Identification of power spectra by reweighted and regularized nuclear norm minimization

Hüseyin Akçay and Semiha Türkay
Department of Electrical and Electronics Engineering, Anadolu University
Eskisehir 26555, Turkey

Abstract—In this paper, we study model order choice in subspace identification algorithms using uniformly spaced spectrum measurements. In these algorithms, model order is determined by singular-value decomposition of a structured matrix constructed from spectrum measurements. This process requires splitting of the two invariant subspaces associated with the causal and the non-causal system poles presumed to be mirror image symmetric with respect to the unit circle. Due to noise, undermodelling, and insufficient amount of data, this symmetry is lost. Recently, a robust model order selection procedure based on the regularized nuclear norm optimization was proposed. We propose a reweighted version of this scheme. Numerical and real-life examples in this paper show that the reweighted nuclear norm minimization makes model order selection easier and results in more accurate models compared to the subspace approaches and the unweighted nuclear norm minimization, in particular at high signal-to-noise ratios.

Index Terms—identification; subspace method; spectrum; frequency-domain; reweighted nuclear norm; uniformly spaced

I. INTRODUCTION

Matrix rank minimization has applications in many fields including control, system identification, signal processing, statistics, and machine learning [1], [2], [3], [4], [5]. This problem is in general hard to solve. One simple and effective heuristic is to minimize nuclear norm instead of rank [6].

In [7], it was suggested to use a weighted nuclear norm to reduce the rank of the solution. It is based on using a non-convex surrogate function \( \log \det (X + \delta I) \) for the rank of \( X \), with \( \delta > 0 \) a fixed number and \( I \) denoting the identity matrix of suitable dimension, and solving the resulting problem via a sequence of convex problems [8]:

\[
\min_{X \in \mathcal{C}} \|W^{(k)}XW^{(k)}\|_*, \quad k \geq 0 \tag{1}
\]

where \( X \) is a symmetric matrix, \( \mathcal{C} \) is a convex set, \( W^{(k)} \) is a sequence of symmetric weight matrices, and \( \|X\|_* \) denotes the nuclear norm of a given matrix \( X \) defined as the sum of the singular values of \( X \). Once the optimal solution \( X^{(k+1)} \) of Eq. (1) is found, \( W^{(k+1)} \) are updated as follows. First, let

\[
W^{(k)}X^{(k+1)}W^{(k)} \sim USU^T \tag{2}
\]

be the reduced order singular value decomposition. Then,

\[
Y^{(k+1)} = [W^{(k)}]^{-1}USU^T[W^{(k)}]^{-1}, \tag{3}
\]

\[
W^{(k+1)} = [Y^{(k)} + \delta I]^{-\frac{1}{2}}. \tag{4}
\]

The update equations Eqs. (3)–(4) together with Eq. (1) describe the called reweighted nuclear norm heuristic [8]. The reader is referred to [8], [9] for the convergence properties of the trace heuristic in Eq. (1). It suffices to say that Eq. (1) can be solved starting with \( W^{(0)} = I \). Selection of the weight matrices is a common theme in compressed sensing.

When the convex set \( \mathcal{C} \) is described by the scalar constraint \( f(X) \leq 0 \), by introducing a Lagrange multiplier \( \lambda > 0 \), Eq. (1) can be written in the regularized form

\[
X^{(k+1)} = \arg \min_{X \in \mathcal{C}} \|W^{(k)}XW^{(k)}\|_* + \frac{\lambda}{2} f(X), \quad k \geq 0 \tag{5}
\]

which is the so-called regularized reweighted nuclear norm heuristic (RRNH). In Eq. (2), the truncation order is determined by inspecting the nonzero entries of \( \Sigma \), i.e., chosen after which there is a sharp drop in the values of nonzero singular values. A threshold procedure was suggested in [2] when a sharp drop is not observed. In this paper, \( X \) will be a structured matrix and \( f(X) \) a quadratic function of \( X \), representing fitting error to a target function.

