An alternating minimization algorithm for structured covariance completion problems

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

***Abstract*— State statistics of linear systems satisfy certain structural constraints that arise from the underlying dynamics and the directionality of input disturbances. These statistics are relevant in understanding the fundamental physics and can be used to develop control-oriented models for large-scale dynamical systems, e.g., stochastically forced linearized Navier-Stokes equations. The problem of completing partially known state statistics via stochastically driven linear time-invariant systems gives rise to a class of structured covariance completion problems. In this, nuclear norm minimization is used to identify forcing models of low complexity. Herein, we develop a customized alternating minimization algorithm (AMA) to solve this optimization problem for large-scale systems. We interpret AMA as a proximal gradient for the dual problem which allows us to prove convergence for the algorithm with fixed step-size.**

***Index Terms*— Alternating minimization algorithm, convex optimization, nuclear norm regularization, state covariances, structured matrix completion problems.**

I. Introduction

Motivated by the necessity for control-oriented modeling of systems with large number of degrees of freedom, e.g., fluid flows, we have developed a framework to account for partially observed second-order statistics using stochastically forced linear models [1]–[5]. These models have shown to preserve the essential dynamics. For instance, linearization of the Navier-Stokes equations around mean-velocity, driven by white-in-time stochastic excitation has shown to qualitatively replicate structural features of wall-bounded shear flows [6]–[9]. However, it has also been recognized that white-in-time stochastic forcing is too restrictive to reproduce all statistical features of the fluctuating velocity field [5], [10], [11]. Building on [12], [13], we depart from white-in-time restriction and consider low-complexity models with colored-in-time excitations that successfully account for all observed second-order statistics that may be available from experiments or numerical simulations.

In our setting, the complexity is quantified by the rank of the correlation structure of excitation sources and we utilize nuclear norm minimization as a surrogate for rank minimization [14], [15]. The resulting convex optimization problem can be cast as a semidefinite program (SDP) which is readily solvable by standard software for small-size problems. However, for larger problems, we have developed customized optimization algorithms based on alternating direction methods [4], [16]. In this presentation, we focus on a customized alternating minimization algorithm (AMA) and show that AMA can be interpreted as a proximal gradient for the dual problem. This interpretation can be used to establish sub-linear convergence of the algorithm [4].

Financial support from the National Science Foundation under Award CMMI 1363266, the University of Minnesota Informatics Institute Transdisciplinary Faculty Fellowship, and the University of Minnesota Doctoral Dissertation Fellowship is gratefully acknowledged.

Armin Zare, Yongxin Chen, 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, chen2468@umn.edu, mihailo@umn.edu, tryphon@umn.edu.

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 definite matrix X qualifies as being the steady-state covariance matrix of the state in (1),

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

if and only if the linear equation,

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

is solvable in terms of $H \in \mathbb{C}^{n\times m}$ [12], [13]. Here, $\mathbf{E}$ is the expectation operator and $*$ denotes the complex conjugate transpose. When u is 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. In contrast, for the case of colored-in-time stochastic excitation of the linear dynamics, the right-hand-side of (2), which we denote as

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

may have both positive and negative eigenvalues.

The covariance completion problem is formulated as,

$$\begin{array}{ll} \underset{X,\,Z}{\text{minimize}} & -\log\det\,(X) \;+\; \gamma\,\|Z\|_* \\ \text{subject to} & \mathcal{A}_1(X) \;+\; Z \;=\; 0 \\ & \mathcal{A}_2(X) \;-\; G \;=\; 0. \end{array} \tag{CC}$$

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

$$\begin{aligned} \mathcal{A}_1(X) \;&:=\; A\,X \;+\; X\,A^* \\ \mathcal{A}_2(X) \;&:=\; (C\,X\,C^*) \circ E. \end{aligned} \tag{3}$$

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 the Hermitian matrix 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. While the logarithmic barrier function in the objective is introduced to guarantee the positive definiteness of X, 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 [14], [15] with the parameter γ indicating the relative weight on the nuclear norm. The symbol $\circ$ denotes elementwise matrix multiplication and E is the structural identity matrix,

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

Convexity of (CC) follows from the convexity of its objective function $J_p(X, Z)$ and convexity of the constraint set [17].

