DATA-DRIVEN SIMULATION USING THE NUCLEAR NORM HEURISTIC

Philippe Dreesen and Ivan Markovsky

Vrije Universiteit Brussel (VUB), Belgium

ABSTRACT

Applications of signal processing and control are classically model-based, involving a two-step procedure for modeling and design: first a model is built from given data, and second, the estimated model is used for filtering, estimation, or control. Both steps typically involve optimization problems, but the combination of both is not necessarily optimal, and the modeling step often ignores the ultimate design objective. Recently, data-driven alternatives are receiving attention, which employ a direct approach combining the modeling and design into a single step. In earlier work, it was shown that data-driven signal processing problems can often be rephrased as missing data completion problems, where the signal of interest is part of an incomplete low-rank mosaic Hankel structured matrix. In this paper, we consider the exact data case and the problem of simulating from a given input, an output trajectory of the unknown data generating system. Our findings suggest that, when using an adequate rescaling of the given data, the exact data-driven simulation problem can be solved by replacing the original structured low-rank matrix completion problem by a convex optimization problem, using the nuclear norm heuristic.

***Index Terms*—** data-driven signal processing, low-rank matrix completion, mosaic Hankel matrix, nuclear norm, convex optimization

1. INTRODUCTION

1.1. Context and motivation

Signal processing and control methods are classically model-based, and proceed in two steps, namely as a 'modeling' step followed by a 'design' step. First a mathematical model is built that explains as well as possible the measured (noisy) signals, exploiting prior information about the data generating system (and possibly noise). Second, the identified model is used for solving a design problem (e.g., Kalman filter).

Both steps typically involve a mathematical optimization problem: In the modeling step, model parameters are optimized as to describe as accurately as possible the measured data. Then the design problem is solved optimally with some purpose in mind, starting from the previously identified model. While each of the steps may separately lead to optimal solutions, the combination of both is no longer necessarily optimal—the identification step does not take into account the subsequent use of the identified model for design.

To account for the above issues, the data-driven paradigm combines the modeling and design steps into a single problem [1]. The data-generating system is assumed linear and time-invariant, admitting a description using low-rank Hankel matrices [2]. In this way, a variety of signal processing and control problems can be reduced to instances of mosaic Hankel low-rank approximation and/or completion tasks. In this paper, we focus on the data-driven approach for the simulation of a system to a given input signal, on the basis of exact measurements of past data. Following [3, 1], the simulation task is formulated as a low-rank block-Hankel completion problem, which we will solve using a convex relaxation heuristic. Our simulation results suggest that, when using an adequate rescaling of the given data, the exact data-driven simulation problem can be solved by replacing the original structured low-rank matrix completion problem by a convex optimization problem, using the nuclear norm heuristic.

1.2. Related work

Low-rank matrix completion, or the recovery of missing entries of low-rank matrices has applications in several scientific and engineering problems, e.g., recommender systems [4], genetic prediction [5], and image processing [6], among others. Generally, the matrix completion problem is NP-hard and no efficient algorithms are known [7, 8]. However, by using the nuclear norm heuristic, introduced in the work of Fazel [9, 10] in the context of rank minimization, the problem is relaxed to a convex one, and can be solved efficiently using semi-definite programming.

The problem of (Hankel) structured low-rank approximation (as opposed to completion) by nuclear norm minimization was studied in [11, 12, 13] and the system identification problem with missing data was studied in [14]. While the nuclear norm heuristic for low-rank matrix completion has been analyzed for unstructured matrices [15, 8], for (Hankel) structured matrices the problem is only recently receiv-

This work was supported in part by Fonds de la Recherche Scientifique (FNRS) and Fonds Wetenschappelijk Onderzoek (FWO Vlaanderen) under Excellence of Science (EOS) Project no 30468160 "Structured low-rank matrix / tensor approximation: numerical optimization-based algorithms and applications", and FWO research projects G.0280.15N and G.0901.17N.

ing attention [16, 17, 18], in particular for sums of exponentials signals. The work of [17] establishes that the nuclear norm heuristic retrieves the original rank minimization solution, provided that a stability criterion is met.

