EUSIPCO 2013 1569743059

SUBSPACE-BASED SPECTRUM ESTIMATION BY NUCLEAR NORM MINIMIZATION

Hüseyin Akçay, 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, linear-time-invariant systems from noisy spectrum samples. In these methods, a critical step is splitting of two invariant subspaces associated with causal and non-causal eigenvalues of some structured matrices built from spectrum measurements via singular-value decomposition in order to determine model order. Mirror image symmetry with respect to the unit circle between the eigenvalue sets of the invariant spaces, required by these algorithms, is lost due to low signal-to-noise ratio, unmodelled dynamics, and insufficient amount of data. Consequently, the choice of model order is not straightforward. In this paper, we propose a robust model order selection scheme based on regularized nuclear norm optimization in combination with a recent subspace algorithm, which uses non-uniformly spaced, in frequencies, spectrum measurements. A simulation example shows the effectiveness of the proposed scheme to large amplitude noise over short data records. Then, the proposed scheme is used to design a linear-shape filter for random road excitations.

Index Terms— spectrum estimation, subspace method, nuclear norm, frequency-domain

1. INTRODUCTION

Subspace methods are popular to obtain low order state-space models of multi-input/multi-output, linear-time-invariant systems from noise corrupted time or frequency-domain measurements. There is an extensive literature on the topic, and the reader is referred to the monograph [1] for a review of its foundations and engineering applications.

The focus of this paper is the estimation of power-spectral density matrix from noise corrupted spectrum samples at non-uniformly spaced frequencies. In [2], a subspace-based identification algorithm was presented. The main idea behind this algorithm is that the range space of a matrix built from frequency-shifted and weighted spectrum samples is exactly the linear span of the extended observability matrix associated with causal and anti-causal components of the power-spectral density matrix in a given state-space realization.

The algorithms proposed in [3, 4] use spectrum measurements on 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 spectrum samples. Under some mild noise assumptions, these algorithms are consistent. An extension to the irregularly spaced frequencies case was reported in [2] where the consistency holds only if corruptions in the measurements have a known bounded covariance function.

In these algorithms, model order is determined by inspecting singular values of a particular matrix used to extract the observability range space. If n is the true order, then the $2n$ most significant singular values and the corresponding left and right singular vectors have to be 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 eigenvalue sets of the causal and the anti-causal invariant spaces of a state-transition matrix. Under this assumption, causal eigenvalues can be obtained by a Jordan decomposition [2, 3, 4].

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 step is inconclusive since the assumed symmetry relation between the eigenvalues of the invariant spaces does not hold. A two-stage identification algorithm was proposed in [5]. The first stage of this algorithm provides an initial estimate to a parametric optimization problem of the second stage by using an asymptotic form of the subspace identification algorithm proposed in [3]. 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].

Nuclear norm optimization methods for structured low-rank matrix approximation have been discussed in several recent papers on system identification [6, 7, 8]. The nuclear norm of a matrix-valued function as a convex heuristic for minimizing its rank was first proposed in [6]. 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 [9]. This approach preserves linear structure in matrix approximation unlike the singular-value decomposition. 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 ap-

proximation problems in time-domain settings. An extension to frequency-domain was recently made in [7]. In this work, the subspace algorithm developed in [10] to identify multivariable systems from measured frequency response at uniformly spaced frequencies was re-examined from a model validation perspective using a nuclear norm heuristic.

The contents of this paper is as follows. In Section 2, power spectrum estimation in frequency-domain via the algorithm proposed in [2] is reviewed. In Section 3, a variation of this algorithm based on the regularized nuclear norm heuristic is presented. In Section 4, first a simulation example is used to demonstrate that the proposed scheme is effective in determining model orders over short data lengths and robust to noise. Then, in an application study, the proposed scheme is used to design a linear-shape filter for random road excitations. Section 5 concludes the paper.

