2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

Alternating direction optimization algorithms for covariance completion problems

Armin Zare, Mihailo R. Jovanović, and Tryphon T. Georgiou

Abstract— **Second-order statistics of nonlinear dynamical systems can be obtained from experiments or numerical simulations. These statistics are relevant in understanding the fundamental physics, e.g., of fluid flows, and are useful for developing low-complexity models. Such models can be used for the purpose of control design and analysis. In many applications, only certain second-order statistics of a limited number of states are available. Thus, it is of interest to complete partially specified covariance matrices in a way that is consistent with the linearized dynamics. The dynamics impose structural constraints on admissible forcing correlations and state statistics. Solutions to such completion problems can be used to obtain stochastically driven linearized models. Herein, we address the covariance completion problem. We introduce an optimization criterion that combines the nuclear norm together with an entropy functional. The two, together, provide a numerically stable and scalable computational approach which is aimed at low complexity structures for stochastic forcing that can account for the observed statistics. We develop customized algorithms based on alternating direction methods that are well-suited for large scale problems.**

Index Terms— **Alternating direction method of multipliers, alternating minimization algorithm, convex optimization, low-rank approximation, nuclear norm regularization, state covariances, structured matrix completion problems.**

I. Introduction

Motivation for this work stems from control-oriented modeling of systems with very large number of degrees of freedom. For example, the Navier-Stokes (NS) equations, which govern the dynamics of fluid flows, are prohibitively complex for purposes of analysis and control design. Thus, it is common practice to investigate low-dimensional models that preserve the essential dynamics. In wall-bounded flows, stochastically driven linearized models of the NS equations have been shown to be capable of qualitatively replicating the structural features of fluid motion [1]–[4]. However, it has also been recognized that white-in-time stochastic forcing is too restrictive to reproduce all statistical features of the nonlinear dynamics [5], [6]. Building on [7], [8], we depart from white-in-time restriction and consider low-complexity dynamical models with colored-in-time excitations that successfully account for the available statistics.

The complexity is quantified by the rank of the correlation structure of excitation sources. This provides a bound on the number of input channels and explains the directionality of input disturbances [9], [10]. The covariance completion problem is formulated by utilizing nuclear norm minimization as a surrogate for rank minimization [11]–[16]. The resulting convex optimization problem can be recast as a semi-definite program (SDP) but cannot be handled by general purpose solvers for large problems. In recent work [17], [18], we provided optimization algorithms for solving the covariance completion problem. We also showed the utility of this framework in explaining turbulent flow statistics [18], [19].

Herein, we consider an optimization criterion that combines the nuclear norm with the logarithmic barrier function as an entropy functional. In order to solve the completion problem for large-scale systems, e.g., fluid flows, we develop customized algorithms based on the Alternating Direction Method of Multipliers (ADMM) and the Alternating Minimization Algorithm (AMA). We demonstrate that AMA, which effectively works as a proximal gradient algorithm on the dual problem is more efficient in handling large problems.

Our presentation is organized as follows. In Section II, we explain the structural constraints imposed on second-order statistics of stochastically forced linear systems. We formulate the covariance completion problem and express the constraint set in a form amenable to alternating optimization methods. In Section III, we present the optimality conditions and derive the dual problem. In Section IV, we present two customized alternating direction algorithms for solving the structured covariance completion problem and present results of numerical experiments in Section V. Finally, we provide concluding thoughts in Section VI.

II. Problem Formulation

Consider a linear time-invariant (LTI) system

$$\dot{x} \;=\; A\,x \,+\, B\,u, \tag{1}$$

where $x(t) \in \mathbb{C}^n$ is the state vector, $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{n\times m}$ are dynamic and input matrices, and $u(t) \in \mathbb{C}^m$ is a stationary zero-mean stochastic process. For a Hurwitz matrix A and controllable pair (A, B), the positive semidefinite matrix X qualifies as being the steady-state covariance matrix of the state in (1),

$$X \;:=\; \lim_{t\,\to\,\infty} \mathbf{E}\left(x(t)\,x^*(t)\right),$$

if and only if the linear equation,

Financial support from the National Science Foundation under Award CMMI 1363266 and the University of Minnesota Informatics Institute Transdisciplinary Faculty Fellowship is gratefully acknowledged. The University of Minnesota Supercomputing Institute is acknowledged for providing computing resources.

Armin Zare, Mihailo R. Jovanović and Tryphon T. Georgiou are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455. E-mails: arminzare@umn.edu, mihailo@umn.edu, tryphon@umn.edu.

$$A\,X \,+\, X\,A^* \;=\; -\,(B\,H^* \,+\, H\,B^*)\,, \qquad (2)$$

is solvable in terms of $H \in \mathbb{C}^{n\times m}$ [7], [8]. Here, $\mathbf{E}$ is the expectation operator and $*$ denotes the complex conjugate transpose. In the case of u being white noise with covariance W, X satisfies the Lyapunov equation

$$A\,X \,+\, X\,A^* \;=\; -B\,W\,B^*.$$

This results from (2) with $H = BW/2$ and yields a negative semi-definite right-hand-side, $-B\,W\,B^*$.

