Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

On the least-squares approximation of structured covariances

Fu Lin and Mihailo R. Jovanović

Abstract— **State covariances of the linear systems satisfy certain constraints imposed by the underlying dynamics. These constraints dictate a particular structure of state covariances. On the other hand, sample covariances (e.g., obtained in experiments) almost always fail to have the required structure. In view of this, it is of interest to approximate sample covariances by positive semi-definite matrices of the required structure. The structured covariance least-squares approximation problem is formulated and the Lyapunov-type matrical linear constraint is converted into an equivalent set of trace constraints. Efficient quasi Newton and generalized Newton methods capable of solving the corresponding unconstrained dual problems with the large number of variables are developed.**

Index Terms— **Convex optimization; least-squares approximation; sample covariances.**

I. Introduction

The use of second order statistics has been widely studied in spectral estimation. State covariances of the linearized equations satisfy certain constraints imposed by the linearized dynamics [1]–[3], which dictate a particular structure. On the other hand, sample covariances (e.g., obtained in experiments or in numerical simulations of nonlinear equations) almost always fail to have the required structure. In view of this, it is relevant to approximate sample covariances by positive semi-definite matrices of the required structure [3].

Our main motivation for this problem stems from fluid mechanics, where the objective is to develop tractable control-oriented models that will reproduce turbulent flow statistics. Over the past 20 years, a variety of experimental and numerical studies have provided an invaluable insight into both structural and statistical characteristics of wall-bounded *turbulent* flows. Although these characteristics of turbulent flows represent the critical factor in evaluation of drag reduction strategies, their utilization in development of control-oriented models has been elusive. In [4], the problem of modeling disturbances in the linearized Navier-Stokes equations by testing the validity−in *quantitative* sense−of a stochastically excited version of this model was addressed. A model for second order statistics of a turbulent channel flow using an associated linear stochastically forced input-output system was developed. These results showed that certain portions of numerically generated flow statistics can be closely matched by the appropriate choice of input forcing covariance. This was done in an *ad hoc* fashion using a variety of excitation force correlations and showing the dependence of the velocity field statistics on them.

The objective of this paper is to develop a convex optimization procedure for least-squares approximation of large-scale sample covariances. Even though our main motivation for this problem stems from fluid mechanics, the theory we propose to develop is generic enough to be applied to the other areas where questions of estimating the power spectrum of the input using state statistics arise [5], [6]. The covariance approximation problem can be cast into a standard semidefinite programming (SDP) problem, which can be solved by available SDP solvers (such as SeDuMi) with modest number of variables, i.e. around 1,500. In this paper, we develop efficient methods for large scale problems. The quasi Newton method that we develop in section III-B is capable of solving problems with around 200,000 variables.

Our paper is organized as follows: in Section II, we set up the problem and provide an equivalent formulation. The dual of the original problem, which is very suitable for optimization, is presented in Section III. We provide an illustrative example and compare the numerical results in Section IV. The paper is concluded with a brief summary and remarks on future work 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 controllable pair (A, B), where $A \in \mathbb{C}^{n\times n}$ is a Hurwitz matrix, and $B \in \mathbb{C}^{n\times m}$ is a full column rank matrix. Under these assumptions, the steady state covariance of x, $X := \lim_{t\to\infty} \mathcal{E}\{x(t)x^*(t)\}$, satisfies the following linear constraint [1]

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

where $H \in \mathbb{C}^{m\times n}$ represents a matrix which depends on the input power spectrum and the pair (A, B). However, the sample covariance

$$\Sigma := \frac{1}{N}\sum_{k=1}^{N} x_k x_k^*$$

almost always fails to have the required structure [1]. It was recently established [1] that a condition for $X = X^* \succeq 0$ to be the state covariance of a linear system (A, B) for some stationary, zero-mean, stochastic input d, is equivalent to solvability of (LC) in terms of H. In view of the above, we formulate the following approximation problem:

- *given a positive semidefinite 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 $H \in \mathbb{C}^{m\times n}$.*

F. Lin and M. R. Jovanović are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA (e-mails: linxx242@umn.edu, mihailo@umn.edu).

