{21} & \bar{\mathbf{V}}_{22} \end{bmatrix}.$$

The total least squares solution is

$$\mathbf{F}_{TLS} = -\bar{\mathbf{V}}_{12}\bar{\mathbf{V}}_{22}^{-1}.$$

A New Approach: Optimal Subspace Estimation

- $\overline{\tilde{\mathbf{U}}}_1$ and $\underline{\tilde{\mathbf{U}}}_1$ define subspaces that are noisy measurements of the true subspace. They are data for a weighted least-squares problem whose solution gives an estimate of the underlying subspace.

- We use a *subspace perturbation expansion* to get a first-order statistical description of the perturbations in the subspaces defined by $\overline{\tilde{\mathbf{U}}}_1$ and $\underline{\tilde{\mathbf{U}}}_1$.

- This first-order description is an approximation to the actual statistical distribution of the perturbations. There is a range of SNR over which the approximation is very good.

A Simulation Example

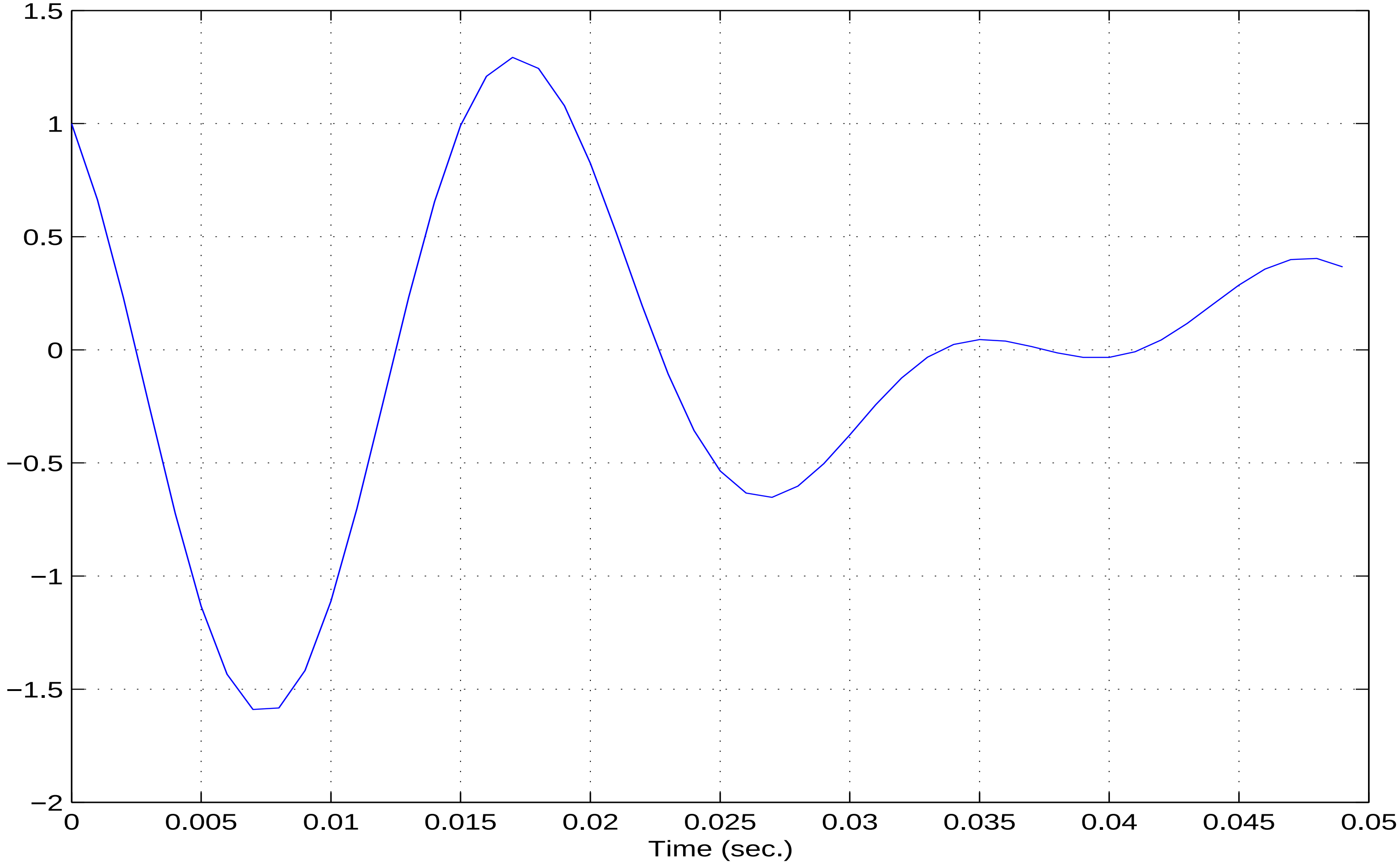
The signal is

$$y(t) = e^{-20t}\cos(2\pi 40t + \frac{\pi}{2}) + e^{-25t}\cos 2\pi 60t$$

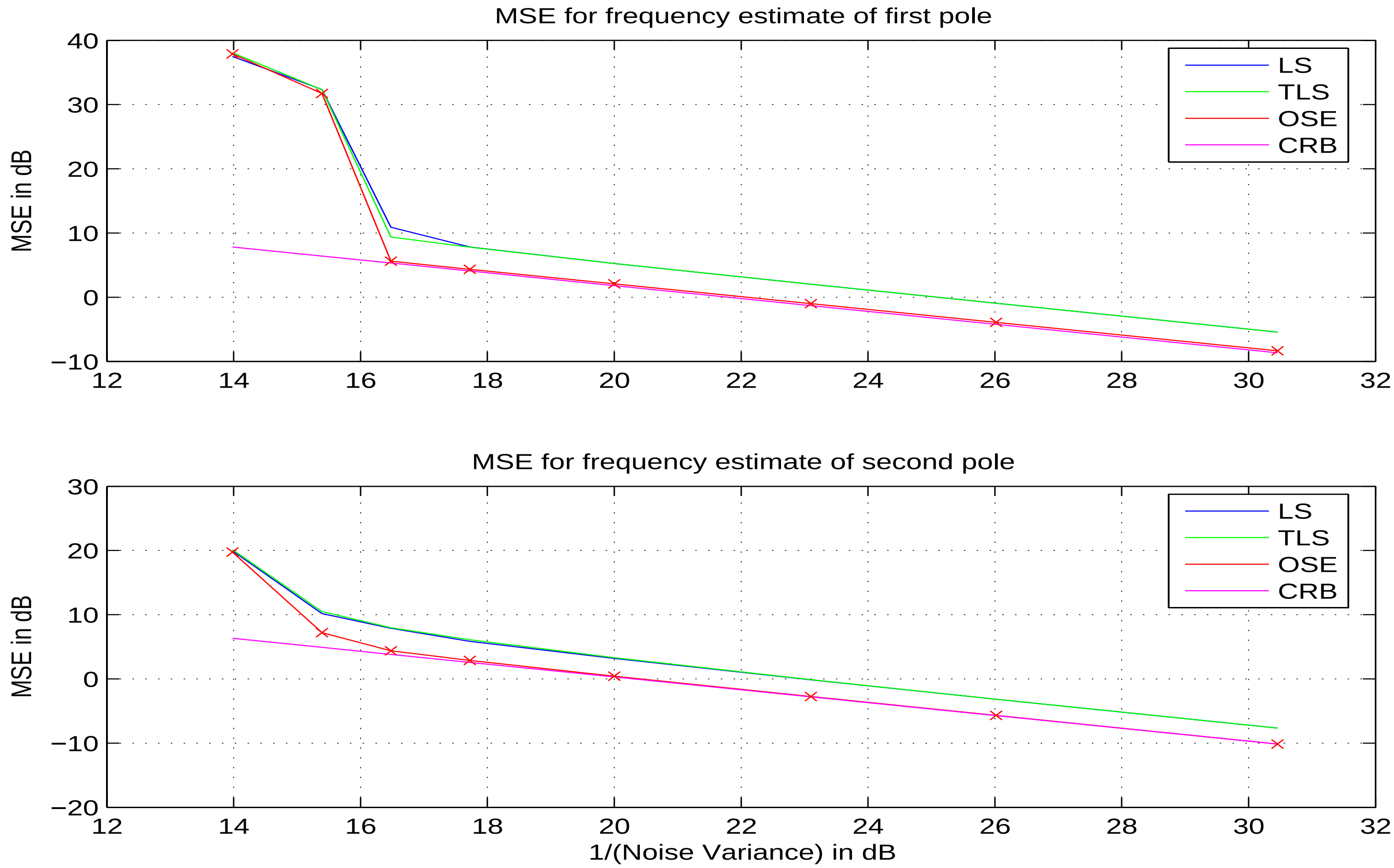
The frequencies are: 40 and 60 Hz
The damping factors are: 20 and 25

The sampling rate is 1,000 Hz and the number of samples is 50.

