an ω-periodic linear map $G(\cdot)$ such that". By this remark, for an arbitrary ω-periodic linear map $G(\cdot) \in \mathbb{G}\left(\mathrm{Im}\ N(\cdot)\mathcal{S}_*^u(\cdot)\right)$ there exists an ω-periodic map $R(\cdot)$ such that

$$\mathrm{Im}\ N(\cdot)\mathcal{S}_*^u(k) = \left\langle \mathcal{L}_k^{RC,A_G} \mid E_k^{A_G} \right\rangle \tag{39}$$

thus $\mathrm{Im}\ N(k-1) \subset_{\mathrm{Im}\ N(\cdot)} \mathcal{S}_*^u(k)$ and this implies that $\mathcal{J}_k^{A_G,N} \subset_{\mathrm{Im}\ N(\cdot)} \mathcal{S}_*^u(k)$. Then

$$\left\langle E_k^{A_G} \mid \mathcal{J}_k^{A_G,N} \right\rangle \subset_{\mathrm{Im}\ N(\cdot)} \mathcal{S}_*^u(k) \tag{40}$$

which verifies (i).

(ii) By condition (21), relation (40) verifies (ii) of FDP.

(iii) Now, by Proposition 7, there exists a $G(\cdot) \in \mathbb{G}\left(\mathrm{Im}\ N(\cdot)\mathcal{S}_*^u(\cdot)\right)$ such that the core spectrum of the induced map $\bar{E}_k^{A_G}$ is the quotient space $\mathcal{X}/_{\mathrm{Im}\mathcal{N}(\cdot)}\mathcal{S}_*^u(k)$ is freely assignable, and this verifies (iii) of FDP. ■

REFERENCES

[1] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Norwell, MA: Kluwer, 1999.

[2] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswieicki, *Diagnosis and Fault-Tolerant Control*. Berlin, Germany: Springer-Verlag, 2003.

[3] P. M. Frank, S. X. Ding, and T. Marcu, "Model-based fault diagnosis in technical processes," *Trans. Inst. Measur. and Control*, vol. 22, no. 1, pp. 57–101, 2000.

[4] E. Y. Chow and A. S. Willsky, "Issues in the development of a general algorithm for reliable failure detection," in *Proc. 19th IEEE Conf. Decision and Control*, Albuquerque, NM, 1980.

[5] M. A. Massoumnia, "A geometric approach to the synthesis of failure detection filters," *IEEE Trans. Automat. Control*, vol. AC-31, no. 9, pp. 839–846, Sep. 1986.

[6] M. Massoumnia, G. C. Verghese, and A. S. Willsky, "Failure detection and identification," *IEEE Trans. Automat. Control*, vol. 34, no. 3, pp. 316–321, Mar. 1989.

[7] O. M. Grasselli and S. Longhi, "Disturbance localization by measurement feedback for linear periodic discrete-time systems," *Automatica (J. IFAC)*, vol. 24, no. 3, pp. 375–385, 1988.

[8] S. Bittanti and P. Colaneri, "Invariant representations of discrete-time periodic systems," *Automatica*, vol. 36, no. 12, pp. 1777–1793, 2000.

[9] A. Varga, "Robust and minimum norm pole assignment with periodic state feedback," *IEEE Trans. Automat. Control*, vol. 45, no. 5, pp. 1017–1022, May 2000.

[10] M. S. Fadali and S. Gummuluri, "Robust observer-based fault detection for periodic systems," in *Proc. Amer. Control Conf*, 2001, pp. 464–469.

[11] M. S. Fadali, P. Colaneri, and M. Nel, "h_2 robust fault estimation for periodic systems," in *Proc. Amer. Control Conf.*, 2003, pp. 2973–2978.

[12] P. Zhang, S. X. Ding, G. Z. Wang, and D. H. Zhou, "Fault detection of linear discrete-time periodic systems," *IEEE Trans. Automat. Control*, vol. 50, no. 2, pp. 239–244, Feb. 2005.

[13] O. M. Grasselli and S. Longhi, "Linear function dead-beat observers with disturbance localization for linear periodic discrete-time systems," *Int. J. Control*, vol. 45, no. 5, pp. 1603–1626, 1987.

[14] O. M. Grasselli and S. Longhi, "Zeros and poles of linear periodic multivariable discrete-time systems," *Circuits, Syst., Signal Process.*, vol. 7, no. 3, pp. 361–380, 1988.

[15] O. M. Grasselli and S. Longhi, "Disturbance localization with dead-beat control for linear periodic discrete-time systems," *Int. J. Control*, vol. 44, no. 5, pp. 1319–1347, 1986.

[16] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*. New York: Springer-Verlag, 1979, vol. 10.

[17] G. Basile and G. Marro, *Controlled and Conditioned Invariants in Linear System Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[18] G. Basile and G. Marro, "Controlled and conditioned invariant subspaces in linear system theory," *J. Optimiz. Theory and Applic.*, vol. 3, no. 5, pp. 306–315, 1969.

[19] O. M. Grasselli and S. Longhi, "Pole placement for nonreachable periodic discrete-time systems," *MCSS. Math. Control, Signals and Syst.*, vol. 4, no. 4, pp. 439–455, 1991.

[20] O. M. Grasselli and S. Longhi, "The geometric approach for linear periodic discrete-time systems," *Linear Algebra and Its Applic.*, vol. 158, pp. 27–60, 1991.

