On a Differential Equation Approach to the Weighted Orthogonal Procrustes Problem

Moody T. Chu[1]

Department of Mathematics

North Carolina State University

Raleigh, North Carolina, 27695-8205

U.S.A.

Nickolay T. Trendafilov[2]

Computer Stochastic Laboratory

Institute of Mathematics, Bulgarian Academy of Sciences

P. O. Box 373, Sofia 1090

BULGARIA

[1] This research was supported in part by National Science Foundation under grant DMS-9422280.

[2] This research was supported in part by Japan Society for the Promotion of Science and by Bulgarian Ministry of Science, Education and Technology under grant MM421.

Abstract. The weighted orthogonal Procrustes problem, an important class of data matching problems in multivariate data analysis, is reconsidered in this paper. It is shown that a steepest descent flow on the manifold of orthogonal matrices can naturally be formulated. This formulation has two important implications: that the constrained regression problem can be solved as an initial value problem by any available numerical integrator and that the first order and the second order optimality conditions can also be derived. The proposed approach is illustrated by numerical examples.

Key words. Constrained regression, Weighted Procrustes, Orthogonal rotation, Projected gradient, Optimality condition, Differential equation.

1. Introduction. Recently there has been an interesting new look at existing numerical algorithms in the area of matrix computation. It has been observed that there is a connection between the dynamics of a certain differential equations and a certain discrete numerical methods. A short list of these connections can be found in a recent review by Chu (1994). The process of using differential equations to realize the solution of a problem is called a continuation method. This approach often furnishes more behavioral information about the solution of the underlying problem. The main contribution of this paper is in the application of one continuous realization process to solve the so called *weighted orthogonal Procrustes problem* (WOPP). In particular, we shall show that a steepest descent flow for the WOPP in the form a matrix differential equation can be derived explicitly. Following this flow becomes a read-made numerical algorithm and will lead us to a desired solution. More significantly, the process makes it possible to completely characterize the necessary and sufficient optimality conditions for the WOPP.

A weighted orthogonal Procrustes problem concerns the following optimization:

$$\text{Minimize} \quad \|AQC - B\| \tag{1}$$

$$\text{Subject to} \quad Q \in R^{p\times p},\ Q^TQ = I_p. \tag{2}$$

where $A \in R^{n\times p}, C \in R^{p\times m}$ and $B \in R^{n\times m}$ are given and fixed, and I_p is the $p \times p$ identity matrix. The well known *orthogonal Procrustes problem* (OPP):

$$\text{Minimize} \quad \|AQ - B\| \tag{3}$$

$$\text{Subject to} \quad Q \in R^{p\times p},\ Q^TQ = I_q, \tag{4}$$

obviously is a special case of (1) with $m = p$ and $C = I_p$. The OPP arises in many areas of applications, including the simple structure rotations in multidimensional scaling and exploratory factor analysis (Gower, 1984; Trendafilov, 1996), and its optimal solution Q_* enjoys a closed form given by

$$Q_* = VU^T, \tag{5}$$

provided V and U are the orthogonal matrices involved in the singular value decomposition of $A^T B = V\Sigma U^T$ (Mulaik, 1972; Golub & Van Loan, 1991). A closed form solution for the more general WOPP does not seem possible. Iterative methods for solving the WOPP have been proposed by Koschat & Swayne (1991) and Mooijaart & Commandeur (1990).

We shall develop a continuous descent flow for the WOPP by following ideas discussed in Chu & Driessel (1990). The same approach also works for the more general case when Q in the above is sought to be a *rectangular* orthonormal matrix. That result is considerably more abstract and is discussed in a separate paper (Chu & Trendafilov, 1996). In this paper, we concentrate on the case when Q is a *square* orthogonal matrix so that we can convey the idea in a more explicit formulation. The most important feature of our approach is that it offers a *globally* convergent method for finding a local solution of the WOPP.

