Nuclear Norm Minimization Algorithms for Subspace Identification from Non-Uniformly Spaced Frequency Data

Conference Paper

Author(s):
Graf Plessen, Mogens; Wood, Tony A.; Smith, Roy

Publication date:
2015

Permanent link:
https://doi.org/10.3929/ethz-b-000111283

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
https://doi.org/10.1109/ECC.2015.7330838
Nuclear Norm Minimization Algorithms for Subspace Identification from Non-Uniformly Spaced Frequency Data

Mogens Graf Plessen, Tony A. Wood and Roy S. Smith

Abstract—The nuclear norm is an effective proxy for matrix rank in a range of minimization problems, including subspace identification. Nuclear norm-based methods are implemented via iterative optimization methods and in problems with very noisy data the quality of the nuclear norm-based estimate may warrant the additional computation cost. We present two methods (based on the dual accelerated gradient projection and the alternating direction method of multipliers) for nuclear norm based subspace identification in the case where the data is given as irregularly spaced frequency samples.

I. INTRODUCTION

The subspace identification method attempts to match a linear, time-invariant model to an experimental data record of input-output signals (see for example [1], [2], [3]). We consider the case where the experimental data is provided as noisy frequency domain transfer function estimates at a set of irregularly spaced frequencies. This is exactly the case considered in [4] using singular value decomposition-based methods. In the case of very noisy data these methods may exhibit problems as truncating singular values modifies the data matrix in such a way that it does not correspond to a linear system realization.

The nuclear norm-based identification problem is formulated [5], [6] as the minimization of the nuclear norm of a matrix constructed from the data. The effects are noise are encoded in the problem in the form of constraints on additive variables. Two assumptions about the noise are considered: a simple 2-norm bound; and a cumulative spectrum linearity condition. For each we provide a first-order iterative computational algorithm suitable for large-scale computation. The methods are based on the dual accelerated gradient projection (DAGP) method and the alternating direction method of multipliers (ADMM). The problems are convex and may be solved by a wide variety of methods—our experience has shown the algorithms presented here to be efficient.

Algorithms for the regularly spaced frequency domain data problem [5] are given in [7]. The contribution of this paper is to develop these algorithms for the irregularly spaced frequency case and demonstrate their applicability on an example. The theoretical basis for this development comes from [6].

II. FREQUENCY DOMAIN SUBSPACE IDENTIFICATION

A. Problem Setting

The problem under consideration is the fitting of a state-space representation to noisy frequency domain estimates—at nonuniformly-spaced frequencies—of the system transfer function matrix. Given data, $G_{\text{data}}(e^{j\omega_n})$, $n = 1, \ldots, M$, we wish to find a low-order realization, $G(e^{j\omega_n}) = C(e^{j\omega_n}I - A)^{-1}B + D$, and a residual noise signal, $W(e^{j\omega_n})$, such that the model plus noise accounts for the data, $G_{\text{data}}(e^{j\omega_n}) = G(e^{j\omega_n}) + W(e^{j\omega_n})$, for all $n = 1, \ldots, M$. The work in [4] is summarized to introduce our notation. In the noise-free frequency domain case we have,

$$
e^{j\omega}X(\omega) = AX(\omega) + BU(\omega),$$
$$Y(\omega) = CX(\omega) + DU(\omega),$$

with $X(\omega) \in \mathbb{C}^{n_x}$, $U(\omega) \in \mathbb{C}^{n_u}$, and $Y(\omega) \in \mathbb{C}^{n_y}$. By applying unit vectors with one in the $n$th position (i.e. sinusoids of frequency $\omega$ in the time domain), $U(\omega) = e_i$, to all input channels, $i = 1, \ldots, n_u$, and stacking the corresponding solutions, $X^i(\omega)$, we define $X^C(\omega) = [X^1(\omega) \ X^2(\omega) \ \cdots \ X^n(\omega)]$, and $U^C(\omega) = I_{n_u}$. With this input we have $G(e^{j\omega}) = CX^C(\omega) + D$. Multiplying this by powers of $e^{j\omega}$ and stacking the rows gives,