Least-Squares Approximation of Structured Covariances

Fu Lin and Mihailo R. Jovanović

***Abstract*—State covariances of linear systems satisfy certain constraints imposed by the underlying dynamics. These constraints dictate a particular structure of state covariances. However, sample covariances almost always fail to have the required structure. The renewed interest in using state covariances for estimating the power spectra of inputs gives rise to the approximation problem. In this note, the structured covariance least-squares problem is formulated and the Lyapunov-type matricial linear constraint is converted into an equivalent set of trace constraints. Efficient unconstrained maximization methods capable of solving the corresponding dual problem are developed.**

***Index Terms*—Convex optimization, least-squares approximation, structured covariances.**

I. INTRODUCTION

The use of second-order statistics has been extensively studied in spectral estimation [1]–[3]. Recently, there has been renewed interest [4], [5] in utilizing state covariances of linear filters to extract information about the power spectra of the input processes. To qualify as a valid state covariance, a positive semi-definite matrix has to satisfy a certain linear constraint imposed by the underlying dynamics. However, the sample covariances, computed from a finite measurement record, almost always fail to have the required structure. Most methods in spectral estimation [2] take sample covariances even though the effect of inaccuracy is not well understood nor analyzed in any detail [6].

In view of the above, it is pertinent to find a nonnegative definite matrix with required structure to approximate the given sample covariance. The natural Euclidean distance gives a least-squares problem which can be solved by standard semi-definite programming (SDP) solvers. For the $n \times n$ covariance matrix, however, the number of optimization variables is of $O(n^2)$, which implies numerical difficulty (computational effort of $O(n^6)$) of the interior-point methods employed in available SDP solvers. In this technical note, we develop an alternative approach to this optimization problem.

Our presentation is organized as follows: we set up the problem and give an equivalent formulation in Section II. We derive the dual problem and present the unconstrained optimization methods in Section III. Then, a numerical example is provided with the computational results presented in Section IV. The technical note is concluded with a brief summary in Section V.

II. PROBLEM FORMULATION

Let a finite dimensional linear system be given by its state equation

$$\dot{x} = Ax + Bd$$

where $d \in \mathbb{C}^m$ is a stationary, zero-mean stochastic process and $x \in \mathbb{C}^n$ is the state vector. The system is characterized by the controllable pair (A, B), where $A \in \mathbb{C}^{n\times n}$ is Hurwitz, and $B \in \mathbb{C}^{n\times m}$ is

Manuscript received May 23, 2008; revised November 07, 2008. Current version published July 09, 2009. This work was supported by the National Science Foundation under CAREER Award CMMI-06-44793. Recommended by Associate Editor Z. Wang.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: mihailo@umn.edu; fu@umn.edu).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2009.2017976

full column rank. Under these assumptions, the steady state covariance $X := \lim_{t \to \infty} \mathcal{E}\{x(t)x^*(t)\}$ satisfies the following linear constraint (cf. [7])

$$AX + XA^* = -(BH + H^*B^*) \tag{LC}$$

where $\mathcal{E}(\cdot)$ is the expectation operator and $(\cdot)^*$ is the complex conjugate transpose. The matrix $H \in \mathbb{C}^{m \times n}$ depends on the input power spectrum and the pair (A, B). It was also established in [7] that the condition for a positive semi-definite matrix X to be the state covariance of a linear system (A, B) for some stationary, zero-mean, stochastic input d, is equivalent to the solvability of (LC) in terms of H. However, the sample covariance

$$\Sigma := \frac{1}{k} \sum_{i=1}^{k} x_i x_i^*$$

computed from k samples almost always fails to satisfy (LC) [6]. In view of this, we formulate the following approximation problem.

- Given a positive semi-definite matrix $\Sigma = \Sigma^* \succeq 0$ and a controllable pair (A, B) with A Hurwitz and B full column rank, find $X = X^* \succeq 0$ that is closest to Σ in the least-squares sense and satisfies (LC) for some $H \in \mathbb{C}^{m \times n}$.

This optimization problem can be formulated as follows:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|X - \Sigma\|_F^2 \\ \text{subject to} \quad & X = X^* \succeq 0 \\ & AX + XA^* = -(BH + H^*B^*) \end{aligned} \tag{P1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and X and H are the optimization variables.

A. Standard SDP Formulation

The primal problem (P1) is a convex optimization problem with a norm objective function and a linear constraint in the positive semi-definite cone $\mathcal{S}_n^+$. By introducing an auxiliary variable κ, (P1) can be cast into an SDP problem [8]

$$\begin{aligned} \text{minimize} \quad & \kappa \\ \text{subject to} \quad & \frac{1}{2}\|X - \Sigma\|_F^2 \le \kappa \\ & X = X^* \succeq 0 \\ & AX + XA^* = -(BH + H^*B^*) \end{aligned}$$

which can be solved by standard primal-dual interior-point methods. However, the number of optimization variables is of $O(n^2)$, which implies the computational complexity $O(n^6)$ of these standard methods.

B. Equivalent Constraints

For a given positive semi-definite matrix X, the solvability of (LC) in terms of H qualifies X to be a valid steady state covariance. However, having H as an optimization variable increases the problem size by $m \times n$, and computations become more expensive as the number of inputs m increases. We note that the Lyapunov-type constraint (LC) implies that X must lie in the range of a certain linear operator $\mathcal{L}$, i.e., $X \in \mathcal{R}(\mathcal{L})$. Namely, the constraint (LC) can be equivalently represented as

$$X = \int_0^\infty \mathrm{e}^{At}(BH + H^*B^*)\mathrm{e}^{A^*t}\,\mathrm{d}t =: \mathcal{L}(H)$$

where $\mathcal{L}$ maps H into X. Equivalently, X must be orthogonal to the null space of the adjoint of $\mathcal{L}$, i.e., $X \perp \mathcal{N}(\mathcal{L}^{ad})$. Next, we determine the basis of $\mathcal{N}(\mathcal{L}^{ad})$.

