

Subspace identification with missing data

Ivan Markovsky

Abstract—The paper presents initial results on a subspace method for exact identification of a linear time-invariant system from data with missing values. The identification problem with missing data is equivalent to a Hankel structured low-rank matrix completion problem. The novel idea is to search systematically and use effectively completely specified submatrices of the incomplete Hankel matrix constructed from the given data. Nontrivial kernels of the rank-deficient completely specified submatrices carry information about the to-be-identified system. Combining this information into a full model of the identified system is a greatest common divisor computation problem. The developed subspace method has linear computational complexity in the number of data points and is therefore an attractive alternative to more expensive methods based on the nuclear norm heuristic.

Index Terms—subspace system identification, missing data, low-rank matrix completion, nuclear norm, realization.

I. INTRODUCTION

Identification of dynamical systems from data with missing values is an important and current undeveloped topic in system identification. A simple heuristic approach to solve the problem is to use interpolation techniques, *e.g.*, splines, in order to complete the data in a preprocessing step and then apply classical identification methods on the interpolated data. Such an approach is theoretically unsatisfactory and may produce bad results in practice.

Special identification problems with missing data can be solved by existing methods. A famous one is the partial realization problem [1] where the data is an impulse response, all missing values are in the “future”, and the given data is exact. Another example is when all input and output variables are missing for at least as many sequential moments of time as the lag of the system. In this case, the identification problem with missing data is equivalent to identification from two independent data sets: the one before the first missing value and the one after the last missing value [2]. Finally, the special case when the missing data is restricted to the output variables is easily handled by classical identification methods.

There are two main classes of methods for addressing the general identification problem with missing data:

- methods based on convex relaxations [3], and
- methods based on local optimization [4].

To the best of our knowledge, the class of the subspace methods [5] has not been extended to deal with missing data. Our purpose is to fulfill this gap. This paper presents

A full version of this paper is in preparation.

Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, Building K, B-1050 Brussels, Belgium; Email: imarkovs@vub.ac.be

a method based on kernel computations of fully specified submatrices of the incomplete Hankel matrix of the data.

We consider an exact (deterministic) identification problem. Throughout the paper we assume that the given data is (a part of) an exact trajectory of the to-be-identified system. In this setup, the identification problem is equivalent to low-rank Hankel structured matrix completion.

It is well known that, in general, the unstructured low-rank matrix completion problem is NP-hard [6]. In the context of linear time-invariant system identification, in addition, the matrix is Hankel structured. These facts show that the identification problem considered is nontrivial.

The paper is organized as follows. We start in Section II with a motivating example. Section III defines the notation. Section IV formally states the identification problem and outlines the proposed subspace method. Challenges and future work is summarized in the conclusion.

II. MOTIVATING EXAMPLE

The identification data w is a $T = 500$ samples long trajectory of $n = 6$ order lightly damped autonomous linear time-invariant system. In total, 215 samples are missing in a periodic pattern, see Figure 1. The precise simulation parameters are specified in the following fragment of the m-file reproducing the presented numerical results:

```
<example>+≡
sys0 = ss(diag([0.8889 + 0.4402i
               0.8889 - 0.4402i
               0.4500 + 0.8801i
               0.4500 - 0.8801i
               0.6368 + 0.7673i
               0.6368 - 0.7673i]),
           [], ones(1, 6), [], -1);
T = 500; n = 6;
w0 = initial(ss(sys0), ones(n, 1), T - 1);
Tm = sort(unique([1:7:T, 3:7:T, 5:7:T]));
w = w0; w(Tm) = NaN;
```

Two identification methods, described in the paper, are applied on the data and the results are validated by the distance between the characteristic polynomials of the true and identified models:

```
<example>+≡
dist = @(sys1, sys2) norm(poly(eig(sys1))...
                        - poly(eig(sys2)));
```

The first method, implemented in the function `hmc_nn`, uses the nuclear norm heuristic (see Appendix). The second method, implemented in the function `mpum_md`, is the new subspace-type method (see Section IV).

```
<example>+≡
tic, sysh_nn = hmc_nn( w(1:Tp), n, 0);
t_nn = toc, e_nn = dist(sys0, sysh_nn)
```