$$
\begin{bmatrix}
G(e^{j\omega}) \\
\vdots \\
e^{j(p-1)\omega}G(e^{j\omega})
\end{bmatrix} = 
\begin{bmatrix}
I_{n_u} \\
\vdots \\
e^{j(p-1)\omega}I_{n_u}
\end{bmatrix},
$$

with $p > 1$ to be chosen later and

$$
\mathcal{O} = 
\begin{bmatrix}
C & CA \\
\vdots & \vdots \\
CA^{p-1}
\end{bmatrix}, \quad 
\Gamma = 
\begin{bmatrix}
D & 0 & \cdots & 0 \\
CB & D & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
CA^{p-2}B & CA^{p-3}B & \cdots & D
\end{bmatrix},
$$

where $\mathcal{O}$ is the extended observability matrix. Formulating at distinct irregularly spaced frequencies, $\omega_n$, $n = 1, \ldots, M$, subsequently stacking the equations column-wise,

MGP was, and TW and RSS are, with the Automatic Control Laboratory, Swiss Federal Institute of Technology, (ETH Zürich), Physikstrasse 3, 8092 Zürich, Switzerland, moyens.plessen@gmx.ch and {woodt, rsmith}@control.ee.ethz.ch.
and defining
\[
\mathcal{G} = \begin{bmatrix}
G(e^{j\omega_1}) & \ldots & G(e^{j\omega_M}) \\
e^{j\omega_1}G(e^{j\omega_1}) & \ldots & e^{j\omega_M}G(e^{j\omega_M}) \\
\vdots & \ddots & \vdots \\
e^{j(p-1)\omega_1}G(e^{j\omega_1}) & \ldots & e^{j(p-1)\omega_1}G(e^{j\omega_M})
\end{bmatrix},
\]
\[
V = \begin{bmatrix}
1 & \ldots & 1 \\
e^{j\omega_1} & \ldots & e^{j\omega_M} \\
\vdots & \ddots & \vdots \\
e^{j(p-1)\omega_1} & \ldots & e^{j(p-1)\omega_M}
\end{bmatrix},
\]
we obtain
\[
\mathcal{G} = OX + \Gamma V_n,
\]
with \(V_n = V \otimes I_{n_x}\) and \(X = \begin{bmatrix} X^C(\omega_1) & \ldots & X^C(\omega_M) \end{bmatrix} \).

Note that \(\mathcal{G} = \Gamma V_n \text{diag} \left( G(e^{j\omega_1}), \ldots, G(e^{j\omega_M}) \right) \), where \(\text{diag}(\cdot)\) denotes a block-diagonal matrix. We assume the system to be minimal, \((A,B)\) is a controllable pair and so (see [4, Lemma 2]) \(X\) is full rank. Thus, for \(p \geq n_x\), we have \(\text{rank}(OX) = n_x\) and \(\text{R}(OX) = \text{R}(O)\), where \(\text{R}(\cdot)\) denotes the range space. We convert \(\mathcal{G}\) into an expression involving only the real valued matrices, i.e.,
\[
\mathcal{G} = OX + \Gamma V_n, \quad \text{where} \quad \mathcal{G} = \begin{bmatrix} \text{Re}(\mathcal{G}) & \text{Im}(\mathcal{G}) \end{bmatrix},
\]
with \(V_n = V \otimes I_{n_x}\) and \(X = \begin{bmatrix} \text{Re}(X) & \text{Im}(X) \end{bmatrix}\) and \(V = \begin{bmatrix} \text{Re}(V) & \text{Im}(V) \end{bmatrix}\). If \(\Gamma V_n\) in (4) were zero, the range space of \(\mathcal{G}\) would equal the range space of \(O\). By multiplying \(\mathcal{G}\) from the right with \(V_n^\perp\), the right nullspace of \(V_n\), we obtain
\[
\mathcal{G}V_n^\perp = OXV_n^\perp \quad \text{and} \quad \text{R}(\mathcal{G})V_n^\perp = \text{R}(O).
\]
The matrix \(V_n^\perp = V^\perp \otimes I_{n_x}\), defined by \(VV^\perp = 0\), can be derived from a QR-decomposition of \(V^T\).