The Lagrange dual of (CC) is given by

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

where Hermitian matrices Y_1, Y_2 are dual variables associated with equality constraints in (CC) and $\mathcal{A}_1^\dagger$ and $\mathcal{A}_2^\dagger$ are the adjoints of the operators defined in (3) [4]. We denote the objective of (D) as $J_d(Y_1, Y_2)$.

III. Alternating minimization algorithm

The logarithmic barrier function in (CC) is strongly convex over any compact subset of the positive definite cone [18]. As a result, problem (CC) is well-suited for the application of AMA, which requires strong convexity of the smooth part of the objective function [19].

The augmented Lagrangian associated with (CC) is

$$\begin{aligned} \mathcal{L}_\rho(X, Z; Y_1, Y_2) \;=\; & -\log\det X \,+\, \gamma\,\|Z\|_* \,+ \\ & \langle Y_1,\, \mathcal{A}_1(X) + Z\rangle \,+\, \langle Y_2,\, \mathcal{A}_2(X) - G\rangle \,+ \\ & \frac{\rho}{2}\,\|\mathcal{A}_1(X) \,+\, Z\|_F^2 \,+\, \frac{\rho}{2}\,\|\mathcal{A}_2(X) \,-\, G\|_F^2 \end{aligned}$$

where ρ is a positive scalar and $\|\cdot\|_F$ is the Frobenius norm.

AMA follows a sequence of iterations,

$$X^{k+1} \;:=\; \underset{X}{\operatorname{argmin}}\; \mathcal{L}_0\left(X,\, Z^k;\, Y_1^k,\, Y_2^k\right) \tag{4a}$$

$$Z^{k+1} \;:=\; \underset{Z}{\operatorname{argmin}}\; \mathcal{L}_\rho\left(X^{k+1},\, Z;\, Y_1^k,\, Y_2^k\right) \tag{4b}$$

$$Y_1^{k+1} \;:=\; Y_1^k \,+\, \rho\left(\mathcal{A}_1(X^{k+1}) \,+\, Z^{k+1}\right) \tag{4c}$$

$$Y_2^{k+1} \;:=\; Y_2^k \,+\, \rho\left(\mathcal{A}_2(X^{k+1}) \,-\, G\right) \tag{4d}$$

which terminate when the duality gap

$$\Delta_{\text{gap}} \;:=\; -\log\det X^{k+1} + \gamma\,\|Z^{k+1}\|_* - J_d\left(Y_1^{k+1}, Y_2^{k+1}\right)$$

and the primal residual Δ_{p} are sufficiently small. In the X-minimization step (4a), AMA leads to a *closed-form expression* for X^{k+1}. This is followed by the Z-minimization step (4b) which amounts to the soft-thresholding operator acting on the singular values of a matrix. Finally, the Lagrange multipliers, Y_1 and Y_2, are updated based on the primal residuals with the step-size ρ. The step-size is obtained by adjusting an initial BB step-size [20] through a backtracking procedure to guarantee positive definiteness of the subsequent iterate of (4a) and sufficient dual ascent; see [16], [21] for additional details.

Our customized AMA is summarized as Algorithm 1. Here $\mathcal{A}^\dagger$ denotes the adjoint of the operator pair $\mathcal{A} := (\mathcal{A}_1, \mathcal{A}_2)$.

Algorithm 1 Customized AMA

input: A, G, $\gamma > 0$, tolerances ϵ_1, ϵ_2, and backtracking constant $\beta \in (0, 1)$.

initialize: $k \,=\, 0$, $\rho_{0,0} \,=\, 1$, $\Delta_{\text{gap}} \,=\, \Delta_{\text{p}} \,=\, 2\epsilon_1$, $Y_2^0 \,=\, \mathrm{O}_{n\times n}$, and choose Y_1^0 such that $\mathcal{A}_1^\dagger(Y_1^0) \,=\, (\gamma/\|Y_1^0\|_2) I_{n\times n}$.