2. POWER SPECTRUM ESTIMATION IN FREQUENCY-DOMAIN BY SUBSPACE METHODS

Suppose we are given N noisy samples $S_k \in \mathbf{C}^{m\times m}$ of the power spectrum $S(z)$ of a linear-time-invariant discrete-time system with m-inputs and m-outputs evaluated at N points on the unit circle:

$$S_k = S(e^{j\theta_k}) + \eta_k, \qquad k = 1, 2, \cdots, N \tag{1}$$

and would like to estimate $\hat{G}(z) = \hat{C}(zI - \hat{A})^{-1}\hat{B} + \hat{D}$ such that the estimated spectrum $\hat{S}(z) = \hat{G}(z)\hat{G}^T(z^{-1})$ evaluated at $e^{j\theta_k}$ is as close to S_k as possible for all k.

The transfer function $\hat{G}(z)$, called the spectral factor of $S(z)$, is required to meet additional criteria. Among them is the absence of poles and zeros outside the open unit disk. Determination of a suitable model order, in particular the lowest possible one, is also part of the estimation problem. Under mild noise and system assumptions, the estimated spectrum should also converge to the true spectrum as $N \to \infty$ when the true spectrum is in the model class. Subspace algorithms with these properties have been proposed [2, 4].

We will briefly outline the steps of the subspace algorithm introduced in [2] relevant to our discussion. Let $\{A, B, C, D\}$ be minimal realization of a square, strictly minimum-phase transfer function matrix $G(z)$ with McMillan degree n. We assume that all the eigenvalues of A are nonzero and distinct.

The identification algorithms in [2, 3, 4] begin by observing that a power spectrum can be written in terms of the so-called spectral summand $H(z) = C(zI_n - A)^{-1}F + \frac{E}{2}$ and its conjugate transpose as $S(z) = H(z) + H^T(z^{-1})$ for a symmetric $E \in \mathbf{R}^{m\times m}$, reducing the problem to identifying a spectral summand directly from the spectrum samples.

Let p be any fixed integer parameter satisfying $p > 2n$. From the data, form the matrices

$$\begin{aligned}
\hat{\mathcal{S}}_{\mathrm{C}} &= \frac{1}{\sqrt{N}}\begin{bmatrix} S_1 & \cdots & S_N \\ \vdots & \ddots & \vdots \\ z_1^{p-1}S_1 & \cdots & z_N^{p-1}S_N \end{bmatrix}, \qquad (2)\\
\mathcal{W}_{\mathrm{C}} &= \frac{1}{\sqrt{N}}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ z_1^{p-1} & \cdots & z_N^{p-1} \end{bmatrix} \otimes I_m,\\
\hat{\mathcal{S}} &= [\mathrm{Re}(\hat{\mathcal{S}}_C)\ \ \mathrm{Im}(\hat{\mathcal{S}}_C)], \qquad (3)\\
\mathcal{W} &= [\mathrm{Re}(\mathcal{W}_C)\ \ \mathrm{Im}(\mathcal{W}_C)]
\end{aligned}$$

where $z_k = e^{j\theta_k}$, $k = 1, 2, \cdots, N$. The extended observability matrix for $H(z) + z^{p-1}H^T(z^{-1})$ is defined by

$$\mathcal{O}_p = \begin{bmatrix} C & F^T(A^T)^{p-1} \\ \vdots & \vdots \\ CA^{p-1} & F^T \end{bmatrix}.$$

Let $\mathcal{W}^\perp = I_{2mN} - \mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W}$.

When the spectrum measurements are noise-free, the frequencies are distinct, and the number of the frequencies satisfies $N \geq (p/2) + n + 1$, the range space of $\hat{\mathcal{S}}\mathcal{W}^\perp$ equals to the range of $\mathcal{O}_p$ [2]. Thus, within a similarity transformation, the matrices C and F of $H(z)$ can be determined from a singular-value decomposition of $\hat{\mathcal{S}}\mathcal{W}^\perp$. A numerically efficient way of forming $\hat{\mathcal{S}}\mathcal{W}^\perp$ is to use the QR-factorization:

$$\begin{bmatrix} \mathcal{W} \\ \hat{\mathcal{S}} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix}\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}. \tag{4}$$

Then, $\hat{\mathcal{S}}\mathcal{W}^\perp = R_{22}Q_2^T$ and it suffices to use R_{22} in the singular-value decomposition. Hence, $R_{22} = \hat{\mathcal{S}}Q_2$.