**B. SVD-based Identification**

A singular value decomposition (SVD) of \(\mathcal{G}V_n^\perp\) yields
\[
\mathcal{G}V_n^\perp = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}.
\]
Thus, since \(U_1\) is a basis for \(O\), there exists an invertible matrix \(T\), such that \(OT = U_1\). The estimates of \(C\) and \(A\) are computed from \(\hat{C} = [I_{n_y} \quad 0] U_1\) and \([I_{n_y} \quad 0] U_1 \hat{A} = [0 \quad I_{n_x}(p-1)] U_1\). These estimates differ from the underlying system by only the similarity transform, \(T\).

If we now assume that additive noise is present the frequency samples are given by
\[
G_{\text{data},k} = G_k + G_{\text{noise},k}, \quad k = 1, \ldots, M,
\]
where \(G_{\text{noise}}\) represents a noise process in the frequency domain. We now, instead of (5), obtain
\[
G_{\text{data}}V_n^\perp = [\hat{U}_1 \quad \hat{U}_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = \hat{U}\Sigma\hat{V}^T.
\]
It is no longer possible to uniquely assign the contribution of the noise and the underlying system to the various singular values. In standard SVD-based subspace methods a model order estimate, \(\hat{n}_x\), is selected, e.g., by inspecting the number of largest singular values above a user-defined threshold \(r_{\text{tol}}\). The first \(\hat{n}_x\) columns of \(U\), corresponding to what is assumed to be the range space of the underlying system, are then used to form the estimates, \(\hat{C}\) and \(\hat{A}\), as above. The estimates \(\hat{B}\) and \(\hat{D}\) are obtained from a least-squares fit to the data in the frequency domain.

**III. NUCLEAR NORM MINIMIZATION PROBLEMS**

The approach taken here is to formulate an optimization objective on \(G\mathcal{V}_n\). Because the mapping between the underlying system, \(G \in \mathbb{C}^{M \times n_u}\), and \(G\mathcal{V}_n\) is linear, this property is shared by additive noise on the data. This allows us to pose the optimization as a search over the noise accounting for the data. Our size or spectral content assumptions on the noise are handled via constraints. In [6] it is shown that \(G\mathcal{V}_n\) can always be factorized as a Hankel matrix multiplied by a full rank matrix, showing that if \(G\mathcal{V}_n\) is low rank then it admits an exact factorization of a lower order system. This is in contrast to the SVD methods which, after truncating singular values and vectors, use an approximation method to obtain a low order system.

To proceed we define the linear operator \(\Omega(\cdot)\) as
\[
\Omega(G) = G\mathcal{V}_n^\perp = \begin{bmatrix} \text{Re}(V_n \text{diag}(G)) & \text{Im}(V_n \text{diag}(G)) \end{bmatrix} V_n^\perp.
\]

The optimization problem to be considered is
\[
\begin{array}{ll}
\text{minimize} & \text{rank} \left( \Omega(G_{\text{data}} - W) \right) \\
\text{subject to} & W \in \mathcal{W},
\end{array}
\]
where \(\mathcal{W}\) is a convex set constraining the noise, \(W\). This formulation mirrors the additive noise effect in [6]. The constrained rank minimization problem (7) is NP-hard [8] and so the rank minimization objective is replaced by the nuclear norm minimization objective, which is a convex heuristic for low rank matrix approximation,
\[
\begin{array}{ll}
\text{minimize} & \|\Omega(G_{\text{data}} - W)\|_* \\
\text{subject to} & W \in \mathcal{W},
\end{array}
\]

The nuclear norm of a matrix \(X \in \mathbb{C}^{a \times b}\) is the sum of the singular values. We will present two optimization problems in the form given in (9), each corresponding to a different assumption on the underlying noise. The noise assumptions turn out to have a significant influence on the underlying optimization problem complexity.

**A. Bounded Noise Assumption**

The first convex constraining set considered is
\[
\mathcal{W} = \left\{ W \left| \|W\|_F^2 \leq \delta^2 \right. \right\}.
\]
This bounds the Euclidean norm of the noise by the parameter \(\delta\). Incorporating (10) as a Lagrangian cost into the objective function of (9) and using the change of variables, \(X = G_{\text{data}} - W\), leads to the regularized least-squares problem,
\[
\begin{array}{ll}
\text{minimize} & \mu \|\Omega(X)\|_* + \frac{1}{2} \|X - G_{\text{data}}\|_F^2
\end{array}
\]
The regularization parameter, $\mu$, allows us to trade off between the closeness of the data fit and the order of the estimated model.