The RRNH was applied in [8] to the problem of low order system identification from input-output data. The numerical experiments in that study demonstrated that the reweighted nuclear norm minimization makes model order selection easier and results in lower order models compared to the interior-point implementation of the unweighted nuclear norm minimization proposed in [2]. More recently, the RRNH was used in [10] to transform non-positive real transfer function matrices into positive real ones. This problem is of engineering interest and arises when a linear-time dynamics is identified by stochastic subspace methods.

Subspace methods are popular to obtain low-order state-space models of multi-input/multi-output, linear-time-invariant systems from noisy time or frequency-domain measurements [11], [12]. The subject of the current paper is spectrum estimation by frequency-domain subspace techniques. This problem arises in many applications, for example, in the design of linear shape filters for noise processes [13], [14], [15]. A closely related problem is estimation of cross-spectrum in frequency-domain.

In [15], [16], [17], frequency-domain subspace-based identification algorithms were presented. The main idea behind these algorithms is to recognize that range space of a matrix built from frequency-shifted and weighted noise-free samples...
of a rational spectrum equals to the linear span of the extended observability matrix associated with the causal and anti-causal components of the power spectral density function matrix in a given state-space realization.

The algorithms proposed in [15], [17] use the spectrum measurements on the uniform grids of frequencies. Furthermore, the matrix used to extract the observability range space has a Hankel structure and it is obtained by the inverse discrete-Fourier transform directly from the spectrum samples. The model order is determined by inspecting singular values of a matrix used to extract the observability range space. A requirement in these algorithm is that there exists a mirror image symmetry with respect to the unit circle between the causal and anti-causal eigenvalues.

When the signal-to-noise ratio is low, the true spectrum is more complex than the assumed one, and when the data record is short, the singular-value decomposition (SVD) step is inconclusive since the assumed symmetry relation between the eigenvalues of the invariant spaces does not hold.

Nuclear norm optimization methods for structured low-rank matrix approximation have been discussed in several recent papers on system identification [6], [2], [8], [18], [19]. 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. These methods have been primarily developed for Hankel structured low-rank approximation problems in time-domain settings. Extensions to frequency-domain were recently made in [18], [20], [21].

The contents of this paper is as follows. In Section 2, frequency-domain spectrum estimation via the algorithm proposed in [17] is reviewed. In Section 3, a variation of this algorithm based on the RRRH is presented. In Section 4, first a simulation example is used to demonstrate that the proposed scheme is effective in determining minimal spectral factors over short data records and large amplitude noise. The second example illustrates application of this scheme to a real-life problem. Section 5 concludes the paper.

II. SPECTRUM ESTIMATION BY A SUBSPACE METHOD

Suppose that we are given $N+1$ noisy samples $S_k \in \mathbb{C}^{m \times m}$ of the power spectrum $S(z)$ of a linear time-invariant, multi-input/multi-output discrete-time system at $N+1$ equidistantly spaced frequencies $\theta_k = \pi k/N$:

$$S_k = S(e^{j\theta_k}) + \eta_k, \quad k = 0, 2, \ldots, N$$

and want to estimate a spectral factor

$$\hat{G}(z) = \hat{C}(zI - \hat{A})^{-1}\hat{B} + \hat{D}$$

such that the deviation of the estimated spectrum

$$\tilde{S}(z) = \hat{G}(z)\hat{G}^T(z^{-1})$$

from $S(z)$ is small. We also require $\hat{G}(z)$ to be stable and minimum phase. Determination of a suitable model order, and identifying the input dimension correctly when unspecified, i.e., determining the row and the column sizes of $\hat{B}$ is also part of the estimation problem.