The problem of completing partially known sampled second-order statistics using stochastically-driven LTI models was introduced in [9], and is formulated by introducing

$$Z \;:=\; -\,(A\,X \,+\, X\,A^*)\,.$$

In contrast to the case of white-in-time excitation, Z may have both positive and negative eigenvalues.

The covariance completion problem considered in this paper combines the nuclear norm together with an entropy functional. This provides an approach which is aimed at low-complexity structures for stochastic forcing and facilitates the construction of a particular class of low-pass filters which generate colored-in-time forcing correlations [7], [8]. Figure 1 shows the interconnection of the spatio-temporal filter and the linearized dynamics which can be used to account for partially observed second-order statistics.

The covariance completion problem can be formulated as,

$$\begin{array}{ll} \underset{X,\,Z}{\text{minimize}} & -\log\det(X) \,+\, \gamma\,\|Z\|_* \\ \text{subject to} & A\,X \,+\, X\!A^* \,+\, Z \;=\; 0 \\ & (CXC^*)\circ E \,-\, G \;=\; 0. \end{array} \qquad \text{(CP)}$$

Here matrices A, C, E, and G denote problem data, and Hermitian matrices X, $Z \in \mathbb{C}^{n\times n}$ are optimization variables. Entries of G represent partially available second-order statistics and C is a matrix that establishes the relationship between entries of the state covariance matrix X and partially available statistics resulting from experiments or simulations. The symbol $\circ$ denotes elementwise matrix multiplication and E is the structural identity matrix,

$$E_{ij} \;=\; \begin{cases} 1, & \text{if } G_{ij} \text{ is available} \\ 0, & \text{if } G_{ij} \text{ is unavailable.} \end{cases}$$

The constraint set in (CP) represents the intersection of two linear subspaces, the Lyapunov-like constraint and the linear constraint which incorporates the available statistics.

In (CP), the nuclear norm, i.e., the sum of singular values of a matrix, $\|Z\|_* := \sum_i \sigma_i(Z)$, is used as a proxy for rank minimization [11], [14]. The parameter γ indicates the relative weight on the nuclear norm objective. On the other hand, the logarithmic barrier function in the objective is introduced to guarantee the positive definiteness of X. The convexity of (CP) follows from the convexity of its objective function, denoted as $J_p(X,Z)$, and the convexity of the constraint set.

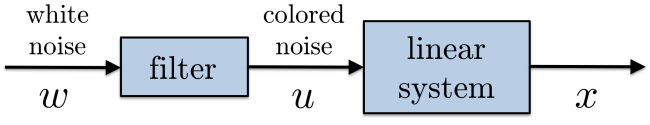


Fig. 1: The cascade connection of an LTI system with a linear filter is used to account for sampled covariance matrix X.

In order to bring (CP) into a convenient form for the alternating direction methods considered in this paper, we reformulate (CP) as

$$\begin{array}{ll} \underset{X,\,Z}{\text{minimize}} & -\log\det(X) \,+\, \gamma\,\|Z\|_* \\ \text{subject to} & \mathcal{A}\,X \,+\, \mathcal{B}\,Z \,-\, \mathcal{C} \;=\; 0, \end{array} \qquad \text{(CP1)}$$

where

$$\mathcal{A} := \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{bmatrix}, \quad \mathcal{B} := \begin{bmatrix} I \\ 0 \end{bmatrix}, \quad \mathcal{C} := \begin{bmatrix} 0 \\ G \end{bmatrix}.$$

Here, $\mathcal{A}_1$, $\mathcal{A}_2 : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ are linear operators, with

$$\begin{aligned} \mathcal{A}_1(X) \;&:=\; A\,X \,+\, X\!A^*, \\ \mathcal{A}_2(X) \;&:=\; (CXC^*)\circ E. \end{aligned}$$

III. Optimality conditions and the dual problem

By splitting Z into positive and negative definite parts,

$$Z \;=\; Z_+ \,-\, Z_-, \qquad Z_+ \;\succeq\; 0, \qquad Z_- \;\succeq\; 0,$$

it can be shown [11, Section 5.1.2] that (CP1) can be cast as an SDP,

$$\begin{array}{ll} \underset{X,\,Z_+,\,Z_-}{\text{minimize}} & -\log\det(X) + \gamma\,(\operatorname{trace}(Z_+) + \operatorname{trace}(Z_-)) \\ \text{subject to} & \mathcal{A}_1(X) \,+\, Z_+ - Z_- \;=\; 0 \\ & \mathcal{A}_2(X) \,-\, G \;=\; 0 \\ & Z_+ \succeq 0, \qquad Z_- \succeq 0. \end{array} \qquad \text{(P)}$$

