Regularized Nuclear Norm Spectrum Estimation in Frequency Domain

Hüseyin Akçay and Semiha Türkay
Department of Electrical and Electronics Engineering, Anadolu University
Eskisehir 26470, Turkey

Abstract—Subspace-based methods have been effectively used to estimate multi-input/multi-output, discrete-time, linear-time invariant systems from spectrum samples. A critical step in these methods is the splitting of causal and noncausal invariant subspaces of a Hankel matrix built from spectrum measurements via singular-value decomposition in order to determine the model order. Quite often, in particular when signal-to-noise ratio is low, unmodelled dynamics is present, and when the number of measurements is small, this step is not conclusive since the assumed mirror image symmetry with respect to the unit circle between the eigenvalues of the invariant spaces is lost. In this paper, we propose a robust model order selection scheme based on the regularized nuclear norm optimization in combination with a particular subspace method.

By a numerical example, efficacy of the proposed scheme is shown for a broad range of signal-to-noise ratio and short data records. Then, in a real-life example, the proposed scheme, integrated into a recently developed subspace-based algorithm, is used to estimate cross-power spectra of induction motors from sound data collected by a microphone array in a test rig.

Index Terms—nuclear norm; spectrum estimation; subspace method; regularization; Hankel structure; acoustic spectrum.

I. INTRODUCTION

Suppose that we are given noisy power spectrum samples of a discrete-time, linear-time invariant, multi-input/multi-output system at equidistantly spaced frequencies:

\[ S_k \in \mathbb{C}^{n \times n}, \quad \omega_k = \frac{\pi k}{M}, \quad k = 0, \ldots, M. \]

We would like to estimate a stable and minimum-phase transfer function \( \hat{G}(z) = \hat{C}(zI-\hat{A})^{-1}\hat{B}+\hat{D} \), called the spectral factor of \( \hat{S}(z) = \hat{G}(z)\hat{G}^\dagger(z^{-1}) \), such that \( \hat{S}(e^{j\theta}) \sim S_k \). Recall that \( \hat{G}(z) \) is stable and minimum-phase if its poles and zeros are inside the unit circle. It is also desired that the spectrum estimate is consistent, i.e., \( \hat{S}(e^{j\theta}) \) converges to the true spectrum \( S(e^{j\theta}) \) as \( M \to \infty \).

This problem arises in many applications, for example, in the design of linear shape filters for noise processes [1], [2]. Subspace methods are popular to obtain low-order state-space models from noisy time or frequency-domain measurements. In [3], a subspace identification algorithm was presented. The main idea behind this algorithm is to recognize that the range space of a Hankel matrix built by the inverse discrete-Fourier transform (IDFT) from the spectrum samples given at the equidistantly spaced frequencies is exactly the span of the extended observability matrix associated with the causal and the anti-causal parts of the power spectrum in a state-space realization. This idea has found applications in the auto and cross-spectrum estimation problems [4] and in the retrieval of transfer functions from phase measurements [5].

The identification algorithms proposed in [2], [3], [4], [5] determine the model order by inspecting the singular values of a Hankel matrix built from the spectrum samples by the IDFT. If \( n \) is the assumed model order, then the \( 2n \) most significant singular values and the corresponding left singular vectors are retained in order to retrieve the observability range space. Implicit in this process is the assumption that there exists a mirror image symmetry with respect to the unit circle between the eigenvalues of the causal and the anti-causal invariant spaces. Under this assumption, the causal eigenvalues are obtained by a Jordan decomposition. See, [2], [6] for details.

Quite often, in particular when signal-to-noise ratio is low, the true spectrum is more complicated than the assumed one and the data record is short, the singular value decomposition (SVD) is not conclusive since the assumed symmetry between the eigenvalues of the invariant spaces is lost. A two-stage identification algorithm was proposed in [7]. The first stage provides initial estimates to a parametric optimization problem in the second stage by using an asymptotic form of the subspace identification algorithm proposed in [2]. The minimum-phase property is guaranteed in the second stage via the solution of a conic linear programming problem. This scheme avoids the need to carry out the numerically sensitive split in [2], [3], [4], [5].