We briefly outline the steps of the subspace algorithm proposed in [17]. The details and technicalities can be found there. As most subspace algorithms [15], [17], this algorithm begins by the observation that a power spectrum can be split into the so-called spectral summands as follows

$$S(z) = H(z) + H^T(z^{-1})$$

where

$$H(z) = C(zI - A)^{-1}F + E \frac{1}{2}$$

with $E = CPC^T + DD^T$, $F = APC^T + BD^T$, and $P$ defined as the unique positive-definite solution of the discrete-time Lyapunov equation $P = AP^T + BB^T$. This splitting transfers the problem of identifying a spectral factor to identifying the spectral summands directly from the spectrum samples.

Define a matrix $\mathcal{H}$ blockwise by

$$\mathcal{H}_{k,l} = \hat{s}_{k l \cdot l - 1}, \quad 1 \leq k,l \leq N$$

where $\hat{s}_k$ denotes the $2N$-point inverse discrete Fourier transform of the spectrum samples

$$\hat{s}_k = \frac{1}{2N} \sum_{l=0}^{2N-1} e^{j2\pi kl/N} S_l, \quad k = 0, \ldots , 2N - 1.$$

Let $0, X^T$ and $X^H$ denote respectively the zero matrix of compatible dimensions, the complex conjugate and the conjugate transpose of a given complex matrix $X$. When $\eta_k = 0$ for all $k$, it was shown in [17] that

$$\mathcal{H} = \mathcal{O} \mathcal{E} \mathcal{A}$$

where

$$\mathcal{O} = \begin{bmatrix} C & F T (A^T)^{N-1} & \vdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

$$\mathcal{A} = \begin{bmatrix} F & \cdots & A^{N-1} F \\ (A^T)^{N-1} C & \cdots & C T \end{bmatrix},$$

$$\mathcal{E} = \frac{1}{2} \begin{bmatrix} (I - A^{2N})^{-1} & 0 \\ 0 & (I - (A^{T})^{2N})^{-1} \end{bmatrix}.$$

With the left-hand side computed from Eq. (11), Eq. (13) is the basic subspace relation. It forms the basis of the identification algorithm proposed in [17]. The block column and row dimensions of $\mathcal{H}$ in Eq. (11) were chosen both equal to $N$ for notational convenience and simplification. They can be selected freely as long as each is greater than twice of the true system order $n$. When they are not equal, a slight modification of Eqs. (1)–(4) is necessary. Several steps of the identification algorithm in [17] are outlined below.

Algorithm 2.1:
1) Expand $S_k$ as $S_{N+k} = S_{N-k}^H, k = 1, \ldots , N - 1$ to obtain signals of lengths $2N$.
2) Compute the discrete Fourier coefficients in Eq. (12).
3) Calculate the SVD of $\mathcal{H}$ in Eq. (11):

$$\mathcal{H} = [U \; \mathcal{U}'] \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma' \end{bmatrix} [V \; \mathcal{V}']$$

where
where $\Sigma$ contains the 2n largest singular values.

4) Determine the model order by inspecting the singular values.

5) With $U$ defined by Eq. (17) and $J^x$ by $J^x = [0 \, I]$, $J^u = [I \, 0]$, calculate $\hat{A} = (J^U)\hat{F}^T J^U$ where $X^\dagger$ denotes the Moore-Penrose pseudoinverse of a given full-column rank matrix $X$ defined by $X^\dagger = (X^H X)^{-1} X^H$.

6) Put $\hat{A}$ into the following Jordan canonical form:

$$\hat{A} = [T^\Sigma \ \Pi^{\Sigma}]^{-1} \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} T^\Sigma & \Pi^{\Sigma} \end{bmatrix}$$

where the eigenvalues of $\Sigma$ lie inside the unit circle.

7) Let $\hat{A} = \Sigma^e$, $\hat{C} = J^U T^\Sigma$ where $J^u = [I \, 0]$.

8) For $\hat{E}$ and $\hat{F}$, solve the least-squares problem

$$\min_{\hat{E}, \hat{F}} \sum_{k=0}^{M} \| \hat{E} \hat{\delta} + \hat{F}^T \hat{\delta} X_k - E_k \|_F^2$$

where $\hat{\delta}_k = \hat{C} (z_k I - \Sigma^e)^{-1}$ and $\|X\|_F = \|tr(X^H X)\|_2^2$ is the Frobenius norms of a given matrix $X$.

