SIAM J. NUMER. ANAL.
Vol. 27, No. 4, pp. 1050-1060, August 1990

014

THE PROJECTED GRADIENT METHOD FOR LEAST SQUARES MATRIX APPROXIMATIONS WITH SPECTRAL CONSTRAINTS*

MOODY T. CHU† AND KENNETH R. DRIESSEL‡

Abstract. The problems of computing least squares approximations for various types of real and symmetric matrices subject to spectral constraints share a common structure. This paper describes a general procedure in using the projected gradient method. It is shown that the projected gradient of the objective function on the manifold of constraints usually can be formulated explicitly. This gives rise to the construction of a descent flow that can be followed numerically. The explicit form also facilitates the computation of the second-order optimality conditions. Examples of applications are discussed. With slight modifications, the procedure can be extended to solve least squares problems for general matrices subject to singular-value constraints.

Key words. least squares approximation, projected gradient, spectral constraints, singular-value constraints

AMS(MOS) subject classifications. 65F15, 49D10

1. Introduction.

Let $S(n)$ denote the subspace of all symmetric matrices in $R^{n\times n}$. Given a matrix $\Lambda \in S(n)$, we define an isospectral surface $M(\Lambda)$ of Λ by

$$M(\Lambda) := \{X \in R^{n\times n} \mid X = Q^T \Lambda Q,\ Q \in O(n)\} \tag{1}$$

where $O(n)$ is the collection of all orthogonal matrices in $R^{n\times n}$. Let Φ represent either a single matrix or a subspace in $S(n)$. For every $X \in S(n)$, the projection of X into Φ is denoted as $P(X)$. If Φ is a single matrix, then $P(X) \equiv \Phi$; otherwise, the projection is taken with respect to the Frobenius inner product. We consider the following matrix least squares problem with spectral constraints.

PROBLEM 1. Find $X \in M(\Lambda)$ that minimizes the function

$$F(X) := \tfrac{1}{2}\|X - P(X)\|^2 \tag{2}$$

where $\|\cdot\|$ means the Frobenius matrix norm.

To be more concrete, we mention below a partial list of problems that can be formulated in the above setting.

PROBLEM A. Given a real symmetric matrix $\hat{A}$, find a least squares approximation of $\hat{A}$ that is still symmetric but has a prescribed set of eigenvalues $\{\lambda_1, \cdots, \lambda_n\}$. In this case, we choose $\Phi \equiv \hat{A}$ and $\Lambda = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$.

PROBLEM B. Construct a symmetric Toeplitz matrix having a set of real numbers $\{\lambda_1, \cdots, \lambda_n\}$ as its eigenvalues. In this case, Φ is the subspace of all symmetric Toeplitz matrices and $\Lambda = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$.

PROBLEM C. Given X_0, find its eigenvalues. In this case, we may choose Φ to be the subspace of all diagonal matrices and $\Lambda = X_0$.

Although some of the above problems may be resolved by some other means (see, for example, [6]-[8], [12], [13]), the general setting as in Problem 1 carries intrinsically some interesting geometric properties. By exploring this geometry, we show in this

* Received by the editors August 22, 1988; accepted for publication (in revised form) October 2, 1989.

† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205. This work was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract W-31-109-Eng-38, while the author was spending his sabbatical leave at Argonne National Laboratory, Argonne, Illinois.

‡ Department of Mathematics, Idaho State University, Pocatello, Idaho 83209-0009.

paper that the projected gradient of the objective function $F(X)$ onto the manifold $M(\Lambda)$ can be calculated explicitly. As a consequence, a vector field on the manifold $M(\Lambda)$ that flows in a descent direction of $F(X)$ can be constructed. As another consequence, the explicit form of the projected gradient facilitates the computation of the second-order optimality conditions. We will see that this information, in turn, offers some new insights into the classification of the stationary points.

Computational efficiency has not been a major concern in the present paper, although our approach does offer a globally convergent numerical method. The vector field defined by the projected gradient can readily be integrated by any available software for initial value problems. But this may well be as slow as the usual steepest descent methods. Since we also know the projected Hessian, convergence certainly can be improved by many other standard techniques [7]. We stress here, however, that our approach is quite flexible in that we may use the subspace Φ to specify any desired (linear) structure on the optimal solution. The Toeplitz structure required in Problem B is such an example. If the subspace Φ does intersect the surface M, then of course the structure is attainable. Otherwise, our approach still finds a point on M that is a least squares approximation to Φ. As another example, we may wish to have an optimal solution that carries a certain specific zero pattern. To our knowledge, very few discrete numerical methods are available for solving this kind of problem. In [2] Chu and Norris have shown (from the matrix decomposition point of view) that a symmetric matrix with any kind of prescribed off-diagonal zero pattern is always reachable by following a specifically formulated isospectral flow.

Suppose now that Σ is a general matrix in $R^{m\times n}$ and Ψ is either a single matrix or a subspace of $R^{m\times n}$. Then analogous to the above notions we may consider the surface $W(\Sigma)$ defined by

$$W(\Sigma) := \{X \in R^{m\times n} \mid X = U^T \Sigma V,\ U \in O(m),\ V \in O(n)\}, \tag{3}$$

and the following optimization problem.