1) Real Field Case: Let us first consider the linear constraint over the field of real numbers. The complex conjugate transpose in (LC) is then replaced by transpose, $AX + XA^T = -(BH + H^TB^T)$. The linear operator $\mathcal{L}(H) := \int_0^\infty \mathrm{e}^{At}(BH + H^TB^T)\mathrm{e}^{A^Tt}\,\mathrm{d}t$ maps $\mathbb{R}^{m \times n}$ to $\mathbb{R}^{n \times n}$. Let symmetric matrix $G \in \mathbb{R}^{n \times n}$ be in the range of $\mathcal{L}$. The unique linear operator $\mathcal{L}^{ad}$ exists and satisfies $\langle G, \mathcal{L}(H)\rangle = \langle \mathcal{L}^{ad}(G), H\rangle$ where the inner product is defined as $\langle M, N\rangle := \mathrm{trace}(M^TN)$. Hence

$$\begin{aligned} &\langle G, \mathcal{L}(H)\rangle \\ &\quad = \mathrm{trace}\left(G^T \int_0^\infty \mathrm{e}^{At}\left(BH + H^TB^T\right)\mathrm{e}^{A^Tt}\,\mathrm{d}t\right) \\ &\quad = \int_0^\infty \mathrm{trace}\left(\mathrm{e}^{A^Tt}G\mathrm{e}^{At}\left(BH + H^TB^T\right)\,\mathrm{d}t\right) \\ &\quad = 2\ \mathrm{trace}\left(\int_0^\infty \mathrm{e}^{A^Tt}G\mathrm{e}^{At}\,\mathrm{d}t\,BH\right). \end{aligned}$$

Thus

$$\mathcal{L}^{ad}(G) = 2B^T\left(\int_0^\infty \mathrm{e}^{A^Tt}G\mathrm{e}^{At}\mathrm{d}t\right) =: 2B^TZ$$

where Z represents the solution to the Lyapunov equation $A^TZ + ZA = -G$. To construct the basis of $\mathcal{N}(\mathcal{L}^{ad})$, we introduce the change of coordinates $\{\tilde{B} = PB,\ \tilde{A} = PAP^{-1}\}$, such that $\tilde{B} = [I_{m \times m}\ O_{(n-m)\times m}]^T$. Under this coordinate transformation, the basis $\tilde{G}$ satisfies

$$\begin{aligned} \mathcal{L}^{ad}(\tilde{G}) = \tilde{B}^T\tilde{Z} &= [\,I \quad O\,]\begin{bmatrix} \tilde{Z}_1 & \tilde{Z}_2^T \\ \tilde{Z}_2 & \tilde{Z}_3 \end{bmatrix} \\ &= [\,\tilde{Z}_1 \quad \tilde{Z}_2^T\,] = [\,O \quad O\,]. \end{aligned}$$

Therefore, any symmetric matrix $\tilde{Z}$ of the form

$$\tilde{Z} = \begin{bmatrix} O & O \\ O & \tilde{Z}_3 \end{bmatrix}$$

gives a member of the $\mathcal{N}(\mathcal{L}^{ad})$ by substituting $\tilde{Z}$ into $\tilde{A}^T\tilde{Z} + \tilde{Z}\tilde{A} = -\tilde{G}$. Matrices $\tilde{G}_i$'s are determined by substituting the basis elements of $\mathcal{S}_{(n-m)\times(n-m)}$ for $\tilde{Z}$. Thus, G_i's in the original coordinates $\{A, B\}$ are recovered by $G_i = P^T\tilde{G}_iP$. Finally, the Gram-Schmidt procedure is employed to orthonormalize G_i's. The number of basis elements, r, is easily determined by the size of $\tilde{Z}_3$, $r = 0.5(n-m)(n-m+1)$.

2) Complex Field Case: When the matrices are defined over the field of complex numbers, the previous inner product procedure fails to give a linear operator $\mathcal{L}^{ad}$, because H^* is not linear with respect to H. To circumvent this difficulty, we note that the bijection between a complex matrix $X = X_r + jX_i$ and a real matrix $\bar{X}$ of the form

$$\bar{X} = \begin{bmatrix} X_r & -X_i \\ X_i & X_r \end{bmatrix}$$

is a ring isomorphism [9]. By mapping $\{A, B, H\}$ into $\{\bar{A}, \bar{B}, \bar{H}\}$, the constraint (LC) transforms to $\bar{A}\bar{X} + \bar{X}\bar{A}^T = -(\bar{B}\bar{H} + \bar{H}^T\bar{B}^T)$. This can be also verified by expanding (LC) and equating the real and imaginary parts on both sides of the resulting equation. Thus, the procedure

of Section II-B.1 can be employed to construct the basis $\bar{G}_i$. Correspondingly, the number of basis elements, r, for the complex case is $r = (n-m)(2n-2m+1)$.

