Semidefinite programming methods for system realization and identification

Zhang Liu and Lieven Vandenberghe

***Abstract*— We describe semidefinite programming methods for system realization and identification. For each of these two applications, a variant of a simple subspace algorithm is presented, in which a low-rank matrix approximation is computed by minimizing the nuclear norm (sum of singular values) of a structured matrix. This technique preserves linear matrix structure in the low-rank approximation, an important advantage over standard approaches based on the singular value decomposition.**

I. Introduction

The basic optimization model used in the paper is the *nuclear norm approximation* problem

$$\text{minimize} \quad \|\mathcal{A}(x) - B\|_*. \tag{1}$$

Here $B \in \mathbf{R}^{p\times q}$ is a given matrix and

$$\mathcal{A}(x) = x_1A_1 + x_2A_2 + \cdots + x_nA_n$$

is a linear mapping from $\mathbf{R}^n$ to $\mathbf{R}^{p\times q}$. The norm $\|\cdot\|_*$ denotes the nuclear norm on $\mathbf{R}^{p\times q}$, *i.e.*, $\|X\|_*$ is the sum of the singular values of X. We will also encounter extensions of (1) with a quadratic regularization term in the objective,

$$\text{minimize} \quad \|\mathcal{A}(x) - B\|_* + (x-\hat{x})^TQ(x-\hat{x}), \tag{2}$$

where $Q \succeq 0$. Problems (1) and (2) are convex optimization problems that can be expressed and solved as semidefinite programs (SDPs).

The nuclear norm approximation problem is of interest as a convex heuristic for the *rank minimization* problem

$$\text{minimize} \quad \mathbf{rank}(\mathcal{A}(x) - B),$$

which is NP-hard in general. This heuristic was first studied by Fazel, Hindi, and Boyd [FHB01], and is based on the fact that the residual $\mathcal{A}(x)-B$ at the optimal solution of (1) typically has low rank. The idea is very useful for a variety of applications that require low-rank approximations of structured matrices; see [FHB01], [Faz02], [CFP03], [FHB04], [RFP07] for surveys of applications.

The nuclear norm of a diagonal matrix is the ℓ_1-norm of its diagonal, and nuclear norm techniques can be viewed as a matrix extension of ℓ_1-norm methods for sparse approximation and cardinality minimization [Don06], [CT05], [CRT06], [Tro06]. Some of the theoretical results that characterize the possibility of exact recovery of sparse signals by ℓ_1-norm methods have recently been extended to nuclear norm methods [RFP07], [CR08], [RXH08].

The purpose of the paper is to examine the effectiveness of the nuclear norm heuristic for rank minimization problems in system realization and identification, and to compare the results with standard subspace algorithms. In section II we compare two methods for *stochastic realization*, the problem of estimating a state-space model of an ARMA process from estimates of the process covariances. The first method is a subspace algorithm that uses the singular value decomposition (SVD) to approximate the Hankel matrix constructed from the covariance estimates by a (non-Hankel) matrix of low rank. In the second method, regularized nuclear norm minimization is used to adjust the covariance estimates to obtain a Hankel matrix of low rank. We present results of an experiment in which the covariance estimates are sample averages for different data lengths. Although we do not observe exact recovery of the covariances for finite data lengths, the nuclear norm method results in more accurate and robust estimates than the subspace method.

In section III we discuss an application in system identification. Again, we compare two methods, a subspace method that uses the singular value decomposition to find a low-rank approximation of a matrix constructed from input-output Hankel matrices, and a method based on regularized nuclear norm approximation. Experiments with benchmark data from [DDDF97] suggest that the nuclear norm method is more accurate and offers a clearer model order selection criterion.

In the numerical experiments we use the interior-point method described in [LV09]. This is a primal-dual semidefinite programming algorithm, customized to exploit problem structure in the SDP formulation of problems (1) and (2). For problem (1), the SDP formulation is

$$\begin{array}{ll} \text{minimize} & (\mathbf{tr}\,U + \mathbf{tr}\,V)/2 \\ \text{subject to} & \left[\begin{array}{cc} U & (\mathcal{A}(x)-B)^T \\ \mathcal{A}(x)-B & V \end{array}\right] \succeq 0, \end{array} \tag{3}$$

with variables $x \in \mathbf{R}^n$, $U \in \mathbf{S}^q$, $V \in \mathbf{S}^p$. (We use $\mathbf{S}^n$ to denote the set of symmetric matrices of order n.) It can therefore be solved by general software for semidefinite programming. Unfortunately the SDP (3) has $n + p(p+1)/2 + q(q+1)/2$ variables and is very difficult to solve by general-purpose solvers, if p and q approach 100. This limits the use of the nuclear norm heuristic in practice. A more efficient interior-point method that exploits the problem structure in (3) and is capable of solving problems with dimensions p, q on the order of several hundred, is described in [LV09]. In this algorithm, the cost per iteration of the