Let Γ be a given block-diagonal *positive-definite* matrix

$$\Gamma = \begin{bmatrix} \Gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Gamma_N \end{bmatrix}$$

with $\Gamma_k \in \mathbf{R}^{m\times m}$, $k = 1, \cdots, N$. By the *Cholesky decomposition*, find a matrix $\Psi \in \mathbf{R}^{pm\times pm}$ satisfying

$$\Psi\Psi^T = \mathrm{Re}(\mathcal{W}_{\mathrm{C}}\Gamma\mathcal{W}_{\mathrm{C}}^H). \tag{5}$$

The square-root factor Ψ is used in [2] as a weight matrix. More precisely, the singular-value decomposition is applied to $\Psi^{-1}R_{22}$. See [2] for the details and the outline of the algorithm. Methods to estimate B and D from $\{A, C, E, F\}$ which guarantee positive-realness of the estimated spectrum are discussed in [2] and the references therein. The algorithm proposed in [2] is strongly consistent if η_k is either zero or a zero-mean complex white-noise process with a known covariance function [2]. In the former case, any positive-definite

sequence Γ_k, $k = 1, \cdots, N$ can be chosen, and in the latter case, the covariance function of the noise process must satisfy

$$\mathbf{E} \begin{bmatrix} \operatorname{Re} \eta_k \\ \operatorname{Im} \eta_k \end{bmatrix} [\operatorname{Re} \eta_s^T \;\; \operatorname{Im} \eta_s^T] = \frac{1}{2} \begin{bmatrix} \Gamma_k & 0 \\ 0 & \Gamma_k \end{bmatrix} \delta_{ks}$$

and the fourth-order moments be bounded: $\mathbf{E}\|\eta_k\|_F^4 < C_\eta$.

When the noise covariance function is known and as above, by pre and post multiplying $\hat{S}$ with Ψ^{-1} and Q_2, we obtain a matrix whose range space coincides with the range space of $\mathcal{O}_p$. Both $\Psi^{-1}\hat{S}Q_2$ and $\Psi^{-1}\hat{S}\mathcal{W}^\perp$ contain the same observability range information; but, the former is preferable since its column size, and as a result, computing time to perform the singular-value decomposition remain constant as N grows.

The model order is determined by inspecting the singular values of $\Psi^{-1}\hat{S}Q_2$ 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

$$\begin{bmatrix} A & 0 \\ 0 & (A^T)^{-1} \end{bmatrix} \tag{6}$$

when the data (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 $(A^T)^{-1}$ in (6) may be destroyed. The nuclear norm minimization of a corrupted sparse signal is known to suppress the singular values originating from the corruptions while maintaining the singular values of the sparse signal. The objective of this paper is to propose a new splitting procedure based on the nuclear norm minimization that is insensitive to noise and undermodelling.

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

The key step in the subspace algorithm is the extraction of the extended observability range space $\mathcal{O}_p$ via the SVD. Although the SVD provides low-rank approximation by minimizing the Frobenius norm of $\Psi^{-1}\hat{\mathcal{S}}Q_2$, it does not preserve the structure of $\hat{\mathcal{S}}_{\mathrm{C}}$ in (2), which is related to $\hat{\mathcal{S}}$ via (3). More specifically, since $\hat{\mathcal{S}}_{\mathrm{C}}$ linearly depends on the data, the corruptions $\eta_1, \cdots, \eta_N$ are distributed in $\hat{\mathcal{S}}_{\mathrm{C}}$ according to the pattern on the right-hand side of (2). It was suggested in [8, 6] to minimize the nuclear norm as a heuristic for low-rank approximation problems, which can not be handled via the SVD, in particular, approximation problems with structured low rank matrices and problems including additional constraints. In this section, we will present a variation of the subspace method in Section 2 based on the regularized nuclear norm heuristic.