Supported in part by the Office of the Dean of the Graduate School of the University of Minnesota Grant-in-Aid of Research, Artistry and Scholarship Award 1546-522-5985.

Thus, we propose to solve the following *primal optimization problem*:

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

where $\|\cdot\|_F$ denotes the Frobenius norm, and X and H are the optimization variables. The solution to this optimization problem can be used to determine the power spectrum of an input to a linear dynamical system whose state covariance approximates a given matrix $\Sigma = \Sigma^* \succeq 0$ in the least-squares sense.

A. Available methods and motivation

The primal problem (P1) is a convex optimization problem with a norm objective function and a linear constraint. By introducing an auxiliary variable κ, it can be cast into an SDP problem [7],

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

This quadratic SDP problem can be solved by standard *primal dual interior point* methods. However, the number of optimization variables is $O(n^2)$, which implies the inefficiency of this method when dealing with large scale problems. The numerical experiments of Section IV verify that SeDuMi runs into numerical problems and fails to give solution when the number of variables is around 1,500, say matrix X of size 30×30 and matrix H of size 30×20. The goal of this paper is to develop efficient methods for solving large scale problems. The quasi Newton method of section III-A can solve problems with matrices A and H of respective sizes 300×300 and 300×295, i.e. the total number of variables around $180,000$. We note that all numerical studies are conducted in MATLAB.

B. Equivalent constraint

For a given positive semidefinite matrix X, the solvability of (LC) in terms of H qualifies X to be a valid 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 grows. We note that the Lyapunov type constraint (LC) implies that X must lie in the range of a certain operator $\mathcal{L}$. The constraint (LC) can be equivalently represented as

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

where $\mathcal{L}$ is an operator that maps H into X, i.e. $X \in \mathcal{R}(\mathcal{L})$. Therefore, X must be orthogonal to the null space of $\mathcal{L}^{ad}$, $\mathcal{N}(\mathcal{L}^{ad})$. By constructing the basis of this space, $\{G_i = G_i^*, i = 1, \ldots, r\}$, (LC) is transformed into the following set of equivalent constraints

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

It can be shown that we need $r = (2n - 2m + 1)(n - m)$ G_i's to span the null space of $\mathcal{L}^{ad}$. This transformation is advantageous for optimization as it:

- eliminates $n \times m$ optimization variables,
- significantly decreases the number of trace constraints for large m's.

The details of constructing the basis of $\mathcal{L}^{ad}$ are to be reported elsewhere.

The primal problem is now formulated as

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

where G_i's determine the orthonormal basis for the null space of $\mathcal{L}^{ad}$. After solving (P2), we can find the least-squares solution H with a very cheap computation, using the procedure given in Appendix.

We note that the primal problem (P2) is the *semidefinite least-squares* (SDLS) problem [8]. Furthermore, Boyd & Xiao [9] studied the *least-squares covariance adjustment problem*, which is an extension of the SDLS problem with trace inequality constraints. Motivated by their work, in the next section, we explore the corresponding dual problem, which is shown to be an unconstrained maximization problem.

III. DUAL PROBLEM

In the sequel, the primal problem (P2) is cast into a dual problem, which can be formulated as an unconstrained maximization problem. To begin with, the Lagrangian is formed by introducing Lagrange multipliers $\nu_i \in \mathbb{C}$ and $Z \in \mathbb{C}^{n\times n}$ with $Z = Z^* \succeq 0$

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

The minimizer of $L(\nu, Z, X)$ over X satisfies

$$\begin{aligned} \frac{\partial L(\nu, Z, X)}{\partial X} &= 0 \\ X_{min} &= \Sigma + Z - \textstyle\sum_{i=1}^{r} \nu_i G_i. \end{aligned}$$

By choosing $X = X_{min}$ and denoting $G_\nu = \sum_{i=1}^{r} \nu_i G_i$ the dual problem is given by

$$\begin{array}{ll} \text{maximize} & g(\nu, Z) = -\frac{1}{2}\|\Sigma + Z - G_\nu\|_F^2 + \frac{1}{2}\|\Sigma\|_F^2 \\ \text{subject to} & Z \succeq 0. \end{array}$$