Thus, (LC) is now transformed into the following equivalent set of constraints:

$$\text{trace } (G_i X) = 0,\ i = 1, 2, \ldots, r. \tag{TC}$$

As already mentioned, the number of G_i's to span $\mathcal{N}(\mathcal{L}^{ad})$ is $r = 0.5(n-m)(n-m+1)$ (in the real case) and $r = (n-m)(2n-2m+1)$ (in the complex case). If $m \geq n$, then (LC) is always satisfied for some H; thus, we assume $m < n$ in the sequel. The transformation of (LC) to (TC) is advantageous for optimization because:

- it eliminates H, which contains $m \times n$ optimization variables;
- the number of corresponding trace constraints decreases as the number of inputs increases.

Examples where the number of inputs is close to the number of states, i.e., $m \approx n$, can be found in spatially distributed systems. A particular example, which is the main motivation for current developments, is encountered in wall-bounded shear flows of incompressible fluids (e.g., boundary layers subject to spatio-temporal excitation in the form of a free-stream turbulence). The algorithms developed in this note are expected to be useful in the study of these problems.

The primal problem is now cast into

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|X - \Sigma\|_F^2 \\ \text{subject to} \quad & X = X^* \succeq 0 \\ & \text{trace } (G_i X) = 0,\ i = 1, 2, \ldots, r \end{aligned} \tag{P2}$$

where G_i's form the orthonormal basis of $\mathcal{N}(\mathcal{L}^{ad})$. After solving (P2), the least-squares solution H can be found by a very cheap computation as follows. Let $D := AX^\star + X^\star A^*$, where $X^\star$ is the unique solution of (P2). Left-multiplication of (LC) with B^* and right-multiplication with B gives

$$B^*B(HB) + (HB)^*B^*B = -B^*DB. \tag{1}$$

With $X \in \mathcal{R}(\mathcal{L})$, there exists H such that HB is Hermitian (see Section III, Remark 1 in [7]). Therefore, H can be computed by

$$H = -(B^*B)^{-1}(B^*D + MB^*)$$

where matrix $M := HB$ is obtained as the solution to the Lyapunov equation (1).

In the sequel, we study optimization problem (P2). The covariance matrix approximation problems have been recently studied by several research groups. Higham first introduced the *nearest correlation matrix* problem and proposed an alternating projection method [10]. Malick studied the *semi-definite least-squares* (SDLS) problem [11], which generalized $\mathcal{S}_n^+$ to any closed convex cone. He proposed a quasi-Newton algorithm and gave a dual interpretation for the alternating projection method as the standard gradient algorithm (see Section 5.2 in [11]). Boyd and Xiao studied the *least-squares covariance adjustment problem* (LSCAP) [12], which is an extension of the SDLS with trace inequality constraints. They proposed a projected gradient method and exploited structure (such as sparsity) to reduce computational expense. The objective functions of the dual problems in both SDLS and LSCAP are not twice continuously differentiable. This implies that the convergence rates of the proposed methods in [11] and [12] are at best linear [13]. Utilizing recent results for strongly semi-smooth functions [14], Qi and Sun developed a generalized Newton method with quadratic convergence rate, which is highly efficient and outperforms quasi-Newton and projected gradient methods (as reported in [13]).

Although fruitful results have been developed for general problems (as evident from the above references), for our specific problem (P2) we derive and cast its dual problem into an unconstrained problem via standard optimization theory. We then implement unconstrained maximization methods proposed in [11]–[13].

III. Dual Problem

In this section, the primal problem (P2) is cast into its dual problem via standard Lagrange multipliers method. It is then converted into an unconstrained maximization problem by projection on $\mathcal{S}_n^+$. To begin with, the Lagrangian [15] is formed by introducing the Lagrange multipliers $\nu_i \in \mathbb{C}$ and $Z \in \mathbb{C}^{n\times n}$

$$L(\nu, Z, X) := \frac{1}{2}\|\Sigma - X\|_F^2 - \text{ trace } (ZX) + \sum_{i=1}^{r} \nu_i \text{ trace } (G_i X)$$

with $Z = Z^* \succeq 0$ corresponding to the inequality constraint. The minimizer of $L(\nu, Z, X)$ over X satisfies $\partial L(\nu, Z, X)/\partial X = 0$, which gives $X_{\min} = \Sigma + Z - \sum_{i=1}^{r} \nu_i G_i$. By choosing $X = X_{\min}$ and denoting $G_\nu := \sum_{i=1}^{r} \nu_i G_i$, we have the dual objective function

$$\begin{aligned} g(\nu, Z) &= -0.5\|Z - G_\nu\|_F^2 - \text{trace } ((Z - G_\nu)\Sigma) \\ &= -0.5\|\Sigma + Z - G_\nu\|_F^2 + 0.5\|\Sigma\|_F^2. \end{aligned}$$

Thus, the dual problem is given by

$$\begin{aligned} \text{maximize} \quad & g(\nu, Z) = -0.5\|\Sigma + Z - G_\nu\|_F^2 \\ & \qquad + 0.5\|\Sigma\|_F^2 \\ \text{subject to} \quad & Z = Z^* \succeq 0. \end{aligned}$$