Decompose the spectrum samples as $S_k = S'_k + S''_k$ where $S'_k = \frac{1}{2}(S_k + \bar{S}_k^T)$ and $S''_k = \frac{1}{2}(S_k - \bar{S}_k^T)$ for $k = 1, \cdots, N$. Then, $S'_1, \cdots, S'_N$ are Hermitian symmetric matrices and can be stacked into a vector $g \in \mathbf{R}^{m^2 N}$. Note that $S''_k = 0$ when $\eta_k = 0$. Omitting the skew-Hermitian parts, $\hat{\mathcal{S}}$ may be considered as a linear parametrization of $S'_1, \cdots, S'_N$ stacked into a parameter vector $\zeta \in \mathbf{R}^{m^2 N}$ formed similarly to g. For fixed Ψ and Q_2 calculated from (4), set $\mathcal{A}(\zeta) = \Psi^{-1}\hat{\mathcal{S}}Q_2$. The following is the basic optimization problem:

$$\text{minimize} \qquad \|\mathcal{A}(\zeta)\|_* + \frac{\lambda}{2}(\zeta - g)^T \mathcal{H}(\zeta - g) \tag{7}$$

to be tackled in this paper where $\mathcal{H}$ is a diagonal matrix with the diagonal entries formed from the sequence $\Gamma_1^{-1}, \cdots, \Gamma_N^{-1}$ similarly to ζ, $\lambda > 0$ is a scalar weight, and the norm on the first term is the nuclear norm defined as the sum of the singular values. The second term is a quadratic penalty on the difference between the measured spectrum and its low rank approximation. Once (7) is solved, $\Psi^{-1}\hat{S}Q_2$ is calculated with the optimal ζ and the rest of the steps in [2] are followed.

The optimization problem (7) will be solved by the alternating direction method of multipliers, a popular method for large scale and distributed convex optimization [11]. It has been demonstrated to be effective for nuclear norm optimization problems arising in system identification and realization.

The regularized nuclear norm optimization (7) may provide accurate results when N is moderate, the Hankel singular values of the true spectral factor decay rapidly, and the signal-to-noise ratio is not high. Under these conditions, the SVD step in the subspace algorithms has been observed to be inconclusive. Besides, the SVD does not respect the structure of $\Psi^{-1}\hat{\mathcal{S}}Q_2$ unlike the explicit parametrization $\mathcal{A}(\zeta)$ used in (7). Numerous studies in the literature have demonstrated that switching to the nuclear norm enforces a sharp transition to a low-rank approximant, thereby making the choice of the model order easier. The purpose of the current paper is to observe a similar behavior.

3.1. Determination of λ

The measurements (1) can be divided into two disjoint sets by putting the measurements and the frequencies with even indices into the estimation data set Λ^{e} and the rest into the validation data set Λ^{v}. For each $\lambda > 0$, the spectral estimate denoted by $\hat{S}_\lambda(z)$ is computed using Λ^{e} and its performance is evaluated by computing the quadratic fit error on Λ^{v}. Then, λ is determined by finding the best value of the quadratic fit error, say for 50 logarithmically-spaced values of λ in the interval 10^{-3} to 10^3. This scheme was suggested in [8] to use in the time-domain subspace identification algorithms.

4. EXAMPLES

In this section, we use a simulation example to evaluate the regularized nuclear norm heuristic. The second example is concerned with the identification of road spectra.

4.1. Simulation example

Let the true system $G(z)$ be a fourth-order system described by the state-space model [4]:

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

$$B = [0.2247 \;\; 0.8989 \;\; 0.0323 \;\; 0.1290]^T,$$

$$C = [0.4719 \;\; 0.1124 \;\; 9.6774 \;\; 1.6129], \quad D = 0.9626.$$

Assume that the corruptions in (1) are given as

$$\eta_k = \varepsilon \frac{0.2z_k^2 - 0.0904z_k + 0.1839}{z_k^2 - 1.1111z_k + 0.8520}\nu_k$$

where ε is the noise amplitude, ν_k are zero-mean, unit-variance, independent, identically distributed complex normal random variables, and θ_k, $k = 1, \cdots, N$ are selected randomly and independently from the intervals $[\pi(k-1)/N, \pi k/N]$.