while: $|\Delta_{\text{gap}}| > \epsilon_1$ and $\Delta_{\text{p}} > \epsilon_2$,

$$\begin{aligned} X^{k+1} \;&=\; (\mathcal{A}^\dagger(Y_1^k, Y_2^k))^{-1} \\ &\text{compute } \rho_k\text{: } \{\beta^j \rho_{k,0}\}_{j=0,1,\ldots} \text{ such that } X^{k+2} \succ 0 \\ &\text{and } Y_1^{k+1} \text{ and } Y_2^{k+1} \text{ achieve sufficient dual ascent} \\ Z^{k+1} \;&=\; \mathcal{S}_{\gamma/\rho}\left(-\mathcal{A}_1(X^{k+1}) \,-\, (1/\rho) Y_1^k\right) \\ Y_1^{k+1} \;&=\; Y_1^k \,+\, \rho\left(\mathcal{A}_1(X^{k+1}) \,+\, Z^{k+1}\right) \\ Y_2^{k+1} \;&=\; Y_2^k \,+\, \rho\left(\mathcal{A}_2(X^{k+1}) \,-\, G\right) \\ k \;&=\; k + 1 \\ \rho_{k,0} \;&=\; \textstyle\sum_{i\,=\,1}^2 \|Y_i^{k+1} - Y_i^k\|_F^2 \;/ \\ &\qquad \textstyle\sum_{i\,=\,1}^2 \left\langle Y_i^{k+1} - Y_i^k,\, \nabla J_d(Y_i^k) - \nabla J_d(Y_i^{k+1})\right\rangle \end{aligned}$$

endwhile

output: ϵ-optimal solutions, X^{k+1} and Z^{k+1}.

Computational complexity: The X- and Z-minimization steps respectively involve a matrix inversion and a singular value decomposition, which require $O(n^3)$ operations each. However, the Z-minimization steps is embedded within an iterative backtracking procedure for selecting the step-size ρ_k. Thus, if this step-size selection takes q inner iterations, the total computational cost for a single iteration of AMA is $O(qn^3)$. In contrast, the worst-case complexity of standard SDP solvers is $O(n^6)$.

IV. AMA as a proximal gradient on the dual

We next show that (4c) and (4d) are equivalent to those obtained by applying the proximal gradient algorithm to (D).

The dual problem (D) takes the following form

$$\underset{Y_1,Y_2}{\text{minimize}}\;\; f(Y_1, Y_2) \,+\, g(Y_1, Y_2) \tag{5}$$

where $f(Y_1, Y_2) \;=\; -\log\det\mathcal{A}^\dagger(Y_1, Y_2) \,-\, \langle G, Y_2\rangle$ and $g(Y_1, Y_2)$ denotes the indicator function

$$\mathcal{I}(Y_1) \;=\; \begin{cases} 0, & \|Y_1\|_2 \le \gamma \\ +\infty, & \text{otherwise.} \end{cases}$$

Both f: $(\mathbb{C}^{n\times n}, \mathbb{C}^{p\times p}) \to \mathbb{R}$ and g: $(\mathbb{C}^{n\times n}, \mathbb{C}^{p\times p}) \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions and f is continuously differentiable. For $Y_1 \in \mathbb{C}^{n\times n}$ and $Y_2 \in \mathbb{C}^{p\times p}$, the proximal operator of g, prox_g: $(\mathbb{C}^{n\times n}, \mathbb{C}^{p\times p}) \to (\mathbb{C}^{n\times n}, \mathbb{C}^{p\times p})$ is given by

$$\text{prox}_g(V_1, V_2) \;=\; \underset{Y_1, Y_2}{\text{argmin}}\; g(Y_1, Y_2) \;+\; \frac{1}{2}\sum_{i=1}^{2} \|Y_i \;-\; V_i\|_F^2$$

where V_1 and V_2 are fixed matrices. For (5), the proximal gradient method [22] determines the updates as

$$(Y_1^{k+1}, Y_2^{k+1}) \;:=\; \text{prox}_{\rho g}(Y_1^k \;-\; \rho\,\nabla_{Y_1} f(Y_1^k, Y_2^k), \\ Y_2^k \;-\; \rho\,\nabla_{Y_2} f(Y_1^k, Y_2^k))$$

where $\rho > 0$ is the step-size. For $\rho \in (0, 1/L]$ this method converges with rate $O(1/k)$ [23]. Here, L represents a bound on the Lipschitz constant.