Nuclear norm optimization methods for structured low-rank matrix approximation have been discussed in several recent papers on system identification [8], [9], [10], [11]. The nuclear norm of a matrix-valued function as a convex heuristic for minimizing its rank was first proposed in [8]. Minimum nuclear norm solutions often have low rank and in certain applications, for example, low-rank matrix completion problems, the quality of the heuristic can be demonstrated analytically [12], [13]. This approach preserves linear structure in matrix approximation unlike the SVD. Convex constraints or regularization terms in the cost function are easily accommodated in this framework.

The content of this paper is as follows. In Section II, frequency-domain power spectrum estimation via the algorithm proposed in [3] is briefly reviewed. In Section III, a variation of this algorithm based on the regularized nuclear norm heuristic is presented. In Section IV, first a simulation
example is used to demonstrate that the proposed scheme is effective in determining true orders for short data records and is robust to noise. Then, the proposed scheme is used in a real-life example. Section V concludes the paper.

II. SPECTRUM ESTIMATION IN FREQUENCY-DOMAIN BY A SUBSPACE METHOD

We will briefly outline the steps of the subspace algorithm introduced in [3] relevant to our discussion. Let the quadruplet \(\{A,B,C,D\}\) be a minimal realization of a square, strictly minimum-phase transfer function \(G(z)\) with Macmillan degree \(n\). This means that all eigenvalues of \(A\) and \(-BD^{-1}C\) lie strictly inside the unit circle and the pairs \(\{A,B\}\) and \(\{C,A\}\) are controllable and observable, respectively.

The subspace algorithms in [3], [2], [6], [7] start by splitting the power spectrum \(S(z) = G(z)G^T(z^{-1})\) into the spectral summands \(H(z) = \overline{f} + C(zI_n - A)^{-1}F\) and \(H^T(z^{-1})\) as \(S(z) = H(z) + H^T(z^{-1})\). The details can be found in these works. This stage reduces the problem to identifying spectral summands from sample spectra.

Next, since \(S(z)\) has a real-valued impulse response, restriction of \(S(e^{j\omega})\) to \([0,\pi]\) can be extended to \([\pi, 2\pi]\) from the Hermitian symmetry property: \(S(e^{-j\omega}) = S^H(e^{j\omega})\) where \(X^T, X^H\) denote respectively the transpose, the complex conjugate transpose of a given matrix \(X\). Hence, we let \(N = 2M\) and \(z_k = e^{j\omega_k}\) where \(\omega_k\) are calculated from Eq. (1) for \(k = M + 1, \ldots, N - 1\).

Let \(p\) and \(r\) be two fixed positive integers satisfying the inequalities \(p > 2n, r \leq p\), and \(p + r \leq N\). Now, let

\[ \mathcal{H} = \mathcal{G}_p \Psi_r + \Psi_p \mathcal{G}_r \Psi_r, \]

where

\[ \mathcal{G}_p = \frac{1}{\sqrt{N}} \left[ \Omega_p(\omega_k) \otimes S_1 \cdots \Omega_p(\omega_{N-1}) \otimes S_{N-1} \right], \]

\[ \mathcal{G}_r = \frac{1}{\sqrt{N}} \left[ \Omega_r(\omega_k) \otimes \tilde{S}_1 \cdots \Omega_r(\omega_{N-1}) \otimes \tilde{S}_{N-1} \right], \]

\[ \Omega_p(\omega_k) = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ z_k^{p-1} & \cdots & z_k^{N-1} \end{bmatrix}, \]

\[ \mathcal{F}_r = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ z_N^{p-1} & \cdots & z_N^{N-1} \end{bmatrix} \otimes I_m, \]

\[ \Psi_k = \begin{bmatrix} 0 & \cdots & I_m \\ \vdots & \ddots & \vdots \\ I_m & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{km \times km}, \]

and \(I_m\), \(0_{k \times 1}\), \(X \otimes Y\) denoting respectively the \(m \times m\) identity matrix, the \(k \times I\) matrix of zeros (omitted when obvious from the context), the Kronecker product of two given matrices \(X \in \mathbb{C}^{m \times n}\) and \(Y \in \mathbb{C}^{p \times q}\) defined as

\[ X \otimes Y = \begin{bmatrix} X_{11}Y & \cdots & X_{1q}Y \\ \vdots & \ddots & \vdots \\ X_{m1}Y & \cdots & X_{mq}Y \end{bmatrix} \in \mathbb{C}^{mp \times nq}. \]