B. Bounded Cumulative Spectrum Assumption

The noise assumption implied by $\{10\}$ often does not capture all of the characteristics we typically associate with noise. Another assumption (which is somewhat stronger) is that the cumulative spectrum is close to linear. The following constraint, introduced in $\{6\}$ captures this requirement. For generality we specify this a channel-wise manner where $\mathcal{M}_i(\cdot)$ extracts all elements of $W$ corresponding to the $i$th input-output channel.

$$W_{\eta_i} = \{ W \mid \| A_m \mathcal{M}_i(W) \|_F^2 \leq \left( \eta + \frac{m}{M} \right) \gamma_i^2, \quad i = 1, \ldots, n_y n_u, \ m = 1, \ldots, M \},$$

(12)

with scalar parameters $\eta$ and $\gamma_i$, $i = 1, \ldots, n_y n_u$, specifying the closeness to linearity and the overall size of the noise respectively.

Note that $\{12\}$ specifies a convex region which is slightly larger than just linear cumulative spectral content. Smaller noise signals with less linear cumulative spectral are also included.

IV. ALGORITHMS

Motivated by the comparative results in $\{9\}$, we develop a DAGP method for the solution of the regularized least-squares problem $\{11\}$. The spectral constraint problem $\{12\}$ will be addressed by an ADMM method (see $\{10\}$). The motivation in this case comes from the observation that the projection onto the spectral constraint can be replaced by a sequence of nested component-wise projection. The overall result is no longer a projection but it is non-expansive and gives a feasible update variable. The benefit is that the calculation of this step is much faster than the exact projection calculation.

A. Bounded Noise Assumption

The dual problem to $\{11\}$ is derived exploiting the same ideas used in $\{9\}$ following its detailed derivation of the dual problem of a general Hankel matrix nuclear norm minimization problem. The final result is

$$\min_{\Lambda \succeq \mu I} \left\{ \frac{1}{2} \| \Omega^* (\Lambda) \|_F^2 + \text{Tr} (\Omega^* (\Lambda)^T G_{\text{data}}) \right\},$$

where $\Lambda$ is the scaled dual variable and the adjoint matrix $\Omega^* (\Lambda)$ is defined by the relation

$$\text{Tr} \left( \Lambda^T \Omega(X) \right) = \text{Tr} \left( \Omega^* (\Lambda)^T X \right).$$

(13)

In the following, we derive $\Omega^* (\Lambda)$, which is a key step within this paper. With the definition in $\{7\}$, we write

$$\Omega(X) = \begin{bmatrix} \text{Re} \left( V_{n_y} \text{diag}(X) \right), \quad \text{Im} \left( V_{n_y} \text{diag}(X) \right) \end{bmatrix} V_{n_y}^\perp$$

$$\quad = \begin{bmatrix} \text{Re} \left( V_{n_y} \right), \quad \text{Im} \left( V_{n_y} \right) \end{bmatrix} \mathcal{L}(X) V_{n_y}^\perp,$$

where

$$\mathcal{L}(X) = \begin{bmatrix} \text{Re} \left( \text{diag}(X) \right), & \text{Im} \left( \text{diag}(X) \right) \\ -\text{Im} \left( \text{diag}(X) \right), & \text{Re} \left( \text{diag}(X) \right) \end{bmatrix}.$$  

(14)

For clarity, the dimensions are $X \in \mathbb{C}^{M n_y \times n_u}$ and $\Omega(X) \in \mathbb{R}^{p n_y \times (2M - p) n_u}$. In terms of the vec($\cdot$)-operator (see $\{11\}$), this gives,

$$\text{Tr} \left( \Lambda^T \Omega(X) \right) = \text{vec}(\Lambda)^T \text{vec}(\Omega(X))$$

$$= \text{vec}(\Lambda)^T (\langle V_{n_y}^\perp \rangle^T \otimes \left[ \text{Re} \left( V_{n_y} \right), \quad \text{Im} \left( V_{n_y} \right) \right]) \text{vec}(\mathcal{L}(X))$$

$$= \text{vec}(\Lambda)^T S (\langle V_{n_y}^\perp \rangle^T \otimes \left[ \text{Re} \left( V_{n_y} \right), \quad \text{Im} \left( V_{n_y} \right) \right]) \mathcal{L}(X).$$