III. DATA AND MODEL CLASS

Data

The input data for the considered identification problem is a set

$$w = \{w^1, \dots, w^N\}$$

of N , q -variate, sequences

$$w^i = (w^i(1), \dots, w^i(T_i)), \quad \text{where } w^i(t) \in \mathbb{R}^q$$

and T_i is the number of data points in the i th sequence w^i . The elements $w^i(t)$ of a sequence w^i are called samples. Note that the sequences w^1, \dots, w^N must have the same number of variables but may have different number of samples. If the data consists of a single sequence, then the superscript index is dropped and w itself is the sequence.

Missing data values are denoted by the symbol NaN (“not a number”). The extended set of real numbers \mathbb{R}_e is the union of the set of the real numbers \mathbb{R} and the symbol NaN:

$$\mathbb{R}_e := \mathbb{R} \cup \text{NaN}.$$

A set of sequences w is parameterized by the column vector

$$\text{vec}(w) = \begin{bmatrix} \text{vec}(w^1) \\ \vdots \\ \text{vec}(w^N) \end{bmatrix}, \quad \text{where } \text{vec}(w^i) = \begin{bmatrix} w^i(1) \\ \vdots \\ w^i(T_i) \end{bmatrix}.$$

The subvector of $\text{vec}(w)$ with indexes in \mathcal{I} is denoted $w|_{\mathcal{I}}$. Similarly, for a matrix M , $M|_{\mathcal{I}}$ is the submatrix of M formed by the rows with indexes in the set \mathcal{I} , and $M|_{\mathcal{I}, \mathcal{J}}$ is the submatrix of M with elements m_{ij} , such that $i \in \mathcal{I}$, $j \in \mathcal{J}$. The set of indexes of missing elements in $\text{vec}(w)$ is denoted by \mathcal{I}_m and the set of the remaining indexes by \mathcal{I}_g .

Linear time-invariant dynamical systems

We are interested in sequences $w \in (\mathbb{R}^q)^{\mathbb{N}}$ that are trajectories of discrete-time linear time-invariant dynamical systems. A linear time-invariant dynamical system \mathcal{B} with q variables is a subspace of the data space $(\mathbb{R}^q)^{\mathbb{N}}$ and can be represented as the kernel of a polynomial operator $R(\sigma)$,

$$\mathcal{B} = \ker(R(\sigma)) = \{w \mid R(\sigma)w = 0\}, \quad (\text{KER})$$

where σ is the backwards shift $(\sigma w)(t) := w(t+1)$

$$\begin{array}{cccccccc} t & \rightarrow & \cdots & 0 & 1 & 2 & 3 & \cdots \\ w(t) & \rightarrow & \cdots & \cdots & w(1) & w(2) & w(3) & \cdots \\ (\sigma w)(t) & \rightarrow & \cdots & w(1) & w(2) & w(3) & \cdots & \cdots \end{array}$$

The minimal natural number ℓ , for which there exists an ℓ th order difference equation representation for \mathcal{B} is an important invariant of the system, called the *lag*.

The variables can always be partitioned element-wise into inputs u and outputs y , *i.e.*, $w = \Pi \begin{bmatrix} u \\ y \end{bmatrix}$, for some permutation matrix Π . The number of inputs m and the number of outputs $p = q - m$ are system invariant. With some loss of generality, we assume that the first m elements of $w(t)$ are inputs and the remaining elements are outputs, *i.e.*, we fix Π to the identity matrix I_q .