Note that any Hermitian matrix can be decomposed as [10]

$$\begin{aligned} X &= X_+ - X_-, \quad X_+X_- = 0, \\ X_+ &= X_+^* \succeq 0, \quad X_- = X_-^* \succeq 0, \end{aligned}$$

with X_+ and X_- respectively being the positive and negative semidefinite parts of X, that is

$$X_+ = \sum_{\lambda_i > 0} \lambda_i x_i x_i^*, \quad X_- = \sum_{\lambda_i < 0} (-\lambda_i) x_i x_i^*.$$

Here $x_1, \ldots, x_n$ denote a set of orthonormal eigenvectors of X with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. It can be shown that $-(1/2)\|\Sigma + Z - G_\nu\|_F^2$ is maximized over all positive semidefinite matrices Z, by the following choice $Z = (\Sigma - G_\nu)_-$. By selecting Z as the negative part of $\Sigma - G_\nu$, the dual problem is finally cast into an *unconstrained dual* problem

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

where $(\cdot)_+$ denotes the projection on the positive semidefinite cone $\mathcal{S}_n^+$. The dual variables are scalars ν_i.

Remark 1: Since *strong duality* is guaranteed by the feasibility of the original convex problem (P1) with linear constraints [11], the optimal of the primal problem can be obtained by solving the dual problem (D).

Remark 2: An advantage of working with dual problem (D) comes from the absence of constraints. Thus, any method for unconstrained maximization can be utilized. Furthermore, the size of the dual problem $r = (2n - 2m + 1)(n - m)$ is only a fraction of the size of the primal problem $(n^2 + n)$ when m is close to n.

A. Unconstrained maximization methods

To implement unconstrained maximization methods, one needs to determine the first or second derivatives of the objective function with respect to dual variables. It is noteworthy that the objective function is not twice continuously differentiable when the matrix $\Sigma - G_\nu$ is singular [9]. However, it was shown by Qi and Sun [12] that a generalized Newton method can be implemented using the fact that the metric projection operator over positive semidefinite cone is *strongly semismooth* [12].

The projection on the positive semidefinite cone brings the difficulty of calculating the derivatives directly from the objective function (D). This is because of the absence of an explicit expression for projection, which prevents us from applying the chain rule to determine derivative. To avoid this problem, we utilize the relationship between the Frobenius norm of a Hermitian matrix and its eigenvalues. Namely, the eigenvalue decomposition of $\Sigma - G_\nu = UDU^*$, yields

$$\begin{aligned}\|\Sigma - G_\nu\|_F^2 &= \text{trace}((\Sigma - G_\nu)^*(\Sigma - G_\nu)) \\ &= \text{trace}(DD) = \textstyle\sum_j d_j^2.\end{aligned}$$

Let $\phi(d) := \max(0, d)$, $d \in \mathbb{R}$, the objective function is rewritten as

$$g(\nu) = -\frac{1}{2}\sum_j \phi^2(d_j) + \frac{1}{2}\|\Sigma\|_F^2.$$

We now employ results from standard perturbation analysis [13] to determine the gradient of $g(\nu)$, i.e.

$$\nabla g(\nu) = \left[\frac{\partial g(\nu)}{\partial \nu_1} \cdots \frac{\partial g(\nu)}{\partial \nu_r}\right]^T.$$

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

$$\frac{\partial \lambda}{\partial \nu} = -u^* G u,$$

where u is the unit eigenvector associated with λ. Therefore, the elements of ∇g are given by

$$\frac{\partial g(\nu)}{\partial \nu_i} = \textstyle\sum_j \lambda_j u_j^* G_i u_j.$$

Furthermore, the entries of ∇g can be compactly rewritten as

$$\begin{aligned}\frac{\partial g(\nu)}{\partial \nu_i} &= \textstyle\sum_j \lambda_j u_j^* G_i u_j \\ &= \textstyle\sum_j u_j^*(\Sigma - G_\nu)_+ G_i u_j \\ &= \textstyle\sum_j \text{trace}(u_j u_j^*(\Sigma - G_\nu)_+ G_i) \\ &= \text{trace}(\textstyle\sum_j u_j u_j^*(\Sigma - G_\nu)_+ G_i) \\ &= \text{trace}((\Sigma - G_\nu)_+ G_i)\end{aligned}$$

using $U^*U = I$ and $\text{trace}(AB) = \text{trace}(BA)$.