2. Steepest Descent Flow. We first introduce a topology for the set

$$\mathcal{O}(p) := \{Q \in R^{p\times p} | Q^T Q = I_p\} \tag{6}$$

of all $p \times p$ orthogonal matrices. It is known that $\mathcal{O}(n)$ forms a smooth manifold of dimension $(p(p-1)/2$ in $R^{p\times p}$. Indeed, any vector tangent to $\mathcal{O}(p)$ at $Q \in \mathcal{O}(p)$ is necessarily of the form QK for some skew-symmetric matrix $K \in R^{p\times p}$. Denote

$$\mathcal{S}(p) := \{\text{all symmetric matrices in } R^{p\times p}\},$$

and introduce the Frobenius inner product of two matrices X and Y:

$$\langle X, Y\rangle := \text{trace}(XY^T).$$

It follows that the tangent space $\mathcal{T}_Q\mathcal{O}(p)$ and the normal space $\mathcal{N}_Q\mathcal{O}(p)$ of $\mathcal{O}(p)$ at any $Q \in \mathcal{O}(p)$ are given, respectively, by:

$$\mathcal{T}_Q\mathcal{O}(p) = Q\mathcal{S}(p)^{\perp} \tag{7}$$

$$\mathcal{N}_Q\mathcal{O}(p) = Q\mathcal{S}(p), \tag{8}$$

where $\mathcal{S}(p)^{\perp}$ is the orthogonal complement of $\mathcal{S}(p)$ with respect to the Frobenius inner product and hence consists of all skew-symmetric matrices in $R^{p\times p}$.

For given matrices $A \in R^{n\times p}, C \in R^{p\times m}$ and $B_{n\times m}$, consider the function

$$F(Z) = \frac{1}{2}\langle AZC - B, AZC - B\rangle \tag{9}$$

defined for all $Z \in R^{p\times p}$. Apparently, the WOPP is equivalent to the minimization of $F(Z)$ over the feasible set $\mathcal{O}(p)$. This is a standard constrained optimization problem, but we shall show below that we can obtain the information about the projected gradient and the projected Hessian without using the conventional Lagrangian multipliers technique.

We first calculate the gradient $\nabla F(Z)$ of the objective function $F(Z)$ to be:

$$\nabla F(Z) = A^T(AZC - B)C^T. \tag{10}$$

Suppose the projection $g(Q)$ of the gradient $\nabla F(Q)$ at a point $Q \in \mathcal{O}(p)$ onto the tangent space $\mathcal{T}_Q\mathcal{O}(p)$ can be computed explicitly. Then the differential equation

$$\frac{dQ}{dt} = -g(Q) \tag{11}$$

naturally defines a flow $Q(t)$ on the feasible manifold $\mathcal{O}(p)$. Along the flow $Q(t)$ the function value $F(Q(t)$ is decreasing most rapidly relative to any other direction. Indeed, we have

$$\begin{aligned} \frac{dF(Q(t))}{dt} &= \langle \nabla F(Q(t)), -g(Q(t))\rangle \\ &= -\|g(Q(t))\|^2 \qquad (12) \end{aligned}$$

due to the fact that $g(Q(t))$ is the orthogonal projection of $\nabla F(Q(t))$ onto $\mathcal{T}_{Q(t)}\mathcal{O}(p)$. It is important to note from (12) that the value of $F(Q(t))$ decreases strictly until $g(Q(t))$ becomes zero, indicating a local minimum has been reached. This descent property is universal regardless where the flow starts. The flow is guaranteed to converge globally.

To obtain this projected gradient $g(Q)$, observe that

$$R^{p\times p} = \mathcal{T}_Q\mathcal{O}(p) \oplus \mathcal{N}_Q\mathcal{O}(p) = Q\mathcal{S}(p)^\perp \oplus Q\mathcal{S}(p). \tag{13}$$

Therefore, any matrix $X \in R^{p\times p}$ has an unique orthogonal decomposition:

$$X = Q\left[\frac{1}{2}(Q^TX - X^TQ)\right] + Q\left[\frac{1}{2}(Q^TX + X^TQ))\right]$$

as the sum of elements from $\mathcal{T}_Q\mathcal{O}(p)$ and $\mathcal{N}_Q\mathcal{O}(p)$. In particular, the projection $g(Q)$ of $\nabla F(Q)$ onto the tangent space $\mathcal{T}_Q\mathcal{O}(p)$ has the form:

$$g(Q) = \frac{Q}{2}\left(Q^T A^T (AQC - B)C^T - C(AQC - B)^T AQ\right) \tag{14}$$