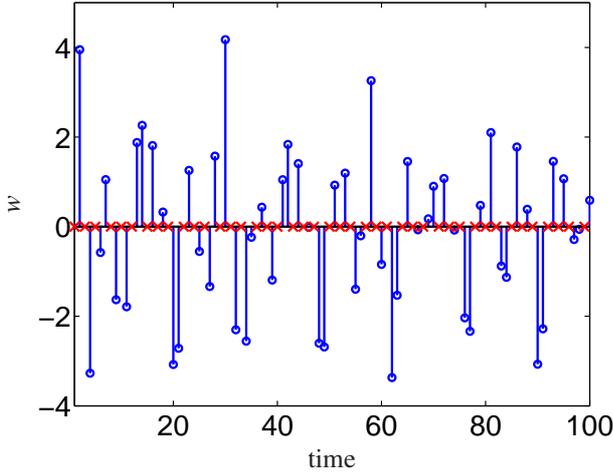


Fig. 1. The first 100 samples of the data. The red crosses denote the location of the missing samples.

```
tic, sysh_ss = mpum_md(w(1:Tp), n, 0);
t_ss = toc, e_ss = dist(sys0, sysh_ss)
```

Zero errors e_{nn} and e_{ss} indicate that the identification methods have recovered exactly the data generating system. In the specific example, the results

```
t_nn =          t_ss =
54.6626          0.3936
e_nn =          e_ss =
4.2057e-10      2.0948e-15
```

show that both methods recover exactly the data generating system. The subspace method `mpum_md`, however, does not require nonlinear minimization and is in the example two orders of magnitude faster than the method `hmc_nn` using the nuclear norm heuristic. Efficient methods for nuclear norm minimization of Hankel structured matrices are currently under active development. Such methods will make possible to use the nuclear norm approach in medium scale system identification problems. With state-of-the-art general purpose semidefinite programming solvers, such as SeDuMi, however, the method is applicable only for small size problems.

In the simulation example presented above, we used only the first 100 samples of the data due to fast growing computation time of `hmc_nn`. The developed subspace identification method has linear computational complexity in the number of data points T and is applicable to medium (and, with some software improvements, large) scale identification problems.

```
t_ss =          e_ss =
1.3768          1.4990e-13
```

Assuming that a discrete-time linear time-invariant dynamical systems \mathcal{B} admits an input/output partitioning $w = \text{vec}(u, y)$, it can be represented in the classical input/state/output form

$$\mathcal{B} = \mathcal{B}_{i/s/o}(A, B, C, D) := \{w = \begin{bmatrix} u \\ y \end{bmatrix} \in (\mathbb{R}^q)^{\mathbb{Z}} \mid \text{there is } x, \text{ such that } \sigma x = Ax + Bu \text{ and } y = Cx + Du\}. \quad (\text{I/S/O})$$

The number of inputs and the lag or the order specify the complexity of the system in the sense that the dimension of the restriction $\mathcal{B}|_{[1, T]}$ of \mathcal{B} to the interval $[1, T]$, where $T \geq \ell$, is bounded by $Tm + \ell(q - m)$. The subset of linear time-invariant systems with at most m inputs, lag at most ℓ , order at most n is denoted by $\mathcal{L}_{m, \ell}^n$. If the order or the lag are not bounded they are skipped from the notation.

Hankel matrices

The mosaic-Hankel matrix with L block rows, constructed from the set of sequences w is defined as

$$\mathcal{H}_L(w) := [\mathcal{H}_L(w^1) \quad \cdots \quad \mathcal{H}_L(w^N)],$$

where

$$\mathcal{H}_L(w^i) := \begin{bmatrix} w^i(1) & w^i(2) & \cdots & w^i(T-L+1) \\ w^i(2) & w^i(3) & \cdots & w^i(T-L+2) \\ w^i(3) & w^i(4) & \cdots & w^i(T-L+3) \\ \vdots & \vdots & \cdots & \vdots \\ w^i(L) & w^i(L+1) & \cdots & w^i(T) \end{bmatrix}.$$

IV. PROBLEM FORMULATION AND SUBSPACE ALGORITHM

It is well known that there is a link between rank deficient Hankel matrices and trajectories of finite dimensional discrete-time linear time-invariant systems [7]. If the Hankel matrix $\mathcal{H}_L(w)$, where

$$L := \left\lceil \frac{T+1}{q+1} \right\rceil,$$

is rank deficient, its rank is equal to $mL + n$, where n is the minimal order of a linear time-invariant system for which w is an exact trajectory.