To derive the dual of the primal problem (P), we introduce the Lagrangian

$$\begin{aligned} \mathcal{L}\,(X,\,Z_\pm;\,Y_1,\,Y_2,\,\Lambda_\pm) \;=\; &-\log\det(X) \;+ \\ &\gamma\operatorname{trace}(Z_+ \,+\, Z_-) \,-\, \langle\Lambda_+,\,Z_+\rangle \,-\, \langle\Lambda_-,\,Z_-\rangle \,+ \\ &\langle Y_1,\,\mathcal{A}_1(X) + Z_+ - Z_-\rangle \,+\, \langle Y_2,\,\mathcal{A}_2(X) - G\rangle\,, \end{aligned}$$

where Hermitian matrices Y_1, Y_2, and $\Lambda_\pm \succeq 0$ are the dual variables, and $\langle\cdot,\cdot\rangle$ represents the standard inner product $\langle M_1, M_2\rangle := \operatorname{trace}(M_1^* M_2)$.

Minimization of $\mathcal{L}$ with respect to the primal variables X and $Z_\pm$ yields the Lagrange dual of (P),

$$\begin{array}{ll} \underset{Y_1,\,Y_2}{\text{maximize}} & \log\det\left(\mathcal{A}_1^\dagger(Y_1) + \mathcal{A}_2^\dagger(Y_2)\right) - \langle G,\,Y_2\rangle + n \\ \text{subject to} & \|Y_1\|_2 \le \gamma, \end{array} \qquad \text{(D)}$$

where the adjoints of the operators $\mathcal{A}_1$ and $\mathcal{A}_2$ are given by

$$\begin{aligned} \mathcal{A}_1^\dagger(Y) \;&=\; A^*\,Y \,+\, Y\!A, \\ \mathcal{A}_2^\dagger(Y) \;&=\; C^*(E\circ Y)\,C. \end{aligned}$$

The dual problem (D) is a convex optimization problem with variables $Y_1, Y_2 \in \mathbb{C}^{n\times n}$ and the objective function

$J_d(Y_1, Y_2)$. These variables are dual feasible if the constraint in (D) is satisfied. This constraint is obtained by minimizing the Lagrangian with respect to Z_+ and Z_-, which leads to

$$\begin{aligned} Y_1 \,+\, \gamma\, I \,\succeq\, \Lambda_+ \,\succeq\, 0, \quad Z_+ \,\succeq\, 0, \\ -Y_1 \,+\, \gamma\, I \,\succeq\, \Lambda_- \,\succeq\, 0, \quad Z_- \,\succeq\, 0. \end{aligned}$$

Therefore, we have that

$$-\gamma\, I \,\preceq\, Y_1 \,\preceq\, \gamma\, I \iff \|Y_1\|_2 \,\leq\, \gamma.$$

In addition, minimizing $\mathcal{L}$ with respect to X yields

$$X^{-1} \,=\, \mathcal{A}_1^\dagger(Y_1) \,+\, \mathcal{A}_2^\dagger(Y_2) \,\succ\, 0. \tag{5}$$

In the case of primal and dual feasibility, any dual feasible pair $(Y_1,\, Y_2)$ gives a lower bound on the optimal value of the primal problem $J_p^\star$. The alternating minimization algorithm of Section IV-B effectively works as a proximal gradient algorithm on the dual problem and is developed to achieve sufficient dual ascent and satisfy (5).

IV. Customized algorithms

We next develop two customized algorithms based on the Alternating Direction Method of Multipliers (ADMM) and the Alternating Minimization Algorithm (AMA). These methods have been effectively employed in low-rank matrix recovery [20], sparse covariance selection [21], image denoising and magnetic resonance imaging [22], sparse feedback synthesis [23], sparse Gaussian graphical model estimation [24], and many other applications [25]–[28]. Our approach exploits the respective structures of the logarithmic barrier function and the nuclear norm, and is well-suited for solving large-scale and distributed optimization problems.

A. Alternating direction method of multipliers

The augmented Lagrangian associated with (CP1) is given by

$$\begin{aligned} \mathcal{L}_\rho(X, Z; Y_1, Y_2) \,&=\, -\log\det(X) \,+\, \gamma\,\|Z\|_* \\ &+\, \langle Y,\, \mathcal{A}X \,+\, \mathcal{B}Z \,-\, \mathcal{C}\rangle \,+\, \frac{\rho}{2}\,\|\mathcal{A}X \,+\, \mathcal{B}Z \,-\, \mathcal{C}\|_F^2, \end{aligned}$$

where $Y := [\, Y_1 \;\; Y_2\,]^*$ with Hermitian $Y_1, Y_2 \in \mathbb{C}^{n\times n}$ is the Lagrange multiplier, ρ is a positive scalar, and $\|\cdot\|_F$ is the Frobenius norm.