Electrical Engineering Department, University of California, Los Angeles. Email: zhang@ee.ucla.edu, vandenbe@ee.ucla.edu. Research supported by NSF under grants ECCS-0524663 and ECCS-0824003 and a Northrop Grumman Ph.D. Fellowship.

interior-point method is reduced to $O(pqn^2)$ operations, if $n \geq \max\{p, q\}$. This is comparable to solving the norm approximation problem in least-squares sense, *i.e.*, to minimizing the Frobenius norm of $\mathcal{A}(x) - B$. More details and a discussion of other algorithms can be found in [LV09].

II. Minimum rank Hankel matrix completion

We will use the notation $H_{(i,j,k)}$ to denote the block Hankel matrix with matrices H_i, ..., H_j in its first block column and H_i, ..., H_k in its first block row:

$$H_{(i,j,k)} = \begin{bmatrix} H_i & H_{i+1} & \cdots & H_k \\ H_{i+1} & H_{i+2} & \cdots & H_{k+1} \\ \vdots & \vdots & & \vdots \\ H_j & H_{j+1} & \cdots & H_{k+j-i} \end{bmatrix}.$$

In this section we discuss two rank minimization problems involving Hankel matrices. The first problem is to find the minimum rank completion of $H_{(1,N,N)}$ if the matrices H_1, ..., H_N are specified, and the matrices H_{N+1}, ..., H_{2N-1} are free variables. The solution is given by the minimum order system realization of the sequence H_1, ..., H_N. In the second part of the section, we discuss an extension in which matrices $\bar{H}_1$, ..., $\bar{H}_N$ are given, and we seek approximations $H_k \approx \bar{H}_k$, $k = 1, \ldots, N$, that have a low-rank Hankel completion. More specifically, we consider the problem

$$\text{minimize} \quad \mathbf{rank}\, H_{(1,N,N)} + \lambda \sum_{k=1}^{N} \|H_k - \bar{H}_k\|_F^2 \qquad (4)$$

with H_1, ..., H_{2N-1} as variables, and λ a positive weight used to trade-off the two objectives. ($\|\cdot\|_F$ denotes Frobenius norm.) In Section II-B.2 we will study the effectiveness of the nuclear-norm heuristic for (4),

$$\text{minimize} \quad \|H_{(1,N,N)}\|_* + \lambda \sum_{k=1}^{N} \|H_k - \bar{H}_k\|_F^2, \qquad (5)$$

in the context of stochastic realization.

A. Minimum partial realization

A fundamental problem in linear system theory is the *minimal partial realization problem:* given a sequence of matrices H_k, $k = 1, \ldots, N$, find a minimal state-space model (A, B, C) such that $H_k = CA^{k-1}B$. The realization problem has been studied since the 1960s, and many different algorithms are described in the literature. These methods are based on checking the rank of Hankel matrices constructed from the sequence H_k [Tet70], [Kal71], [Sil71]; see also [Wil86, page 689] and the survey paper [Sch00].

The minimum partial realization problem can also be formulated as a rank minimization problem [FHB03]. Suppose H_1, ..., $H_N \in \mathbf{R}^{p\times m}$ are given, and let $n^\star$ be the optimal value of the rank minimization problem

$$\text{minimize} \quad \mathbf{rank}\, H_{(1,N,N)}, \qquad (6)$$

with variables H_{N+1}, ..., H_{2N-1}. In this problem, the top left block-triangular part of an $N \times N$ block Hankel matrix is specified; the problem is to complete the bottom part so that the rank is minimized. It can be shown that $n^\star$ is equal to the order of the minimal realization of H_1, ..., H_N. This was proved for $p = m = 1$ in [FHB03]. For completeness we include a general proof below.