In Figure 1, the singular values of $\Psi^{-1}\hat{S}Q_2$ for Monte Carlo simulations of the subspace algorithm and (7) over 100 noise realizations with $\lambda = 1$ (determined by the procedure in subsection 3.1), $\varepsilon = 10$, $N = 250$, and $p = 125$ are plotted. Observe that the singular values with indices larger than 4 are mixed in the subspace algorithm whereas the two bands containing the first 8 largest singular values are separated from the rest in the regularized nuclear norm heuristic. This mixing influences the eigenvalue distribution shown in Figure 2.

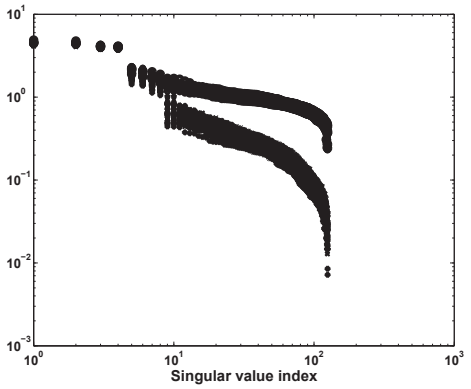

Fig. 1. Monte Carlo simulations showing the singular values of $\Psi^{-1}\hat{S}Q_2$: "o" subspace algorithm; "x" (7).

In Table 1, Monte Carlo simulations comparing the average value of the quadratic errors over 100 noise realizations are shown. The effectiveness of the proposed scheme is clearly visible for all $\varepsilon \geq 1$ even for such a small N.

4.2. Identification of road spectra

In this subsection, we consider a practical application of the regularized nuclear nuclear norm optimization. In Figure 3,

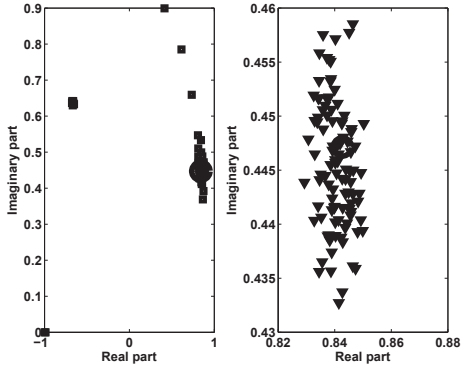

Fig. 2. Monte Carlo simulations showing the maximum modulus eigenvalues of $\hat{A}$ for the subspace algorithm (left) and (7) (right). The maximum modulus eigenvalue of A are shown by a circle.

Table 1. Monte Carlo simulations over 100 noise realizations comparing $\|\hat{S} - S\|_{\mathrm{m},2}$ for the subspace algorithm and (7).

ε	0.1	1	5	10	15
Subspace	0.0066	0.0662	0.3499	1.5691	3.1876
Heuristic	0.0055	0.0547	0.2723	0.5543	1.2929

the spectral density of a typical road and its split power law approximation [12]:

$$S_c(2\pi\Omega) = \begin{cases} \gamma|\Omega/\Omega_0|^{-2\delta_1}, & 0 < |\Omega| < \Omega_0; \\ \gamma|\Omega/\Omega_0|^{-2\delta_2}, & \Omega_0 \leq |\Omega| < \infty \end{cases}$$

obtained by trial and error for $\gamma = 0.76 \times 10^{-5}$ m^2/cycle/m, $\Omega_0 = 0.15708$ cycles/m, $\delta_1 = 1.6$, and $\delta_2 = 1.1$ are plotted. In the figure, we also show the integrated white noise approximation to the data: $\gamma|\Omega/\Omega_0|^{-2}$, which is commonly used in the stochastic road modeling studies. Clearly, the fit by the integrated white-noise is rather poor; in particular at the frequencies below Ω_0 and the problem with the split power approximation is that it can not be generated by rational linear shape-filters. Hence, it is not suitable for simulating stochastic response of a vehicle. Besides, it is unbounded at $\Omega = 0$.

We seek a low order linear-shape filter whose output spectrum matches the spectral data plotted in Figure **??** as closely as possible when driven by white noise. In Figure 4, the road spectra estimated by the subspace method and the regularized nuclear norm optimization scheme are plotted. They show that the road spectrum can accurately be modeled as the outputs of rational linear-shape filters driven by white noise. In particular, the shape filter designed by the proposed scheme is more accurate than the shape filter designed by the algorithm in [2]; yet, its complexity is less. In the Cholesky decomposition, S_k was substituted in place of Γ_k for all k. A linear-shape filter can then be used, for example, to study the response of the vehicle to random road inputs [13].