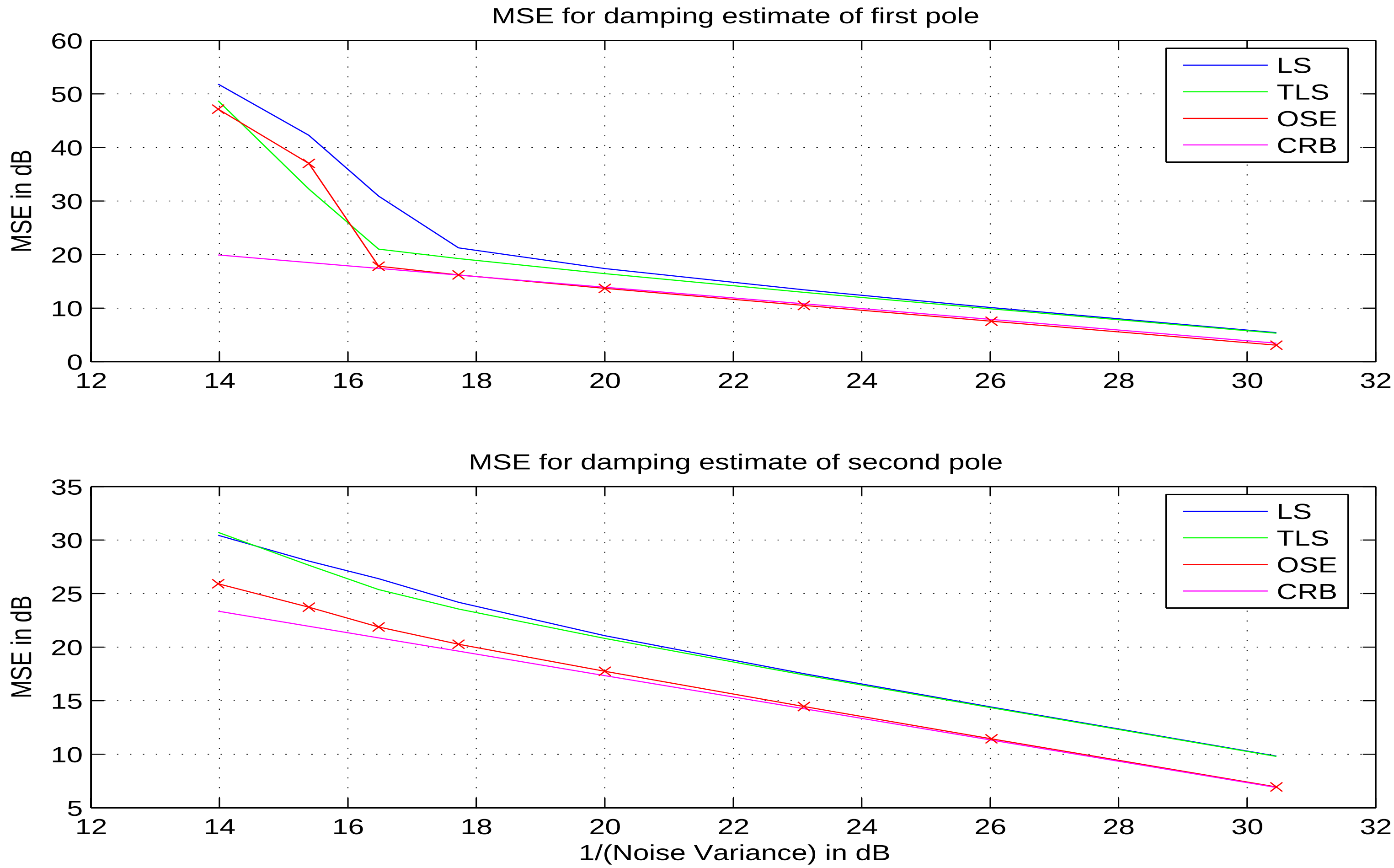
The Test Signal



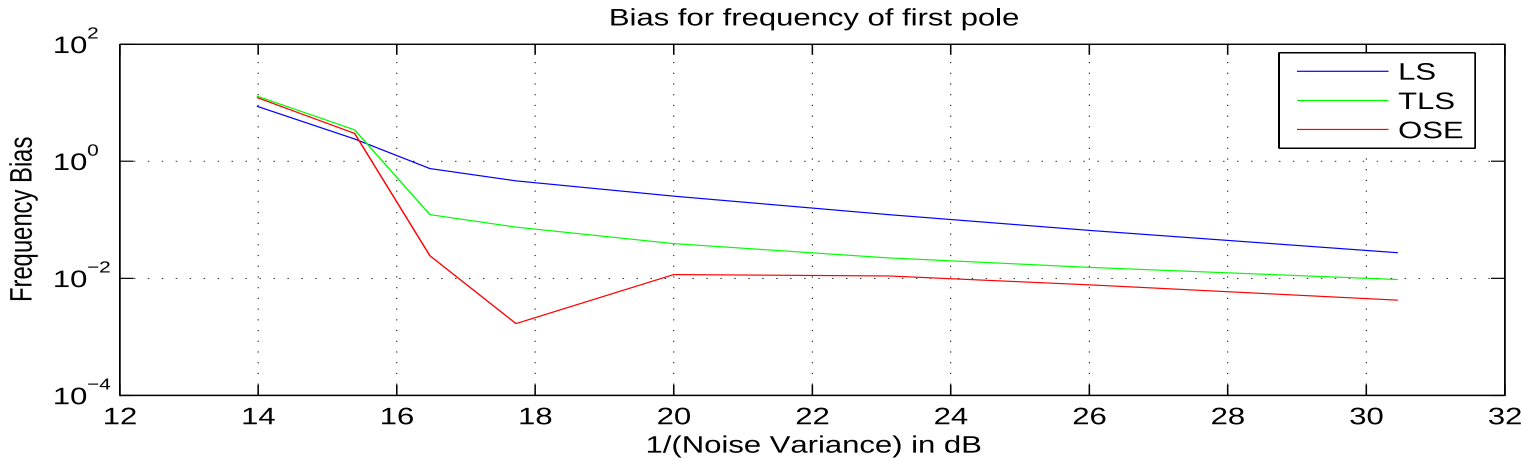
MSE in Frequency Estimates vs. SNR

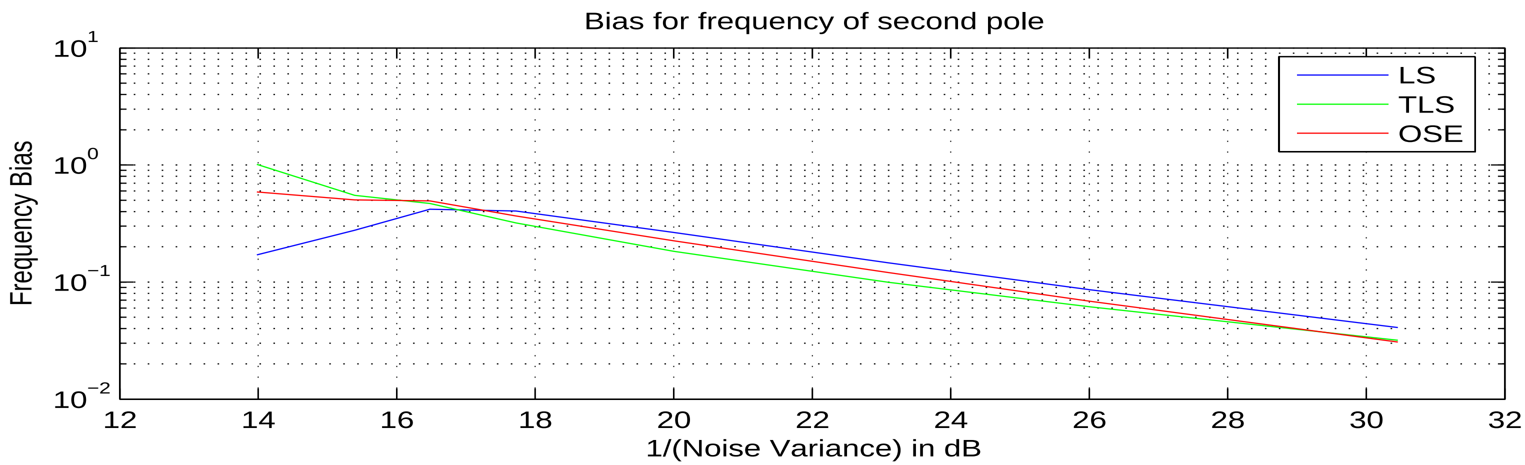


MSE in Damping Estimates vs. SNR

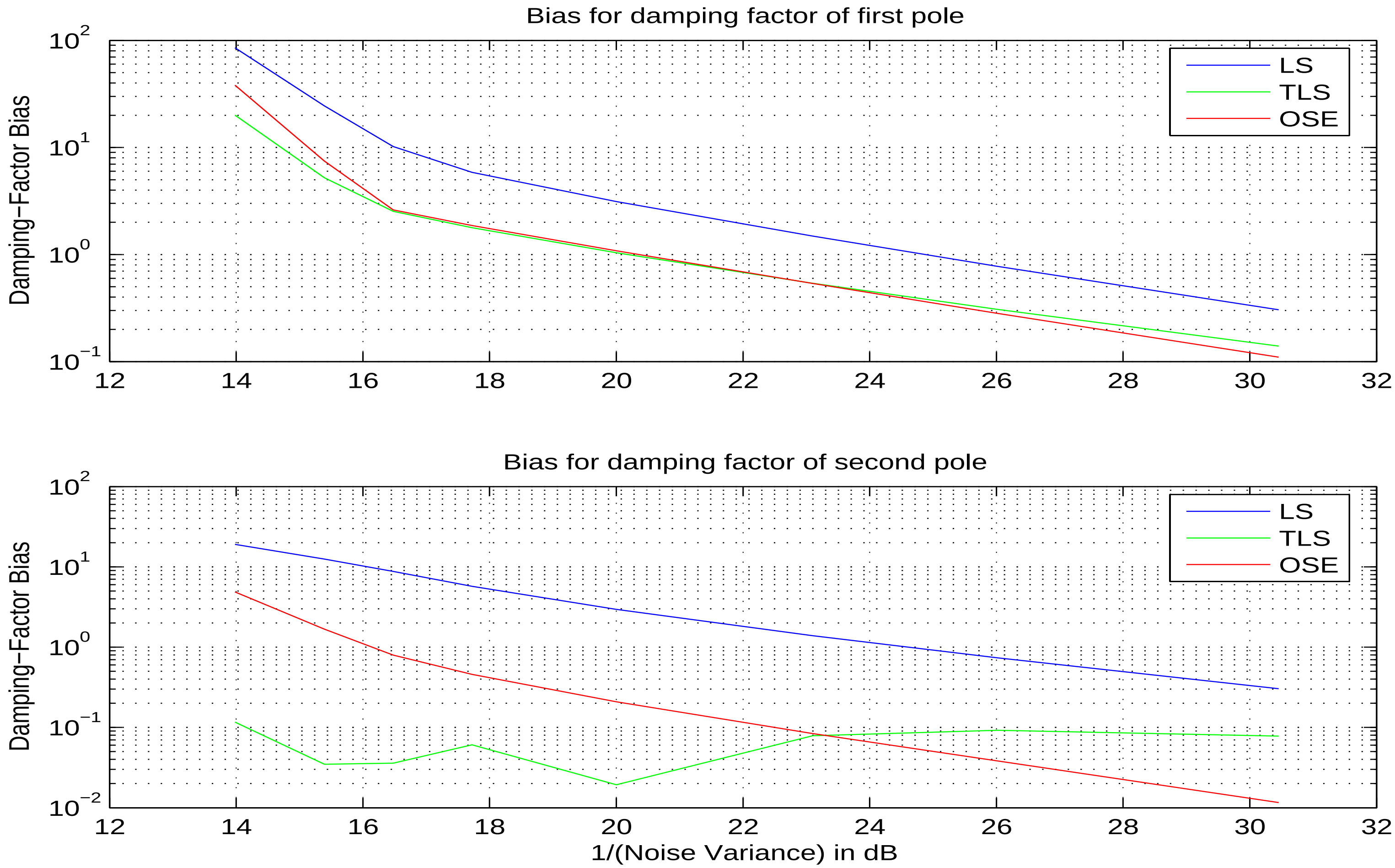


Bias in Frequency Estimates vs. SNR





Bias in Damping vs. SNR



Subspace Perturbation Expansion

The noisy matrix and its SVD are given by

$$\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{N}$$

$$\tilde{\mathbf{H}} = [\,\tilde{\mathbf{U}}_1 \quad \tilde{\mathbf{U}}_2\,] \begin{bmatrix} \tilde{\mathbf{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_1^H \\ \tilde{\mathbf{V}}_2^H \end{bmatrix}$$

The perturbed subspace is col($\tilde{\mathbf{U}}_1$). How is it related to col($\mathbf{U}_1$)? The answer is given by the following approximate basis for col($\tilde{\mathbf{U}}_1$):

$$\tilde{\mathbf{X}}_1 \stackrel{1}{=} \mathbf{U}_1 + \mathbf{U}_2 \mathbf{U}_2^T \mathbf{N} \mathbf{V}_1 \mathbf{\Sigma}_1^{-1}$$

Optimal Subspace Estimation (OSE)

The matrix $\tilde{\mathbf{U}}_1$ is be partitioned into top and bottom pieces. These pieces would span the same subspace in the absence of noise. Because we are forming projection matrices, we can use any basis. We can use $\tilde{\mathbf{U}}_1$ (for numerical calculation) or $\tilde{\mathbf{X}}_1$ (for calculating the covariance matrix of the error terms).

Let

$$\overline{\tilde{\mathbf{P}}} = \overline{\tilde{\mathbf{U}}}_1(\overline{\tilde{\mathbf{U}}}_1^T\overline{\tilde{\mathbf{U}}}_1)^{-1}\overline{\tilde{\mathbf{U}}}_1^T$$