First, we can note that no realization of order less than $n^\star$ exists. If $H_k = CA^{k-1}B$, $k = 1, \ldots, N$, with A of order less than $n^\star$, then taking $H_k = CA^{k-1}B$ for $k = N+1, \ldots, 2N-1$ gives a completion $H_{(1,N,N)}$ with rank less than $n^\star$. This contradicts the definition of $n^\star$.

The less obvious part of the theorem is that there exists a realization with order $n^\star$. Such a realization can be constructed as follows. For $i = 1, \ldots, p$, define $n_i = 0$ if row i of $H_{(1,1,N)}$ is linearly dependent on rows $1, \ldots, i-1$ of $H_{(1,1,N)}$. Otherwise define n_i as the largest $k \in \{1, \ldots, N\}$ such that row i of block row k of $H_{(1,k,N-k+1)}$ is linearly independent of all the rows of $H_{(1,k,N-k+1)}$ that precede it. (Note that all the Hankel matrices in this definition depend only on H_1, ..., H_N, and hence are completely specified.) Define

$$n = n_1 + \cdots + n_p.$$

Then $n^\star \geq n$, because in any completion of $H_{(1,N,N)}$, the n rows $j + (k-1)p$, $j = 1, \ldots, p$, $k = 1, \ldots, n_j$ are linearly independent. To show that $n^\star = n$, we construct a completion with rank n. Without loss of generality we will assume that $n_i > 0$ for $i = 1, \ldots, p$. Following the notation of [Gui81, §3], define

$$n_{ij} = \begin{cases} \min\{n_i + 1, n_j\} & j < i \\ n_i & j = i \\ \min\{n_i, n_j\} & j > i. \end{cases}$$

If $n_i < N$, then, by definition of the n_i's, we can express row i of block row $n_i + 1$ of $H_{(1,n_i+1,N-n_i)}$ as a linear combination of rows $j + p(k-1)$, $j = 1, \ldots, p$, $k = 1, \ldots, n_{ij}$. Therefore there exist coefficients α_{ijk}, $i, j = 1, \ldots, p$, $k = 1, \ldots, n_{ij}$, that satisfy

$$\begin{aligned} & e_i^T \begin{bmatrix} H_{n_i+1} & \cdots & H_N \end{bmatrix} \\ & \quad = \sum_{j=1}^{p} \sum_{k=1}^{n_{ij}} \alpha_{ijk} e_j^T \begin{bmatrix} H_k & \cdots & H_{k+N-n_i-1} \end{bmatrix}, \end{aligned}$$

where e_i is the ith unit vector in $\mathbf{R}^p$. If $n_i = N$, the coefficients α_{ijk}, $i, j = 1, \ldots, p$, $k = 1, \ldots, n_{ij}$, are chosen arbitrarily. Now define the following model (A, B, C) of order n:

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1p} \\ \vdots & & \vdots \\ A_{p1} & \cdots & A_{pp} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix},$$

$$C = \begin{bmatrix} C_1 & \cdots & C_p \end{bmatrix}$$

where

$$B_j = \begin{bmatrix} e_j^T H_1 \\ e_j^T H_2 \\ \vdots \\ e_j^T H_{n_j} \end{bmatrix} \in \mathbf{R}^{n_j \times m},$$

$$C_j = \begin{bmatrix} e_j & 0 & \cdots & 0 \end{bmatrix} \in \mathbf{R}^{p \times n_j},$$

$$A_{ii} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_{ii1} & \alpha_{ii2} & \alpha_{ii3} & \cdots & \alpha_{iin_i} \end{bmatrix} \in \mathbf{R}^{n_i \times n_i},$$

and

$$A_{ij} = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \alpha_{ij1} & \cdots & \alpha_{ijn_{ij}} & 0 & \cdots & 0 \end{bmatrix} \in \mathbf{R}^{n_i \times n_j}$$

for $i \neq j$. It can be verified that $H_k = CA^{k-1}B$ for $k = 1, \ldots, N$. Defining $H_k = CA^{k-1}B$ for $k = N+1, \ldots, 2N-1$ gives a Hankel completion of rank n. This shows that there exists a realization of $H_1, \ldots, H_N$ with order equal to n and that $n^\star = n$.