The considered problem is defined as follows.

Problem 1. Given a set of time series w , possibly with missing values, and a model class $\mathcal{L}_{m, \ell}^n$,

$$\text{find } \mathcal{B} \text{ and } \hat{w} \text{ such that } \hat{w}|_{\mathcal{I}_g} = w|_{\mathcal{I}_g} \text{ and } \hat{w}^i \in \hat{\mathcal{B}}|_{[1, T_i]} \in \mathcal{L}_{m, \ell}^n, \text{ for } i = 1, \dots, N.$$

Theorem 2. Problem 1 is equivalent to the mosaic-Hankel structured low-rank matrix completion problem:

$$\text{find } \hat{w} \text{ such that } \hat{w}|_{\mathcal{I}_g} = w|_{\mathcal{I}_g} \text{ and } \text{rank}(\mathcal{H}_L(\hat{w})) \leq Lm + n. \quad (\text{HMC})$$

The method presented next is related to the class of the subspace identification methods [5]. The procedure is based on linear algebra operations, such as solution of a system of linear equations and does not require nonlinear optimization.

First, we illustrate the main idea on a simple example. The following section summarized the algorithm for the general identification problem with missing data.

Example

Consider the data

$$w = (1, 2, \text{NaN}, 4, 5, \text{NaN}, 7, 8, \text{NaN}, 10, 11).$$

The to-be-identified data generating system has lag $\ell = 2$. Therefore, there is a nonzero vector $R \in \mathbb{R}^{1 \times 3}$, such that

$$R\mathcal{H}_3(\bar{w}) = 0,$$

where, \bar{w} is the unknown complete trajectory of the system. The R vector is a parameter of a kernel representation of the system and can be found from the left kernel of $\mathcal{H}_3(\bar{w})$. The Hankel matrix $\mathcal{H}_3(w)$ of the given data

$$\begin{bmatrix} 1 & 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} \\ 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & 10 \\ \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & 10 & 11 \end{bmatrix},$$

however, does not allow us to compute its left kernel due to unspecified entries in every column of the matrix. (In this sense, the periodic pattern of the missing values with a period $\ell + 1$ or less is the worse distribution of the missing data.)

The main idea behind the proposed method is to consider the *extended Hankel matrix*

$$\mathcal{H}_4(w) = \begin{bmatrix} 1 & 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 \\ 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} \\ \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & 10 \\ 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & 10 & 11 \end{bmatrix}$$

and select the two submatrices of $\mathcal{H}_4(w)$

$$\tilde{H}^1 = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ \text{NaN} & \text{NaN} & \text{NaN} \\ 4 & 7 & 10 \end{bmatrix} \quad \text{and} \quad \tilde{H}^2 = \begin{bmatrix} 2 & 5 & 8 \\ \text{NaN} & \text{NaN} & \text{NaN} \\ 4 & 7 & 10 \\ 5 & 8 & 11 \end{bmatrix},$$

which have a single row of NaN's. The matrices H^1 and H^2 , obtained from \tilde{H}^1 and \tilde{H}^2 , respectively, by removing the rows of NaN's have nontrivial left kernels

$$\underbrace{\begin{bmatrix} 1 & -3/2 & 1/2 \end{bmatrix}}_{R^1} \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 4 & 7 & 10 \end{bmatrix} = 0$$

and

$$\underbrace{\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}}_{R^2} \begin{bmatrix} 2 & 5 & 8 \\ 4 & 7 & 10 \\ 5 & 8 & 11 \end{bmatrix} = 0.$$

Inserting zeros in the R^1 and R^2 vectors, we obtain vectors \tilde{R}^1 and \tilde{R}^2 in the kernels of \tilde{H}^1 and \tilde{H}^2 , respectively:

$$\underbrace{\begin{bmatrix} 1 & -3/2 & 0 & 1/2 \end{bmatrix}}_{\tilde{R}^1} \tilde{H}^1 = 0$$

and

$$\underbrace{\begin{bmatrix} 1 & 0 & -3 & 2 \end{bmatrix}}_{\tilde{R}^2} \tilde{H}^2 = 0.$$

(By definition $0 \times \text{NaN} = 0$, *i.e.*, the value of the NaN is irrelevant.)