Note that any Hermitian matrix can be decomposed as $X = X_+ + X_-$ with X_+ and X_-, respectively, being the positive and negative semi-definite parts of X, i.e., $X_+ = \sum_{\bar{\lambda}_i > 0} \bar{\lambda}_i \bar{u}_i \bar{u}_i^*$ and $X_- = \sum_{\bar{\lambda}_i < 0} \bar{\lambda}_i \bar{u}_i \bar{u}_i^*$. Here $\bar{u}_1, \ldots, \bar{u}_n$ denote a set of orthonormal eigenvectors of X with the corresponding eigenvalues $\{\bar{\lambda}_1, \ldots, \bar{\lambda}_n\}$. Hence, the eigenvalue decomposition of X gives

$$\|X\|_F^2 = \text{ trace } (\bar{U}\bar{\Lambda}\bar{U}^*\bar{U}\bar{\Lambda}\bar{U}^*) = \sum_{i=1}^{n} \bar{\lambda}_i^2.$$

Thus, for given vector ν, the choice $Z^\star = -(\Sigma - G_\nu)_-$ eliminates the negative eigenvalues of $\Sigma - G_\nu$. It follows that $Z^\star$ is the maximizer over $\mathcal{S}_n^+$. Hence, the dual problem is finally cast into the following *unconstrained maximization* problem

$$\text{maximize } g(\nu) = -\frac{1}{2}\|(\Sigma - G_\nu)_+\|_F^2 + \frac{1}{2}\|\Sigma\|_F^2 \tag{D}$$

where $(\Sigma - G_\nu)_+$ is the corresponding positive semi-definite part of $\Sigma - G_\nu$. The operation $(\cdot)_+$ on a Hermitian matrix can be interpreted as the projection on the positive semi-definite cone $\mathcal{S}_n^+$. Specifically, one can compute the eigenvalue decomposition and replace the negative eigenvalues by zero to obtain the nonnegative definite part. The dual variables are scalars ν_i's.

For the convex primal problem with *linear constraints*, feasibility is sufficient to guarantee strong duality [15, p. 504]. The primal problem constraints are easily seen feasible (for instance, $X = 0$). Therefore, there is no duality gap and the optimal can be obtained by solving the unconstrained dual problem (D). With the solution from dual problem given by ν_i^*'s, the optimal solution of the primal problem is determined by $X^\star = (\Sigma - G_{\nu^\star})_+$. One advantage of the dual formulation is that

the number of the variables $r = 0.5(n-m)(n-m+1)$ is only a fraction of the size of the primal problem (P2), $0.5(n^2+n)$, when $m \approx n$. Furthermore, due to the absence of constraints any method for unconstrained maximization problem can be utilized.

A. Unconstrained Maximization Methods

We derive the gradient of the objective function using standard perturbation theory and borrow available results from [13] to obtain generalized second-order derivative. Let λ_j be the eigenvalue of $\Sigma - G_\nu$ with the associated orthonormal eigenvector u_j. Define $\phi(\lambda) := \max(0, \lambda)$. Then, the objective function is given by

$$g(\nu) = -\frac{1}{2}\sum_{j=1}^{n}\phi^2(\lambda_j) + \frac{1}{2}\|\Sigma\|_F^2.$$

We now employ results from standard perturbation analysis [16] to determine the gradient of $g(\nu)$. For simplicity, consider a matrix F perturbed by νG, where G and F are Hermitian matrices and ν is a scalar. The derivative of an isolated eigenvalue λ of the resulting matrix $F - \nu G$ with respect to ν is given by $\partial\lambda/\partial\nu = -u^*Gu$, where u is the unit eigenvector associated to λ. Therefore, the ith entry of ∇g is determined by $\partial g(\nu)/\partial\nu_i = \sum_{\lambda_j>0}\lambda_j u_j^* G_i u_j$, and it can be rewritten compactly as

$$\begin{aligned}\frac{\partial g(\nu)}{\partial \nu_i} &= \sum_{\lambda_j>0}\lambda_j u_j^* G_i u_j \\ &= \sum_{\lambda_j>0} u_j^*(\Sigma - G_\nu)_+ G_i u_j \\ &= \operatorname{trace}\left(\sum_{\lambda_j>0} u_j u_j^*(\Sigma - G_\nu)_+ G_i\right) \\ &= \operatorname{trace}\,(UU^*(\Sigma - G_\nu)_+ G_i) \\ &= \operatorname{trace}\,((\Sigma - G_\nu)_+ G_i).\end{aligned}$$

The quasi-Newton method utilizes the gradients of two consecutive steps to construct the approximation of the second-order derivative (Hessian). The Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme is employed due to its efficiency for general problems. Starting with negative identity, the negative definite matrix H_s in our maximization problem is updated using the following BFGS scheme (see [15, p. 150])

$$H_s^+ = H_s + \frac{yy^T}{y^T s} - \frac{H_s s s^T H_s}{s^T H_s s}$$

where $y := \nabla g(\nu^+) - \nabla g(\nu)$ and $s := \nu^+ - \nu$ with ν^+ and ν denoting the current and previous step variables, respectively. The BFGS direction is determined by $v_{BFGS} = -H_s^{-1}\nabla g(\nu)$. Given ν_0, the first step ν_1 can be obtained by a gradient method which is necessary for the BFGS iterations to proceed.