When the spectrum samples are noise-free, it was shown in [3] that Eq. (2) can be written as

\[ \mathcal{H} = \mathcal{G}_p \delta_{N,p,r} \Delta_r \]

where

\[ \mathcal{G}_p = \begin{bmatrix} C & F^T(A^T)^{-1} \\ \vdots & \vdots \\ CA^{-1} & F^T \end{bmatrix}, \]

\[ \Delta_r = \begin{bmatrix} F & \cdots & A^T \\ (A^T)^{-1}C & \cdots & C \end{bmatrix}, \]

\[ \delta_{N,p,r} = \begin{bmatrix} I_n + A^{N-p-r} & 0 \\ 0 & I_n + (A^T)^{N-p-r} \end{bmatrix}, \]

\[ \Delta_r^{-1} = \begin{bmatrix} (I_n - A^{N})^{-1} & 0 \\ 0 & (I_n - (A^T)^{N})^{-1} \end{bmatrix}. \]

Since the eigenvalues of \(A\) are inside the unit circle, \(\delta_{N,p,r}\) is always nonsingular. Hence, the range spaces of \(\mathcal{G}_p\) and \(\mathcal{H}\) are equal when the data are noise-free. Moreover, in this case \(\mathcal{H} \rightarrow \mathcal{G}_p \Delta_r\) as \(N \rightarrow \infty\). Notice that \(p\) and \(r\) can be chosen freely subject to the inequalities \(p > 2n, r \geq 2n; p + r \leq N\).

The block entries of \(\mathcal{H}\) satisfy the relation [3]:

\[ \mathcal{H}_{kl} = \hat{s}_{k+l-1} + \hat{s}_{N-p-r+k+l-1} \]

where \(\hat{s}_k\) denotes the \(N\)-point IDFT of the spectrum samples

\[ \hat{s}_k = \frac{1}{N} \sum_{i=0}^{N-1} e^{j2\pi k i} S_i, \quad k = 0, \ldots, M - 1. \]

Eq. (3) with the left-hand side computed from Eq. (4) is the basic subspace relation. The reader is referred to [3] for the outline of the identification algorithm. This algorithm is consistent if noise in the spectrum samples is a zero-mean complex white-noise process with a bounded covariance function.

As mentioned in Section I, the model order is determined by inspecting the singular values of \(\mathcal{H}\) in its SVD, and an estimate of \(A\) in the Jordan form is obtained from the Jordan decomposition of a matrix, which is similar to

\[ \begin{bmatrix} A & 0 \\ 0 & (A^T)^{-1} \end{bmatrix} \]

when the data in Eq. (1) are noise-free and have been generated by the true system. Due to insufficient amount of data, noise, and undermodelling, the presumed mirror image symmetry between the eigenvalues of \(A\) and \(A^T\) may be destroyed. The objective of this paper is to propose a robust model order selection procedure.

III. FREQUENCY-DOMAIN SPECTRUM ESTIMATION BY REGULARIZED NUCLEAR NORM OPTIMIZATION

Let

\[ x_{k+l-1} = \mathcal{H}_{kl}, \quad k = 1, \ldots, p; l = 1, \ldots, r. \]

From Eq. (5) and \(S(e^{-j\omega}) = S^H(e^{j\omega})\), \(\forall \omega \in \mathbb{R}\), note that \(x_k \in \mathbb{R}^{m \times m}\) are symmetric matrices for all \(k\). Then, \(\mathcal{H}\) is a
block Hankel matrix given by
\[
H = \begin{bmatrix}
x_1 & x_2 & \cdots & x_r \\
x_2 & x_3 & \cdots & x_{r+1} \\
\vdots & \vdots & \ddots & \vdots \\
x_p & x_{p+1} & \cdots & x_{p+r-1}
\end{bmatrix}.
\]

Stack \(x_1, \ldots, x_{r+p-r-1}\) into a long vector \(x\). Then, \(H\) is a linear function of \(x\) denoted by \(H(x)\). In this parametrization, the Hankel structure of \(H\) is respected. We omit the derivations due to space constraints.