We therefore obtain the differential equation

$$\frac{dQ}{dt} = \frac{Q}{2}\left(C(AQC - B)^T AQ - Q^T A^T (AQC - B)C^T\right) \tag{15}$$

that defines a steepest descent flow on the manifold $\mathcal{O}(p)$ for the objective function F in (9). In the case of OPP, the differential equation reduces to a much simpler form:

$$\frac{dQ}{dt} = \frac{Q}{2}(Q^T A^T B - B^T AQ). \tag{16}$$

Starting with any point in $\mathcal{O}(p)$, say $Q(0) = I$, we may follow the flow defined by (15) by any available initial value problem solver. The flow eventually will converge to a local solution for the WOPP. This is the ready-made numerical algorithm referred to above for the WOPP.

3. Optimality Conditions. The explicit formulation of the projected gradient (14) provides additional information about the first order optimality condition for a stationary point:

THEOREM 3.1. *A necessary condition for $Q \in \mathcal{O}(p)$ to be a stationary point of the WOPP is that the matrix $C(AQC - B)^T AQ$ is symmetric.*

Proof. Obviously Q is a stationary point only if $g(Q) = 0$. The assertion then follows from (14). □

Corollary 3.2. *A necessary condition for $Q \in \mathcal{O}(p)$ to be a stationary point of the OPP is that the matrix $B^T AQ \in R^{p\times p}$ be symmetric.*

We also can derive a second order optimality condition to further classify the stationary points. For completion, we first outline below a technique for obtaining the projected Hessian. More detailed development of this idea can be found in (Chu & Driessel, 1990). Suppose we are given the following equality constrained optimization problem:

$$\text{Minimize} \quad f(x) \tag{17}$$

$$\text{Subject to} \quad c(x) = 0, \tag{18}$$

where $x \in R^n, f : R^n \longrightarrow R$ and $c : R^n \longrightarrow R^k$ with $k < n$ are sufficiently smooth functions. Suppose the constraint $c(s) = [c_1(x), \ldots c_k(x)]^T$ is regular in the sense that vectors $\nabla c_i(x), i = 1, \ldots, k$, are linearly independent. Then the feasible set

$$\mathcal{M} := \{x \in R^n | c(x) = 0\}.$$

forms an $(n-k)$-dimensional smooth submanifold in R^n. We may express the vector $\nabla f(x)$ as the sum of its projection $p(x)$ onto the tangent space $\mathcal{T}_x\mathcal{M}$ of $\mathcal{M}$ at x and a linear combination of vectors from the normal space, i.e., we may write

$$p(x) = \nabla f(x) - \sum_{i=1}^{k} \lambda_i(x) \nabla c_i(x) \tag{19}$$

for some appropriate scalar functions $\lambda_i(x)$. Suppose $p(x)$ is smooth in x. Then for every $x, v \in R^n$, we have

$$v^T p'(x) v = v^T(\nabla^2 f(x) - \sum_{i=1}^{k} \lambda(x) \nabla^2 c_i(x)) v - v^T(\sum_{i=1}^{k} \nabla c_i(x)(\nabla \lambda_i(x))^T) v. \tag{20}$$

In particular, if $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$, then (20) is reduced to

$$v^T p'(x) v = v^T(\nabla^2 f(x) - \sum_{i=1}^{k} \lambda_i(x) \nabla^2 c_i(x)) v \tag{21}$$

since $v \perp \nabla c_i(x)$. We note from (21) that the condition $v^T p'(x) v \geq 0$ for every $v \in T_x\mathcal{M}$ is precisely the well known second order necessary optimality condition for the problem (17). The above technique works only if the projected gradient $p(x)$ is explicitly known, which is our case. The advantage is that it significantly cuts short the work of what would be if going through the classical notion of Lagrangian multipliers.