The ADMM algorithm uses a sequence of iterations to find the minimizer of the constrained optimization problem (CP1),

$$X^{k+1} \;:=\; \underset{X}{\operatorname{argmin}}\; \mathcal{L}_\rho\,(X,\, Z^k,\, Y^k) \tag{6a}$$

$$Z^{k+1} \;:=\; \underset{Z}{\operatorname{argmin}}\; \mathcal{L}_\rho\,(X^{k+1},\, Z,\, Y^k) \tag{6b}$$

$$Y^{k+1} \;:=\; Y^k \,+\, \rho\,\big(\mathcal{A}X^{k+1} + \mathcal{B}Z^{k+1} - \mathcal{C}\big). \tag{6c}$$

ADMM iterations terminate when conditions on primal and dual residuals are satisfied [26, Section 3.3],

$$\begin{aligned} \|\mathcal{A}X^{k+1} \,+\, \mathcal{B}Z^{k+1} - \mathcal{C}\|_F \;&\leq\; \epsilon, \\ \|\rho\,\mathcal{A}_1^\dagger\,\big(Z^{k+1} - Z^k\big)\,\|_F \;&\leq\; \epsilon. \end{aligned}$$

1) Solution to the X-minimization problem (6a): For fixed $\{Z^k, Y^k\}$, minimizing the augmented Lagrangian with respect to X amounts to

$$\underset{X}{\text{minimize}} \;\; -\log\det(X) + \frac{\rho}{2}\,\|\mathcal{A}X \,-\, U^k\|_F^2$$

where $U^k := -\big(\mathcal{B}Z^k - \mathcal{C} + (1/\rho)Y^k\big)$. We use a proximal gradient approach [29] to solve this sub-problem. By linearizing the quadratic term around the current inner iterate X_i and adding the quadratic penalty on the difference between X and X_i, X_{i+1} is obtained as the minimizer of

$$-\log\det(X) \,+\, \rho\,\big\langle \mathcal{A}^\dagger\,\big(\mathcal{A}X_i - U^k\big),\, X\big\rangle \,+\, \frac{\mu}{2}\|X - X_i\|_F^2. \tag{7}$$

Here, the parameters ρ and μ satisfy $\mu \,\geq\, \rho\,\lambda_{max}(\mathcal{A}^\dagger\mathcal{A})$, where power iteration is used to compute $\lambda_{max}(\mathcal{A}^\dagger\mathcal{A})$. By taking the variation of (7) with respect to X, we obtain the first order optimality condition

$$\mu\, X \,-\, X^{-1} \,=\, \big(\mu\, I - \rho\,\mathcal{A}^\dagger\mathcal{A}\big)\, X_i \,+\, \rho\,\mathcal{A}^\dagger\,\big(U^k\big). \tag{8}$$

The solution to (8) is given by

$$X_{i+1} \,=\, V\,\mathrm{diag}\,(g)\,V^*,$$

where g is a vector with the jth entry,

$$g_j \,=\, \frac{\lambda_j}{2\mu} \,+\, \sqrt{\left(\frac{\lambda_j}{2\mu}\right)^2 \,+\, \frac{1}{\mu}}.$$

Here, λ_j's are the eigenvalues of the matrix on the right-hand-side of (8) and V is the matrix of the corresponding eigenvectors. As typically done in proximal gradient algorithms [29], starting with $X_0 \,:=\, X^k$, we obtain X^{k+1} by repeating inner iterations until the desired accuracy is reached.

2) Solution to the Z-minimization problem (6b): For fixed $\{X^{k+1},\, Y^k\}$, the augmented Lagrangian is minimized with respect to Z,

$$\underset{Z}{\text{minimize}} \quad \gamma\,\|Z\|_* \,+\, \frac{\rho}{2}\|Z - V^k\|_F^2 \tag{9}$$

where $V^k \;:=\; -\big(\mathcal{A}_1\,\big(X^{k+1}\big) + (1/\rho)Y_1^k\big)$. The solution to (9) is obtained by singular value thresholding [30],

$$Z^{k+1} \,=\, \mathcal{S}_{\gamma/\rho}(V^k).$$

For this purpose we first compute the singular value decomposition of the symmetric matrix $V^k \,=\, U\,\Sigma\,U^*$, where Σ is the diagonal matrix of the singular values σ_i. The soft-thresholding operator $\mathcal{S}_\tau$ is defined as

$$\mathcal{S}_\tau(V^k) \;:=\; U\,\mathcal{S}_\tau(\Sigma)\,U^*, \quad \mathcal{S}_\tau(\Sigma) \,=\, \mathrm{diag}\,\big((\sigma_i \,-\, \tau)_+\big),$$

and $a_+ = \max\{a, 0\}$. Thus, the optimality condition in (6b) is satisfied by applying the soft-thresholding operator $\mathcal{S}_{\gamma/\rho}$ on the singular values of V^k.