and

$$\underline{\tilde{\mathbf{P}}} = \underline{\tilde{\mathbf{U}}}_1(\underline{\tilde{\mathbf{U}}}_1^T\underline{\tilde{\mathbf{U}}}_1)^{-1}\underline{\tilde{\mathbf{U}}}_1^T$$

Then

$$\begin{bmatrix} \overline{\tilde{\mathbf{P}}} \\ \underline{\tilde{\mathbf{P}}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} \mathbf{P} + \begin{bmatrix} \Delta\overline{\mathbf{P}} \\ \Delta\underline{\mathbf{P}} \end{bmatrix}$$

OSE (cont.)

We now "compress" the projection matrix estimation equation by multiplying with a fixed matrix $\mathbf{Z}$ whose columns are a basis for the true underlying subspace.

$$\begin{bmatrix} \overline{\tilde{\mathbf{P}}}\mathbf{Z} \\ \underline{\tilde{\mathbf{P}}}\mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} \mathbf{P}\mathbf{Z} + \begin{bmatrix} \Delta\overline{\mathbf{P}}\mathbf{Z} \\ \Delta\underline{\mathbf{P}}\mathbf{Z} \end{bmatrix}$$

$$\begin{bmatrix} \tilde{\mathbf{Z}}_1 \\ \tilde{\mathbf{Z}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{m-1} \\ \mathbf{I}_{m-1} \end{bmatrix} \mathbf{Z} + \begin{bmatrix} \Delta\mathbf{Z}_1 \\ \Delta\mathbf{Z}_2 \end{bmatrix}.$$

In practice, we use $\mathbf{Z} = \text{orth}(\overline{\tilde{\mathbf{U}}})$.

OSE (cont.)

We vectorize the upper and lower halves of the previous equation, making use of the following formula to vectorize $\Delta\mathbf{Z}_1$ and $\Delta\mathbf{Z}_2$ into $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively.