The estimates of $B$ and $D$ are extracted from $\hat{A}, \hat{F}, \hat{C}, \hat{E}$ by solving a Riccati equation. Due to noise and undermodelling, the causal spectral summand estimate

$$\hat{H}(z) = \hat{C} (z I - \hat{A})^{-1} \hat{F} + \hat{E}$$

may not be positive definite. In this case, the Riccati equation can not be solved. This problem can be circumvented by modifying $\hat{E}$ and $\hat{F}$ matrices via the algorithms in [10], [25].

When $N$ is not large enough and the spectrum measurements are significantly corrupted by noise and undermodelling, the splitting in Step 6 may be problematic since with respect to the unit circle, the eigenvalues in $\Sigma_{eg}$ do no longer mirror images of the eigenvalues in $\Sigma_e$ due to corruptions. The nuclear norm minimization, on the other hand, suppresses noise singular values and enforces a sharp transition to a low rank approximant, thereby making the choice of the model order easier and resulting in more accurate models. The purpose of this paper is to demonstrate through examples that the RRHN has also these properties.

III. FREQUENCY-DOMAIN SPECTRUM ESTIMATION BY THE RRHN

From Eq. (12) and the fact that $S(z^{-1}) = S^H(z)$ on the unit circle, observe that $\mathcal{H}$ defined in Eq. (11) is a real symmetric block Hankel matrix when the corruptions are absent. In this case, insert $x_k$ in place of $\hat{\delta}_k$. Then, with the introduction of the Kronecker delta defined by $\delta_{kl} = 1$ when $k = l$ and 0 otherwise, the basis vectors in $\mathbb{R}^m$ denoted by $e_{ms}$ for $s = 1, \ldots, m$, and the Kronecker product $X \otimes Y$ of two given matrices $X \in \mathbb{C}^{m \times m}$ and $Y \in \mathbb{C}^{p \times q}$ defined as

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

$\mathcal{H}$ can be written as

$$\mathcal{H}(x) = \sum_{s=1}^{2N-1} \sum_{t=1}^{m} x_{i,s,t} E^{i,s,t}$$

where $x \in \mathbb{R}^{m(n+1)(2N-1)/2}$ is obtained by concatenating the distinct components of the symmetric matrices $x_1, \ldots, x_{2N-1}$ while preserving the lexicographic ordering of the triplets $\{i,s,t\}$ and

$$E^{i,s,t} = \begin{cases} E^r \otimes e_r e_r^T, & s = t \\ E^c \otimes (e_r e_r^T + e_c e_c^T), & t > s, \end{cases}$$

and $E^c$ is defined componentwise by

$$E^c_{i,j} = \delta_{i+j+1}, \quad 1 \leq s, t \leq N.$$
IV. EXAMPLES

A. Numerical example

Let the true system $G(z)$ be a fourth-order system described by the state-space model [17]:

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

$$B = \begin{bmatrix}
0.2247 & 0.8989 & 0.0323 & 0.1290 \\
\end{bmatrix}^T,$$

$$C = \begin{bmatrix}
0.4719 & 0.1124 & 9.6774 & 1.6129 \\
\end{bmatrix},$$

$$D = 0.9626.$$  

and assume that the corruptions in Eq. (6) are given as

$$\eta_k = \epsilon \frac{0.2z_k^2 - 0.0904z_k + 0.1839}{z_k^2 - 1.111z_k + 0.8520} v_k,$$

where $v_k$ are zero-mean, unit-variance, independent, and identically distributed complex normal random variables and $\epsilon$ is a scaling factor to adjust the amplitude of the noise.

We will examine accuracy of the identification algorithms by performing Monte Carlo simulations where the quality of the models is assessed by computing the quadratic norm of the misfit $\|S - \hat{S}\|_m^2 = N^{-1} \sum_{k=1}^{N} \|\hat{S}(z_k) - S(z_k)\|^2$ with $\hat{S}(z)$ defined by Eq. (8) and averaging it over noise realizations.