We also consider an accelerated variant of the ADMM algorithm. The accelerated algorithm is simply ADMM with a Nesterov-type (predictor-corrector) acceleration step. Due to weak convexity of the objective function, a restart rule

is required [27]. Although a global convergence rate cannot be guaranteed, the restart rule reduces the oscillating behavior which is often encountered in first-order iterative methods [27], [31].

Accelerated ADMM algorithm is given by,

$$\begin{aligned}
X^k &:= \underset{X}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(X,\, \hat{Z}^k,\, \hat{Y}^k\right)\\
Z^k &:= \underset{Z}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(X^k,\, Z,\, \hat{Y}^k\right)\\
Y^k &:= \hat{Y}^k + \rho\left(\mathcal{A}\left(X^k\right) + \mathcal{B}Z^k - \mathcal{C}\right)\\
c_k &:= \tfrac{1}{\rho}\left\|Y^k - \hat{Y}^k\right\|_F^2 + \rho\left\|Z^k - \hat{Z}^k\right\|_F^2\\
&\text{if } c_k < \eta\, c_{k-1},\\
&\quad \alpha_{k+1} := (1+\sqrt{1+4\,\alpha_k^2})/2\\
&\quad \hat{Z}^{k+1} := Z^k + \tfrac{\alpha_k - 1}{\alpha_{k+1}}(Z^k - Z^{k-1})\\
&\quad \hat{Y}^{k+1} := Y^k + \tfrac{\alpha_k - 1}{\alpha_{k+1}}(Y^k - Y^{k-1})\\
&\text{else}\\
&\quad \alpha_{k+1} = 1,\ \hat{Z}^{k+1} = Z^{k-1},\ \hat{Y}^{k+1} = Y^{k-1}\\
&\quad c_k \leftarrow \eta^{-1}\, c_{k-1}
\end{aligned}$$

Following [27], the algorithm is initialized with $Z^{-1} = \hat{Z}^0$, $Y^{-1} = \hat{Y}^0$, $\rho > 0$, $\alpha_1 = 1$, and terminated using similar criteria as in the ADMM algorithm.

B. Alternating minimization algorithm

The alternating minimization algorithm which was originally developed by Tseng in [32] involves simpler steps than ADMM, but requires strong convexity of the smooth part of the objective function. Since the logarithmic barrier function in (CP) is strongly convex over any compact subset of the positive definite cone [33], we can use AMA to solve (CP).

AMA follows a sequence of iterations,

$$X^{k+1} := \underset{X}{\operatorname{argmin}}\ \mathcal{L}_0\left(X,\, Z^k,\, Y^k\right) \tag{10a}$$

$$Z^{k+1} := \underset{Z}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(X^{k+1},\, Z,\, Y^k\right) \tag{10b}$$

$$Y^{k+1} := Y^k + \rho\left(\mathcal{A}X^{k+1} + \mathcal{B}Z^{k+1} - \mathcal{C}\right) \tag{10c}$$

which terminate when the duality gap

$$\Delta_{\text{gap}} := -\log\det\left(X^{k+1}\right) + \gamma\left\|Z^{k+1}\right\|_* - J_d\left(Y^{k+1}\right),$$

and the primal residual

$$\Delta_{\text{p}} := \left\|\mathcal{A}\,X^{k+1} + \mathcal{B}\,Z^{k+1} - \mathcal{C}\right\|_F,$$

are sufficiently small, i.e., $|\Delta_{\text{gap}}| \le \epsilon$, and $\Delta_{\text{p}} \le \epsilon$. In the X-minimization step, AMA minimizes the Lagrangian $\mathcal{L}_0$ to obtain a closed form expression for X^{k+1}. This is in contrast to ADMM which aims at minimizing the augmented Lagrangian in both X- and Z-minimization steps. It can be shown that AMA works as a proximal gradient on the dual function, which allows us to select the step-size ρ in order to achieve sufficient dual ascent.

1) Solution to the X-minimization problem (10a)*:* At the kth iteration of AMA, minimizing the Lagrangian $\mathcal{L}_0$ with respect to X for fixed $\{Z^k,\, Y^k\}$ yields

$$X^{k+1} = \left(\mathcal{A}^\dagger\left(Y^k\right)\right)^{-1}. \tag{11}$$

2) Solution to the Z-minimization problem (10b)*:* This step is identical to (6b) in the ADMM algorithm.