The following is the basic optimization problem
\[
\text{minimize} \quad \|H(x)\|_s + \lambda \|x - g\|_2^2, \quad \lambda > 0 \quad (7)
\]
to be solved in this paper where the first term in the objective is the nuclear norm defined as the sum of the singular values of \(H(x)\), \(g\) is the column vector formed from the sequence \(\hat{g}_1, \ldots, \hat{g}_{p+r-1}\) with \(\hat{g}_k = \hat{s}_k + \hat{s}_{N-p-r+k}\) similarly to \(x\). The optimization variable is the vector \(x\). The second term is a quadratic penalty on the difference between the computed impulse response and the impulse response estimated by the inverse DFT in Eq. (5); but, only the first through the last \(p + r - 1\) coefficients are used.

The optimization problem in Eq. (7) will be solved by the alternating direction method of multipliers (ADMM), a popular method for large scale and distributed convex optimization [11]. The key step in the subspace algorithm is the extraction of the extended observability range space \(O_p\) via SVD. Although SVD provides low-rank approximation, it does not preserve the shift-invariance structure of the Hankel matrix \(H\). In [9], it was suggested to minimize regularized or non-regularized nuclear norm as a heuristic for low-rank approximation problems, which can not be handled via an SVD, in particular, approximation problems with structured low rank matrices and problems including additional constraints.

The structure of the optimization problem in Eq. (7) is similar to the constrained nuclear norm optimization problem studied in [10] if the role played by the regularization term is recognized as the role played by the constraint \(\|x - \hat{g}\|_2^2 \leq \lambda^{-1}\). As \(\lambda\) is varied in Eq. (7), a range of Pareto optimal solutions is obtained as in [10]. Observe that \(\hat{g}\) converges to the impulse response of \(S(z)\) w.p.1 as \(N \to \infty\). When \(\lambda\) is large, the quadratic penalty term can be made small by letting the model order be greater than or equal to \(n\). Since \(\|H(x)\|_s\) increases with the model order, the minimizer of Eq. (7) will yield the right order. Hence, asymptotically the tendency will be towards the right order when \(\lambda\) is large. If \(\lambda\) is small and \(N\) is large, the minimizer of Eq. (7) will yield an order smaller than \(n\) and a bias error will be present. It was suggested in [11] to determine \(\lambda\) by model validation.

The regularized nuclear norm optimization in Eq. (7) may provide accurate results when \(M\) is moderate and the Hankel singular values of the true system decay rapidly. In fact, a high order system with fast decaying impulse response is hard to recover if corruptions are persistent. Under these conditions, the SVD step in the subspace algorithm is inconclusive. Switching to the nuclear norm enforces a sharp transition to low-rank approximant, making the model order choice easier. These claims will be verified by a numerical example in the next section.

IV. EXAMPLES

In this section, first we use a simulation example to evaluate the regularized nuclear norm heuristic in Eq. (7) as implemented by the ADMM scheme in combination with the subspace identification method proposed in [3]. The second example concerns modeling of acoustic power spectra for detecting faults in induction motors.

A. Simulation example

Let the true system \(G(z)\) be a fourth-order system described by the state-space model [3]:

\[
A = \begin{bmatrix}
0.8876 & 0.4494 & 0 & 0 \\
-0.4494 & 0.7978 & 0 & 0 \\
0 & 0 & -0.6129 & 0.0645 \\
0 & 0 & -6.4516 & -0.7419
\end{bmatrix},
\]

\[
B = \begin{bmatrix}
0.2247 & 0.8989 & 0.0323 & 0.1290 \\
0.4719 & 0.1124 & 9.6774 & 1.6129
\end{bmatrix},
\]

\[
D = 0.9626.
\]

Now, assume that the corruptions in Eq. (1) are given as

\[
\hat{s}_k = \frac{0.2z_k^2 - 0.0904z_k + 0.1839}{z_k^2 - 1.1111z_k + 0.8520} v_k,
\]

i.e., \(S_k = S(z_k) + \tilde{s}_k, k = 0, \ldots, M\) where \(v_k\) are zero-mean, unit-variance, independent, and identically distributed complex normal random variables.