PROBLEM 2. Find $X \in W(\Sigma)$ that minimizes the function

$$F(X) := \tfrac{1}{2}\|X - P(X)\|^2 \tag{4}$$

where $P(X)$ means the natural projection of X into Ψ.

The following problems of practical interest are special cases of Problem 2.

PROBLEM D. Given a matrix $\hat{A} \in R^{m\times n}$ (say, $m \geqq n$), find a least squares approximation of $\hat{A}$ that has a prescribed set of singular values $\{\sigma_1, \cdots, \sigma_n\}$. In this case, we choose $\Psi \equiv \hat{A}$ and $\Sigma = \operatorname{diag}\{\sigma_1, \cdots, \sigma_n\}$, where diag is understood to be an $m \times n$ matrix with extra rows filled by zeros.

PROBLEM E. Given a matrix $\hat{A} \in R^{m\times n}$ (say, $m \geqq n$), find the singular values of $\hat{A}$. In this case, we may choose Ψ to be the subspace of all diagonal matrices in $R^{m\times n}$ and $\Sigma = \hat{A}$.

Again, Problem 2 can be understood from its intrinsic geometric properties. We will see that with slight modifications, the procedures developed for Problem 1 can easily be extended to Problem 2.

This paper is organized as follows. In the next section we begin with a brief review of the optimization theory. We point out in particular how, without forming the Lagrangian function, the derivative of any generalization of the projected gradient gives rise to the quadratic form of the projected Hessian. The central framework for calculating the projected gradient and forming isospectral flow for Problem 1 is discussed in § 3. Its application to Problems A, B, and C are further detailed in § 4. We show how the framework can be extended to answer Problems D and E in § 5.

2. Preliminaries. For completeness, we first review some important facts from the optimization theory. Consider the following basic equality constrained optimization problem:

$$\begin{aligned} &\text{Minimize} \quad F(x) \\ &\text{subject to} \quad C(x)=0 \end{aligned} \tag{5}$$

where $x\in R^n$, $F:R^n\to R$, and $C:R^n\to R^k$ with $k<n$ being sufficiently smooth functions. Let

$$M:=\{x\in R^n \mid C(x)=0\}. \tag{6}$$

We will assume that M is a regular surface, that is, for all $x\in M$, the set $\{\nabla C_i(x)\mid i=1,\cdots,k\}$ of vectors is linearly independent. Therefore, M is a smooth $(n-k)$-dimensional manifold [9]. Furthermore, for any $x\in M$, the space tangent to M at an x is given by

$$T_xM=\{y\in R^n \mid C'(x)y=0\}. \tag{7}$$

It is a fundamental fact [7] that for $\hat{x}$ to be optimal, it is necessary that the gradient vector $\nabla F(\hat{x})$ is perpendicular to the manifold M. Let $Z(x)\in R^{n\times(n-k)}$ denote a matrix whose columns form an orthonormal basis for T_xM. Then the projection $g(x)$ of $\nabla F(x)$ onto the tangent space T_xM is given by $g(x):=Z(x)Z(x)^T\nabla F(x)$. Note that $-g(x)$ also represents the "steepest descent" direction of F on the manifold M. Obviously, a necessary condition for $\hat{x}$ to be optimal is that

$$Z(\hat{x})^T\nabla F(x)=0. \tag{8}$$

For each $x\in M$, we may rewrite

$$g(x)=\nabla F(x)-\sum_{i=1}^{k}\lambda_i(x)\nabla C_i(x) \tag{9}$$

for some appropriate scalar functions $\lambda_i(x)$, since the second term on the right-hand side of (9) represents the component of $\nabla F(x)$ normal to T_xM. We now suggest a rather simple way of deriving the quadratic form of the projected Hessian. This shortcut may not work for general nonlinear optimization problems, but it proves convenient and valid for our consideration. Suppose the function g can be smoothly extended to the entire space R^n; that is, suppose the function

$$G(x):=\nabla F(x)-\sum_{i=1}^{k}\lambda_i(x)\nabla C_i(x) \tag{10}$$

is defined for every $x\in R^n$ and is smooth. Then for every x, $v\in R^n$, we have

$$v^TG'(x)v=v^T\left(\nabla^2F(x)-\sum_{i=1}^{k}\lambda_i(x)\nabla^2C_i(x)\right)v-v^T\left(\sum_{i=1}^{k}\nabla C_i(x)(\nabla\lambda_i(x))^T\right)v. \tag{11}$$

In particular, if $x\in M$ and $v\in T_xM$, then (11) is reduced to

$$v^TG'(x)v=v^T\left(\nabla^2F(x)-\sum_{i=1}^{k}\lambda_i(x)\nabla^2C_i(x)\right)v \tag{12}$$

since $v\perp\nabla C_i(x)$. We note from (12) that the condition $v^TG'(x)v\geqq 0$ for every $v\in T_xM$ is precisely the well-known second-order necessary optimality condition for problem (5).