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$$

The result is

$$\begin{bmatrix} \tilde{\mathbf{z}}_1 \\ \tilde{\mathbf{z}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

Correlation Structure of OSE

The covariance matrix of the noise terms in the subspace estimation equation is:

$$E\left\{\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}\left[\,\mathbf{e}_1^T \quad \mathbf{e}_2^T\,\right]\right\} = \sigma_n^2\begin{bmatrix}\mathbf{B}_1\\ \mathbf{B}_2\end{bmatrix}\left[\,\mathbf{B}_1^T \quad \mathbf{B}_2^T\,\right]$$

The error covariance matrix is computed using the subspace perturbation expanion. It is a rank deficient matrix, and Paige's method is used to solve the rank-deficient, weighted least-squares problem.

Multichannel Data

$$y_k(t) = \sum_{i=1}^{q} A_{ik} e^{-\delta_i t} \cos(2\pi f_i + \phi_{ik}), \;\; k = 1, \cdots, C$$

- There are C data channels. The covariance matrix of the noise samples is known up to a scale factor.
- The amplitudes (A_{ik}) and phases (ϕ_{ik}) are different for each channel
- The frequencies (f_i) and damping factors (δ_i) are the same for each channel

Multichannel Data Simulation Parameters

- There are two channels containing complex-valued signals. The length of the signals is 25 samples.
- The sampling interval is $T = 1$ second. The frequencies are $f_1 = 0.2$, $f_2 = 0.22$. The damping factors are $\delta_1 = 0.02$, $\delta_2 = 0.02$.
- The amplitudes and phases of the signals were generated randomly.