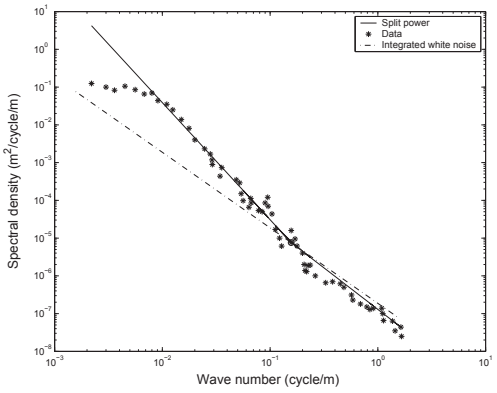

Fig. 3. The road power spectrum [12] and its approximate modeling by the split power law and the integrated white noise.

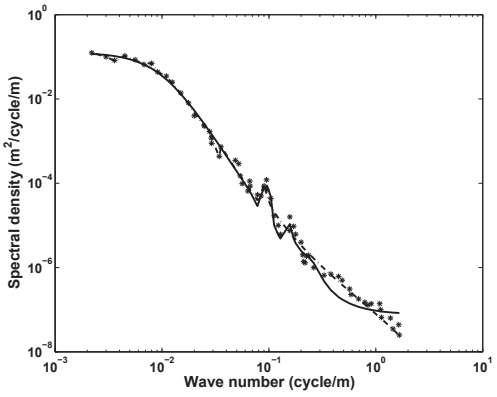

Fig. 4. The spectral data and its modeling by rational spectra: (*) the road data; (-.) the 10^{th} order model estimated by the subspace algorithm; (-) the 6^{th} order model estimated by (7).

5. CONCLUSIONS

In this paper, we studied identification of multi-input/multi-output, discrete-time, linear-time-invariant systems by a recent subspace algorithm from spectrum measurements at irregularly spaced frequencies. A numerically sensitive step in this algorithm and other subspace algorithms was the splitting of the invariant subspaces associated with the causal and anti-causal components of the power 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 effectiveness of the proposed scheme to large amplitude noise over short data records. In an application study, the proposed scheme was used to design a linear-shape filter for random road excitations.

6. REFERENCES

[1] T. Katayama, *Subspace Methods for System Identification*, Springer-Verlag, London, 2005.

[2] H. Akçay and S. Türkay, "Frequency domain subspace-based identification of discrete-time power spectra from nonuniformly spaced measurements," *Automatica*, vol. 40, pp. 1333–1347, 2004.

[3] P. Van Overschee, B. De Moor, W. Dehandschutter, and J. Swevers, "A subspace algorithm for the identification of discrete-time frequency domain power spectra," *Automatica*, vol. 33, pp. 2147–2157, 1997.

[4] H. Akçay, "Frequency domain subspace-based identification of discrete-time power spectra from uniformly spaced measurements", *Automatica*, vol. 47, pp. 363–367, 2011.

[5] K. Hinnen, M. Verhaegen, and N. Doelman, "Robust spectral factor approximation of discrete-time frequency domain power spectra," *Automatica*, vol. 41, pp. 1791–1798, 2005.

[6] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system identification," in: Proc. *ACC*, 2001, pages 4734–4739.

[7] R. S. Smith, "Nuclear norm minimization methods for frequency domain subspace identification," in: Proc. *ACC*, 2012, pages 2689–2694.

[8] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *SCL*, vol. 62, pp. 605–612, 2013.

[9] E. J. Candes and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2053–2080.

[10] T. McKelvey, H. Akçay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE TAC*, vol. 41, pp. 960–979, 1996.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Found. Trends Machine Learn.*," vol. 3, pp. 1–122, 2010.

[12] C. J. Dodds and J. D. Robson, "The description of road surface roughness," *JSV*, vol. 31, pp. 175–183, 1973.

[13] S. Türkay and H. Akçay, "A study of random vibration characteristics of the quarter-car model," *JSV*, vol. 282, pp. 111–124, 2005.