To compute generalized second derivative of the objective function, $\nabla^2 g(\nu)$, we follow [12] and construct the symmetric matrix Ω as follows,

$$\Omega_{ij} = \begin{cases} (\phi(d_i) - \phi(d_j))/(d_i - d_j) & \text{if } d_i \neq d_j \\ \phi'(d_i) & \text{if } d_i = d_j \end{cases}$$

Then for any infinitesimal variation $\Delta\nu \in \mathbb{R}^r$, we have

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

where $\circ$ denotes the Schur product, i.e., entrywise multiplication. The solution to

$$\nabla^2 g(\nu)(\Delta\nu) = -\nabla g(\nu)$$

yields the generalized Newton direction $\Delta\nu$. The approximate solution to this equation can be obtained using conjugate gradient (CG) method. Alternatively, the direct Hessian whose i, jth entry is given by

$$H_{ij} = -\text{trace}(U(\Omega \circ U^*(G_j)U)U^*G_i),$$

can be used to solve this equation. The computation expense of each method is discussed in III-C.

B. Implementation

Three standard unconstrained methods, i.e. gradient method, quasi Newton method (e.g. Broyden-Fletcher-Goldfarb-Shanno (BFGS)), and generalized Newton method are implemented next. More details about these algorithms can be found in [10]–[12]. The basic principles of all of these three methods are the same; given an initial point, after projecting $\Sigma - G_\nu$ onto the positive semidefinite cone $\mathcal{S}_n^+$, one calculates the step direction, chooses a step size, updates iterative point and then repeats the procedure until the optimal point with specified accuracy is obtained.

The difference between these methods is the ascending direction v, i.e.

$$\begin{aligned} v_g &= \nabla g(\nu), \\ v_{BFGS} &= H_s^{-1}\nabla g(\nu), \\ v_{NT} &= (\nabla^2 g(\nu))^{-1}\nabla g(\nu), \end{aligned}$$

for gradient, quasi Newton, and Newton methods respectively. The quasi Newton method utilizes the gradients of

two consecutive steps to approximate the Hessian. Starting with identity, the positive definite matrix H_s is updated using the following BFGS scheme [11],

$$H_s^+ \;=\; H_s \;+\; \frac{yy^T}{y^Ts} \;+\; \frac{H_s s s^T H_s}{s^T H_s s},$$

where $y = \nabla g(x^+) - \nabla g(x),\ s = x^+ - x$ and symbols $(\cdot)^+$ denote the current values. We choose the standard backtracking line search [10] with parameters $\alpha = 0.3, \beta = 0.5$ as the method for selecting step size.

Algorithm

Start with initial point $\nu_0 = [0, 0, \cdots, 0]$, and at each step k

repeat:

1) project $\Sigma - G_\nu$ on $\mathcal{S}_n^+$,
2) compute ascending direction v_k,
3) use the backtracking line search to obtain step size γ,
4) update $\nu_{k+1} = \nu_k + \gamma v_k$.

until: stopping criterion is reached

C. Complexity analysis

The cost estimation of each algorithm is studied in this section. At each step of the gradient method, the main cost is the computation of a gradient direction, which is of $O(n^3 r)$ operations. It takes $O(2n^2 r)$ flops to form the matrix G_ν, whereas the eigenvalue decomposition of $\Sigma - G_\nu$ requires $O(n^3)$ operations. The total cost of each gradient step is of $O((r+1)n^3)$. As BFGS method utilizes gradient direction information, the expense of constructing H_s is of the same order as for the gradient method. The extra cost comes from the computation of the inverse of H_s, which is of $O(r^3)$. Thus, the cost of each BFGS step is $O(\max(r^3, (r+1)n^3))$.

To compute the generalized Newton direction, if conjugate gradient method is employed, then in each CG step, it takes $O(\max(n^3, rn^2))$ to compute $U(\Omega \circ U^*(G_{\Delta\nu})U)U^*$. The number of CG steps depends heavily on problem size r. When r is not too large and the problem is well preconditioned, the number of steps is of order $O(r)$. Therefore, each Newton step needs $O(\max(rn^3, r^2n^2))$ flops. On the other hand, the cost of direct computation is more expensive. It takes $O(n^3)$ operations to compute each entry of the Hessian and the inverse needs $O(r^3)$ flops. Therefore, each Newton step takes $O(\max(r^2n^3, r^3))$. However, direct computation method outperforms CG method in most experiments due to the slow convergence of CG method.