3. Central framework. We now show how to calculate the projected gradient and to construct a continuous steepest descent flow on the manifold $M(\Lambda)$ for Problem 1. Note that Problem 1 is equivalent to the following.

PROBLEM 3.

$$\begin{aligned} &\text{Minimize} && F(Q) := \tfrac{1}{2}\langle Q^T\Lambda Q - P(Q^T\Lambda Q),\ Q^T\Lambda Q - P(Q^T\Lambda Q)\rangle \\ &\text{subject to} && Q^TQ = I \end{aligned} \tag{13}$$

since $X = Q^T\Lambda Q$. In (13), $\langle A, B\rangle$ denotes the Frobenius inner product of two matrices $A,\ B \in R^{n\times n}$ and is defined by

$$\langle A, B\rangle := \text{trace}\,(AB^T) = \sum_{i,j} a_{ij}b_{ij}. \tag{14}$$

Also, without causing ambiguity, we have used the same notation F for the objective function.

It is well known (and easy to prove) that under the Frobenius inner product the orthogonal complement of $S(n)$ is given by

$$S(n)^\perp = \{\text{all skew-symmetric matrices}\}. \tag{15}$$

Regarding the feasible set $O(n)$ in Problem 3 as the zero set of the function $C(X) := \frac{1}{2}(X^TX - I)$, we obtain from (7) that the tangent space of $O(n)$ at any orthogonal matrix Q is given by

$$\begin{aligned} T_QO(n) &= \{H \mid H^TQ + Q^TH = 0\} \\ &= \{H \mid Q^TH \text{ is skew-symmetric}\} \\ &= QS(n)^\perp \text{ (since } Q^T = Q^{-1}). \end{aligned} \tag{16}$$

It follows that the orthogonal complement of $T_QO(n)$ in $R^{n\times n}$ is given by

$$N_QO(n) = QS(n). \tag{17}$$

This is the space normal to $O(n)$ at Q.

Note that the objective function F in (13) is well defined for every general matrix $A \in R^{n\times n}$. Let $\alpha(A) := \frac{1}{2}\langle A, A\rangle$ and $\beta(A) := A^T\Lambda A - P(A^T\Lambda A)$. By the chain rule and the product rule, it is not difficult to show that for every $A,\ B \in R^{n\times n}$, the Fréchet derivative of F at A acting on B is given by

$$\begin{aligned} F'(A)B &= \alpha'(\beta(A))(\beta'(A)B) \\ &= \langle \beta(A), \beta'(A)B\rangle \\ &= \langle \beta(A), A^T\Lambda B - P'(A^T\Lambda A)A^T\Lambda B + B^T\Lambda A - P'(A^T\Lambda A)B^T\Lambda A\rangle \\ &= 2\langle \beta(A), A^T\Lambda B - P'(A^T\Lambda A)A^T\Lambda B\rangle \\ &\qquad\qquad \text{(since } \beta(A) \text{ is symmetric)} \\ &= 2\langle \beta(A), A^T\Lambda B\rangle \\ &\qquad\qquad \text{(since either } P' \equiv 0, \text{ or } P' = P; \text{ and } \beta(A) \perp \Phi) \\ &= 2\langle \Lambda A\beta(A), B\rangle. \end{aligned} \tag{18}$$

Equation (18) suggests that with respect to the Frobenius inner product, the gradient of F at a general matrix A can be interpreted as the matrix

$$\nabla F(A) = 2\Lambda A\beta(A). \tag{19}$$

With (19) on hand, we can identify a stationary point.

LEMMA 3.1. *A necessary condition for $Q \in O(n)$ to be a stationary point for Problem 3 is that the matrix $X := Q^T\Lambda Q$ commutes with its own projection $P(X)$.*

Proof. From (8) we know that Q is a stationary point for F only if $\nabla F(Q)$ is perpendicular to $T_Q O(n)$. By (17) and (19), this condition is equivalent to $\Lambda Q\beta(Q) \in N_Q O(n) = QS(n)$. Since $Q^{-1} = Q^T$, it follows that $X\beta(Q) = X(X - P(X)) \in S(n)$. Thus it must be that $XP(X) = P(X)X$.

We now calculate the projected gradient of $F(Q)$ on the manifold $O(n)$. We have seen that

$$R^{n\times n} = T_Q O(n) \oplus N_Q O(n) = QS(n)^{\perp} \oplus QS(n). \tag{20}$$

Therefore any matrix $X \in R^{n\times n}$ has a unique orthogonal splitting

$$X = Q\{\tfrac{1}{2}(Q^T X - X^T Q)\} + Q\{\tfrac{1}{2}(Q^T X + X^T Q)\} \tag{21}$$

as the sum of elements from $T_Q O(n)$ and $N_Q O(n)$. Accordingly, the projection $g(Q)$ of $\nabla F(Q)$ into the tangent space $T_Q O(n)$ can be calculated explicitly as follows:

$$\begin{aligned} g(Q) &= \tfrac{1}{2}Q\{(Q^T 2\Lambda Q\beta(Q) - 2\beta(Q)Q^T \Lambda Q)\} \\ &= Q\{X\beta(Q) - \beta(Q)X\} \\ &= Q[X, \beta(Q)] \\ &= Q[P(X), X]. \end{aligned} \tag{22}$$

In the above we have adopted the notation $X := Q^T \Lambda Q$ and the Lie bracket $[A, B] := AB - BA$.