1.3. Organization of the paper

This paper is organized as follows. Section 2 introduces the preliminaries and notation used throughout the paper. Section 3 discusses how the simulation problem can be phrased as a low-rank mosaic Hankel matrix completion problem, for which we consider a convex relaxation using the nuclear norm heuristic. Section 4 contains numerical experiments illustrating the use of the nuclear norm completion. Section 5 draws the conclusions of this work and points out open problems.

2. PRELIMINARIES AND NOTATION

2.1. Linear dynamical systems

Willems' behavioral system theory [19] defines a system $\mathcal{B}$ as the set of its admissible trajectories w. A trajectory w of a discrete-time q-variate system $\mathcal{B}$ (in shorthand notation '$w \in \mathcal{B}$') is a sequence $w = (\, w(1), \ldots, w(T)\,)$, with $w(t) \in \mathbb{R}^q$ for $t = 1, \ldots, T$. In this paper we consider the class of q-variate linear time-invariant (LTI) systems $\mathcal{L}^q$. An LTI system $\mathcal{B} \in \mathcal{L}^q$ with m inputs and p outputs (with $q = p + m$) has a kernel representation

$$\mathcal{B} = \big\{ w \mid R_0 w(t) + \cdots + R_\ell w(t+\ell) = 0, \text{ for } t \geq 1 \big\}, \quad (1)$$

where $R = \begin{bmatrix} R_0 & R_1 & \ldots & R_\ell \end{bmatrix}$ with $R_i \in \mathbb{R}^{p \times q}$ is a kernel parameter that specifies the system. This formulation can be viewed as a difference equation describing the admissible trajectories $w \in \mathcal{B}$. The minimal value for ℓ for which (1) holds, is an invariant of the system $\mathcal{B}$ and is called the *lag*.

2.2. The block-Hankel matrix

A central tool in the remainder of the paper is the block-Hankel matrix, which captures into the language of linear algebra, the linear time-invariance of $\mathcal{B}$. The block-Hankel matrix $H_L(w)$, with $L < T$ block rows, built from a trajectory $w = (\, w(1), \ldots, w(T)\,)$, is defined as

$$H_L(w) = \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ \vdots & \vdots & \ddots & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}, \quad (2)$$

having size $qL \times T - L + 1$.

For a trajectory $w \in \mathcal{B}$, the block-Hankel matrix $H_L(w)$ with $L \geq \ell + 1$ is (row) rank-deficient, since $R H_{\ell+1}(w) = 0$. More precisely, the rank of the block-Hankel matrix $H_L(w)$ is [2]

$$\operatorname{rank} H_L(w) \leq mL + p\ell, \text{ for } L \geq \ell + 1. \quad (3)$$

2.3. On the initial conditions

A trajectory $w \in \mathcal{B}$ admits a partitoning into inputs u and outputs y. Let $w = (u, y)$ denote such a partitioning. The output sequence $y = (\, y(1), \cdots, y(T)\,)$ is uniquely determined by the input sequence $u = (\, u(1), \cdots, u(T)\,)$ (free variables) and initial conditions w_i ('i' for initial conditions), with

$$w_i = (\, w_i(-\ell+1), w_i(-\ell+2), \ldots, w_i(-1), w_i(0)\,), \quad (4)$$

such that $w_i \wedge w \in \mathcal{B}$, with '$\wedge$' denoting the concatenation of trajectories. Remark that if we assume that the initial conditions are zero, i.e., $w_i \equiv 0$, the specification of initial conditions is done by prepending the trajectory w with ℓ zeros.