For a given problem with number of states n, the cost of the above algorithms relies heavily on the number of dual variables. From the construction of this basis, we observe that the number of inputs m dictates the size of r. It is thus expected that if the number of inputs is close to the number of states, i.e. m close to n, presented algorithms should outperform standard interior point method (which requires $O(n^6)$ computation in each iterative step) in terms of computation efficiency.

IV. Numerical experiment

For illustrative purpose, we consider a nonlinear heat equation on $y \in [-1, 1]$ with Dirichlet boundary conditions, i.e. $\psi(y = \pm 1, t) = 0$

$$\psi_t(y,t) = \psi_{yy}(y,t) - 0.1\psi^3(y,t) + f(y)u(y,t). \qquad (1)$$

It is assumed that the zero-mean white stochastic process $u(y,t)$ enters the equation through

$$f(y) = \frac{1-y^2}{2\sqrt{0.5\pi}} \exp(-\frac{(y+0.9)^2}{2}).$$

After linearizing the above equation around the origin, we obtain

$$\psi_t = \psi_{yy} + f(y)u(t,y). \qquad (2)$$

To approximate this partial differential equation, we discretize the spatial operator at each point of the Gauss-Lobatto grid, using a Chebyshev collocation scheme [14]. The weight function $f(y)$ is also discretized at the same points and the input matrix B is built by putting the discrete values of $f(y)$ on the diagonal. By neglecting entries less than a specific tolerance (10^{-4}) , the matrix B is truncated into a skinny matrix. We employ the 4th order Runge-Kutta method to approximate the solution to (1) as time propagates. With zero initial conditions and stochastic inputs, a large number of samples are collected and the sampled covariance Σ is computed. It turns out that Σ does not satisfy linear constraint (LC). Thus, we want to compute a covariance matrix of (2) that approximates Σ in the least-squares sense.

As already discussed, the original primal problem (P1) is equivalent to primal problem (P2), which is cast into the dual problem (D). The first two formulations can be solved by available SDP solvers such as SeDuMi. We develop unconstrained methods for the dual problem. As there are three different but equivalent formulations, it is relevant to consider their numerical efficiency in terms of computation time and solution accuracy. The following experiments are performed in MATLAB on a personal computer with 3.2GHz CPU and 2.5GB RAM. Both the primal problems (P1) and (P2) are solved using YALMIP, a MATLAB interface running SeDuMi as its SDP solver.

A. Performance of BFGS method

To compare with standard solvers, we mainly present the results from BFGS method for the dual formulation. The discussion of Newton method performance is postponed to next section. From complexity analysis presented in section III-C, the cost of BFGS method at each step is of order $\max\{r^3, (r+1)n^3\}$, where r is the number of G_i's and n is the number of states. The number of inputs determines the number of G_i's, hence dictates BFGS performance. From results listed in Table I, the time required to construct the basis is negligible compared to the optimization time. Moreover, the basis can be calculated offline and stored for future utilization. The time in all of the tables below is in *seconds*. The stopping criterion for BFGS method is $\|\nabla g(\nu)\| \le 10^{-6}$.

Two sets of optimization experiments are carried out with difference in the number of inputs m with respect to the number of states n. The first set of computations restricts the input number to one and increases the number of states (discretization size in y). The results are shown in Table II. Three different formulations give very close optimal solutions, i.e. up to at least five significant digits. In terms of solution time, the two primal formulations have similar performances. On the other hand, BFGS (as well as gradient and Newton method), has the difficulty in solving problems

	N=10 M=1	N=20 M=1	N=30 M=1
Basis Time	0.066	0.23	2.3
BFGS Time	1.4	114.8	1418.5

	N=30 M=28	N=50 M=47	N=100 M=93	N=300 M=297
Basis Time	0.06	0.06	0.28	1.16
BFGS Time	0.2	1	67.7	121

TABLE I
COMPARISON OF OPTIMIZATION TIME AND BASIS CONSTRUCTION TIME.