In Figure 1, the singular values of \(H(x)\) for Monte Carlo simulations of the proposed scheme over 100 noise realizations with \(\lambda = 5\) (determined by model validation), \(\varepsilon = 0.01, M = 500, p = r = 250\) are plotted. The singular values do not scatter, at least for indices up to 40, over all noise realizations. As a result, the eigenvalues of the 4th order identified models are all packed around the true values. The Monte Carlo simulation results repeated for the same parameter values but \(\varepsilon = 10\) are shown in Figure 2. Although amplitude-wise noise was as large as the true spectrum, the regularized nuclear norm heuristic is extremely robust to noise and the true order can be chosen unambiguously over all noise realizations.

The results shown in Figure 3 were obtained with the subspace algorithm in [3] using the same data set and the parameter values above for \(\varepsilon = 0.01\). For large noise amplitudes, in particular for \(\varepsilon = 10\), and the subspace algorithm in [3], the singular values of \(H(x)\) fluctuate wildly over noise realizations as shown in Figure 4. In this case, a unique model order valid for all noise realizations can not be found.

The Monte Carlo simulation results illustrating dependence of the eigenvalue estimates of \(A\) and the power spectrum estimates on the noise amplitude for both the regularized nuclear norm heuristic and the subspace algorithm in [3] are again omitted due to space constraints. Observe that the components of \(x - g\) in Eq. (7) are equally penalized.
The next example shows that if the measured data contain spikes, then it will be necessary to premultiply $x - g$ by a suitable weighting matrix $W^{-1/2}$ in order to best describe the measured data.

B. An application to modeling of acoustic spectra

Our long term goal is to carry out diagnostic experiments on induction motors in order to detect faults. For this purpose, we conducted tests on both healthy and faulty motors. Defects in the inner race, the outer race, and the rolling balls of bearings and bearing eccentricities were the mechanical faults studied. As electrical faults, we considered short circuits in the stator windings and defects in the rotor cage bars. An array of five directional cardioid type microphones were hemispherically placed around the motors. The microphones connected to an amplifier, and then to a professional sound card were sampled at 44.1 kHz approximately 30 seconds and the measurements were recorded after the motors reached steady state. The experiments were performed in a reverberant and noisy room to simulate the real working environment. The motors were coupled to a single phase generator with permanent magnet, which provided six different stator currents ranging from 3.6 Amperes to 5.4 Amperes. For each current value, three experiments were performed. The test rig used in the experiments is shown in Figure 5.

In Figure 6, the sound pressure levels at Microphones 1 and 2 from a healthy motor at 3.6 Amperes stator current are shown for $t \in [0, 11.34]$ seconds. We divided the data equally into the two disjoints sets: the estimation and the validation data sets, and performed the model estimation and the validation over these sets of cardinality 250,000 each. It was observed that over the segments of this size the data
were stationary, and the measurements at Microphones 2–5 were nearly correlated versions of the measurements at Microphone 1. In fact, this correlation is evident in Figure 6.

Recall that this information yields the entire correlation structure of Microphones 1–5. A two-stage procedure proposed in [4] and further studied in [14] can be used to estimate the discrete-time cross-power spectrum of Microphones 1 and 2 based on the sound data plotted in Figure 6. The minimizer of Eq. (7) yields the cross-power spectrum estimate plotted in Figure 7 for a rectangular window. In the first stage, the sound data are divided into \( L = 50 \) blocks, each of which containing \( M = 5000 \) equidistantly spaced frequencies in \([0, 2\pi]\). Then, the modified periodograms are computed and averaged over these blocks. The result is the so-called Welch cross-power spectrum estimate plotted in Figure 7 for a rectangular window.