3. METHOD

3.1. Problem formulation

The simulation of a system for a given input signal can be stated as follows. Given a system $\mathcal{B} \in \mathcal{L}^q$, an input u_s, and initial conditions w_i as in (4), find the output y_s such that $w_i \wedge w_s = w_i \wedge (u_s, y_s) \in \mathcal{B}$. This is a basic problem in system theory and is studied in various formulations.

In *data-driven simulation*, the system $\mathcal{B}$ is defined implicitly by a given trajectory $w_d = (\, w_d(1), \ldots, w_d(T_d)\,) \in \mathcal{B}$ ('d' for data). In this context, the simulation of the output $y_s = (\, y_s(1), \ldots, y_s(T_s)\,)$ for a given input $u_s = (\, u_s(1), \ldots, u_s(T_s)\,)$ ('s' for simulation) is done without first identifying the system $\mathcal{B}$. We assume that the input u_d of the given trajectory $w_d = (u_d, y_d)$ is persistently exciting, so w_d completely specifies the system $\mathcal{B}$ [20]. The data-driven simulation problem can then formally be stated as follows.

Problem 1 (Exact data-driven simulation)**.**
Given *a trajectory $w_d = (u_d, y_d) \in \mathcal{B} \in \mathcal{L}^q$, an input u_s, and initial conditions w_i,* ***find*** *the output y_s, such that $w_i \wedge w_s = w_i \wedge (u_s, y_s) \in \mathcal{B}$.*

3.2. Data-driven simulation as matrix completion

The rank-deficiency of the Hankel matrix $H_L(w)$, for $L \geq \ell + 1$, is closely related to the kernel representation of the system $\mathcal{B}$ and its interpretation in terms of difference equations. The mosaic Hankel matrix $\begin{bmatrix} H_L(w') & H_L(w'') \end{bmatrix}$ built from two different trajectories $w', w'' \in \mathcal{B}$ has the same rank as $H_L(w)$, since w' and w'' satisfy the same difference equations, so $\operatorname{rank} \begin{bmatrix} H_L(w') & H_L(w'') \end{bmatrix} \leq mL + p\ell$.

This observation is crucial for formulating the data-driven simulation problem as a matrix completion problem: Consider the mosaic Hankel matrix $\begin{bmatrix} H_L(w_d) & H_L(w_s) \end{bmatrix}$, where the leftmost Hankel block $H_L(w_d)$ contains the given trajectory w_d, and the rightmost Hankel block $H_L(w_s)$ contains the simulation trajectory w_s, which is only partially

known—the inputs u_s are given, but the outputs y_s are unknown. Hence, the unknown trajectory y_s should be determined such that

$$\text{rank}\begin{bmatrix} H_L(w_d) & H_L(w_s) \end{bmatrix} \leq mL + p\ell, \text{ for } L \geq \ell + 1, \tag{5}$$

implying that both trajectories belong to the same system $\mathcal{B}$.

Remark that again we need to specify initial conditions to uniquely determine the simulation output y_s. For simplicity, we will henceforth assume that the simulations start from zero initial conditions $w_i \equiv 0$. In this case, the specification of initial conditions is done by prepending the to-be-simulated trajectory w_s with $L-1$ zeros. If we denote by $w'_s = w_i \wedge w_s$ the simulation trajectory with prepended zeros, we thus have

$$\begin{aligned} u'_s &= (\underbrace{0, \ldots, 0}_{L-1 \text{ zeros}}, \underbrace{u_s(1), \ldots, u_s(T_s)}_{\text{simulation input}}), \text{ and} \\ y'_s &= (\underbrace{0, \ldots, 0}_{L-1 \text{ zeros}}, \underbrace{y_s(1), \ldots, y_s(T_s)}_{\text{simulation output}}). \end{aligned} \tag{6}$$

The data-driven simulation problem can then be phrased as the following block-Hankel matrix completion problem.