B. Stochastic realization

We now turn to the regularized minimum rank Hankel completion problem (4) and the convex heuristic (5) for it, applied to a stochastic realization problem. Consider a state space model of an autoregressive moving-average (ARMA) process $y(t) \in \mathbf{R}^p$:

$$\begin{array}{rcl} x(t+1) & = & Ax(t) + Be(t), \\ y(t) & = & Cx(t) + e(t), \end{array} \tag{7}$$

where $x(t) \in \mathbf{R}^n$ and $e(t)$ is white noise with covariance Q. The process covariances $H_k = \mathbf{E}(y(t+k)y(t)^T)$ are given by

$$H_0 = CPC^T + Q, \qquad H_k = CA^{k-1}D, \quad k \geq 1, \tag{8}$$

where $D = APC^T + BQ$, and $P = \mathbf{E}(x(t)x(t)^T)$ satisfies the Lyapunov equation

$$P = APA^T + BQB^T$$

(see [SM97], [Bal95]). In a stochastic realization problem, we are given estimates $\bar{H}_0, \bar{H}_1, \ldots, \bar{H}_N$ of the covariance matrices H_k (for example, from sample averages of a sequence of observations $y(t)$), and are asked to estimate the model parameters n, A, B, C, Q.

1) Subspace algorithm: We first briefly review a subspace method for stochastic realization [MSM00]. Subspace methods for stochastic realization are based on the fact that the block Hankel matrix of process covariances $H_{(1,r,s)}$ has rank n, if $r \geq n$, $s \geq n$, and that it can be factored as

$$H_{(1,r,s)} = FG, \tag{9}$$

where

$$F = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix}, \quad G = \begin{bmatrix} D & AD & \cdots & A^{s-1}D \end{bmatrix}.$$

Given exact covariances H_k, we can therefore compute a state-space model by first factoring the Hankel matrix as in (9), and then computing A, C, D from F and G. The matrices Q and B can be found by solving the algebraic Riccati equation

$$P = APA^T + (D - APC^T)(H_0 - CPC^T)^{-1}(D - APC^T)^T \tag{10}$$

and taking

$$Q = H_0 - CPC^T, \qquad B = (D - APC^T)Q^{-1}.$$

Now suppose we are given a sequence of approximate covariances $\bar{H}_0, \ldots, \bar{H}_{r+s-1}$. In a subspace algorithm the singular value decomposition of the block-Hankel matrix with blocks $\bar{H}_k$ is used to estimate the model order and make a low-rank approximation

$$\bar{H}_{(1,r,s)} \approx \hat{F}\hat{G}. \tag{11}$$

We take $\hat{C} = \hat{F}_1$ (the first block of $\hat{F}$), and compute $\hat{A}$ from the least-squares problem

$$\text{minimize} \quad \|\hat{F}_{(1,r-1,1)}\hat{A} - \hat{F}_{(2,r,1)}\|_F^2. \tag{12}$$

If necessary, the solution $\hat{A}$ is corrected to make it stable, typically by inverting the unstable eigenvalues. An alternative method proposed in [MSM00] computes an approximate stable $\hat{A}$ by solving an SDP. The matrix $\hat{D}$ can be computed from the least-squares problem, by minimizing the norm of

$$\begin{bmatrix} \bar{H}_1 \\ \bar{H}_2 \\ \vdots \\ \bar{H}_{r+s-1} \end{bmatrix} - \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \\ \hat{C}\hat{A}^{r+s-2} \end{bmatrix} \hat{D}. \tag{13}$$

The state covariance $\hat{P}$ can be computed from the Riccati equation (10), with $\hat{A}$, $\hat{C}$, $\hat{D}$, $\bar{H}_0$, $\hat{P}$ substituted for A, C, D, H_0, P. Lastly, we can compute $\hat{Q}$ and $\hat{B}$ from

$$\hat{Q} = \bar{H}_0 - \hat{C}\hat{P}\hat{C}^T, \qquad \hat{B} = (\hat{D} - \hat{A}\hat{P}\hat{C}^T)\hat{Q}^{-1}.$$

The algebraic Riccati equation (10) may fail to have a positive definite solution, and [MSM00] proposed a remedy to deal with such situations.

2) Nuclear norm optimization: The key step in the subspace algorithm is the low-rank approximation (11). As an alternative which preserves Hankel structure, we can use the regularized nuclear norm approximation problem (5). From the optimal solution H^* of (5), we can then find model parameters $\hat{A}$, $\hat{B}$, $\hat{C}$, and $\hat{Q}$ as in the subspace method.