(15)

The linear mappings $S(\cdot)$ and $\mathcal{L}(\cdot)$ are described below. Matrix $\{14\}$ is sparse, i.e., composed of four block diagonal matrices with $\text{Re} \left( \text{diag}(X) \right)$ and $\text{Im} \left( \text{diag}(X) \right)$ both occurring twice. The operator $\mathcal{L}(\cdot)$ extracts the nonzero elements of $\mathcal{L}(X)$ and stacks them into a column vector. We define

$$\mathcal{L}(X) = \text{vec} \left( \begin{bmatrix} \text{Re}(X^1), \ldots, \text{Re}(X^M), \quad \text{Im}(X^1), \ldots, \text{Im}(X^M) \end{bmatrix} \right),$$

(16)

where $X^j$, $j = 1, \ldots, M$, denotes the $j$th block-element of dimension $n_y \times n_u$ of $X \in \mathbb{C}^{M n_y \times n_u}$. The operator $S : \mathbb{R}^{(2M - p)p n_y \times n_u} \rightarrow \mathbb{R}^{(2M - p)pn_y \times p n_u}$ manipulates $\langle V_{n_y}^\perp \rangle^T \otimes \left[ \text{Re}(V_{n_y}), \quad \text{Im}(V_{n_y}) \right]$ correspondingly. This is computationally efficient. For the following, we abbreviate

$$S = \langle V_{n_y}^\perp \rangle^T \otimes \left[ \text{Re}(V_{n_y}), \quad \text{Im}(V_{n_y}) \right].$$

So far, the lefthand side of $\{15\}$ has been considered. The righthand side can be reformulated as

$$\text{Tr} \left( \Omega^* (\Lambda)^T X \right) = \mathcal{L}(\Omega^* (\Lambda))^T \mathcal{L}(X).$$

(17)

Comparing $\{15\}$ and $\{17\}$, we obtain

$$\mathcal{L}(\Omega^* (\Lambda)) = S^T \text{vec}(\Lambda),$$

from which we reconstruct $\Omega^* (\Lambda)$ by inverting the definition of operator $\mathcal{L}(\cdot)$ given in $\{16\}$. Thus, we have derived a method to compute $\Omega^* (\Lambda)$. For clarity, the dimensions are $\Lambda \in \mathbb{R}^{p n_y \times (2M - p) n_u}$ and $\Omega^* (\Lambda) \in \mathbb{C}^{M n_y \times n_u}$.

To solve $\{11\}$, we are now able to use the DAGP method proposed in $\{9\}$ adjusted to our problem and summarized in Table $\{II\}$. It is denoted by DAGP-regLS. The Lipschitz constant $L_D$ on the gradient of the dual objective function is derived from the relation

$$\| \nabla d(\Lambda_1) - \nabla d(\Lambda_2) \|_F \leq L_D \| \Lambda_1 - \Lambda_2 \|_F,$$

as $L_D = \| S \|^2_2$.

B. Bounded Cumulative Spectrum Assumption

ADMM can be efficient in problems where the objective and/or constraints can be written as simply calculated functions. In this case we will exploit the structure of the cumulative spectrum constraint. To obtain the required problem format, we rewrite $\{9\}$ with constraint set $\{12\}$ by introducing the indicator function, $I_{\mathcal{W}, \eta}(W) = 0$, if $w \in \mathcal{W}_{\eta},$ and
We then rewrite (18) as

\[ \text{minimize}_{X, W} \| \Omega(X) \|_* + \hat{I}_{V_{n-1}}(W) \]
subject to \( X + W = G \text{data} \).