By construction

$$\begin{bmatrix} \tilde{R}^1 \\ \tilde{R}^2 \end{bmatrix} \mathcal{H}_4(\bar{w}) = 0,$$

so that, the polynomial matrix

$$\tilde{R}(z) = \begin{bmatrix} \tilde{R}^1(z) \\ \tilde{R}^2(z) \end{bmatrix} = \begin{bmatrix} z^0 - 3/2z^1 + 1/2z^3 \\ z^0 - 3z^2 + 2z^3 \end{bmatrix}$$

is a (nonminimal) kernel representation of the data generating system. A minimal kernel representation [8] can be obtained from \tilde{R} by computing the greatest common divisor of the polynomials \tilde{R}^1 and \tilde{R}^2

$$R(z) := \text{GCD}(\tilde{R}^1(z)\tilde{R}^2(z)) = z^0 - 2z^1 + z^2.$$

Once the model is identified, it is trivial to fill the missing data. In the example, we iterate the difference equation

$$w(t) = 2w(t-1) - w(t-2),$$

starting from the given initial condition $w(1) = 1$ and $w(2) = 2$

$$\bar{w} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \dots)$$

Note 3. The shortest length data sequence from which the system is identifiable by the proposed method is

$$w = (1, 2, \text{NaN}, 4, 5, \text{NaN}, 7, 8).$$

In this case the matrices H_1 and H_2 are 3×2 and still have the correct kernels

$$\{\alpha [1 \quad -3/2 \quad 1/2] \mid \alpha \in \mathbb{R}\}$$

and

$$\{\alpha [1 \quad -3 \quad 2] \mid \alpha \in \mathbb{R}\},$$

respectively.

General method

The generalization of the procedure, used in the example is summarized in Algorithm 1. An implementation of the algorithm in Matlab is available from the author's webpage:

<http://homepages.vub.ac.be/~imarkovs>

V. CONCLUSIONS AND DISCUSSION

We have presented a novel subspace-type exact linear time-invariant system identification method for data with missing values. The main idea of the method is to extract complete submatrices of the incomplete mosaic-Hankel matrix, constructed from the data, and compute their left kernels. The collection of the kernel parameters of the submatrices, extended with zeros at the location of the missing values in the full mosaic-Hankel matrix, gives a nonminimal representation of the data generating system. The reduction of the computed nonminimal representation to a minimal one is a standard problem and can be implemented by a greatest common divisor computation.

Algorithm 1 Subspace algorithm for linear time-varying system identification with missing data.

Input: Set of sequence $w \in (\mathbb{R}_e^q)^{T_1} \times \dots \times (\mathbb{R}_e^q)^{T_N}$ and natural numbers ℓ and m .

1: Select real valued submatrices

$$H^i \in \mathbb{R}^{m_i \times n_i}, \quad \text{with } m_i \geq n_i - 1,$$

of the mosaic-Hankel matrix $\mathcal{H}_T(w_{\text{ext}})$, where

$$w_{\text{ext}} = (w, \underbrace{\text{NaN}, \dots, \text{NaN}}_T).$$

2: Compute bases R^i for the left kernels of H^i , *i.e.*, full row rank matrices $R^i \in \mathbb{R}^{g_i \times m_i}$ of maximum row dimension g_i , such that $R^i H^i = 0$.

3: Extend $R^i \in \mathbb{R}^{g_i \times m_i}$ to $\tilde{R}^i \in \mathbb{R}^{g_i \times Tq}$ by inserting zero columns at the location of the rows removed from $\mathcal{H}_T(w_{\text{ext}})$ in the selection of H^i .

4: Compute the greatest common divisor R of $\tilde{R}^1, \dots, \tilde{R}^K$, where K be the number of complete submatrices.

Output: Minimal kernel representation $R(z)$ of the data generating system.