From (22) it is clear that the vector field defined by the system

$$\frac{dQ(t)}{dt} = Q(t)[Q(t)^T \Lambda Q(t), P(Q(t)^T \Lambda Q(t))] \tag{23}$$

defines a (steepest) descent flow on the manifold $O(n)$ for the objective function $F(Q)$. Let $X(t) := Q(t)^T \Lambda Q(t)$. Then $X(t)$ is governed by the ordinary differential equation

$$\begin{aligned} \frac{dX(t)}{dt} &= \frac{dQ(t)^T}{dt} \Lambda Q(t) + Q(t)^T \Lambda \frac{dQ(t)}{dt} \\ &= -[X(t), P(X(t))]X(t) + X(t)[X(t), P(X(t))] \\ &= [[P(X(t)), X(t)], X(t)]. \end{aligned} \tag{24}$$

Note that by definition the flow $X(t)$ defined by (24) stays on the isospectral surface $M(\Lambda)$ for any initial value $X(0) = X_0 \in M(\Lambda)$. Furthermore, the value of the objective function $F(X)$ in (2) is guaranteed to be nonincreasing along the forward flow $X(t)$. (Indeed, it decreases in the steepest direction for most of the time.) Problem 1, therefore, may be solved simply by integrating the initial value problem

$$dX/dt = [[P(X), X], X], \qquad X(0) = X_0 \in M(\Lambda). \tag{25}$$

The explicit formula of the projected gradient $g(Q)$ in (22) may be used to calculate the second-order derivative condition for the objective function in the same way as we mentioned in the preceding section. We first extend the function g to the function $G: R^{n\times n} \to R^{n\times n}$ by defining

$$G(Z) := Z[P(Z^T \Lambda Z), Z^T \Lambda Z]. \tag{26}$$

By the product rule, it is easy to see that for any $Z, H \in R^{n\times n}$,

$$\begin{aligned} G'(Z)H = {} & H[P(Z^T \Lambda Z), Z^T \Lambda Z] + Z[P(Z^T \Lambda Z), Z^T \Lambda H + H^T \Lambda Z] \\ & + Z[P'(Z^T \Lambda Z)(Z^T \Lambda H + H^T \Lambda Z), Z^T \Lambda Z]. \end{aligned} \tag{27}$$

Consider the case when $Z = Q \in O(n)$ and $H \in T_Q O(n)$. Then $H = QK$ for some $K \in S(n)^{\perp}$. Let $X := Q^T \Lambda Q$. Upon substitution, we have

$$\begin{aligned}\langle G'(Q)QK, QK\rangle &= \langle QK[P(X), X] + Q[P(X), [X, K]] \\ &\quad + Q[P'(X)[X, K], X], QK\rangle \\ &= \langle K[P(X), X], K\rangle + \langle [P(X), [X, K]], K\rangle \\ &\quad + \langle [P'(X)[X, K], X], K\rangle.\end{aligned} \tag{28}$$

At a stationary point, $[P(X), X] = 0$. So (28) becomes

$$\langle G'(Q)(QK), QK\rangle = \langle [P(X), K] - P'(X)[X, K], [X, K]\rangle. \tag{29}$$

Note that P' is either P itself or identically zero. So (29) can be further simplified and thus provides additional information for the stationary points. We will demonstrate how these formulas can be used in the next section.

4. Applications. In this section we provide more computational details by applying the framework established earlier to Problems A, B, and C, respectively.

PROBLEM A. The projection mapping is the constant $P(X) \equiv \hat{A}$. Let $\Lambda = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$. According to (22), the projected gradient is given by

$$g(Q) = Q[\hat{A}, Q^T \Lambda Q]. \tag{30}$$

The solution $X(t)$ to the initial value problem

$$dX/dt = [[\hat{A}, X], X], \qquad X(0) = \Lambda \tag{31}$$

determines an isospectral flow that converges to a stationary solution of the least squares problem.

Let us now consider the second-order condition. For simplicity, we assume that

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n \tag{32}$$

and that the eigenvalues of $\hat{A}$ are ordered as

$$\mu_1 > \mu_2 > \cdots > \mu_n. \tag{33}$$

Let Q be a stationary point of F on $O(n)$. We define $X := Q^T \Lambda Q$ and

$$E := Q\hat{A}Q^T. \tag{34}$$

By Lemma 3.1 we should have $[\hat{A}, X] = 0$. It follows that E must be a diagonal matrix since E commutes with the diagonal matrix Λ [11]. (Assumption (32) is used here.) Since E and $\hat{A}$ are similar, the diagonal elements $\{e_1, \cdots, e_n\}$ of E must be a permutation of $\{\mu_1, \cdots, \mu_n\}$. We now use the second-order sufficient condition to check the type of the stationary point. Equation (29) at the stationary point Q becomes