We then rewrite (18) as

\[ \text{minimize}_{X, V_1, \ldots, V_M} \| \Omega(X) \|_* + \sum_{n=1}^{M} \hat{I}_V(V_n) \]
subject to \( X + V_n = G \text{data}, \ n = 1, \ldots, M \),

where for all \( n = 1, \ldots, M \), we define

\[ V_n = \left\{ V_n \left| \| A_n M_i(V_n) \|_2 \leq \left( \eta + \frac{n}{M} \right) \gamma_i, i = 1, \ldots, n_y n_n \right. \right\} . \]

At this point the derivation follows [7] and so we present only the resulting scaled ADMM update equations,

\[ X^{k+1} = \text{argmin}_X \| \Omega(X) \|_* + \rho \sum_{n=1}^{M} \| X + V_n - G \text{data} + U_n^k \|_F^2 \] (19)

\[ V_n^{k+1} = \text{argmin}_{V_n} \hat{I}_V(V_n) + \rho \| X^{k+1} + V_n - G \text{data} + U_n^k \|_F^2, \]

\[ V_n^{k+1} = \hat{I}_V(V_n) \]

\[ U_n^{k+1} = U_n^k + X^{k+1} + V_n^{k+1} - G \text{data}, \]

\[ \forall n = 1, \ldots, M. \] (20)

The \( X^{k+1} \) update (19) can be reformulated in terms of the means \( \bar{V}^k = 1/M \sum_{n=1}^{M} V_n^k \) and \( \bar{U}^k = 1/M \sum_{n=1}^{M} U_n^k \), as

\[ X^{k+1} = \text{argmin}_X \frac{1}{\rho M} \| \Omega(x) \|_* + \frac{1}{2} \| X + \bar{V}^k - G \text{data} + \bar{U}^k \|_F^2 \] (22)

Table I

The algorithm for the solution of (11).

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Choose any ( \Lambda ) satisfying ( \Lambda^T \Lambda \leq \mu I ).</td>
</tr>
<tr>
<td>2.</td>
<td>Initialize ( \Lambda_{old} = \Lambda, \theta_{old} = 1 ) and ( \theta = 1 ).</td>
</tr>
<tr>
<td>3.</td>
<td>Compute ( \varphi = \Lambda + \frac{\rho}{\theta_{old}} (\Lambda - \Lambda_{old}) ).</td>
</tr>
<tr>
<td>4.</td>
<td>Compute a SVD: ( \varphi = \Gamma^T (\Gamma^T \varphi + G_{data}) = USV^T ).</td>
</tr>
<tr>
<td>5.</td>
<td>Update: ( \Lambda_{old} = \Lambda, \theta_{old} = \theta ), ( \Lambda = \text{min} (S, \mu I) V^T ), ( \theta = \sqrt{\sigma + \theta^2} ).</td>
</tr>
<tr>
<td>6.</td>
<td>Check stopping criterion: ( X = \Omega^T(\Lambda) + G_{data}, p(X) = | \Omega^T(\Lambda) |<em>* + \frac{\rho}{\theta} | X - G</em>{data} |_F^2 ), ( d(\Lambda) = \frac{1}{2} | \Omega^T(\Lambda) |<em>F^2 + | \Omega^T(\Lambda), G</em>{data} | ) ( \frac{p(x) + d(\Lambda)}{1 + d(\Lambda)} &lt; \epsilon ).</td>
</tr>
<tr>
<td>If not met, go to step 3. Else, return ( x ).</td>
<td></td>
</tr>
</tbody>
</table>

Table II

The algorithm for the solution of (9) with constraint set \( \mathcal{C} \).

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Initialize ( k = 0, V^0 = 0 ) and ( U_n^0 = 0, \forall n = 1, \ldots, M ).</td>
</tr>
<tr>
<td>2.</td>
<td>Compute ( X^{k+1} ), see (22), via the ADMM method.</td>
</tr>
<tr>
<td>3.</td>
<td>Update ( V_n^{k+1}, \forall n = 1, \ldots, M ), see (20), via channel-wise projections.</td>
</tr>
<tr>
<td>4.</td>
<td>Compute ( U_n^{k+1}, \forall n = 1, \ldots, M ), see (21).</td>
</tr>
<tr>
<td>5.</td>
<td>Check the stopping criterion (22). If not met: ( k = k+1 ) and go to step 2. Else: return ( W = \bar{V} ).</td>
</tr>
</tbody>
</table>

Equation (22) is similar to (11) with \( 1/\rho M \) and \( G_{data} - \bar{V}^k - \bar{U}^k \) replacing \( \mu \) and \( G_{data} \).