Problem 2 (Data-driven simulation via Hankel completion)**.** ***Given*** *a trajectory $w_d = (u_d, y_d) \in \mathcal{B} \in \mathcal{L}^q$, an input u_s, and zero initial conditions $w_i \equiv 0$,* ***find*** *the output y_s from the following minimization problem.*

$$\underset{y_s}{\text{minimize}} \quad \text{rank}\begin{bmatrix} H_L(w_d) & H_L(w'_s) \end{bmatrix}, \tag{7}$$

where $w'_s = (u'_s, y'_s)$ is defined in (6).

3.3. Relaxing the structured matrix completion problem

The rank minimization of a (block-Hankel) matrix is in general an NP-hard problem for which no efficient solutions are known. Here we consider the *nuclear norm* minimization, which is a convex relaxation of the matrix rank minimization. The *nuclear norm* of a matrix X, denoted by $\|X\|_\star$, is the sum of its singular values $\sigma_i(X)$, i.e., $\|X\|_\star = \sum_i \sigma_i(X)$, and serves as a convex proxy for the rank function. The appeal of nuclear norm minimization is that it can be reduced to a semi-definite programming problem and solved globally and efficiently. The data-driven simulation problem can hence be relaxed using the nuclear norm heuristic as follows.

Problem 3 (Data-driven simulation via nuclear norm minimization)**.**
Given *a trajectory $w_d \in \mathcal{B} \in \mathcal{L}^q$ and an input u_s,*
find *the output y_s from the following minimization problem.*

$$\underset{y_s}{\text{minimize}} \quad \left\| \begin{bmatrix} H_L(w_d) & H_L(w'_s) \end{bmatrix} \right\|_\star, \tag{8}$$

where $w'_s = (u'_s, y'_s)$ is defined in (6).

3.4. On the choice of the number L of block rows

So far we have not commented on the number L of block rows of the block-Hankel matrix $\begin{bmatrix} H_L(w_d) & H_L(w'_s) \end{bmatrix}$, except that rank-deficiency occurs as of $L \geq \ell + 1$. Since the lag ℓ is not known *a priori*, an appropriate choice for L is not straightforward. The choice of L should firstly ensure that $L \geq \ell + 1$ for the rank-deficiency of the block-Hankel matrix. Further, the choice of L determines the dimensions of the block-Hankel matrix, and hence has an influence on the subsequent (nuclear norm) minimization problem. As a result, the choice of L has an effect on the performance of the nuclear norm minimization result. In Section 4 we will illustrate, by means of numerical experiments, the influence of the choice of L on the simulation results.

4. NUMERICAL EXPERIMENTS

The numerical experiments were done using CVX, a MATLAB package for specifying and solving convex optimization problems [21, 22].

4.1. The effect of rescaling the data w_d

We generated a random SISO ($m = p = 1$) LTI system using MATLAB's `drss` function of order 10 (i.e., $\ell = 10$), and generated a standard normal zero-mean white Gaussian input sequence u_d of length $T_d = 50$ and the corresponding output y_d. The simulation input u_s was also generated as a standard normal zero-mean white Gaussian sequence of length $T_s = 30$, and the true simulation output is denoted by $\bar{y}_s$. Assuming ℓ is known, we set $L = \ell + 1$, and replaced in Problem 3 the trajectory $w_d = (u_d, y_d)$ by γw_d. We investigated the results for values of γ in the interval $\left[10^{-3}, 10^3\right]$.

Notice that rescaling the trajectory w_d does not alter the rank-deficiency of the (block-)Hankel matrix. Since we assume that the data is exact, $w_d \in \mathcal{B}$ implies that $\gamma w_d \in \mathcal{B}$, for all γ. In this setup u_s is not rescaled, so the simulated output y_s obtained by nuclear norm minimization (Problem 3 with rescaled w_d) returns the simulation output in the original scale, and can immediately be compared to $\bar{y}_s$.