3) Lagrange multiplier update: The expressions for X^{k+1} and Z^{k+1} can be used to bring (10c) into the following form

$$\begin{aligned}
Y_1^{k+1} &= \mathcal{T}_\gamma\left(Y_1^k + \rho\,\mathcal{A}_1\left(X^{k+1}\right)\right)\\
Y_2^{k+1} &= Y_2^k + \rho\left(\mathcal{A}_2\left(X^{k+1}\right) - G\right).
\end{aligned}$$

For Hermitian matrix M with singular value decomposition $M = U\,\Sigma\,U^*$, we have

$$\begin{aligned}
\mathcal{T}_\tau(M) &:= U\,\mathcal{T}_\tau(\Sigma)\,U^*\\
\mathcal{T}_\tau(\Sigma) &= \operatorname{diag}\left(\min\left(\max(\sigma_i, -\tau), \tau\right)\right)
\end{aligned}$$

where the saturation operator $\mathcal{T}_\tau$ restricts the singular values of M between $-\tau$ and τ. This guarantees dual feasibility of the update, i.e., $\|Y_1^{k+1}\|_2 \le \gamma$ at each iteration, and justifies the choice of stopping criteria in ensuring primal feasibility of the solution.

4) Choice of step-size for the dual update (10c)*:* We follow an enhanced variant of AMA [24] which utilizes an adaptive Barzilia-Borwein step-size selection [34] in (10b) and (10c) to guarantee sufficient dual ascent and positive definiteness of X. Our numerical experiments indicate that this provides substantial acceleration relative to the use of a fixed step-size. Since the standard Barzilia-Borwein step-size may not always satisfy the feasibility or the sufficient ascent conditions, we determine an appropriate step-size through backtracking.

At the kth iteration of AMA, an initial step-size,

$$\rho_{k,0} = \frac{\left\langle Y^{k+1} - Y^k,\, Y^{k+1} - Y^k\right\rangle}{\left\langle Y^{k+1} - Y^k,\, \nabla J_d(Y^k) - \nabla J_d(Y^{k+1})\right\rangle},$$

is adjusted through a backtracking procedure to guarantee positive definiteness of the subsequent iterate of (10a) and sufficient ascent of the dual function,

$$\mathcal{A}^\dagger\left(Y^{k+1}\right) \succ 0 \tag{12a}$$

$$\begin{aligned}
J_d\left(Y^{k+1}\right) \ge\ & J_d\left(Y^k\right) + \left\langle \nabla J_d(Y^k),\, Y^{k+1} - Y^k\right\rangle -\\
& \frac{1}{2\rho_k}\|Y^{k+1} - Y^k\|_F^2.
\end{aligned} \tag{12b}$$

Here, ∇J_d is the gradient of the dual function. Condition (12a) guarantees the positive definiteness of X^{k+1}, cf. (11), and the right hand side of (12b) is a local quadratic approximation of the dual objective around Y^k.

C. Computational complexity

The X-minimization step involves an eigenvalue decomposition in ADMM, and a matrix inversion in AMA, which costs $O(n^3)$ operations in both cases (n is the number of states). Both methods have Z-minimization steps that amount to a singular value decomposition, and require $O(n^3)$ operations. Thus, the total computational cost for a single iteration

TABLE I: Comparison of different algorithms (in seconds) for different number of masses and $\gamma = 10$.

N	CVX	ADMM	Fast ADMM + Restart	AMA
10	67.04	6.80	3.10	7.00
20	942.23	105.55	65.52	36.97
50	–	3930.4	3492.5	625.92
100	–	40754	34420	5429.8

of our customized algorithms is $O(n^3)$. In contrast, standard SDP solvers have a worst-case complexity of $O(n^6)$.

V. Numerical experiments

We present an illustrative example to demonstrate the effectiveness of our customized algorithms. Consider a mass-spring-damper (MSD) system subject to stochastic disturbances that are generated by a low-pass filter,

$$\text{low-pass filter: } \dot{\zeta} = -\zeta + d \tag{13a}$$

$$\text{MSD system: } \dot{x} = Ax + B\zeta \tag{13b}$$

The state vector $x = [\, p^* \;\; v^* \,]^*$, contains position and velocity of masses, and d represents a zero-mean unit variance white process. State and input matrices are

$$A = \begin{bmatrix} O & I \\ -T & -I \end{bmatrix}, \quad B_\zeta = \begin{bmatrix} 0 \\ I \end{bmatrix},$$

where O and I are zero and identity matrices and T is a symmetric tridiagonal Toeplitz matrix with 2 on the main diagonal and -1 on the first upper and lower sub-diagonals.