$$\begin{aligned}\langle G'(Q)(QK), QK\rangle &= \langle [\hat{A}, K], [X, K]\rangle \\ &= \langle Q^T EQK - KQ^T EQ, Q^T \Lambda QK - KQ^T \Lambda Q\rangle \\ &= \langle E\hat{K} - \hat{K}E, \Lambda\hat{K} - \hat{K}\Lambda\rangle\end{aligned} \tag{35}$$

where the matrix $\hat{K} = QKQ^T$ is still skew symmetric. Let $\hat{k}_{ij}$ denote the (i, j)-component of the matrix $\hat{K}$. It is easy to see that (35) can be expressed as

$$\langle G'(Q)(QK), QK\rangle = 2 \sum_{i<j} (\lambda_i - \lambda_j)(e_i - e_j)\hat{k}_{ij}^2. \tag{36}$$

From (36) and assumption (32) we see that the second-order optimality condition has the following equivalent statements:

$$
\begin{aligned}
&\langle G'(Q)QK, QK\rangle > 0 \quad \text{for every } K \in S(n)^{\perp},\\
&\Leftrightarrow (\lambda_i - \lambda_j)(e_i - e_j) \geqq 0 \quad \text{for all } i<j\\
&\Leftrightarrow e_1 > e_2 > \cdots > e_n\\
&\Leftrightarrow e_i = \mu_i \quad \text{for every } i.
\end{aligned} \tag{37}
$$

Putting together (34) and (37), we have proved the following theorem.

THEOREM 4.1. *Under assumptions* (32) *and* (33), *a stationary point* Q *is a local minimizer of* F *on* $O(n)$ *if and only if the columns* $q_1, \cdots, q_n$ *of the matrix* Q^T *are the normalized eigenvectors of* $\hat{A}$ *corresponding, respectively, to* $\mu_1, \cdots, \mu_n$. *The solution to Problem* A *is unique* (*hence, is the global minimizer*) *and is given by*

$$X = \lambda_1 q_1 q_1^T + \cdots + \lambda_n q_n q_n^T. \tag{38}$$

Remark. The above theorem can be generalized with slight modifications for the multiple eigenvalue case. The only difference is that the least squares solution X is not necessarily unique if the matrix $\hat{A}$ has multiple eigenvalues. The details are discussed in [3].

Remark. Theorem 4.1 may be regarded as a reproof of the well-known Wielandt–Hoffman theorem [10], [15]. That is, let A, $A+E$ and $E \in S(n)$ have eigenvalues $\mu_1 > \cdots > \mu_n$, $\lambda_1 > \cdots > \lambda_n$ and $\tau_1 > \cdots > \tau_n$, respectively. Then

$$\sum_{i=1}^{n} (\lambda_i - \mu_i)^2 \leqq \sum_{i=1}^{n} \tau_i^2. \tag{39}$$

Obviously, the equality in (39) holds when the matrix $X = A+E$ is given by (38), where the Frobenius norm of the perturbation matrix E is minimized. We think the proof, being different from both the original proof of [10] and the one given in [15], is of interest in its own right.

PROBLEM B. For this problem Φ is the n-dimensional subspace of all symmetric Toeplitz matrices and $\Lambda := \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$. Note that Φ has a natural orthonormal basis $\{E_1, \cdots, E_n\}$ where $E_k := (e_{ij}^{(k)})$ and

$$e_{ij}^{(k)} := \begin{cases} 1/\sqrt{2(n-k+1)} & \text{if } 1 < k \leqq n \text{ and } |i-j| = k-1,\\ 1/\sqrt{n} & \text{if } k = 1 \text{ and } i = j,\\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

Thus the projection P is easy to compute and is given by

$$P(X) = \sum_{k=1}^{n} \langle X, E_k\rangle E_k. \tag{41}$$

Note that any diagonal matrix is necessarily a stationary point. So the initial value $X(0)$ of the differential equation (25) cannot be chosen to be just Λ. This restriction is not serious.

To our knowledge, the existence question of a solution to the inverse Toeplitz eigenvalue problem has not yet been settled [12]. Our descent flow approach, nevertheless, offers a globally convergent method of computation. By using the subroutine ODE in [14] as the integrator, for example, we have never failed to get convergence in our numerical experimentation. Occasionally we did experience cases of convergence to a stable stationary point that is not Toeplitz (the limit point, nevertheless, is always

persymmetric). By picking up a different initial value, we can easily change the course and converge to another stationary point. Our numerical experience seems to suggest that the inverse Toeplitz eigenvalue problem might have multiple solutions in general.

The second-order condition (see (29))

$$\langle G'(Q)QK, QK\rangle = \langle [P(X), K] - P[X, K], [X, K]\rangle \tag{42}$$

for Problem B becomes more involved now. For the time being we have not tried to use (42) to classify the stationary points. It seems plausible that by classifying all stationary points we could answer the theoretical existence question for the inverse Toeplitz eigenvalue problem. This direction certainly deserves further exploration.