Figure 1 contains a typical result of this analysis, where the parameter γ is varied from 10^{-3} to 10^3 and we show the relative simulation error $\|y_s\|/\|\bar{y}_s\|$ with $\bar{y}_s$ the true simulation output. For $\gamma \ll 1$, the simulation result is (close to) a zero sequence, resulting in a relative error of one. Around $\gamma \approx 1$, a transition occurs from large relative error to small relative error (the specific value at which this occurs varies from one setup to the other). For $\gamma \gg 1$, the nuclear norm minimization is able to perfectly recover the simulation output, resulting in a relative error of zero. We observed the same (qualitative) result on all tested systems, over a wide range of data lengths T_d and T_s, data generating system lags ℓ.

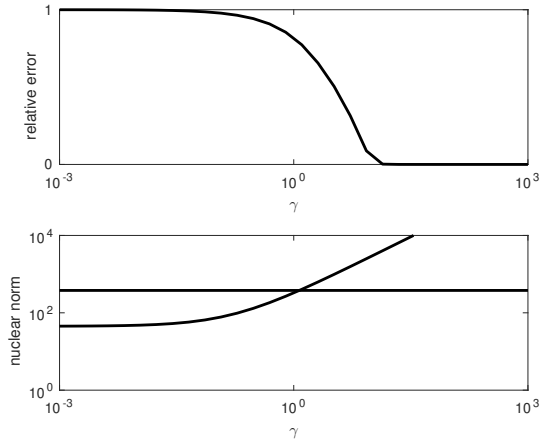

Fig. 1. The top plot shows that rescaling the given trajectory w_d by a factor γ results in a relative simulation error that ranges from one down to zero. The bottom plot shows the nuclear norm of the completed (rescaled) Hankel matrix, which indicates when the transition to exact reconstruction occurs—as a reference, the nuclear norm of the true underlying (unscaled) Hankel matrix is shown.

4.2. The number L of block rows

If the system lag ℓ is not known in advance, an appropriate choice for L should be made. We consider the same simulation setup as in Section 4.1, but vary the number L of block rows. In Figure 2, we show how the choice of L influences the results. For $L < \ell + 1$, we observe that the data-driven simulation does not attain zero error, while for $L \geq \ell + 1$ the transition to zero error is observed—typically, an 'optimal' value is observed for which the transition is sharpest, around the value for L at which $H_L(w_d)$ is square, i.e., $L = 17$.

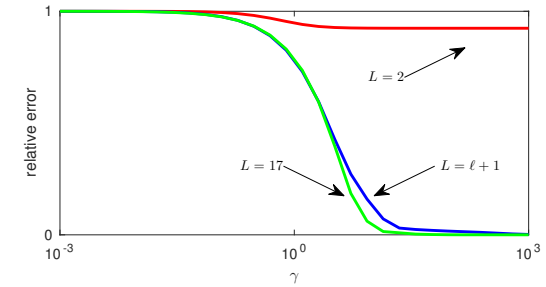

Fig. 2. The number L of block rows of the block-Hankel matrix has an influence on the simulation results. For $L < \ell + 1$ (red), the simulation error does not reach zero. For $L \geq \ell + 1$ (blue), the simulation error reaches zero for sufficiently large values of γ, and our experiments suggest there is an optimal value for which the transition is sharpest (green).

4.3. Exploring the noisy case

We consider the setup of Section 4.1 with additive zero-mean white Gaussian (measurement) noise on $w_d = \bar{w}_d + \tilde{w}_d$ with $\bar{w}_d \in \mathcal{B}$ denoting the true trajectory, and $\tilde{w}_d$ is a standard normal zero-mean white Gaussian noise sequence with SNR of 30 dB on input and output measurements.

We compare the method of the current paper to a model-based approach. The model-based approach proceeds as follows. Assuming that the system lag ℓ is known, we construct the block-Hankel matrix $H_{\ell+1}(w_d)$, from which a numerical basis for the left null space (of dimension one) is computed by the singular value decomposition. The (approximate) left null space is then viewed, as in (1), as the difference equations governing the system, and used to generate the output sequence y_s for the simulation input sequence u_s.