3) Numerical experiment: We generate data using an ARMA process with transfer function

$$\frac{z^5 + 1.7z^4 - 4.0z^3 + 2.4z^2 - 0.86z + 0.27}{z^5 - 1.4z^4 + 0.66z^3 - 0.16z^2 + 0.023z + 0.012},$$

and Gaussian noise with input covariance $Q = I$. The sample averages

$$\bar{H}_k = \frac{1}{M}\sum_{t=1}^{M-k} y(t+k)y(t)^T. \tag{14}$$

are used as covariance estimates. Table I shows the results for the different methods. All values in the table are averaged

M	Sample est. error	SVD $P_{\rm us}$	SVD error	Nuclear norm error
1000	0.622	54%	0.526	0.282
5000	0.266	48%	0.232	0.158
20000	0.136	32%	0.115	0.088

TABLE I

COMPARISON OF STOCHASTIC REALIZATION METHODS

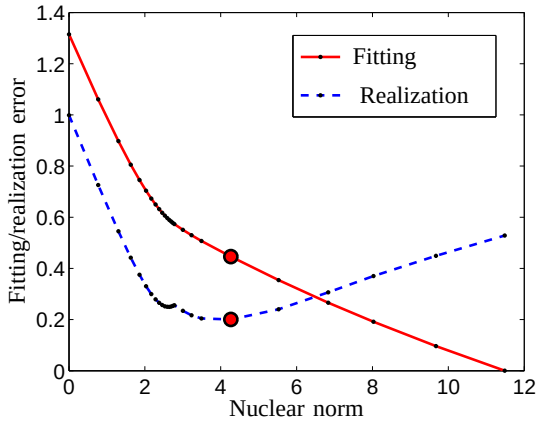

Fig. 1. Trade-off curve between the fitting/realization error and the nuclear norm of $H_{(1,N,N)}$ computed by solving (5) for a range of different λ.

over 50 simulations. M is the length of the sample.

The error in column 2 is the normalized average of $\sum_{k=1}^{10}(\bar{H}_k - H_k^{\rm true})^2$, where $\bar{H}_k$ are the sample covariances (14), and $H_k^{\rm true}$ are the true covariances. The error in column 4 is the normalized average of $\sum_{k=1}^{10}(\hat{H}_k - H_k^{\rm true})^2$, where $\hat{H}_k$ are the covariances computed from the state-space realization $\hat{A}$, $\hat{C}$, $\hat{D}$ returned by the SVD subspace method, with $r = s = 10$. $P_{\rm us}$ is the percentage of the 50 instances where $\hat{A}$ from (12) was unstable. In these cases, the matrix was stabilized by inverting its unstable eigenvalues. All system orders are selected to be 5 for a fair comparison. The last column, labeled 'Nuclear norm' shows the error for the algorithm based on nuclear norm optimization in section II-B.2, with $N = 10$. In this experiment the method based on the nuclear norm optimization always produced a stable matrix $\hat{A}$. We also note that the errors in the covariance estimates were smaller than in the SVD method. Figures 1 and 2 show plots for an instance with $M = 1000$. Figure 1 shows the trade-off curves between the fitting/realization errors and the nuclear norm of the Hankel matrix, obtained by solving (5) for a range of different values of λ. The fitting error is the error between the optimized covariances and the sample covariances. The realization error is the error between the optimized covariances and the true covariances. Figure 2 shows the singular values of the optimal Hankel matrix $H_{(1,N,N)}$ in (5), for the value of λ marked with a dot on the trade-off curve in figure 1. We observe a sharp transition of the singular values, corresponding to a model order of 5. For comparison, we also show the singular values of the Hankel matrix $\bar{H}_{(1,N,N)}$ based on the sample covariances.

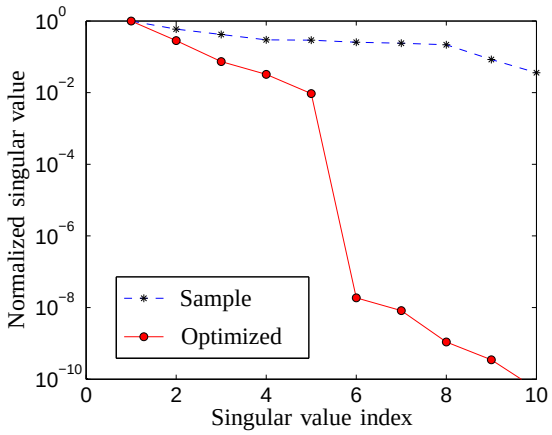

Fig. 2. Singular values of the Hankel matrix $\bar{H}_{(1,N,N)}$ constructed from the sample covariances $\bar{H}_k$ and the Hankel matrix $H_{(1,N,N)}$ of nuclear-norm optimized covariances.