The objective function is not twice continuously differentiable when $\Sigma - G_\nu$ has zero eigenvalues [13]. This implies that the classical Hessian needs to be generalized. The detailed discussion about the smoothness property of the objective function and the derivation of generalized Newton direction can be found in [13]. To compute generalized $\nabla^2 g(\nu)$, one constructs the symmetric matrix Ω as follows,

$$\Omega_{ij} = \begin{cases} \frac{(\phi(\lambda_i)-\phi(\lambda_j))}{(\lambda_i-\lambda_j)} & \text{if } \lambda_i \neq \lambda_j \\ 1 & \text{if } \lambda_i = \lambda_j > 0 \\ 0 & \text{if } \lambda_i = \lambda_j \leq 0. \end{cases}$$

Then for any $\Delta\nu \in \mathbb{R}^r$, the generalized Hessian acting on $\Delta\nu$ is given by

$$\nabla^2 g(\nu)(\Delta\nu) = \begin{pmatrix} \operatorname{trace}\,(-U(\Omega \circ (U^* G_{\Delta\nu} U))U^* G_1) \\ \vdots \\ \operatorname{trace}\,(-U(\Omega \circ (U^* G_{\Delta\nu} U))U^* G_r) \end{pmatrix}$$

where $\circ$ denotes the Hadamard product, i.e., entrywise multiplication. By solving $\nabla^2 g(\nu)(v_{NT}) = -\nabla g(\nu)$ using conjugate gradient (CG) method [13], [15], we obtain the generalized Newton direction v_{NT}.

B. Implementation

With the three ascending directions $\nabla g, v_{BFGS}$ and v_{NT} determined above, the algorithm for the unconstrained problem is given next. We choose the standard backtracking line search as the step size method, with parameters $\alpha = 0.3$, $\beta = 0.5$ (see [17, p. 464]).

Algorithm:

Start with initial point ν_0 (e.g., $\nu_0 = 0$) and at each step k

repeat:

1) project $\Sigma - G_{\nu_k}$ onto $\mathcal{S}_n^+$, then compute ascending direction v_k;
2) use the backtracking line search to determine step size t, then update $\nu_{k+1} = \nu_k + tv_k$.

until: stopping criterion $\|\nabla g(\nu)\|_2 < \epsilon$ is reached.

The convergence to the global optimal is guaranteed by the convexity of the problem.

C. Complexity Analysis

The computational effort of each algorithm is studied in this section. At each step of the gradient method, the computational effort is $O(\max\,(n^3, rn^2))$ operations, where $O(n^3)$ operations are required for the eigenvalue decomposition of $\Sigma - G_\nu$ and $O(n^2)$ operations are required for computation of the matrix inner product $\operatorname{trace}((\Sigma - G_\nu)_+ G_i)$ for each $i = 1, 2, \ldots, r$. As BFGS method uses gradient directions to form H_s, it requires the same amount of operations as the gradient method. The extra effort comes from computing the inverse of H_s, which requires $O(r^3)$ operations. Thus, each BFGS step costs $O(\max\,(n^3, rn^2, r^3))$ operations. To compute the generalized Newton direction, it takes $O(\max\,(n^3, rn^2))$ operations to determine $U(\Omega \circ (U^* G_{\Delta\nu} U))U^*$ in each CG step, where $O(rn^2)$ operations are required to form the sum $G_{\Delta\nu}$ and $O(n^3)$ operations are required for matrix multiplications. Thus, the cost of each CG step is $O(\max\,(n^3, rn^2))$ operations. The number of CG steps is $O(r)$ provided that the problem is well pre-conditioned [13], [15]. Therefore, each Newton step costs $O(\max\,(rn^3, r^2n^2))$ operations assuming $O(r)$ CG steps.

For a given problem with n states, the cost of the above algorithms relies heavily on the number of dual variables. From the construction of the basis in Section II-B, the number of inputs m dictates the size of $r = 0.5(n-m)(n-m+1)$. Thus, if the number of inputs is close to the number of states, i.e., $m \approx n$, presented algorithms are expected to be computationally efficient.

IV. Numerical Experiments

We present a mass-spring-damper example and compare the computation results for different formulations of the problem. As discussed in Section II, the original primal problem (P1) is equivalent to primal problem (P2), which is then cast into the dual problem (D) in Section III. The first two formulations can be solved by available SDP solvers such as SeDuMi. Following [12], [13], we develop unconstrained maximization methods for the dual formulation. The

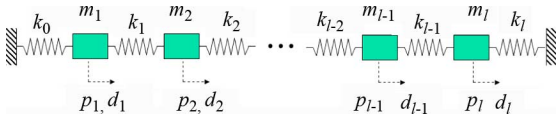

Fig. 1. Mass-spring-damper system.

following experiments are performed in Matlab on a personal computer with 3.2 GHz CPU and 2.5 GB RAM. Primal problems (P1) and (P2) are solved using Yalmip [18] with SeDuMi as its SDP solver.

A. Mass-Spring-Damper Example

We consider a mass-spring-damper system consisting of l masses and $l+1$ springs and dampers on a line as in Fig. 1. (The dampers are not shown in the figure.) The dynamics of the ith mass m_i are given by

$$m_i\ddot{p}_i + b_{i-1}(\dot{p}_i - \dot{p}_{i-1}) + b_i(\dot{p}_i - \dot{p}_{i+1}) + k_{i-1}(p_i - p_{i-1}) + k_i(p_i - p_{i+1}) = d_i$$

where p_i represents the displacement from a reference position of the ith mass, b_i is the damping coefficient of the ith damper and k_i is the spring constant of the ith spring. We assign unit values to $\{m_i, b_i, k_i\}$, and assume that a stationary Gaussian white stochastic process, d_i, with zero-mean and unit variance is introduced to the ith mass. The first and the last masses are connected to fixed boundaries; hence, $\dot{p}_0 \equiv 0, p_0 \equiv 0, \dot{p}_{l+1} \equiv 0, p_{l+1} \equiv 0$. By selecting the state variables $x_1 := \mathrm{col}\{p_i\}$ and $x_2 := \mathrm{col}\{\dot{p}_i\}$, the state-space representation is determined by

$$A = \begin{bmatrix} O & I \\ T & T \end{bmatrix},\ B = \begin{bmatrix} O \\ I \end{bmatrix}$$

where $T := \mathrm{toeplitz}([-2\ 1\ 0\ \cdots\ 0])$, I is $l \times l$ identity matrix, O is $l \times l$ zero matrix, and input $d := \mathrm{col}\ \{d_i\}$. To demonstrate the performance with respect to different number of inputs, we assume the first $l-m$ components in d to be identically equal to zero; the definition of input matrix B should be changed correspondingly in this case.