	Method	Time	Optimal Value
N = 10	BFGS	1.4	0.05014808
M = 1	P1	0.6	0.05014818
	P2	0.7	0.05014818
N = 20	BFGS	114.8	0.39130384
M = 1	P1	3.6	0.39130563
	P2	4.0	0.39130563
N = 30	BFGS	1418.5	0.39854288
M = 1	P1	26.5	0.39854725
	P2	24.6	0.39854726

TABLE II
COMPUTATION RESULTS FOR EXAMPLES WITH A SINGLE INPUT.

with a small number of inputs. As discussed in Section III-C, the reason is that the cost of algorithms for dual problem is dictated by the number of inputs compared with the number of states.

In second set of computations, the number of inputs is close to the number of states. The results are shown in Table III. Similarly, the optimal solutions obtained from the three formulations are very close to each other. In terms of efficiency, BFGS method shows very attractive feature in computation time. This is because the number of dual variables is small when m is close to n. Between the two primal formulations, (P2) can be solved with approximately half of time it takes to solve (P1) and more importantly, SeDuMi is more numerically stable when solving (P2). When dealing with (P1), SeDuMi usually runs into numerical problems and fails to give solutions when the problem size is larger than 1,500 ('NA' denotes that numerical problems have occurred).

From the numerical experiments, we observe that the transformation of primal problem (P1) to primal problem (P2) is advantageous for optimization. For large scale problems, the dual unconstrained formulation can be solved efficiently by developed algorithms when the number of inputs is close to the number of states (which is, for example, of interest in fluids problems).

	Method	Time	Optimal Value
N = 10	BFGS	0.04	0.0052754
M = 9	P1	1.6	0.0052754
	P2	1.3	0.0052754
N = 30	BFGS	0.2	0.0662776
M = 28	P1	40.7	NA
	P2	12.4	0.0662776
N = 50	BFGS	1.0	0.9997168
M = 47	P1	489.4	NA
	P2	243.8	0.9997167
N = 100	BFGS	67.7	16.344947
M = 93			
N = 300	BFGS	121	158.50262
M = 297			

TABLE III
COMPUTATION RESULTS FOR EXAMPLES WITH m CLOSE TO n.

B. Performance of generalized Newton method

It is proved by Qi and Sun [12] that generalized Newton method is quadratically convergent provided that ν is sufficiently close to its optimal value ν^*. In our computations, $\|\nabla g(\nu)\|$ decreases almost always by *half* in each Newton step before quadratic convergence occurs (see Fig. 1 for illustration).

The conjugate gradient method is also implemented to solve for the Newton direction. It almost always takes r steps, i.e. the size of the problem to obtain a sufficiently good solution. Another observation is that the Gram-Schmidt orthonormalization of the G_i's tremendously speeds up the convergence rate of CG method compared to the case with the original G_i's. The results are listed in Table IV, where CG and Dir denote the conjugate gradient method and direct method of computing Newton directions, respectively.

V. CONCLUDING REMARKS

The sample state covariances almost always fail to satisfy linear constraint dictated by the underlying dynamics. The consistency with such dynamics is crucial in addressing the problem of characterizing the input power spectrum. Hence, we formulate the structured covariance least-squares problem and convert the matrical linear constraint into an equivalent set of trace constraints. We develop quasi Newton and generalized Newton methods capable of solving the corresponding unconstrained dual problems with large number of variables.

With a different notion of distance, one can formulate an alternative version of covariance approximation problem [3]. The quantum relative entropy is a metric used to quantify information and uncertainty in quantum systems. Georgiou

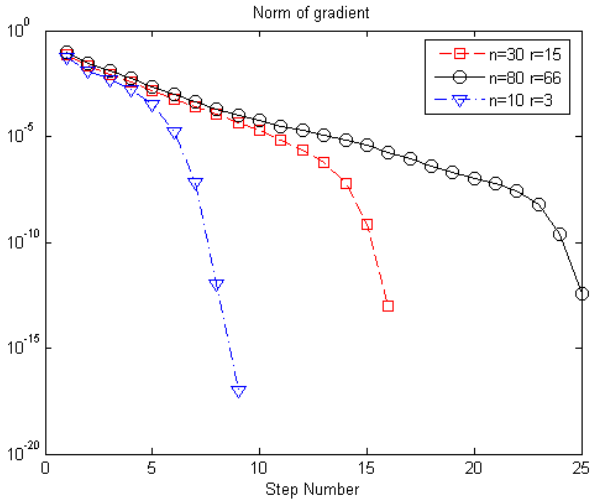

Fig. 1. The norm of gradient is plotted for different problem sizes using direct Newton method. The norm decreases approximately by half before quadratic convergence occurs.