TLS Approach (Papy, DeLathauwer, Van Huffel 2006)

1. Put each signal into a Hankel matrix, $\mathbf{H}_k, k = 1, \cdots, C$

2. Compute SVDs: $\mathbf{H}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^T,\ k = 1, \cdots, C$

3. Aggregate the left singular vectors from all channels: $\mathbf{U}_{tot} = [\, \mathbf{U}_1 \quad \mathbf{U}_2 \quad \cdots \quad \mathbf{U}_k \,]$

4. Compress into a p-dimensional subspace using SVD of $\mathbf{U}_{tot}$ to get $\widehat{\mathbf{U}}$.

5. Extract frequencies and dampings using TLS (SVD of $[\, \bar{\widehat{\mathbf{U}}} \quad \underline{\widehat{\mathbf{U}}} \,]$

OSE Approach to Multichannel Signal Modeling

1. Form vector sequence $\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_C(t) \end{bmatrix}$, $t = kT, k = 1, \cdots, N$

2. Form block Hankel matrix

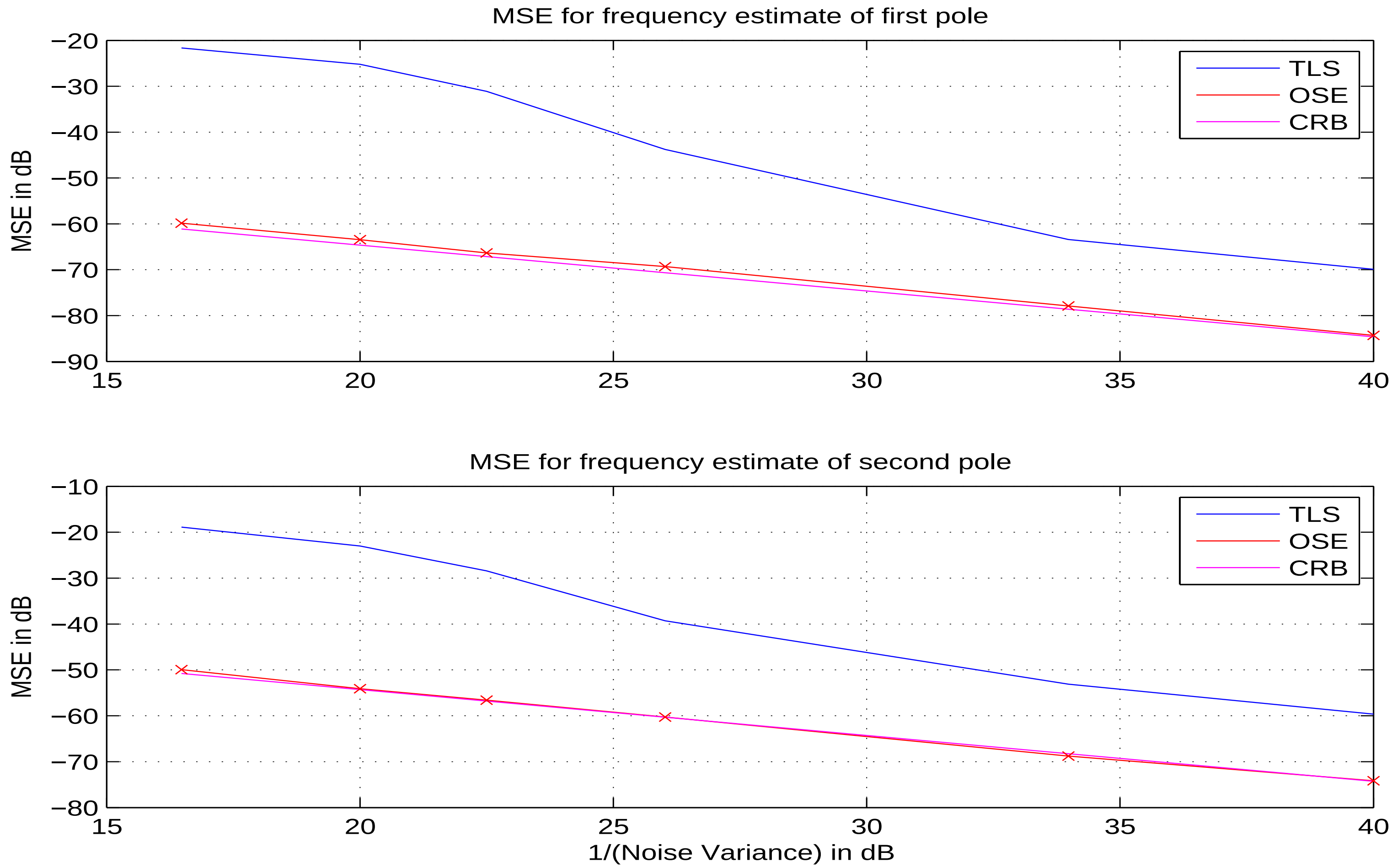
$$\mathbf{H}_{m\times n} = \begin{bmatrix} \mathbf{y}[1] & \mathbf{y}[2] & \cdots & \mathbf{y}[N-m+1] \\ \mathbf{y}[2] & \mathbf{y}[3] & \cdots & \mathbf{y}[N-m+2] \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}[m] & y[m+1] & \cdots & \mathbf{y}[N] \end{bmatrix}$$

3. Calculate SVD: $\mathbf{H} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T = (\mathbf{U}_1)(\mathbf{\Sigma}_1\mathbf{V}_1^T)$