We now apply the above technique to calculate the projected Hessian. We claim that

THEOREM 3.3. *The action of the projected Hessian of F at a stationary point $Q \in \mathcal{O}(p)$ on a tangent vector QK where K is skew-symmetric is given by*

$$\langle g'(Q)QK, QK \rangle = \left\langle C(AQC - B)^T AQ, K^2 \right\rangle + \left\langle Q^T A^T AQKCC^T, K \right\rangle. \tag{22}$$

Proof. From (14), observe that the Fréchet derivative of g at Q on a general H is given by

$$\begin{aligned} g'(Q)H \; := \; & \frac{H}{2}\left(Q^T A^T (AQC - B) C^T - C(AQC - B)^T AQ\right) \\ & + \frac{Q}{2}\left(H^T A^T (AQC - B) C^T + Q^T A^T AHCC^T \right. \\ & \left. - CC^T H^T A^T AQ - C(AQC - B)^T AH\right). \end{aligned}$$

At a stationary point, by Theorem 3.1, the quantity in the first big parentheses in the above is zero. The Hessian action on a tangent vector $H = QK$ can be calculated as follows:

$$\begin{aligned}\langle g'(Q)QK, QK\rangle &= \left\langle \frac{K^TQ^TA^T(AQC-B)C^T - \left(K^TQ^TA^T(AQC-B)C^T\right)^T}{2}, K\right\rangle \\ &\quad + \left\langle \frac{Q^TA^TAQKCC^T - \left(Q^TA^TAQKCC^T\right)^T}{2}, K\right\rangle \\ &= \left\langle K^TQ^TA^T(AQC-B)C^T, K\right\rangle + \left\langle Q^TA^TAQKCC^T, K\right\rangle.\end{aligned}$$

The assertion follows from the adjoint property $\langle XY, Z\rangle = \langle Y, X^TZ\rangle$ and the fact that $K^T = -K$. □

The following characterization is a standard result in optimization theory. See, for example, (Gill, Murray, & Wright (1981), 3.4.)

COROLLARY 3.4. *A second order sufficient (necessary) condition for a stationary point $Q \in \mathcal{O}(p)$ to be a minimizer the WOPP is that*

$$\left\langle C(AQC-B)^TAQ, K^2\right\rangle + \left\langle Q^TA^TAQKCC^T, K\right\rangle > (\geq) 0 \tag{23}$$

for all nonzero $K \in \mathcal{S}(p)^\perp$.

Let the singular value decomposition of the skew-symmetric matrix K be denoted by

$$K = U\Sigma W^T, \tag{24}$$

where $U, W \in \mathcal{O}(p)$, and $\Sigma = \text{diag}\{\sigma_1, \ldots, \sigma_p\}$ contains singular values. It follows that

$$K^2 = -U\Sigma^2U^T \tag{25}$$

which in fact is the spectral decomposition of K^2. We know from Theorem 3.1 that $C(AQC-B)^TAQ$ is necessarily symmetric at any stationary point Q. Let

$$C(AQC-B)^TAQ = V\Lambda V^T, \tag{26}$$

and similarly,

$$A^T A = T\Phi T^T$$

$$CC^T = S\Psi S^T$$

denote, respectively, the spectral decomposition of the corresponding matrix. Note that all entries in $\Phi = \operatorname{diag}\{\phi_1, \ldots, \phi_p\}$ and $\Psi = \operatorname{diag}\{\psi_1, \ldots, \psi_q\}$ are nonnegative. Using (25), we can rewrite

$$\begin{aligned} \langle g'(Q)QK, QK\rangle &= -\left\langle V\Lambda V^T, U\Sigma^2 U^T\right\rangle + \langle \Phi R\Psi, R\rangle \\ &= -\sum_{i=1}^{p} \lambda_i \left(\sum_{t=1}^{p} p_{it}^2 \sigma_t^2\right) + \sum_{j=1}^{p} \phi_j \left(\sum_{s=1}^{q} r_{js}^2 \psi_s\right) \end{aligned} \tag{27}$$

where $P = (p_{it}) := V^T U$ and $R = (r_{js}) := T^T QKS$. We can make the following claim:

THEOREM 3.5. *Suppose $Q \in \mathcal{O}(p)$ is a stationary point and suppose $C(AQC - B)^T AQ \neq 0$. Then a sufficient condition for Q to be a solution for the WOPP is that the matrix $C(AQC - B)^T AQ$ be negative semi-definite.*

Proof. The negative semi-definiteness of $C(AQC - B)^T AQ$ implies that $\lambda_i \leq 0$ for all i and that at least one λ_i is negative. Observe that $\sum_{t=1}^{p} p_{it}^2 \sigma_t^2 > 0$ for all t unless $K = 0$. Observer also that every factor in the second term of (27) is nonnegative. It follows that $\langle g'(Q)QK, QK\rangle > 0$ for any nonzero skew-symmetric matrix K. □

For the OPP (where $C = I_p$), the projected Hessian (22) is reduced to

$$\langle g'(Q)QK, QK\rangle = -\left\langle B^T AQ, K^2\right\rangle. \tag{28}$$

The necessary condition in the following result is known in the literature (Gower, 1984) and (ten Berge & Knol, 1984), but we now prove that it is also sufficient.

Corollary 3.6. *Suppose $B^T A \neq 0$. Then a second order necessary and sufficient condition for a stationary point $Q \in \mathcal{O}(p)$ to be a solution of the OPP is that the matrix $B^T AQ$ be positive semi-definite.*

Proof. Recall from Corollary 3.2 that at a stationary point Q the matrix $B^T AQ$ is necessarily symmetric. Without causing ambiguity, we use the same notation as in (26) to denote the spectral decomposition

$$B^T AQ = V \Lambda V^T.$$

Then from (25) and (28) we find that

$$\langle g'(Q)QK, QK \rangle = \sum_{i=1}^{p} \lambda_i \left(\sum_{t=1}^{p} p_{it}^2 \sigma_t^2 \right) \tag{29}$$

with $P = (p_{it}) = V^T U$. Observe that P is an arbitrary orthogonal matrix and that, by choosing K appropriately, σ_i can be made arbitrarily small. In order to have $\langle g'(Q)QK, QK \rangle \geq 0$ for all nonzero skew-symmetric K, it is necessarily $\lambda_i \geq 0$ for all i. On the other hand, $B^T AQ$ must have at least one positive eigenvalue since otherwise $B^T A = 0$. That one positive eigenvalues is sufficient to guarantee $\langle g'(Q)QK, QK \rangle > 0$ since the second sum in (29) is positive for all i. ☐

We remark here that, in contrast to Corollary 3.6, we fall short to claim that the negative semi-definiteness of $C(AQC - B)^T AQ$ is also a necessary condition in Theorem 3.5. The difficulty is due to the two terms in (27) that are of the same order $O(\|K\|^2)$. At this point we simply cannot provide a rigorous argument to show that $\lambda_i \geq 0$ for all i is necessary to guarantee $\langle g'(Q)QK, QK \rangle \geq 0$ for all nonzero skew-symmetric K, although it is so tempting to think this is true. The matter of fact is that we can argue the opposite by considering the special case $m = p$ and $C = I_p$. We claim that the sufficient condition in Theorem 3.5 is more restrictive than that in Corollary 3.6. It therefore cannot be a necessary condition. To see our point, observe that

$$(AQ - B)^T AQ \preceq 0 \Longleftrightarrow Q^T A^T AQ \preceq B^T AQ$$

where we have used the partial ordering $\preceq$ to indicate $X \preceq Y$ if and only if $X - Y$ is negative semi-definite. Since $0 \preceq Q^T A^T AQ$ is obvious, $(AQ - B)^T AQ \preceq 0$ certainly implies $0 \preceq B^T AQ$, but not the converse.

4. Numerical Experiment. In this section, we report some of our numerical experiments with the application of (15) to the WOPP. The computation is carried out by MATLAB 4.2a on an ALPHA 3000/300LX workstation. We choose to use **ode113** and **ode15s** from the MATLAB ODE SUITE (Shampine & Reichelt, 1995) as the integrators for the initial value problem. These codes are available from the network. The code **ode113** is a PECE implementation of Adams-Bashforth-Moulton methods for non-stiff systems. The code **ode15s** is a quasi-constant step size implementation of the Klopfenstein-Shampine family of the numerical differential formulas for stiff systems. More details of these codes can be found in the document prepared by Shampine & Reichelt (1995). Our reason for using theses codes is simply for convenience and illustration, especially when the equation (15) involves only matrix computations. Any other ODE solvers can certainly be used instead.