Knowledge of the dynamics of the linear system and covariance Q of d can be used to obtain the steady state covariance X by solving the Lyapunov equation

$$AX + XA^T = -BQB^T.$$

However, we consider a situation in which only limited observed sample data is available to estimate state covariance and infer the second-order statistics of the inputs. In the numerical experiments, we take 1000 state samples (uniformly sampled in time from 0 to 10 seconds) and compute the sample covariance Σ. Invariably, Σ fails to satisfy (LC) and we utilize approximation algorithms to find least-squares estimate of Σ.

B. Performance Comparison of Three Formulations

As discussed in Section III-C, the computational effort of all unconstrained maximization algorithms depends heavily on the number of dual variables determined by $r = 0.5(n-m)(n-m+1)$ (in this example, $n = 2l$). Hence, two sets of optimization experiments are carried out with difference in the number of inputs m relative to the number of states n ($m = 0.1n$ in the first set and $m = 0.5n$ in the second set with the results shown in Tables I and II, respectively). The time for all computations is given in seconds. We only present results of BFGS method for the dual formulation because it generally outperforms gradient and generalized Newton methods. It is observed in this example that the conjugate gradient method usually runs into difficulty when $r \geq 100$. However, in numerical examples reported in [13], the generalized Newton method outperformed BFGS method significantly. The purpose of our experiments is not to provide comparison between different unconstrained maximization methods but rather to compare the different formulations of the problem in terms of their computational efficiency. For simplicity, the initial condition for BFGS method is set to be a zero vector and the stopping criterion is $\|\nabla g(\nu)\|_2 \leq 10^{-5}$.

TABLE I
PERFORMANCE COMPARISON FOR $m = 0.1n$

		Time(s)	It. No.	$\|X^\star - \Sigma\|_F^2$	$\frac{\|X^\star-\Sigma\|_F^2}{\|\Sigma\|_F^2}$
$r = 378$	BFGS	25.1	238	3.6280e-4	13.5%
$n = 30$	P1	7.7	21	3.6288e-4	13.5%
	P2	13.0	21	3.6288e-4	13.5%
$r = 666$	BFGS	108.3	298	4.4958e-4	13.5%
$n = 40$	P1	34.6	23	4.4965e-4	13.5%
	P2	64.1	22	4.4965e-4	13.5%
$r = 1035$	BFGS	1182.8	1104	3.2386e-3	12.7%
$n = 50$	P1	127.1	28	3.2388e-3	12.7%
	P2	281.6	28	3.2388e-3	12.7%

TABLE II
PERFORMANCE COMPARISON FOR $m = 0.5n$

		Time(s)	It. No.	$\|X^\star - \Sigma\|_F^2$	$\frac{\|X^\star-\Sigma\|_F^2}{\|\Sigma\|_F^2}$
$r = 120$	BFGS	0.1	5	6.3172e-3	5.3%
$n = 30$	P1	24.7	24	6.3172e-3	5.3%
	P2	4.5	24	6.3172e-3	5.3%
$r = 210$	BFGS	0.2	5	7.2892e-3	5.9%
$n = 40$	P1	117.3	24	7.2893e-3	5.9%
	P2	17.9	24	7.2893e-3	5.9%
$r = 325$	BFGS	0.5	5	6.7127e-2	7.4%
$n = 50$	P1	413.5	23	6.7127e-2	7.4%
	P2	62.4	24	6.7127e-2	7.4%

TABLE III
BFGS AND BASIS TIME FOR $m = 0.1n$

	$n = 30$ $r = 378$	$n = 40$ $r = 666$	$n = 50$ $r = 1035$
Basis time(s)	9.3	34.6	219.2
BFGS time(s)	25.1	108.3	1182.8

As evident from Tables I and II, three different formulations give very close optimal solutions. For $m = 0.1n$, the primal formulation (P1) can be solved more efficiently than formulation (P2) by standard SDP solver. The BFGS (as well as gradient and generalized Newton methods) has difficulty in solving the dual formulation. For $m = 0.5n$, however, the unconstrained formulation can be solved very efficiently and the BFGS method significantly outperforms standard SDP solvers. Also, note that even for the same SDP solver, formulation (P2) is much easier to handle than formulation (P1). Another aspect of the dual formulation is the construction of the basis. From results listed in Table III, the time required to construct the basis is actually comparable to the optimization time. However, the basis can be computed off-line and stored for future computations.

V. Concluding Remarks

The state sample covariances almost always fail to satisfy linear constraint imposed by the underlying dynamics. The consistency with such dynamics is crucial in addressing the problem of characterizing the input power spectra. We formulate the structured covariance least-squares problem and convert the matricial linear constraint into an equivalent set of trace constraints. The corresponding dual problem can be solved efficiently by unconstrained maximization methods when the number of inputs is close to the number of states.