III. SYSTEM IDENTIFICATION

In this section we present results for a system identification method based on nuclear norm minimization. Subspace algorithms use the SVD of matrices constructed from observed inputs and outputs to estimate the system order, and to compute low-rank approximations from which the model parameters are determined [Lar90], [Ver94], [OD94], [Vib95]. Nuclear norm minimization offers an alternative method to compute low rank matrix approximations.

A. System identification by nuclear norm optimization

Suppose we are given a sequence of inputs $u(t) \in \mathbf{R}^m$ and outputs $y_{\rm meas}(t) \in \mathbf{R}^p$, $t = 0, \ldots, N$. Our objective is to find a low-order discrete-time linear time-invariant state-space model

$$\begin{array}{rcl} x(t+1) & = & Ax(t) + Bu(t), \\ y(t) & = & Cx(t) + Du(t), \end{array} \qquad (15)$$

that satisfies $y(t) \approx y_{\rm meas}(t)$. The proposed method is based on the fact that if $u(t)$, $y(t)$ are related by (15), then

$$Y = GX + HU,$$

where

$$Y = \begin{bmatrix} y(0) & y(1) & \cdots & y(N-r) \\ y(1) & y(2) & \cdots & y(N-r+1) \\ \vdots & \vdots & & \vdots \\ y(r) & y(r+1) & \cdots & y(N) \end{bmatrix},$$

$$U = \begin{bmatrix} u(0) & u(1) & \cdots & u(N-r) \\ u(1) & u(2) & \cdots & u(N-r+1) \\ \vdots & \vdots & & \vdots \\ u(r) & u(r+1) & \cdots & u(N) \end{bmatrix},$$

$$G = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^r \end{bmatrix}, \quad X = \begin{bmatrix} x(0) & x(1) & \cdots & x(N-r) \end{bmatrix},$$

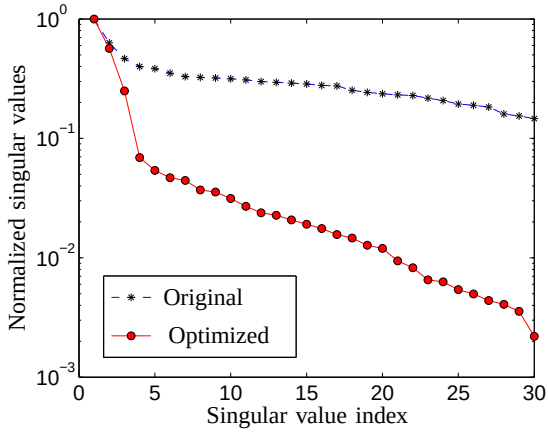

Fig. 3. Largest normalized singular values of $Y_{\rm meas}U^\perp$ ('original') and $YU^\perp$ ('optimized').

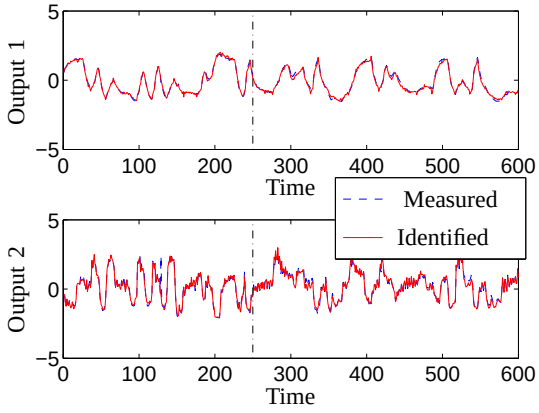

Fig. 4. Measured and identified outputs.