PROBLEM C. Numerous algorithms have already been developed for solving the matrix eigenvalue problem. It is not, of course, our intention to claim that we have a new and effective method. Just as the Toda flow is a continuous realization of the QR algorithm [2], we want to show here that the Jacobi method also has a continuous analogue. Recall that the main idea behind the Jacobi method is to systematically reduce the norm of the off-diagonal elements. Let $\Lambda = X_0$ be the matrix whose eigenvalues are to be found. We choose Φ to be the subspace of all diagonal matrices. Since the projection $P(X) = \text{diag}\,(X)$ is just the diagonal matrix of X, we see that the objective of Problem 1 is now the same as that of the Jacobi method. The gradient flow (see (24)) defined by the initial value problem

$$dX/dt = [[\text{diag}\, X, X], X], \qquad X(0) = X_0, \tag{43}$$

therefore, may be regarded as a continuous analogue of the iterates generated by the Jacobi method.

The necessary condition for X to be a stationary point, by Lemma 3.1 is

$$[\text{diag}\, X, X] = 0. \tag{44}$$

The second-order sufficient condition for optimality at a stationary point, according to (29), is

$$\langle G'(Q)(QK), QK\rangle = \langle [\text{diag}\, X, K] - \text{diag}\,[X, K], [X, K]\rangle > 0 \tag{45}$$

for every skew symmetric matrix K. By using (44) and (45) we are able to classify all stationary points as follows.

THEOREM 4.2. *Let X be a stationary point for Problem* 1 *where Λ and Φ are defined as for Problem* C.

(1) *If X is a diagonal matrix, then X is an isolated global minimizer.*

(2) *If X is not a diagonal matrix but* diag X *is a scalar matrix (that is,* diag $X = cI$ *for some scalar c), then X is a global maximizer.*

(3) *If X is not a diagonal matrix and* diag X *is not a scalar matrix, then X is a saddle point.*

Proof. Readers are referred to [4, pp. 33-36], for detailed proofs.

We finally remark that the gradient flow (43) is moving by its own nature in a descent direction of the function $F(X)$. So the existence of the latter two cases in Theorem 4.2 should not cause any annoyance in the computation.

5. Extensions. The framework discussed in § 3 can be easily extended to Problem 2. The key to our approach is to define an inner product on the product space $R^{m\times m} \times R^{n\times n}$ through the induced Frobenius inner product:

$$\langle (A_1, A_2), (B_1, B_2)\rangle := \langle A_1, B_1\rangle + \langle A_2, B_2\rangle. \tag{46}$$

Regarding the feasible set $O(m)\times O(n)$ as the zero set of the function $C(A_1, A_2) := (\frac{1}{2}(A_1^TA_1 - I), \frac{1}{2}(A_2^TA_2 - I))$, we can show that the tangent space and the normal space of $O(m)\times O(n)$ at a point (Q_1, Q_2) are given, respectively, by

$$T_{(Q_1,Q_2)}O(m)\times O(n) = Q_1S(m)^{\perp}\times Q_2S(n)^{\perp} \tag{47}$$

and

$$N_{(Q_1,Q_2)}O(m)\times O(n) = Q_1S(m)\times Q_2S(n). \tag{48}$$

Without repeating too much, we now demonstrate how Problems D and E can be solved.

PROBLEM D. It is easy to see that Problem D is equivalent to the following formulation:

$$\begin{aligned} &\text{Minimize} && F(Q_1, Q_2) := \tfrac{1}{2}\|\Sigma - Q_1^T\hat{A}Q_2\|^2 \\ &\text{subject to} && Q_1^TQ_1 = I, \qquad Q_2^TQ_2 = I. \end{aligned} \tag{49}$$

Analogous to (18), we find that the Fréchet derivative of F at a general point (A_1, A_2) acting on (B_1, B_2) is given by

$$\begin{aligned} F'(A_1, A_2)(B_1, B_2) &= \langle \Sigma - A_1^T\hat{A}A_2, -B_1^T\hat{A}A_2 - A_1^T\hat{A}B_2\rangle \\ &= \langle \Sigma - A_1^T\hat{A}A_2, -B_1^T\hat{A}A_2\rangle + \langle \Sigma - A_1^T\hat{A}A_2, -A_1^T\hat{A}B_2\rangle \\ &= \langle -(\Sigma - A_1^T\hat{A}A_2)A_2^T\hat{A}^T, B_1^T\rangle + \langle -\hat{A}^TA_1(\Sigma - A_1^T\hat{A}A_2), B_2\rangle. \end{aligned} \tag{50}$$

Therefore, with respect to the inner product (46), we may interpret that the gradient of F at (A_1, A_2) is given by the pair

$$\nabla F(A_1, A_2) = (-\hat{A}A_2(\Sigma - A_1^T\hat{A}A_2)^T, -\hat{A}^TA_1(\Sigma - A_1^T\hat{A}A_2)). \tag{51}$$

A necessary condition for $(Q_1, Q_2)\in O(m)\times O(n)$ to be a stationary point of F is $\nabla F(Q_1, Q_2)\perp T_{(Q_1,Q_2)}O(m)\times O(n)$. This is equivalent to $\hat{A}Q_2(\Sigma - Q_1^T\hat{A}Q_2)\in Q_1S(m)$ and $\hat{A}^TQ_1(\Sigma - Q_1^T\hat{A}Q_2)\in Q_2S(n)$. Let

$$X := Q_1^T\hat{A}Q_2. \tag{52}$$

It is not difficult to see that the above necessary condition is equivalent to

$$X\Sigma^T = \Sigma X^T \quad \text{and} \quad X^T\Sigma = \Sigma^T X. \tag{53}$$