First, we study the convergence behavior of the RRNH defined by Eq. (24). We computed the estimation error for $N = 32$, $\lambda = 20$, $\epsilon = 0.1$, $\delta = 0.1$, and one noise realization only. Figure 1 shows that the RRNH converges only in three iterations. The first iteration yields the unweighted regularized nuclear norm heuristic studied in [20]. By iterating this heuristic two more times, we are able to reduce the estimation error by 15%. The same convergence behavior has been observed with different choices of $\delta$ and noise realizations more than one.

![Fig. 1. Monte Carlo simulations in Example 1 for one noise realization only comparing $|S - \hat{S}|_m^2$ as a function of the number of iterations for the RRNH in Eq. (24) with $N = 32$, $\lambda = 20$, $\epsilon = 0.1$, and $\delta = 0.1$.](image)

Next, we examine the influence of the surrogate function parameter $\delta$ on the convergence behavior of the RRNH while fixing the number of iterations to three. The rest of the parameters are chosen as before. From Figure 2, observe that the estimation error drops rapidly first and changes little with $\delta$ for all $\delta \geq 200$. Again, this behavior is irrespective of the number of iterations and noise realizations.

![Fig. 2. Monte Carlo simulations in Example 1 for one noise realization only comparing $|S - \hat{S}|_m^2$ as a function of $\delta$ for the RRNH in Eq. (24) with $N = 32$, $\lambda = 20$, $\epsilon = 0.1$, and three iterations.](image)

In Table 1, the averaged quadratic-norm errors for the subspace algorithm [17], the regularized nuclear norm heuristic (RNH) [20] and the RRNH in Eq. (24) calculated over 100 estimated models are shown for $\epsilon$ ranging from $10^{-4}$ to $10^{-1}$ and $\lambda = 2000$, $N = 32$, $\delta = 100$. The number of iterations is three for each estimated model. The RRNH outperforms the other two competitors at high signal-to-noise ratios (SNR). The numerical study in [17] showed that the subspace algorithm is more effective than the RNH in suppressing noise at high SNRs while the latter prevails at low SNRs. To recapitulate [20] and the above observations, the RNH appears to be more accurate and robust than the subspace estimator at low SNRs while at high SNRs by iterating the RNH a few times, an estimator which outperforms the subspace estimator is derived.

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$10^{-4}$</th>
<th>$10^{-3}$</th>
<th>$10^{-2}$</th>
<th>$10^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subspace</td>
<td>0.00092</td>
<td>0.00094</td>
<td>0.00180</td>
<td>0.01520</td>
</tr>
<tr>
<td>RNH</td>
<td>0.00160</td>
<td>0.00160</td>
<td>0.00220</td>
<td>0.01500</td>
</tr>
<tr>
<td>RRNH</td>
<td>0.00090</td>
<td>0.00092</td>
<td>0.00180</td>
<td>0.01520</td>
</tr>
</tbody>
</table>

In Table 1, the averaged quadratic-norm errors for the subspace algorithm [17], the regularized nuclear norm heuristic (RNH) [20] and the RRNH in Eq. (24) calculated over 100 estimated models are shown for $\epsilon$ ranging from $10^{-4}$ to $10^{-1}$ and $\lambda = 2000$, $N = 32$, $\delta = 100$. The number of iterations is three for each estimated model. The RRNH outperforms the other two competitors at high signal-to-noise ratios (SNR). The numerical study in [17] showed that the subspace algorithm is more effective than the RNH in suppressing noise at high SNRs while the latter prevails at low SNRs. To recapitulate [20] and the above observations, the RNH appears to be more accurate and robust than the subspace estimator at low SNRs while at high SNRs by iterating the RNH a few times, an estimator which outperforms the subspace estimator is derived.