In the case of exact data without missing values, there are sufficient identifiability conditions in the literature [9]. Generalization of these conditions to (or derivation of new conditions) in the case of missing data is an open problem. It is well known that minor modifications of exact subspace identification methods result are very effective methods for approximate and stochastic system identification. The modification of the method proposed in this paper for the noisy case, however, is nontrivial due to the reduction step of the computed (highly) nonminimal kernel representation to a minimal one. An approximate version of this step is a hard problem (approximate greatest common divisor [10] computation). The uncertainty related to the kernel parameters \tilde{R}^i may vary due to the fact that different amounts of data are used for their computations. These and other issues are a topic of current research.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

REFERENCES

- [1] R. E. Kalman, "On partial realizations, transfer functions, and canonical forms," *Acta Polytechnica Scandinavica*, vol. 31, pp. 9–32, 1979.
- [2] J. Schoukens, G. Vandersteen, Y. Rolain, and R. Pintelon, "Frequency response function measurements using concatenated subrecords with arbitrary length," *IEEE Trans. on Instr. and Measurement*, vol. 61, no. 10, pp. 2682–2688, 2012.
- [3] Z. Liu, A. Hansson, and L. Vandenbergh, "Nuclear norm system identification with missing inputs and outputs," *Control Lett.*, vol. 62, pp. 605–612, 2013.

- [4] R. Pintelon and J. Schoukens, "Frequency domain system identification with missing data," *IEEE Trans. Automat. Control*, vol. 45, no. 2, pp. 364–369, Feb. 2000.
- [5] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory, implementation, applications*. Boston: Kluwer, 1996.
- [6] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is NP-hard," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1149–1165, 2011.
- [7] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012.
- [8] J. C. Willems, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Control*, vol. 36, no. 3, pp. 259–294, 1991.
- [9] J. C. Willems, P. Rapisarda, I. Markovsky, and B. D. Moor, "A note on persistency of excitation," *Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.
- [10] I. Markovsky and S. Van Huffel, "An algorithm for approximate common divisor computation," in *Proc. 17th Symp. on Math. Theory of Networks and Systems*, Kyoto, Japan, 2006, pp. 274–279.
- [11] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Elec. Eng. Dept., Stanford University, 2002.
- [12] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," 2008. [Online]. Available: stanford.edu/~boyd/cvx

APPENDIX

A related problem to (HMC) is the mosaic-Hankel structured rank minimization problem:

$$\begin{aligned} & \text{minimize} && \text{over } \widehat{w} \in (\mathbb{R}^q)^T && \text{rank}(\mathcal{H}_L(\widehat{w})) \\ & \text{subject to} && \widehat{w}|_{\mathcal{S}_g} = w|_{\mathcal{S}_g}. \end{aligned}$$

The number of block-rows L of the mosaic-Hankel matrix is a design parameter. Our empirical experience shows that best results are obtained for nearly squares matrix, so that L is set to

$$L := \left\lceil \frac{T+1}{q+1} \right\rceil.$$

Replacing the minimization of the rank by minimization of the nuclear norm [11], we obtain a convex optimization problem:

$$\begin{aligned} & \text{minimize} && \text{over } \widehat{w} \in (\mathbb{R}^q)^T && \|\mathcal{H}_L(\widehat{w})\|_* \\ & \text{subject to} && \widehat{w}|_{\mathcal{S}_g} = w|_{\mathcal{S}_g}, \end{aligned}$$

which, in turn, is equivalent to the following semidefinite programming problem

$$\begin{aligned} & \text{minimize} && \text{over } \widehat{w}, U, \text{ and } V && \text{trace}(U) + \text{trace}(V) \\ & \text{subject to} && \begin{bmatrix} U & \mathcal{H}_L(\widehat{w})^\top \\ \mathcal{H}_L(\widehat{w}) & V \end{bmatrix} \succeq 0 && \text{and } \widehat{w}|_{\mathcal{S}_g} = w|_{\mathcal{S}_g}. \end{aligned}$$

The semidefinite relaxation can be solved globally by existing methods. We use the CVX package [12] for this purpose.