In our experiments, the tolerance for both absolute error and relative error is set at 10^{-12}. This criterion is used to control the accuracy in following the solution path. The high accuracy we required here has little to do with the dynamics of the underlying vector field, and perhaps is not needed for practical applications in data analysis. It is used only for illustration to accurately follow the flow. Lower accuracy requirement in the calculation certainly can save some CPU time, but not significantly since our calculation is fast already. We also discover that **ode15s** is more expensive but faster (due to larger step size taken) than **ode113** while the latter usually performs reasonably well. The output values at time interval of 10 are examined. The integration terminates automatically when the relative improvement of $F(Q)$ between two consecutive output points is less than 10^{-10}, indicating local minimizer has been found. This stopping criterion can be modified if desires to do so.

We should make one more comment concerning the implementation. Note that in theory the flow defined by (15) should automatically stay on the manifold $\mathcal{O}(p)$. In numerical calculation, however, round-off errors and truncations errors can easily throw the computed $Q(t)$ off the manifold of constraint. This can be remedied by an additional non-linear projection scheme suggested by Gear (1986): Suppose Q is an approximate solution to (15) satisfying

$$Q^T Q = I + O(h^r)$$

where r represents the order of the numerical method. Let $Q = \tilde{Q}R$ be the unique QR decomposition of Q with $\mathrm{diag}(R) > 0$. Then

$$\tilde{Q} = Q + O(h^r) \tag{30}$$

and $\tilde{Q} \in \mathcal{O}(p)$. The condition $\mathrm{diag}(R) > 0$ is important to ensure the transition of $Q(t)$ is smooth in t. In our code, the matrix Q on the right-hand side of (25) is replaced by the corresponding $\tilde{Q}$.

We have experimented with many tests where the problem data are generated randomly. Because of the global convergence property of our method, all tests have similar dynamical behavior. The following example represent a typical run although the length it takes to reach convergence may vary from data to data. So as to fit the data comfortably in the running text, we display all numbers only with five digits. All codes used in this experiment are available upon request.

Consider the case where

$$A = \begin{bmatrix} 0.9795 & 0.1703 & 0.3155 & 0.7757 \\ 0.3289 & 0.6185 & 0.4219 & 0.9204 \\ 0.1669 & 0.4626 & 0.4884 & 0.7317 \\ 0.2024 & 0.6667 & 0.1514 & 0.0655 \\ 0.4467 & 0.2071 & 0.0708 & 0.6308 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.8429 & 0.0162 & 0.6133 \\ 0.2509 & 0.4534 & 0.4681 \\ 0.7911 & 0.2829 & 0.3588 \\ 0.9180 & 0.3291 & 0.1564 \end{bmatrix}.$$

Consider also the random permutation matrix

$$Q_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We define $B := AQ_0C$ so that the underlying WOPP has a global solution at Q_0. Suppose the flow (15) starts with the random orthogonal matrix

$$Q(0) = \begin{bmatrix} 0.3662 & 0.1100 & 0.8964 & -0.2241 \\ 0.3662 & 0.8484 & -0.3098 & -0.2241 \\ 0.7731 & -0.5060 & -0.3098 & -0.2241 \\ 0.3662 & 0.1100 & 0.0674 & 0.9216 \end{bmatrix}.$$

Figure 1 records the history of the changes of the objective value $2F(Q(t)) = \|AQ(t)C - B\|$ where $Q(t)$ is determined by integrating the differential equation (15). Clearly, the global solution is obtained in this case.

Also recorded in Figure 1 is the history of the function

$$\Omega(Q(t)) := \|I_p - Q(t)^T Q(t)\| \qquad (31)$$

that measures the deviation of $Q(t)$ from the manifold of constraint $\mathcal{O}(p)$. It is seen that $Q(t)$ is well kept within the local tolerance.