Application of the proximal gradient method to the dual problem (5) yields

$$Y_1^{k+1} \;:=\; \underset{Y_1}{\text{argmin}}\; \left\langle \nabla_{Y_1}(-\log\det \mathcal{A}^\dagger(Y_1^k, Y_2^k)), Y_1 \right\rangle \;+ \\ \mathcal{I}\,(Y_1) \;+\; \frac{L}{2}\,\|Y_1 \;-\; Y_1^k\|_F^2 \quad \text{(6a)}$$

$$Y_2^{k+1} \;:=\; \underset{Y_2}{\text{argmin}}\; \left\langle \nabla_{Y_2}(-\log\det \mathcal{A}^\dagger(Y_1^k, Y_2^k)), Y_2 \right\rangle \;+ \\ \langle G, Y_2\rangle \;+\; \frac{L}{2}\,\|Y_2 \;-\; Y_2^k\|_F^2 \quad \text{(6b)}$$

The gradient in (6a) is determined by

$$\nabla_{Y_1}(-\log\det \mathcal{A}^\dagger(Y_1^k, Y_2^k)) \;=\; -\mathcal{A}_1(\mathcal{A}^\dagger(Y_1^k, Y_2^k)^{-1})$$

and we thus have

$$Y_1^{k+1} \;:=\; \underset{Y_1}{\text{argmin}}\; \mathcal{I}(Y_1) \;+ \\ \frac{L}{2}\,\|Y_1 \;-\; (Y_1^k \;+\; \frac{1}{L}\,\mathcal{A}_1(\mathcal{A}^\dagger(Y_1^k, Y_2^k)^{-1}))\|_F^2. \quad (7)$$

Since $X^{k+1} = \mathcal{A}^\dagger(Y_1^k, Y_2^k)^{-1}$, it follows that the dual update Y_1^{k+1} given by (4c) solves (7) with $\rho = 1/L$. This is because the saturation operator $\mathcal{T}_\gamma$ represents the proximal mapping for the indicator function $\mathcal{I}\,(Y_1)$ [22]. Finally, from the first order optimality conditions for (6b), the dual update

$$Y_2^{k+1} = Y_2^k + \frac{1}{L}(\mathcal{A}_2(\mathcal{A}^\dagger(Y_1^k, Y_2^k)^{-1}) - G)$$

is equivalent to (4d) with $\rho = 1/L$.

The equivalence between AMA and the proximal gradient algorithm can be used to prove convergence of customized AMA [4]. For this, we establish Lipschitz continuity of the gradient of the logarithmic barrier in the dual objective function over a convex domain. In addition, we show that the dual iterates are bounded within this domain. Therefore, a bound on the step-size ρ can be sought that guarantees convergence at a sub-linear rate that is no worse than $O(1/k)$. While AMA with a constant step-size cannot achieve a linear convergence rate, in practice we observe that a heuristic step-size selection (BB step-size initialization followed by backtracking) can improve its convergence properties [4].

REFERENCES

[1] 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*, Florence, Italy, 2013, pp. 1702–1707.

[2] 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.

[3] 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.

[4] A. Zare, Y. Chen, M. R. Jovanovic, and T. T. Georgiou, "Low-complexity modeling of partially available second-order statistics: theory and an efficient matrix completion algorithm," *IEEE Trans. Automat. Control*, 2016, conditionally accepted; also arXiv:1412.3399.

[5] A. Zare, M. R. Jovanovic, and T. T. Georgiou, "Color of turbulence," *J. Fluid Mech.*, 2016, submitted; also arXiv:1602.05105.

[6] 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.

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

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

[9] 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.

[10] 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*, Orlando, FL, 2001, pp. 4944–4949.

[11] M. R. Jovanović and T. T. Georgiou, "Reproducing second order statistics of turbulent flows using linearized Navier-Stokes equations with forcing," in *Bull. Am. Phys. Soc.*, Long Beach, CA, November 2010.

[12] 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.

[13] 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.

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

[15] 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.

[16] A. Zare, M. R. Jovanović, and T. T. Georgiou, "Alternating direction optimization algorithms for covariance completion problems," in *Proceedings of the 2015 American Control Conference*, Chicago, IL, 2015, pp. 515–520.

[17] S. Boyd and L. Vandenberghe, *Convex optimization*. New York: Cambridge Univ Press, 2004.

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

[19] 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.

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

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

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

[23] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.