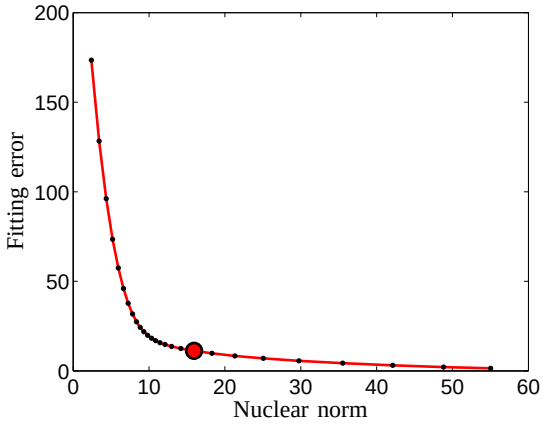

Fig. 5. Trade-off between the fitting error $\sum_{t=0}^{N}\|y(t)-y_{\rm meas}(t)\|_2^2$ and the nuclear norm $\|YU^\perp\|_*$.

and H is a lower-triangular block-Toeplitz matrix constructed from A, B, C, D. Therefore

$$YU^\perp = GXU^\perp, \tag{16}$$

where $U^\perp$ denotes a matrix whose columns span the nullspace of U. If X has rank n and there is no rank cancellation in the product $XU^\perp$, then we can find the model order from the rank of the matrix $YU^\perp$. From the column space of $YU^\perp$, one can compute G, A, and C; and once A and C are computed, the rest of the state-space model follows easily [DMVV88], [Lju99, Section 10.6].

This leads us to the following method. We use $y(t)$, $t = 0,\ldots,N$, as optimization variables, and compute $y(t)$ by solving the convex optimization problem

$$\text{minimize} \quad \|YU^\perp\|_* + \lambda\sum_{t=0}^{N}\|y(t)-y_{\rm meas}(t))\|_2^2\,, \tag{17}$$

for a positive weighting parameter λ, This gives a point on the trade-off curve between the nuclear norm of $YU^\perp$ and the deviation between the sequences $y(t)$ and $y_{\rm meas}(t)$. From the optimal solution $y(t)$ of (17), we compute the singular value decomposition of $YU^\perp$, and a realization of G, A, C and the rest of the state space model; see [LV09] for details.

B. Experimental results

We apply the method of section III-A to a benchmark problem from [DDDF97]. The system is an industrial winding process with five inputs and two outputs (data set 97-003). The number of data points used in the identification experiment is $N = 200$. $N_{\rm V} = 600$ points were used for model validation. (This includes the N data points used in the identification.)

Figure 3 shows the singular values of $YU^\perp$, where Y is the solution of the optimization problem (17). The parameter λ was selected by examining different points and choosing the value that gives approximately the smallest identification error. The figure also shows the singular values of $Y_{\rm meas}U^\perp$, where $Y_{\rm meas}$ is the Hankel matrix constructed from the output data. As can be seen, the optimized matrix $YU^\perp$ is closer to low rank (rank three) than the original matrix $Y_{\rm meas}U^\perp$. In figure 4 we compare the measured outputs with the model predictions based on the estimated third-order model.

Figure 5 shows the trade-off curve between the two terms in the cost function of (17). The dot marks the point used in figures 3 and 4. Figure 6 shows the identification and validation errors for the models computed from the solutions on the trade-off curve in figure 5. The error $e_{\rm V}$ is computed as

$$e_{\rm V} = \left(\frac{\sum_{t=0}^{N_{\rm V}-1}\|y_{\rm meas}(t)-\hat{y}(t)\|_2^2}{\sum_{t=0}^{N_{\rm V}-1}\|y_{\rm meas}(t)-\bar{y}_{\rm V}\|_2^2}\right)^{1/2}, \tag{18}$$

where $\bar{y}_{\rm V} = (1/N_{\rm V})\sum_{t=0}^{N_{\rm V}-1} y_{\rm meas}(t)$, $y_{\rm meas}(t)$ are the given output data, and $\hat{y}(t)$ is the output of the identified state-space model, starting at the estimated initial state. $e_{\rm I}$ is calculated using the same formula with sums ranging from $t = 0$ to $N-1$. At the point marked on the curves the validation error is $e_{\rm V} = 0.17$.

We also compared the results with the subspace identification algorithm implemented in the Matlab Identification

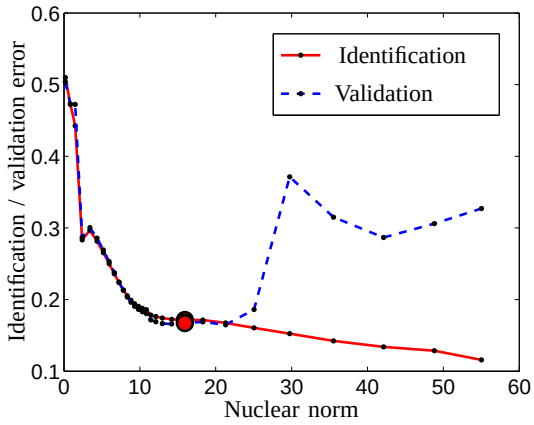

Fig. 6. Trade-off between the identification/validation error ($e_{\mathrm{I}}/e_{\mathrm{V}}$) and the nuclear norm $\|YU^{\perp}\|_*$.