As in the uniform case [7], the \( V_n^{k+1} \)-update in (20) can be calculated analytically. First, compute for all \( n = 1, \ldots, M \) and for all \( i = 1, \ldots, n_y n_n \):

\[ A_n M_i(V_n^{k+1}) = \sqrt{\eta + \frac{n}{M} \gamma_i \| A_n M_i(F_n^{k+1}) \|_F^2} . \]

if \( \| A_n M_i(F_n^{k+1}) \|_F^2 > \sqrt{\eta + \frac{n}{M} \gamma_i} \), or

\[ A_n M_i(V_n^{k+1}) = A_n M_i(F_n^{k+1}) \]

otherwise; and \( B_{M-n} M_i(V_n^{k+1}) = B_{M-n} M_i(F_n^{k+1}) \), where \( B_{M-n} = \left[ 0_{M-n \times M} I_{M-n} \right] \) and \( F_n^{k+1} = G_{data} - X^{k+1} - U_n^k \). Then, combine the channel-wise solutions into \( V_n^{k+1} \).

To specify a stopping criteria (see [10]) we define the primal and dual residuals, \( r_n^{k+1} \) and \( s_n^{k+1} \), as

\[ r_n^{k+1} = X_n^{k+1} + \bar{V}^{k+1} - G_{data} \]
and \( s_n^{k+1} = \rho (\bar{V}^{k+1} - \bar{V}^{k}) \).

The ADMM algorithm is terminated when

\[ \| r_n^{k+1} \|_F < \epsilon_{\text{ADMM,r}} \quad \text{and} \quad \| s_n^{k+1} \|_F < \epsilon_{\text{ADMM,s}} \] (23)

where \( \epsilon_{\text{ADMM,r}} \) and \( \epsilon_{\text{ADMM,s}} \) are scalar parameter choices.

The ADMM algorithm (denoted by ADMM-SC) is summarized in Table II.

V. ALGORITHM IMPLEMENTATION

In both problem formulations the resulting model order is a function of a scalar parameter. For the spectral constraint case in (12) the nuclear norm (and in practice the model order) forms a Pareto optimal curve with respect to \( \gamma \). It is up to the user to decide the appropriate point on the trade-off curve between assumed noise size and model order. In practice one can start with a value, \( \gamma_{\max, i} = c \| M_i(G_{data}) \|_2 \) (with \( c > 1 \), and by decreasing \( \gamma \) identify a series of models of increasing order.

As the problem is resolved with multiple \( \gamma \) values, warm starting the initial estimates (\( V^0 \) and \( U_n^0 \) in Table II) reduces the number of iterations.

The convergence of the ADMM algorithm can be improved by variable \( \rho \) and the corresponding scaled dual variables (see [10]). The reader is referred to an identical approach in our previous work [7] for details. For reference...
in the Section VI example, the tuning parameters determining the $\rho$-update dynamics are denoted by $\tilde{\mu}$ and $\tilde{\gamma}$.

In case of ADMM-SC, after the optimal solution, $W$, to (9) with constraint set (12) is obtained, state-space estimates can be reconstructed similarly as in Section III-B starting from a SVD of $\Omega(G_{\text{data}} - W)$. Analogously, $\Omega(\hat{X})$ with $X$ the optimal decision variable to (11) is the starting point for DAGP-regLS.

In case of (11), analogously, a $\mu$-iteration with the option of early termination is implemented with parameters $\mu_{\text{max}}$, $\mu_{\text{min}}$ and $n_{\mu}$. In addition, $\Lambda$ is warm-started.

VI. A SIMULATION EXAMPLE

A SISO simulation is constructed to illustrate the application of these algorithms. The nominal plant, $G$, is of order 6 and has an $L_\infty$-norm 1. There are 100 frequency samples non-uniformly spaced between 0 and $\pi$ [rad/s]. The frequency domain noise, $G_{\text{noise}}$, is applied additively, as in (9).

For comparison with a more standard subspace identification method we use the MATLAB System Identification toolbox function n4sid. In this case the target model orders are set from 1 to 9.

For the SVD-based identification, the number of rows, $p$, is chosen to be 10. The model order estimate is varied between 1 and 9.