The positivity of the estimated spectra above was ensured by using the recently developed scheme in [10]. This scheme
implements a constrained version of Eq. (1) with the hyperparameter values fixed as \( \delta = 0.1 \) and \( \lambda = 10,000 \).

\[ \delta = 0.1 \quad \lambda = 10,000 \]

B. Modeling of acoustic spectra

In this example, we consider modeling of acoustic noise spectra for fault detection and health monitoring in induction motors. See, [23] and the references therein for details. A two-stage algorithm was proposed in [24] to estimate a discrete-time cross or auto power spectral density function matrix from time domain measurements. In the first stage, the data are divided into \( L \) blocks of \( M \) measurements each. Then, the modified periodograms are computed and averaged over these blocks for \( M/2 + 1 \) equidistantly spaced frequencies in \( [0, \pi] \). The resulting auto power spectral density estimator is the so-called Welch estimator. In the second stage, a rational model in the state-space form is extracted from the samples of the Welch estimator by using the identification algorithm developed in [17].

We would like to use the RRNH to transform a non-positive acoustic noise spectrum estimated by the two-stage algorithm using the sound measurements from one microphone into a positive one. In Figure 3, the Welch spectrum estimate and the spectral summand estimated by the subspace algorithm are plotted for \( M = 640, L = 400, p = 320, \) and \( \Gamma = I \). The order of the non-positive real spectral summand is 9. It was modified by using the RRNH with \( \delta = 0.1 \) and \( \lambda = 10,000. \) The number of iterations was three. In Figure 4, the modified power spectrum is plotted.

When applied to this spectral summand estimate, the semi-definite programming (SDP) method proposed in [25] yields gross estimation error as displayed in Figure 5. Motor faults can be detected by monitoring changes in the estimated power spectrum around the peak frequencies. In structural vibration monitoring, a plethora of fault detection methods is based on monitoring changes in eigenstructures.

On the same data set, the RRN\( \text{H} \) was also used to estimate a spectral summand and a spectral factor for \( M = 160, L = 1600, p = 80, \) and \( \Gamma = \text{diag} \left( \delta^0, \ldots, \delta^p \right) \). The order of the non-positive spectral summand estimates was 4 and the hyper-parameters chosen in the modification process were \( \delta = 0.1 \) and \( \lambda = 10,000 \). In estimating a spectral summand for the RRNH, \( \delta = 0.1 \) and \( \lambda = 2,000 \) were used as hyperparameter values. As expected, the RRNH and the subspace algorithm deliver almost identical estimates since the misfit error is heavily penalized. In Figures 6 and 7, the estimation results are plotted for the RRNH. Observe that even if the user defined parameters and the peak amplitudes change, the locations of the peak frequencies remain fixed in all figures.

V. Conclusions

In this paper, we studied identification of multi-input/multi-output, discrete-time, linear time-invariant systems from uniformly spaced spectrum measurements by subspace algorithms. A critical step in these algorithms is splitting of invariant spaces in order to determine model order and the observability range space. We proposed a new model order selection and the splitting procedure based on the RRNH. A numerical example demonstrated the efficacy of the proposed procedure over the popular SVD-based approaches and the more recent regularized nuclear norm heuristic, in particular, for high signal-to-noise ratios situations. In a real-life example, where the state-of-art SDP method failed, the RRNH performed remarkably well in enforcing positivity of the estimated spectrum. Parallel conclusions with the numerical example were drawn with respect to accuracy of the estimated spectra.

ACKNOWLEDGMENTS

This work was supported by the Scientific & Technological Research Council of Turkey under Grant 112E264.

REFERENCES

Fig. 5. The power spectrum estimate of the microphone in Example 2, derived from the spectral summand estimate using the SDP method in [25].

Fig. 6. The spectral summand estimates of the microphone in Example 2: “dot” the Welch method; “rectangle” the RRNH. The latter was derived from the spectral summand estimate using the RRNH.

Fig. 7. The power spectrum estimates of the microphone in Example 2: “dot” the Welch method; “triangle” the RRNH.