For simplicity, let us assume that

$$\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0 \tag{54}$$

and that the singular values of $\hat{A}$ are ordered (in the generic case) as

$$\mu_1 > \mu_2 > \cdots > \mu_n > 0. \tag{55}$$

Then the two equations in (53) imply that the $m\times n$ matrix X must be a diagonal matrix where the extra rows are filled with zeros. We know, therefore, that the diagonal elements, say, $e_1, \cdots, e_n$ of X, must be a permutation of singular values of $\hat{A}$.

The projection of $\nabla F(Q_1, Q_2)$ into the tangent space $T_{(Q_1,Q_2)}O(m)\times O(n)$ can be calculated according to the same principle as in (22). We claim that the projection of $\nabla F(Q_1, Q_2)$ is given by

$$\begin{aligned} g(Q_1, Q_2) &= (\tfrac{1}{2}\{Q_1\Sigma Q_2^2\hat{A}^TQ_1 - \hat{A}Q_2\Sigma^T\}, \tfrac{1}{2}\{Q_2\Sigma^TQ_1^T\hat{A}Q_2 - \hat{A}^TQ_1\Sigma\}) \\ &= (\tfrac{1}{2}Q_1(\Sigma X^T - X\Sigma^T), \tfrac{1}{2}Q_2(\Sigma^TX - X^T\Sigma)). \end{aligned} \tag{56}$$

Readers are invited to furnish the proof by themselves. As is suggested in § 1, we may define a flow $X(t)$ by

$$\begin{aligned} &dX/dt = \tfrac{1}{2}(\Sigma X^T X - X\Sigma^T X + XX^T\Sigma - X\Sigma^T X), \\ &X(0) = \hat{A}, \end{aligned} \tag{57}$$

which will move in a descent direction of F. Furthermore, we may extend the function g to

$$G(Z_1, Z_2) := (\tfrac{1}{2}\{Z_1\Sigma Z_2^T\hat{A}^T Z_1 - \hat{A}Z_2\Sigma^T\}, \tfrac{1}{2}\{Z_2\Sigma^T Z_1^T\hat{A}Z_2 - \hat{A}^T Z_1\Sigma\}) \tag{58}$$

for general matrices $(Z_1, Z_2) \in R^{m\times m} \times R^{n\times n}$ and take its derivative. In particular, we claim that at a stationary point (Q_1, Q_2) the projected Hessian of F acting on a tangent vector (Q_1K_1, Q_2K_2), where $K_1 \in S(m)^\perp$ and $K_2 \in S(n)^\perp$ is given by

$$\langle (Q_1K_1, Q_2K_2), G'(Q_1, Q_2)(Q_1K_1, Q_2K_2)\rangle = \langle K_1\Sigma - \Sigma K_2, K_1X - XK_2\rangle. \tag{59}$$

Again, readers are invited to fill in the details. We note here the similarity between (59) and (36). Let $k_{ij,1}$ and $k_{ij,2}$ denote the (i,j)-components of the skew matrices K_1 and K_2, respectively. Then (59) can be expressed as

$$\begin{aligned} &\langle (Q_1K_1, Q_2K_2), G'(Q_1, Q_2)(Q_1K_1, Q_2K_2)\rangle \\ &= \sum_{i=1}^{n}\sum_{p=n+1}^{n} e_i\sigma_i k_{pi,1}^2 \\ &\quad + \sum_{i\neq k} \{(e_i\sigma_i + e_k\sigma_k)k_{ik,1}^2 + (e_i\sigma_i + e_k\sigma_k)k_{ik,2}^2 - 2(e_i\sigma_k + e_k\sigma_i)k_{ik,1}k_{ik,2}\} \end{aligned} \tag{60}$$

since $k_{ii,j} = 0$ and $k_{ik,j} = -k_{ki,j}$ for all $1 \leqq i,\ k \leqq n$ and $j = 1$ or 2. The second-order optimality condition has the following equivalent statements:

$$\begin{aligned} &\langle (Q_1K_1, Q_2K_2), G'(Q_1, Q_2)(Q_1K_1, Q_2K_2)\rangle > 0 \\ &\quad \text{for every } K_1 \in S(m)^\perp,\ K_2 \in S(n)^\perp \\ &\Leftrightarrow (e_i\sigma_i + e_k\sigma_k)k_{ik,1}^2 + (e_i\sigma_i + e_k\sigma_k)k_{ik,2}^2 - 2(e_i\sigma_k + e_k\sigma_i)k_{ik,1}k_{ik,2} > 0 \\ &\quad \text{for every } k_{ik,1},\ k_{ik,2} \in R \\ &\Leftrightarrow \text{The discriminant } (e_i + e_k)(\sigma_i + \sigma_k)(e_i - e_k)(\sigma_i - \sigma_k) > 0 \text{ for every } i \text{ and } k \\ &\Leftrightarrow e_1 > e_2 > \cdots > e_n \\ &\Leftrightarrow e_i = \mu_i \quad \text{for every } i. \end{aligned} \tag{61}$$

In summary, we have proved the following theorem.