With a different notion of distance, one can formulate an alternative covariance approximation problem [6]. The quantum relative entropy is an interesting distance function used to quantify information and uncertainty in quantum systems. A solution to this problem using homotopy-based approach was recently provided in [19]. We intend to explore large-scale covariance approximation problems with this distance measure in our future work. An interesting question may be to compare the optimal covariances from both the relative entropy distance and the Euclidean least-squares distance.

Appendix
Results of Numerical Experiments

In the following tables, the time is given in seconds. In Tables I and II, the optimal value $\|X^\star - \Sigma\|_F^2$ and the ratio $\|X^\star - \Sigma\|_F^2/\|\Sigma\|_F^2$ are reported. The time for (P1) and (P2) is the time required to run the SDP solver SeDuMi. We note that the interface Yalmip requires more time than solver SeDuMi. The stopping criterion for BFGS method is $\|\nabla g(\nu)\|_2 \leq 10^{-5}$. The number of dual variables is given by $r = 0.5(n-m)(n-m+1)$.

Acknowledgment

The authors would like to thank Prof. D. Sun for his generous suggestions about generalized Newton method, and Prof. T. T. Georgiou and Prof. Z.-Q. Luo for stimulating discussions.

References

[1] S. Haykin, *Nonlinear Methods of Spectral Analysis*. New York: Springer-Verlag, 1979.
[2] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[3] T. T. Georgiou, "Spectral estimation via selective harmonic amplification," *IEEE Trans. Automat. Control*, vol. 46, no. 1, pp. 29–42, Jan. 2001.
[4] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3189–3205, Nov. 2000.
[5] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. Automat. Control*, vol. 45, no. 6, pp. 822–839, Jun. 2001.
[6] T. T. Georgiou, "Structured covariances and related approximation questions," in *Directions in Mathematical Systems Theory and Optimization*, A. Rantzer and C. I. Byrnes, Eds. New York: Springer-Verlag, 2003, pp. 135–140.
[7] T. T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Automat. Control*, vol. 47, no. 7, pp. 1056–1066, Jul. 2002.
[8] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.
[9] L. J. Goldstein, *Abstract Algebra; A First Course*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
[10] N. Higham, "Computing the nearest correlation matrix—a problem from finance," *IMA J. Numer. Anal.*, vol. 22, pp. 329–343, 2002.
[11] J. Malick, "A dual approach to semidefinite least-squares problems," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 1, pp. 272–284, 2004.
[12] S. Boyd and L. Xiao, "Least-squares covariance matrix adjustment," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 2, pp. 532–546, 2005.
[13] H. Qi and D. Sun, "A quadratically convergent Newton method for computing the nearest correlation matrix," *SIAM J. Matrix Anal. Appl.*, vol. 26, pp. 360–385, 2006.
[14] L. Qi and J. Sun, "A nonsmooth version of Newton's method," *Math. Programm.*, vol. 58, pp. 353–367, 1993.
[15] D. P. Bertsekas, *Nonlinear Programming*. New York: Athena Scientific, 1999.
[16] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators*. New York: Springer-Verlag, 1982.
[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[18] J. Löberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, R.O.C., 2004 [Online]. Available: http://control.ee.ethz.ch/~joloef/yalmip.php
[19] T. T. Georgiou, "Relative entropy and the multi-variable multi-dimensional moment problem," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

Decentralized Learning in Finite Markov Chains: Revisited

Hyeong Soo Chang, *Senior Member, IEEE*

***Abstract*—The convergence proof in the paper "Decentralized learning in finite Markov chains," published in the *IEEE Transactions on Automatic Control*, vol. AC-31, no. 6, pp. 519–526, 1986, is incomplete. This note first provides a sufficient condition for the existence of a unique optimal policy for infinite-horizon average-cost Markov decision processes (MDPs), making the convergence result established by Wheeler and Narendra preserved with the condition. We then present a novel simulation-based decentralized algorithm, called "sampled joint-strategy fictitious play for MDP" for average MDPs based on the recent study by Garcia *et al.* of a decentralized approach to discrete optimization via fictitious play applied to games with identical payoffs. We establish a stronger almost-sure convergence result than Wheeler and Narendra's, showing that the sequence of probability distributions over the policy space for a given MDP generated by the algorithm converges to a unique optimal policy with probability one.**

***Index Terms*—Controlled Markov chain, decentralized learning, fictitious play, learning automata, Markov decision process.**

I. Introduction

Wheeler and Narendra [27] studied a decentralized learning approach to solving ergodic controlled Markov chains, i.e., Markov decision processes (MDPs) [21] with finite state and finite action spaces. In their approach, each state is associated with a learning automaton (LA). At each time step, only the LA associated with the currently visited state updates its probability distribution over its available actions based on a response from the system and the LA samples an action according to its updated probability distribution, which will take to a next state at the next time step and this overall

Manuscript received October 20, 2008; revised February 24, 2009. First published June 23, 2009; current version published July 09, 2009. This work was supported by the Korea Science and Engineering Foundation (KOSEF), funded by the Korea government under (MOST) Grant R01-2007-000-10511-0 and by Sogang University (200811037). Recommended by Associate Editor I. Paschalidis

The author is with the Department of Computer Science and Engineering, Sogang University, Seoul 121-742, Korea (e-mail: hschang@sogang.ac.kr).

Digital Object Identifier 10.1109/TAC.2009.2017977