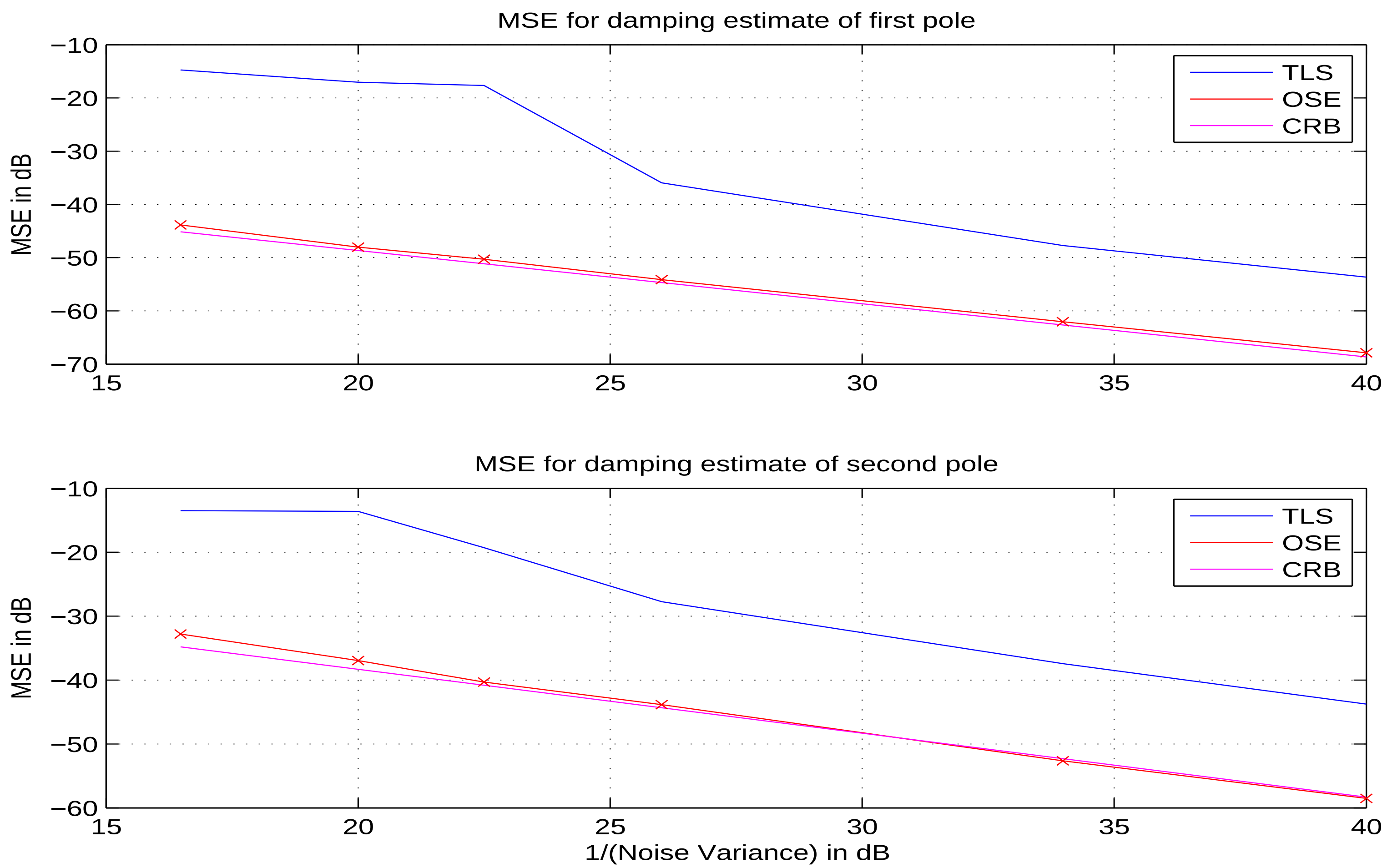
4. Perform OSE using block shift:

$$\bar{\mathbf{U}}_1 = \text{rows 1 to } m - c \text{ of } \mathbf{U}_1$$
$$\underline{\mathbf{U}_1} = \text{rows } C + 1 \text{ to } M \text{ of } \mathbf{U}_1$$

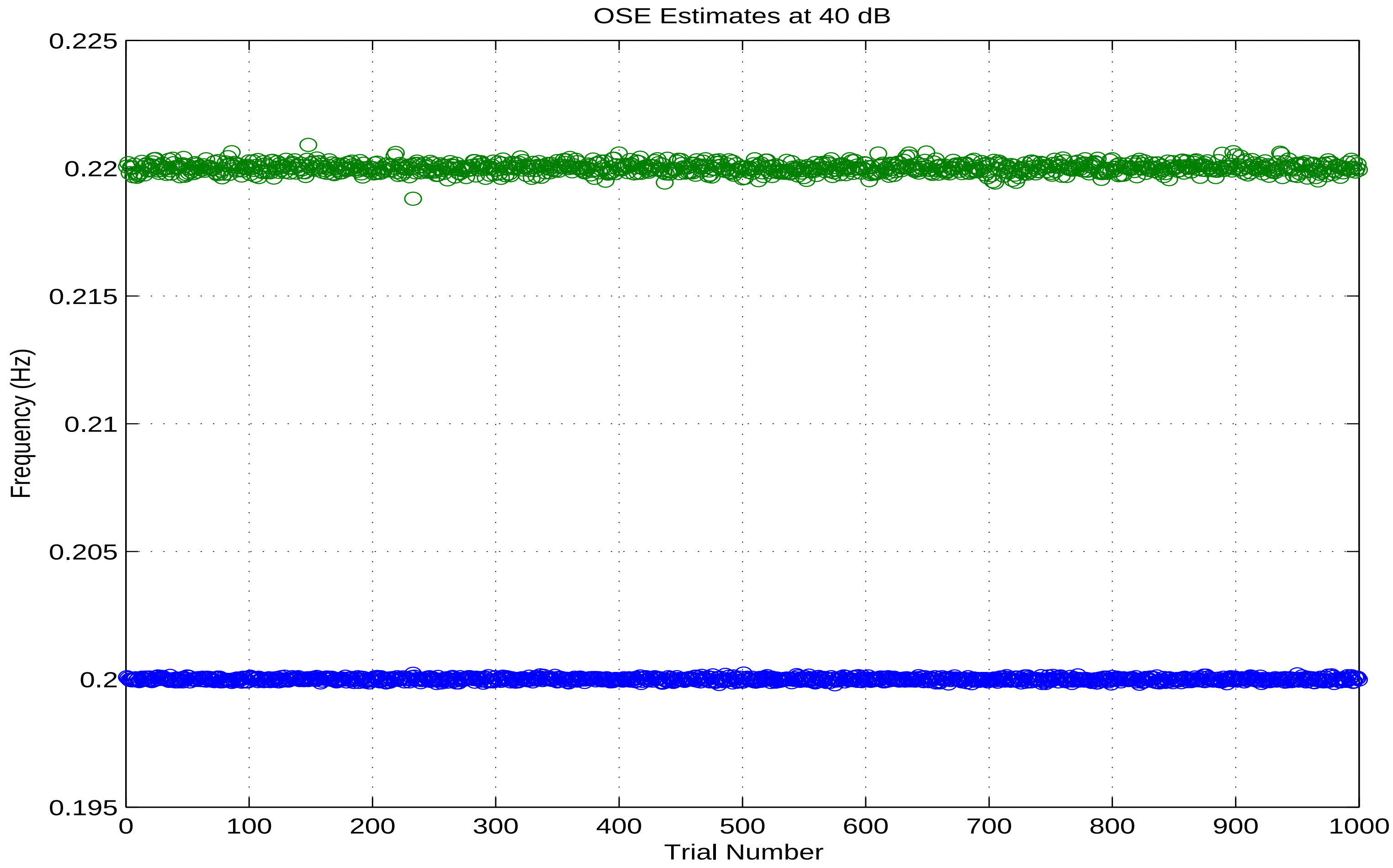
MSE in Frequency Estimates vs. SNR for 2-Channel Data