	NT_ CG	NT_ Dir	BFGS	
N = 10	9	10	40	Iter No.
r = 6	0.08	0.03	0.04	Time(Sec)
N = 50	10	10	45	
r = 6	0.78	0.40	0.43	
N = 100	11	10	47	
r = 6	5.03	2.12	2.62	
N = 20	13	14	166	Iter No.
r = 21	1.54	0.35	0.47	Time(Sec)
N = 40	38	14	345	
r = 45	68.79	4.60	3.92	
N = 80	30	14	475	
r = 78	825.90	68.90	30.15	

TABLE IV
PERFORMANCES OF GENERALIZED NEWTON AND BFGS METHOD

recently provided a solution to this problem using homotopy based approach in [15]. We intend to explore large-scale covariance approximation problems with this metric in a future work. An interesting question may be to compare the optimal covariances from both the relative entropy distance and the Euclidean least-squares distance.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Defeng Sun and two anonymous reviewers for their suggestions about generalized Newton method. The authors would also like to thank Prof. Zhi-Quan Luo and Prof. Tryphon T. Georgiou for stimulating discussions.

APPENDIX

To compute the least-squares H after finding optimal X, one approach is to vectorize (LC) and solve the resultant linear equations of size $nm \times nm$. An alternative way developed in [1] is described as follows. Denote $D = AX + XA^*$ for notation simplicity,

$$\begin{aligned} BH + H^*B^* &= -D \\ B^*(BH + H^*B^*)B &= -B^*DB \\ B^*B(HB) + (HB)^*B^*B &= -B^*DB. \end{aligned}$$

With $X \in \mathcal{R}(\mathcal{L})$, there exists H such that HB is Hermitian. Therefore, solving this Lyapunov equation, one obtains HB and computes H as follows,

$$\begin{aligned} B^*(BH + H^*B^*) &= -B^*D \\ B^*BH + B^*H^*B^* &= -B^*D \\ H &= -(B^*B)^{-1}(B^*D + HBB^*). \end{aligned}$$

The cost of this computation is much smaller than the cost of solving the linear equations of dimension $nm \times nm$.

REFERENCES

[1] 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.

[2] ——, "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.

[3] ——, "Structured covariances and related approximation questions," in *Directions in mathematical systems theory and optimization*, A. Rantzer and C. I. Byrnes, Eds. Springer-Verlag, 2003, pp. 135–140.

[4] 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.

[5] 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, 2000.

[6] ——, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. Autom. Control*, vol. 45, no. 6, pp. 822–839, 2001.

[7] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

[8] J. Malick, "A dual approach to semidefinite least-squares problems," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 1, pp. 272–284, 2004.

[9] S. Boyd and L. Xiao, "Least-squares covariance matrix adjustment," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 2, pp. 532–546, 2005.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[11] D. P. Bertsekas, *Nonlinear programming*. Belmont, MA: Athena Scientific, 1999.

[12] H. Qi and D. Sun, "A quadratically convergent newton method for computing the newarest correlation matrix," *SIAM J. Matrix Analysis and Application.*, 2005.

[13] T. Kato, *A short Introduction to Perturbation Theory for Linear Operators*. Springer-Verlag, 1982.

[14] J. Weideman and S. Reddy, "A matlab differentiation matrix suite," *ACM Trans. on Mathematical Software*, vol. 26, no. 4, pp. 465–519, 2000.

[15] T. T. Georgiou, "Relative entropy and the multi-variable multi-dimensional moment problem," *IEEE Trans. on Information Theory*, vol. 52, no. 3, pp. 1052–1066, March 2006.