Note that due to the nonlinear nature of the WOPP, the initial value determines the destination the descent flow converges to. Suppose the flow starts with this orthogonal matrix

$$Q(0) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

We find that the flow converges to the orthogonal matrix

$$Q_* = \begin{bmatrix} 0.2079 & 0.3435 & -0.3968 & 0.8254 \\ 0.3238 & 0.8509 & 0.2877 & -0.2974 \\ -0.7844 & 0.2512 & 0.4692 & 0.3186 \\ 0.4865 & -0.3081 & 0.7346 & 0.3588 \end{bmatrix}$$

that is only a local solution to the WOPP. The test results are presented in Figure 2.

Define the matrix

$$W_* := C(AQ_*C - B)^T AQ_* = \begin{bmatrix} 0.0857 & 0.0310 & 0.0611 & 0.0558 \\ 0.0310 & 0.0112 & 0.0221 & 0.0202 \\ 0.0611 & 0.0221 & 0.0436 & 0.0398 \\ 0.0558 & 0.0202 & 0.0398 & 0.0363 \end{bmatrix}.$$

According to our theory, Theorem 3.1, W_* should be symmetric. We check that $\|W - W^T\| = 1.9283 \times 10^{-7}$, indicating a severe loss of symmetry during the integration. It is interesting to note that the computed eigenvalues of W_* are

$$\left\{1.7687 \times 10^{-1},\ -3.9217 \times 10^{-7},\ 1.7485 \times 10^{-7},\ 2.5429 \times 10^{-18}\right\}.$$

The last eigenvalue perhaps is a machine zero, but the middle two eigenvalues have the ambiguity of being either round-off approximations of the zero or truly numbers with small magnitude. If the latter, then this example would be a case where at a local minimizer Q_* the corresponding W_* need not be negative semi-definite. But given the fact that W_* itself might carry an error of the order 10^{-7}, we simply cannot be conclusive.