The steady-state covariance of system (13) can be found as the solution to the Lyapunov equation,

$$\tilde{\mathcal{A}}\,\Sigma + \Sigma\,\tilde{\mathcal{A}}^* + \tilde{\mathcal{B}}\,\tilde{\mathcal{B}}^* = 0,$$

where

$$\tilde{\mathcal{A}} = \begin{bmatrix} A & B \\ O & -I \end{bmatrix}, \quad \tilde{\mathcal{B}} = \begin{bmatrix} 0 \\ I \end{bmatrix},$$

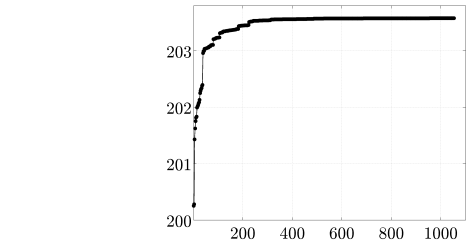
$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{x\zeta} \\ \Sigma_{\zeta x} & \Sigma_{\zeta\zeta} \end{bmatrix}.$$

The sub-covariance Σ_{xx} denotes the state covariance of the MSD system, and is partitioned as

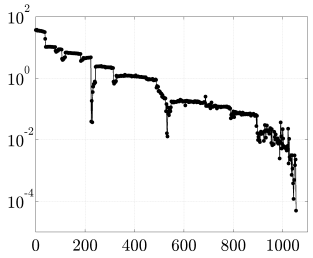
$$\Sigma_{xx} = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pv} \\ \Sigma_{vp} & \Sigma_{vv} \end{bmatrix}.$$

We assume knowledge of one-point correlations of the position and velocity of masses, i.e., the diagonal elements of Σ_{pp}, Σ_{vv}, and Σ_{pv}. In order to account for these available statistics, we solve (CP) for a state covariance X of the MSD system which agrees with the available statistics.

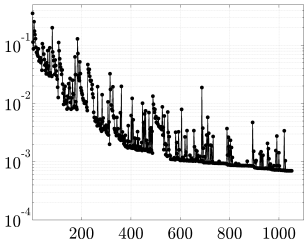
Numerical experiments were conducted for $N = 10$, 20, 50, and 100 masses and for various values of γ. Iterations were run for each method until an iterate achieves a certain distance from optimality, i.e., $\|X^k - X^\star\|/\|X^\star\| < \epsilon_1$ and $\|Z^k - Z^\star\|/\|Z^\star\| < \epsilon_2$. For $\gamma = 10$, Table I compares various methods based on run times (sec). For $N = 50$ and $N = 100$ CVX [35] ran out of memory. The choice of $\epsilon_1, \epsilon_2 = 10^{-3}$ and $\gamma = 10$, guarantees sufficient optimality in matching primal constraints and achieving low-rank solutions which bring the objective to within 0.1% of $J_p(X^\star, Z^\star)$. Clearly, AMA outperforms ADMM for large problems.



(a) $J_d(Y_1, Y_2)$



(b) Δ_{gap}



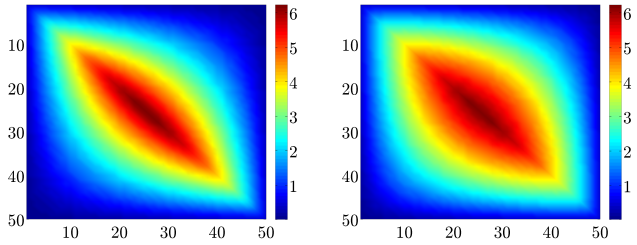
(c) Δ_{p}

Fig. 2: Performance of the customized AMA for the MSD system with 50 masses, $\gamma = 2.2$, and $\epsilon = 10^{-3}$. (a) The dual objective function $J_d(Y_1, Y_2)$ of (CP); (b) the duality gap, $|\Delta_{\text{gap}}|$; and (c) the primal residual, Δ_{p}.

We now focus on the performance of customized alternating minimization algorithm. For 50 masses and $\gamma = 2.2$ we use the algorithm discussed in Section IV-B to solve (CP) with $\epsilon = 10^{-3}$. Figure 2a illustrates the monotonic increase of the dual objective function. The absolute value of the duality gap, $|\Delta_{\text{gap}}|$, and the primal residual, Δ_{p} are displayed in Fig. 2, thereby demonstrating convergence.

In (CP), the parameter γ determines the importance of the nuclear norm relative to the logarithmic barrier function. While larger values of γ result in lower rank solutions for Z, they may fail to achieve a reasonable completion of the "ideal" state covariance Σ_{xx}. When the true values of all entries of the covariance matrix are not known, γ is typically chosen on an empirical basis or by cross-validation. In our example, however, the true state covariance Σ_{xx} is known. Minimum error is obtained at $\gamma = 1.4$, but this value of γ does not yield a low-rank input correlation Z. For $\gamma = 2.2$ reasonable matching is obtained (82.7% matching) and the resulting Z displays a clear-cut in its singular values with 62 of them being nonzero. Figure 3b shows the recovered covariance matrix of mass positions,

X_{pp}. We observe reasonable correspondence with the true covariance Σ_{pp} in Fig. 3a.