For the data-driven approach, we set $L = \ell + 1$, and let $\gamma = 1000$ to ensure a sufficiently large rescaling of w_d. We then generated 100 Monte Carlo experiments with randomly generated systems of order $\ell = 10$ and random realizations of the input and output data sequences.

The results suggest that data-driven simulation using a nuclear norm heuristic is a competitive alternative for the model-based method. In 52 out of 100 experiments, the data-driven approach resulted in the lowest simulation error. Obviously, the model-based approach performs poorly when a limited number of (estimation) data points is available. However, it should also be remarked that this is a very naive application of the data-driven method, where ideally also the noisy data w_d should be corrected to restore rank-deficiency.

5. DISCUSSION AND PERSPECTIVES

We found that, in the exact case, a rescaling of the given trajectory w_d by a sufficiently large factor γ always allows for an exact recovery using the nuclear norm minimization. We investigated the effect of the number L of block rows of the Hankel matrix, suggesting an optimal value at around the point where the matrix $H_L(w_d)$ is square. Finally, we explored a naive extension to a noisy measurement setup, where again a suitable scaling of the data led to promising results.

We are currently investigating how the value of γ at which the transition occurs, is related to the ratio between nuclear norm of the matrix completion versus nuclear norm of the underlying block-Hankel matrix. Future work is concerned with deriving a theoretical basis explaining the observed behavior,and a thorough analysis of the noisy setup, where both w_d is corrected and the missing elements in w_s are computed.

6. ACKNOWLEDGMENTS

The authors thank prof. Marc Van Barel (KU Leuven) for the suggestion to investigate the effect of the amplitude of the to-be-completed signal on the nuclear norm solution.

7. REFERENCES

[1] I. Markovsky, "A missing data approach to data-driven filtering and control," *IEEE Trans. Automat. Contr.*, vol. 62, no. 4, pp. 1972–1978, April 2017.

[2] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, Applications*, Springer, second edition edition, 2019.

[3] I. Markovsky and K. Usevich, "Structured low-rank approximation with missing data," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 2, pp. 814–830, 2013.

[4] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proc. 22nd Int. Conf. Machine Learning (ICML 2005)*, 2005, pp. 713–719.

[5] N. Natarajan and I. S. Dhillon, "Inductive matrix completion for predicting genedisease associations," *Bioinformatics*, vol. 30, no. 12, pp. i60–i68, 2014.

[6] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, Calif, USA, 2010, pp. 1791–1798.

[7] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, April 2009.

[8] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[9] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," 2001, vol. 6 of *Proceedings American Control Conference*, pp. 4734–4739.

[10] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.

[11] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl*, vol. 34, no. 3, pp. 946–977, 2013.

[12] I. Markovsky, "How effective is the nuclear norm heuristic in solving data approximation problems?," in *Proc. 16th IFAC Symposium on System Identification (SYSID 2012)*, Brussels, 2012, pp. 316–321.

[13] L. Vandenberghe, "Convex optimization techniques in system identification," in *Proc. 16th IFAC Symposium on System Identification (SYSID 2012)*, Brussels, Belgium, 2012, vol. 45(16) of *IFAC Proceedings Volumes*, pp. 71–76.

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

[15] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[16] L. Dai and K. Pelckmans, "On the nuclear norm heuristic for a Hankel matrix completion problem," *Automatica*, vol. 51, pp. 268–272, 2015.

[17] K. Usevich and P. Comon, "Hankel low-rank matrix completion: Performance of the nuclear norm relaxation," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 637–646, June 2016.

[18] W. Liao and A. Fannjiang, "MUSIC for single-snapshot spectral estimation: stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[19] J. W. Polderman and J. C. Willems, *Introduction to mathematical systems theory: a behavioral approach*, Springer, New York, 1998.

[20] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor, "A note on persistency of excitation," *Syst. Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[22] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx`, Mar. 2014.