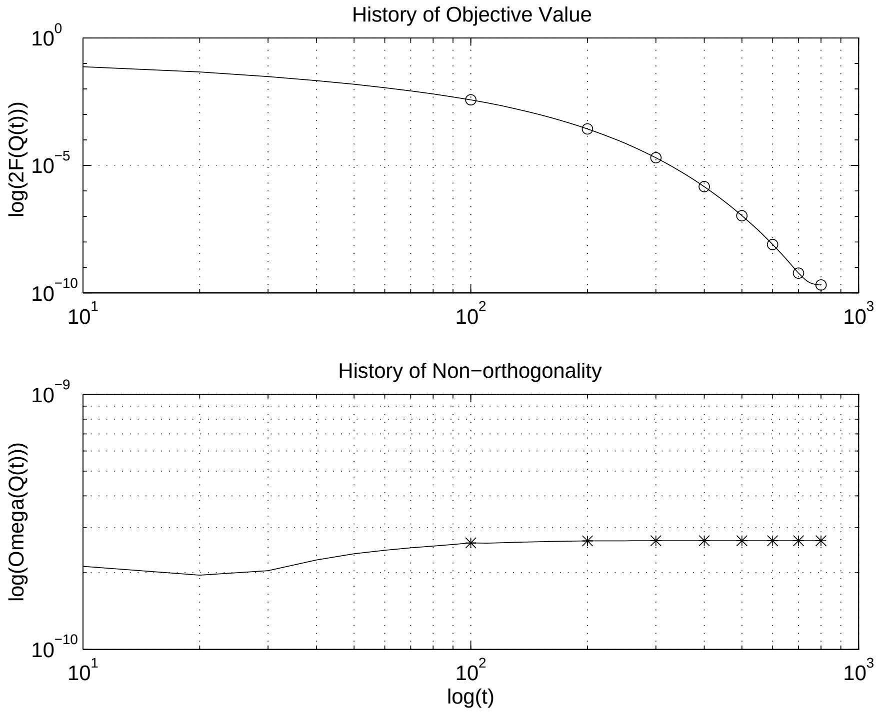

FIG. 1. *A semi-log plot of $F(Q(t))$ and $\Omega(Q(t))$ with global minimizer.*

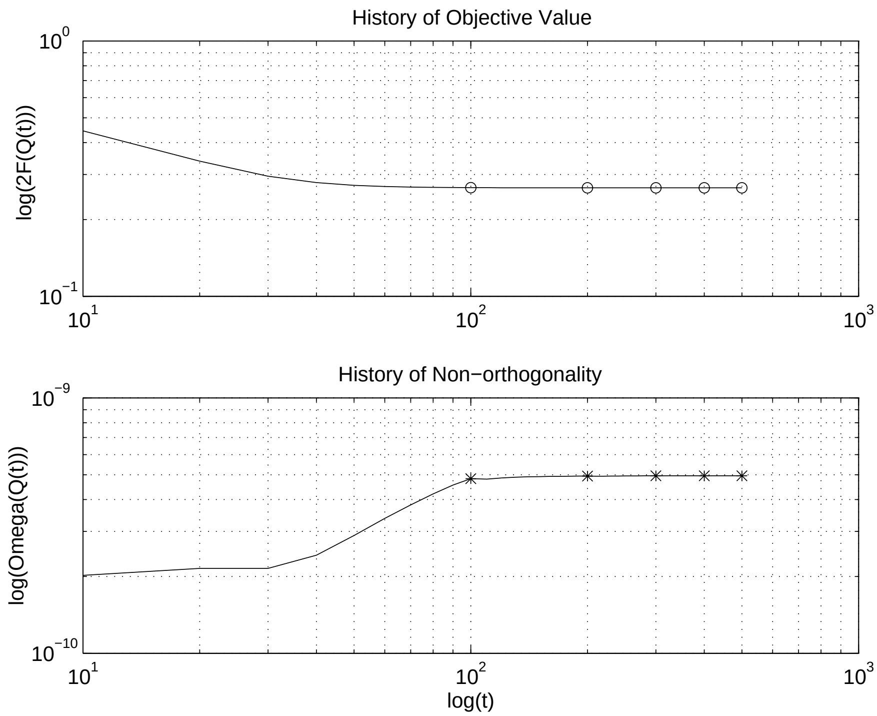

FIG. 2. *A semi-log plot of $F(Q(t))$ and $\Omega(Q(t))$ with local solution.*

REFERENCES

[1] Chu, M. T. & Driessel, K. R. (1990). The projected gradient method for least squares matrix approximations with spectral constraints, *SIAM J. Numer. Anal., 27,* 1050-1060.

[2] Chu, M. T. (1994). A list of matrix flows with applications, *Fields Institute Communications, 3,* 87-97.

[3] Chu, M. T. & Trendafilov, N. (1996). The orthogonally constrained regression revisited. Submitted for publication.

[4] Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimization.* Florida: Academic Press.

[5] Golub, G. H. & Van Loan, C. F. (1991). *Matrix Computation (2nd ed.).* Baltmore: The Johns Hopkins University Press.

[6] Gower, J. C. (1984). Multivariate analysis: Ordination, multidimensional scaling and allied topics. In Emlyn Lloyd (Ed.), *Handbook of Applicable Mathematics, Volume IV: Statistics, part B.* New York: John Wiley & Sons.

[7] Koschat, M. A. & Swayne, D. F. (1991). A weighted Procrustes criterion. *Psychometrika, 56,* 229-239.

[8] Mooijaart, A. & Commandeur, J.J.F. (1990). A general solution of the weighted orthonormal Procrustes problem. *Psychometrika, 55,* 657-663.

[9] Mulaik, S. A. (1972) *The Foundations of Factor Analysis.* New York: McGraw-Hill.

[10] Shampine, L. F. & Reichelt, M. W. (1995). *The MATLAB ODE suite*, preprint, (available from *ftp.mathworks.com:/pub/mathworks/toolbox/matlab/funfun.*)

[11] ten Berge, J. M. F. & Knol, D. L. (1984). Orthogonal rotations to maximal agreement for two or more matrices of different column orders, *Psychometrika, 49,* 49-55.

[12] Trendafilov, N. T. (1996). Iterative majorizing rotation to orthogonal simple structure solution, to appear in *Multivariate Behavioral Research.*