Toolbox. The Matlab command `n4sid` with the default settings gives a system model of order 10, and a validation error $e_{\mathrm{V}} = 0.18$. If we fix the order to three, `n4sid` returns a model with $e_{\mathrm{V}} = 0.17$.

These results and the results for other benchmark problems from [LV09] indicate that the identification errors for the nuclear norm optimization algorithm are comparable to `n4sid`. The main advantage of the nuclear norm technique is that it makes the selection of an appropriate model order easier (see figure 3).

IV. Conclusion

We have described identification and system realization methods based on regularized nuclear norm minimization, a convex heuristic for computing low-rank approximations of structured matrices. Experiments with simulated data indicate that the stochastic realization subspace method based on nuclear norm minimization improves the accuracy of the covariance estimation. Similarly, the experimental results of the system identification method suggest that the quality of the models obtained by this method is comparable with state-of-the-art subspace identification software. An advantage of the nuclear norm approach for system identification is that it provides a sharper criterion for the selection of the system order.

References

[Bal95] A. V. Balakrishnan. *Introduction to Random Processes in Engineering*. John Wiley & Sons, Inc., New York, 1995.

[CFP03] M. T. Chu, R. E. Funderlic, and R. J. Plemmons. Structured low rank approximation. *Linear Algebra and its Applications*, 366:157–172, 2003.

[CR08] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. 2008. Preprint available at `arXiv.org` (0805.4471).

[CRT06] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[CT05] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[DDDF97] B. De Moor, P. De Gersem, B. De Schutter, and W. Favoreel. DAISY: A database for the identification of systems. *Journal A*, 38(3):4–5, Sep. 1997.

[DMVV88] B. De Moor, M. Moonen, L. Vandenberghe, and J. Vandewalle. A geometrical approach for the identification of state space models with the singular value decomposition. In *Proceedings of the 1988 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2244–2247, 1988.

[Don06] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[Faz02] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[FHB01] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, pages 4734–4739, 2001.

[FHB03] M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In *Proceedings of the American Control Conference*, pages 2156–2162, 2003.

[FHB04] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. In *Proceedings of American Control Conference*, pages 3273–3278, 2004.

[Gui81] R. P. Guidorzi. Invariants and canonical forms for systems structural and parametric identification. *Automatica*, 17(1):117–133, 1981.

[Kal71] R. E. Kalman. On minimal partial realizations of a linear input/output map. In N. DeClaris, editor, *Aspects of Network and System Theory*, pages 385–407. Holt, Rinehart and Winston, New York, 1971.

[Lar90] W. E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the 29th IEEE Conference on Decision and Control*, volume 2, pages 596–604, 1990.

[Lju99] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, New Jersey, second edition, 1999.

[LV09] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 2009. To appear.

[MSM00] J. Mari, P. Stoica, and T. McKelvey. Vector ARMA estimation: a reliable subspace approach. *IEEE Transaction on signal processing*, 48(7):2092–2104, 2000.

[OD94] P. Van Overschee and B. De Moor. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.

[RFP07] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. 2007. Submitted to *SIAM Review*.

[RXH08] B. Recht, W. Xu, and B. Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 3065–3070, 2008.

[Sch00] B. De Schutter. Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics*, 121(1-2):331–354, 2000.

[Sil71] L. M. Silverman. Realization of linear dynamical systems. *IEEE Transactions on Automatic Control*, AC-16(6):554–567, 1971.

[SM97] P. Stoica and R. L. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[Tet70] A. J. Tether. Construction of minimal linear state-variable models from finite input-output data. *IEEE Transactions on Automatic Control*, 15(4):427–436, 1970.

[Tro06] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.

[Ver94] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.

[Vib95] M. Viberg. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 31(12):1835–1851, 1995.

[Wil86] J. C. Willems. From time series to linear system — Part II. Exact modelling. *Automatica*, 22(6):675–694, 1986.