THEOREM 5.1. *Under the assumptions* (54) *and* (55), *a pair of matrices* (Q_1, Q_2) *is a local minimizer of* F *on* $O(m)\times O(n)$ *if and only if the columns of* Q_1 *and the columns of* Q_2 *are, respectively, the left and right singular vectors of* $\hat{A}$. *In this case the unique least squares approximation to* $\hat{A}$ *subject to the singular values constraints is given by*

$$X = Q_1\Sigma Q_2^T. \tag{62}$$

Remark. Using the above theorem, we can easily prove an analogue of the Wielandt–Hoffman theorem for singular values [8]. That is, let A and $A+E$ have singular values $\mu_1 > \mu_2 > \cdots > \mu_n > 0$ and $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$, respectively. Then

$$\sum_{i=1}^{n} (\sigma_i - \mu_i)^2 \leqq \|E\|^2. \tag{63}$$

Remark. We note here that the initial value problem (57) defines a descent flow $X(t)$ regardless of (54) and (55). Indeed, it is always the case in our approach that

we can define a descent flow without any knowledge of the second-order derivative. Of course, the descent path terminates when it hits a stationary point. So our approach finds a local minimum only if it starts at a suitable point.

PROBLEM E. This problem can be handled in a similar way as Problem C. We choose Ψ to be the subspace of all diagonal matrices and consider a Jacobi-type flow from the following optimization problem:

$$\begin{aligned} &\text{Minimize} && F(Q_1, Q_2) := \tfrac{1}{2}\| Q_1^T \hat{A} Q_2 - \operatorname{diag}(Q_1^T \hat{A} Q_2)\|^2 \\ &\text{subject to} && Q_1^T Q_1 = I, \qquad Q_2^T Q_2 = I. \end{aligned} \tag{64}$$

We can formulate the projected gradient of F explicitly. In particular, we claim that the initial value problem

$$\begin{aligned} &dX/dt = \tfrac{1}{2}\{((\operatorname{diag} X)X^T - X(\operatorname{diag} X)^T)X - X((\operatorname{diag} X)^T X - X^T(\operatorname{diag} X))\}, \\ &X(0) = \hat{A} \end{aligned} \tag{65}$$

defines a descent flow on the manifold $W(\hat{A})$ for the function F. As before, we can further classify the equilibrium points of (65) by means of the projected Hessian form. The details can be found in [5] and we will mention only the major result without proof.

THEOREM 5.2. *Let $\hat{A} \in R^{m\times n}$ have distinct, nonzero singular values, and let X be an equilibrium point of the differential equation* (65). *Then X is stable if and only if X is a diagonal matrix.*

Acknowledgment. A referee has suggested that the work of Arnold [1, p. 240 ff.] may be related to the results of § 3.

REFERENCES

[1] V. I. ARNOLD, *Geometrical Methods in the Theory of Ordinary Differential Equations*, Second edition, Springer-Verlag, New York, 1988.

[2] M. T. CHU AND L. K. NORRIS, *Isospectral flows and abstract matrix factorizations*, SIAM J. Numer. Anal., 25 (1988), pp. 1383-1391.

[3] M. T. CHU, *Least squares approximation by real normal matrices with specified spectrum*, SIAM J. Matrix Anal. Appl., to appear.

[4] K. R. DRIESSEL, *On finding the eigenvalues and eigenvectors of a matrix by means of an isospectral gradient flow*, Tech. Report 541, Department of Mathematical Sciences, Clemson University, Clemson, NC, 1987.

[5] ———, *On finding the singular values and singular vectors of a matrix by means of an isospectral gradient flow*, Tech. Report 87-01, Department of Mathematics, Idaho State University, Pocatello, Idaho, 1987.

[6] S. FRIEDLAND, J. NOCEDAL, AND M. L. OVERTON, *The formulation and analysis of numerical methods for inverse eigenvalue problems*, SIAM J. Numer. Anal., 24 (1987), pp. 634-667.

[7] P. E. GILL, W. MURRAY, AND M. H. WRIGHT, *Practical Optimization*, Academic Press, Florida, 1981.

[8] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Second edition, The Johns Hopkins University Press, Baltimore, MD, 1989.

[9] V. GUILLEMIN AND A. POLLACK, *Differentiable Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[10] A. J. HOFFMAN AND H. WIELANDT, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953), pp. 37-39.

[11] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices*, Second edition, Academic Press, Florida, 1985.

[12] D. P. LAURIE, *A numerical approach to the inverse Toeplitz eigenproblem*, SIAM Sci. Statist. Comput., 9 (1988), pp. 401-405.

[13] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[14] L. F. SHAMPINE AND M. K. GORDON, *Computer Solution of Ordinary Differential Equations, The Initial Value Problem*, W. H. Freeman, San Francisco, CA, 1975.

[15] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.