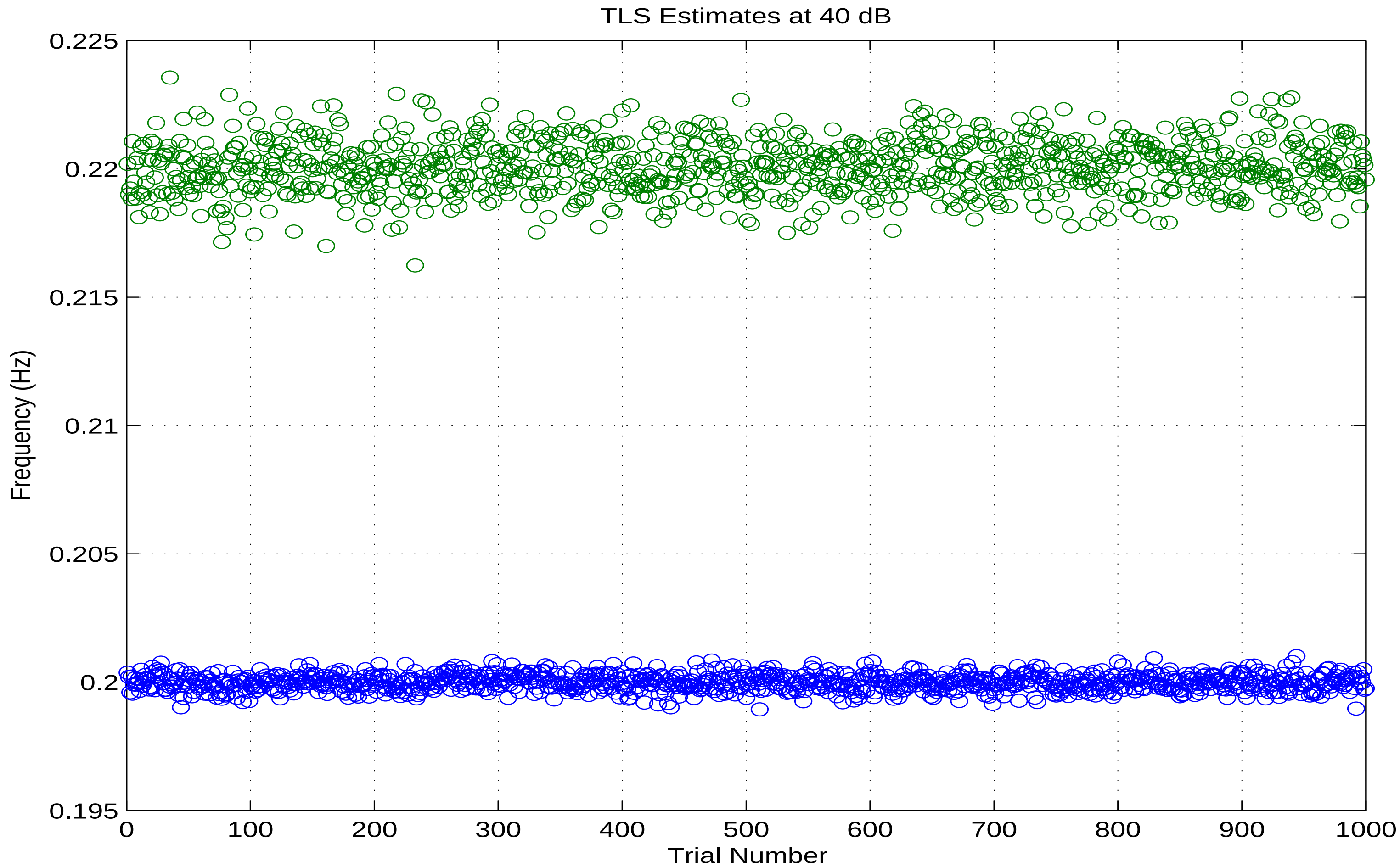
MSE in Damping Estimates vs. SNR for 2-Channel Data