In the second stage, the subspace algorithm outlined in [4] extracts a rational model \( \hat{S}_{12}(z) \) from the Welch estimate in the state-space form. This two-stage procedure turns out to be strongly consistent. We would like to evaluate the regularized nuclear norm heuristic in Eq. (7) as implemented by the ADMM scheme and adapted to the two-stage subspace-based cross-power spectrum estimation algorithm, which requires Eq. (6) be replaced with

\[
\begin{bmatrix}
A_1 & 0 \\
0 & A_2^{-1}
\end{bmatrix}
\]

with the spectral radii of \( A_1 \in \mathbb{R}^{n_1 \times n_2} \) and \( A_2 \in \mathbb{R}^{n_2 \times n_2} \) being less than unity and Eq. (7) be replaced with

\[
\minimize ||H(x)||_s + \lambda ||\mathcal{W}^{-\frac{1}{2}}(x-g)||_2^2, \quad \lambda > 0. \tag{8}
\]

We choose \( \mathcal{W}^{-\frac{1}{2}} \) as a diagonal matrix containing \( g \) on the diagonal. In Eq. (8), the first term tries to minimize the \( n_1 + n_2 \) singular values of \( \mathcal{H}(x) \) while the quadratic term attempts to minimize the weighted two norm of the mismatch. In Figures 8 and 9, the cross-power spectrum identification results on the estimation and the validation data sets using the nuclear norm heuristic in Eq. (7) in combination with the algorithm in [4] and the subspace algorithm alone are shown for \( \lambda = 5 \), \( p = r = 200 \), and \( n_1 + n_2 = 24 \).

The identification results plotted in Figures 7–9 on the estimation and the validation data sets are nearly overlapping; meaning that the claimed stationarity over the data segments of length 5.67 seconds during the experiments has prevailed. The minimizer of Eq. (7) yields the cross-power spectrum estimates almost coinciding with the subspace cross-power spectrum estimates. This is hardly surprising since the mismatch was heavily penalized in Eq. (8) by means of the weight \( \mathcal{W}^{-\frac{1}{2}} \). In this example, we traded the robustness of the regularized nuclear norm heuristic with the high performance of the subspace methods.

If \( f_s \) is the sampling rate of the continuous-time processes obtained by zero-order hold equivalence from the time-domain measurements plotted in Figure 6, i.e., 44.1 kHz,
then the modal natural frequencies of the continuous-time cross-power spectrum are calculated as

\[ \Omega_1(k) = \frac{f_c}{L} |\log(\hat{\lambda}_1(k))|, \quad k = 1, \ldots, n_1, \]

\[ \Omega_2(k) = \frac{f_c}{L} |\log(\hat{\lambda}_2(k))|, \quad k = 1, \ldots, n_2 \]

where \( \hat{\lambda}_1(1), \ldots, \hat{\lambda}_1(n_1) \) and \( \hat{\lambda}_2(1), \ldots, \hat{\lambda}_2(n_2) \) are respectively the estimates of the eigenvalues of \( A_1 \) and \( A_2 \).

Motor faults may be detected by monitoring changes in the estimated cross-power spectrum around these frequencies. In structural vibration monitoring, a plethora of fault detection methods is based on monitoring changes in eigenstructures. With few exceptions, the estimated poles were almost on the unit circle. This is an example to the auto-regressive moving-average modeling of harmonic signals. Our aim is to use the proposed scheme or the subspace method in [4] to locate the frequencies of the harmonics for fault detection and condition monitoring of induction motors.

V. CONCLUSIONS

In this paper, we studied identification of MIMO, discrete-time, LTI systems from uniformly spaced power spectrum measurements by a particular subspace-based algorithm. A numerically sensitive step in this method was the splitting of the invariant subspaces associated with the causal and anti-causal eigenvalues of the spectrum to determine the model order. To avoid this step, we proposed a robust model order selection criterion based on the regularized nuclear norm optimization. A numerical example showed the robustness of the proposed scheme to noise. A real-life example was also presented. The non-uniformly spaced frequencies case was recently treated by the authors in [15].

ACKNOWLEDGMENT

The authors would like to thank Prof. E. Germen for providing them with the measured acoustic data.

REFERENCES