(a) Σ_{pp} (b) X_{pp}

Fig. 3: The matrix Σ_{pp} of the MSD system and the covariance X_{pp} which comes as a solution to (CP).

VI. Concluding remarks

We are interested in explaining partially known second-order statistics that originate from experimental measurements or simulations using stochastic linear models. We assume that the linearized approximation of the dynamical generator is known whereas the nature and directionality of disturbances that can explain partially observed statistics are unknown. This inverse problem can be formulated as a convex optimization problem, which utilizes nuclear norm minimization to identify noise parameters of low rank and to complete unavailable covariance data. We also constrain the solution set to positive definite state covariances by including a logarithmic barrier function in the objective. To efficiently solve covariance completion problems of large size we develop customized algorithms based on alternating direction methods. Numerical experiments show that the enhanced variant of AMA performs much better than strategies based on ADMM, especially for large-scale problems.

References

[1] B. F. Farrell and P. J. Ioannou, "Stochastic forcing of the linearized Navier-Stokes equations," *Phys. Fluids A*, vol. 5, no. 11, pp. 2600–2609, 1993.

[2] B. Bamieh and M. Dahleh, "Energy amplification in channel flows with stochastic excitation," *Phys. Fluids*, vol. 13, no. 11, pp. 3258–3269, 2001.

[3] M. R. Jovanović and B. Bamieh, "Componentwise energy amplification in channel flows," *J. Fluid Mech.*, vol. 534, pp. 145–183, 2005.

[4] R. Moarref and M. R. Jovanović, "Model-based design of transverse wall oscillations for turbulent drag reduction," *J. Fluid Mech.*, vol. 707, pp. 205–240, September 2012.

[5] M. R. Jovanović and B. Bamieh, "Modelling flow statistics using the linearized Navier-Stokes equations," in *Proceedings of the 40th IEEE Conference on Decision and Control*, 2001, pp. 4944–4949.

[6] M. R. Jovanović and T. T. Georgiou, "Reproducing second order statistics of turbulent flows using linearized Navier-Stokes equations with forcing," in *Bulletin of the American Physical Society*, Long Beach, CA, November 2010.

[7] T. T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1056–1066, 2002.

[8] T. T. Georgiou, "Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1811–1823, 2002.

[9] Y. Chen, M. R. Jovanović, and T. T. Georgiou, "State covariances and the matrix completion problem," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 1702–1707.

[10] A. Zare, Y. Chen, M. R. Jovanović, and T. T. Georgiou, "Low-complexity modeling of partially available second-order statistics via matrix completion," *IEEE Trans. Automat. Control*, 2014, submitted; also arXiv:1412.3399v1.

[11] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

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

[13] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. on Inform. Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[14] 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.

[15] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[16] R. Salakhutdinov and N. Srebro, "Collaborative filtering in a non-uniform world: Learning with the weighted trace norm," *Advances in Neural Information Processing Systems 24*, 2011.

[17] F. Lin, M. R. Jovanović, and T. T. Georgiou, "An ADMM algorithm for matrix completion of partially known state covariances," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 1684–1689.

[18] A. Zare, M. R. Jovanović, and T. T. Georgiou, "Completion of partially known turbulent flow statistics," in *Proceedings of the 2014 American Control Conference*, Portland, OR, 2014, pp. 1680–1685.

[19] A. Zare, M. R. Jovanović, and T. T. Georgiou, "Completion of partially known turbulent flow statistics via convex optimization," in *Proceedings of the 2014 Summer Program*, Center for Turbulence Research, Stanford University/NASA, 2014, pp. 345–354.

[20] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 1, pp. 57–81, 2011.

[21] X. Yuan, "Alternating direction method for covariance selection models," *J. Sci. Comput.*, vol. 51, no. 2, pp. 261–273, 2012.

[22] T. Goldstein and S. Osher, "The split Bregman method for ℓ_1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[23] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2426–2431, 2013.

[24] O. A. Dalal and B. Rajaratnam, "G-AMA: Sparse Gaussian graphical model estimation via alternating minimization," *arXiv preprint arXiv:1405.3034*, 2014.

[25] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented Lagrangian methods for semidefinite programming," *Math. Program. Comp.*, vol. 2, no. 3-4, pp. 203–230, 2010.

[26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[27] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM J. Imaging Sci.*, vol. 7, no. 3, pp. 1588–1623, 2014.

[28] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Operator splitting for conic optimization via homogeneous self-dual embedding," *arXiv preprint arXiv:1312.3039*, 2013.

[29] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.

[30] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[31] B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, pp. 1–18, 2013.

[32] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM J. Control Optim.*, vol. 29, no. 1, pp. 119–138, 1991.

[33] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, 2008.

[34] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

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