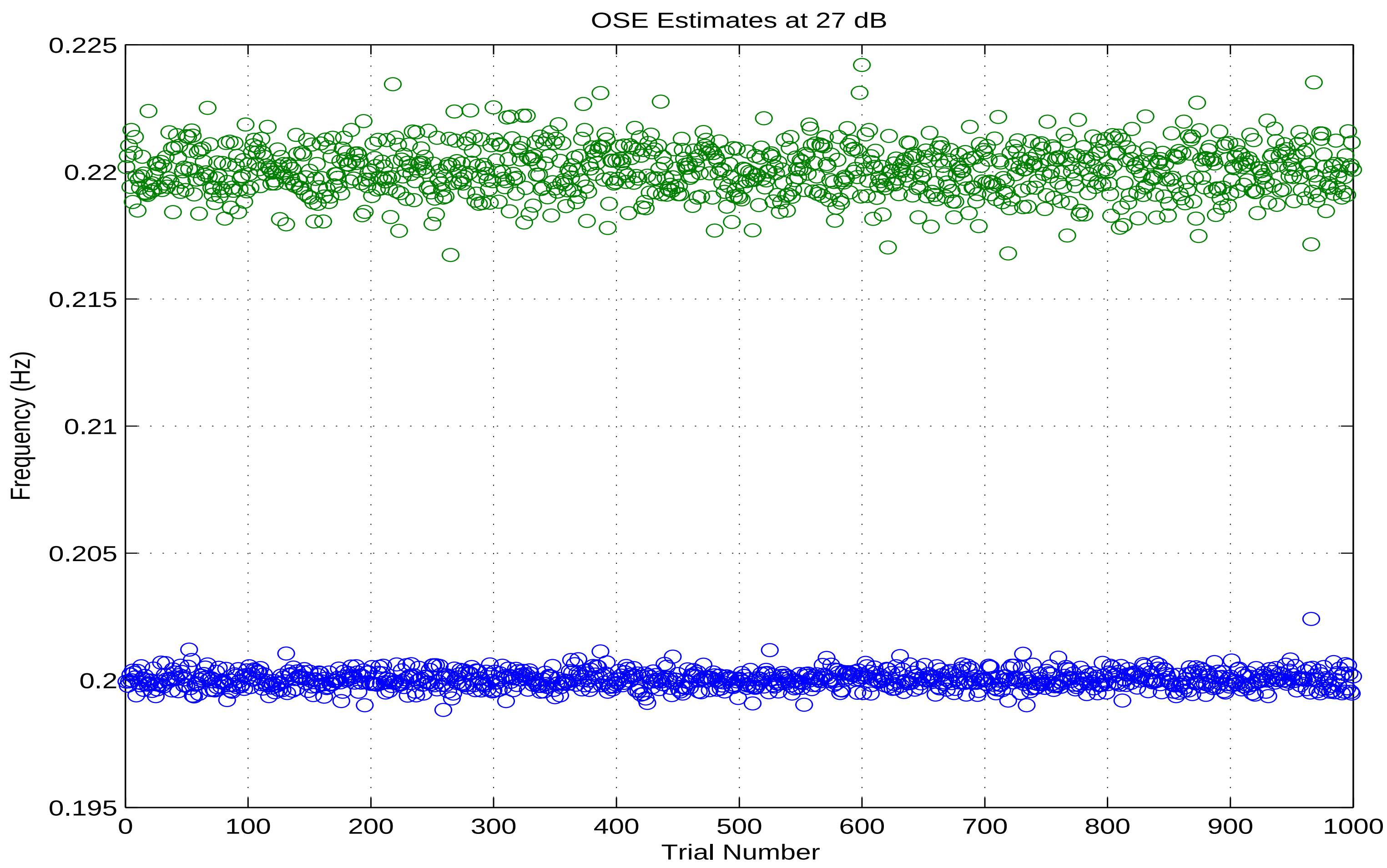
2-Channel OSE Estimates at 40 dB



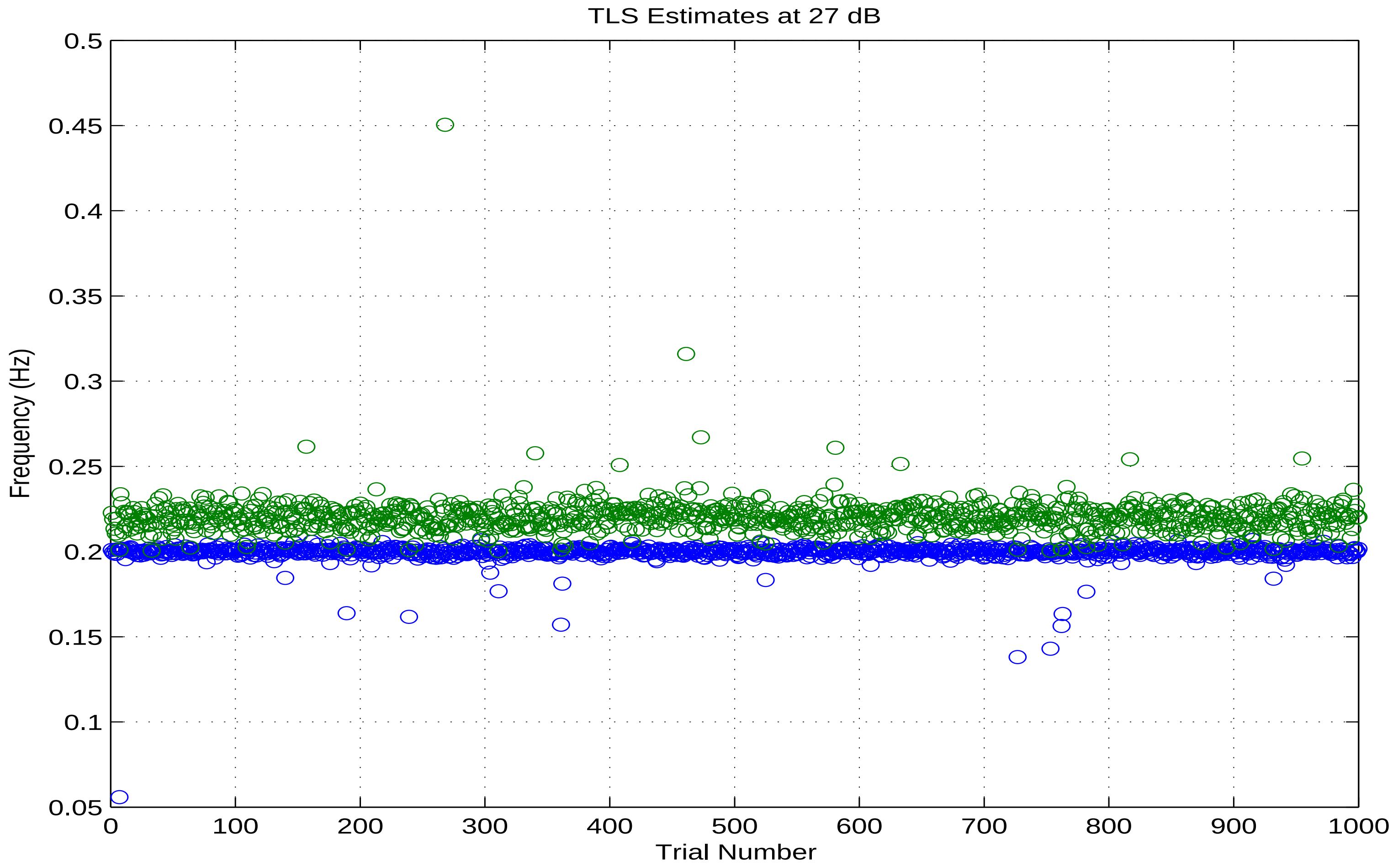
2-Channel TLS Estimates at 40 dB



2-Channel OSE Estimates at 27 dB



2-Channel TLS Estimates at 27 dB



Conclusions

- Parameter estimation problems involving shift-invariant subspaces result in the need to solve noisy linear equations with a special structure.

- The structure can be accounted for by using the structured-TLS approach, but then an initial guess must be obtained and one has to worry about the convergence of an iterative algorithm.

- The first-order statistical information about the error structure provided by the subspace perturbation expansion is adequate to obtain statistically efficient estimates with a non-iterative algorithm.