The parameter choices for ADMM-SC and DAGP-regLS are summarized in Table III. The selections for $f_{\text{DAGP}}$ of 1 and 10 for ADMM-SC and DAGP-regLS yielded fastest computation times respectively. The maximum allowable number of ADMM iterations is indicated by $k_{\text{max,ADMM}}$. The allowable gap between primal and dual residuals before $\rho$ is updated was set to $\tilde{\mu} = 1000$. In this case this had the effect of keeping $\rho$ constant at $\rho = 1$.

Figure 1 shows the results of the $\mu$-iteration for DAGP-regLS. In this example $\mu_{\text{max}} = 2$ yields $\hat{n}_x = 0$ and the $\mu$-iteration was terminated when $\hat{n}_x = \hat{n}_{x,\text{max}}$ was met. Higher model orders, $\hat{n}_x$, were found as $\mu$ was decreased. The improvement from warm-starting is evident as $\mu$ values corresponding to changes in $\hat{n}_x$ (for which we expect no benefit from warm-starting) require significantly more computation time. This observation also applies to the $\gamma$ parameter and warm-starting in the ADMM-SC problem case. The solutions corresponding to the smallest $\mu$ (or $\gamma$) associated to a particular $\hat{n}_x$ are used for the reconstruction of the state-space estimates.

Figure 2 compares the Bode plots of the “true system”, the noisy data and one example of the estimated transfer function, with $n_x = 3$, using the ADMM-SC method. The cumulative spectrum corresponding to this data fit is shown in Figure 3.

To assess the accuracy of the estimate, we evaluate the RMS error of difference between the data and the noise-free
The performance of the different methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>$|\hat{G}<em>x - G</em>{data, 3}|_F$</th>
<th>$\epsilon_{tr}$</th>
<th>$\epsilon_{d}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n4sid</td>
<td>0.581/0.355</td>
<td>0.322/0.362</td>
<td>0.528/0.375</td>
</tr>
<tr>
<td>SVD-based</td>
<td>0.372/0.362</td>
<td>0.372/0.362</td>
<td>0.528/0.375</td>
</tr>
<tr>
<td>ADMM-SC</td>
<td>0.344/0.208</td>
<td>0.450/0.235</td>
<td>0.528/0.375</td>
</tr>
<tr>
<td>DAGP-regLS</td>
<td>0.374/0.049</td>
<td>0.372/0.059</td>
<td>0.528/0.375</td>
</tr>
<tr>
<td>$n_x$ = 3</td>
<td>0.371/0.086</td>
<td>0.371/0.086</td>
<td>0.528/0.375</td>
</tr>
<tr>
<td>$n_x$ = 4</td>
<td>0.371/0.127</td>
<td>0.372/0.060</td>
<td>0.528/0.375</td>
</tr>
</tbody>
</table>

**Table V**

Computation times to solve the corresponding nuclear norm minimization problems.

<table>
<thead>
<tr>
<th>Method</th>
<th>$T_{avg}/T_{CPU}/T_{RM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMM-SC</td>
<td>1.87/12.23/0.01</td>
</tr>
<tr>
<td>DAGP-regLS</td>
<td>0.15/0.63/0.01</td>
</tr>
</tbody>
</table>

Two iterative first-order algorithms (based on DAGP and ADMM methods) have been developed for subspace identification on irregularly spaced frequency domain data. The algorithms extend our earlier work for the uniform frequency spacing case and share similar computational properties.

The DAGP method uses a nuclear norm regression formulation to obtain low order solutions. It is well suited to the case where we are searching for the model with a small 2-norm error with respect to the data. Additional noise constraints can be included and the ADMM-based formulation is applicable to the case where there are also cumulative spectrum constraints on the noise. The nested nature of these constraints is easily cast as a series of simple and efficient projection calculations which fit within the ADMM iteration.

Standard SVD based methods (such as N4SID or similar) are well suited to small to moderate experimental noise cases. The approaches considered here appear to outperform the standard methods when the noise level is close to that of the data. In the simulation example presented here the singular values of the noisy data all lie within two orders of magnitude making it difficult to use the usual SVD-based heuristics for model order determination.

Future work in this area will look at applying these methods to experimental systems to evaluate their effectiveness in practice.

**References**


