On the Convergence Rate of a Quasi-Newton Method for Inverse Eigenvalue Problems

Raymond H. Chan* Shu-fang Xu† Hao-min Zhou‡

April 19, 2001

Abstract

In this paper, we first note that the proof of the quadratic convergence of the quasi-Newton method as given in Friedland, Nocedal and Overton [1] is incorrect. Then we give a correct proof of the convergence.

Keywords. Newton method, inverse eigenvalue problem.

1 Introduction

Let $\{A_j\}_{j=1}^n$ be n real symmetric $n \times n$ matrices. For any vector $c = (c_1, c_2, \cdots, c_n)^T$ in $\mathbb{R}^n$, we define

$$A(c) \equiv \sum_{j=1}^{n} c_j A_j. \tag{1}$$

We denote the eigenvalues of $A(c)$ by $\{\lambda_i(c)\}_{i=1}^n$ with $\lambda_1(c) \leq \cdots \leq \lambda_n(c)$, and their corresponding normalized eigenvectors by $\{q_i(c)\}_{i=1}^n$. The inverse eigenvalue problem we consider is: Given n real numbers $\{\lambda_i^*\}_{i=1}^n$, which are ordered as $\lambda_1^* \leq \cdots \leq \lambda_n^*$, find a vector $c^* \in \mathbb{R}^n$ such that $\lambda_i(c^*) = \lambda_i^*$ for $i = 1, \cdots, n$. This problem can be posed as a problem of solving the nonlinear system

$$f(c) = 0, \tag{2}$$

where

$$f(c) = (\lambda_1(c) - \lambda_1^*, \cdots, \lambda_n(c) - \lambda_n^*)^T. \tag{3}$$

*Department of Mathematics, Chinese University of Hong Kong, Shatin, Hong Kong. Research supported in part by HK Research Grant Council grant no. CUHK316/94E and CUHK Summer Research Grant.

†Department of Mathematics, Peking University, Beijing, 100871, P. R. China.

‡Department of Mathematics, University of California, Los Angeles, CA 90095, USA. Research supported in part by ONR-N00017-96-1-0277 and NSF ASC-92-01266.

To simplify the discussion, we will assume that the given eigenvalues are distinct, i.e.

$$\lambda_1^* < \lambda_2^* < \cdots < \lambda_n^*. \tag{4}$$

and that the Jacobian $J(c^*)$ of $f(c)$ at the true solution c^* is nonsingular. In Friedland, Nocedal and Overton [1], the nonlinear system (2) is solved by different Newton-type methods. The first one considered was the Newton method and the second one was a quasi-Newton method based on the inverse power method. It was proved that both methods converge quadratically.

In this paper, we first note that the proof of the quadratic convergence of the second method given in [1] is incorrect and then we give a correct proof of the convergence.

2 The Algorithms

Since the inverse eigenvalue problem is equivalent to the problem of solving the nonlinear system (2), one can use Newton-type methods to solve it. In this section, we recall two methods discussed in [1].

We note that by using assumption (4) and results on matrix perturbation theory [3, pp.66–68], one can show that the eigenvalues and eigenvectors of $A(c)$ are differentiable functions with respect to c for c sufficiently close to c^*, see for instance [2, Theorem 2.3].

Lemma 1 *Let $A(c) \in \mathbb{R}^{n\times n}$ be an analytic symmetric matrix-valued function defined on $\mathbb{R}^n$. For any given vector $c^* \in \mathbb{R}^n$, if $A(c^*)$ has n distinct eigenvalues, then there exist a scalar $\epsilon_0 > 0$, n analytic scalar functions $\{\lambda_i(c)\}_{i=1}^n$ and n analytic vector-valued functions $\{q_i(c)\}_{i=1}^n$, such that for all c with $\|c - c^*\| < \epsilon_0$, we have*

$$A(c)q_i(c) = \lambda_i(c)q_i(c), \qquad i = 1, \ldots, n \tag{5}$$

and

$$q_i(c)^T q_i(c) = 1, \qquad i = 1, \ldots, n. \tag{6}$$

According to (6), we have

$$\frac{\partial q_i(c)^T}{\partial c_j} q_i(c) = 0, \quad 1 \le i, j \le n. \tag{7}$$

Clearly, from the definition of $A(c)$ in (1), we have $\partial A(c)/\partial c_j = A_j$, for $j = 1, \ldots, n$. Therefore, by (5) and (7), we have

$$\frac{\partial \lambda_i(c)}{\partial c_j} = q_i(c)^T \frac{\partial A(c)}{\partial c_j} q_i(c) = q_i(c)^T A_j q_i(c), \quad 1 \le i, j \le n.$$

Thus the Jacobian $J(c)$ of the function $f(c)$ defined in (3) is given by

$$[J(c)]_{i,j} = \left[\frac{\partial f(c)}{\partial c}\right]_{i,j} = q_i(c)^T A_j q_i(c), \quad 1 \le i, j \le n. \tag{8}$$

Using (1), (5) and (6), we have, for any given vector c,

$$\begin{aligned} [J(c)c]_i &= \sum_{j=1}^{n} q_i(c)^T A_j q_i(c) c_j = q_i(c)^T A(c) q_i(c) \\ &= \lambda_i(c) q_i(c)^T q_i(c) = \lambda_i(c), \quad i = 1, \ldots, n. \end{aligned} \tag{9}$$

Thus,

$$J(c)c = (\lambda_1(c), \cdots, \lambda_n(c))^T. \tag{10}$$

Recall that the Newton method for $f(c) = 0$ is defined by

$$c^{k+1} = c^k - [J(c^k)]^{-1} f(c^k), \quad k = 1, 2, \ldots.$$

By (10) and (3), this becomes

$$J(c^k) c^{k+1} = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)^T, \quad k = 1, 2, \ldots. \tag{11}$$

Thus the Newton method for solving the inverse eigenvalue problem (2) is as follows:

Method I
Choose a starting vector c^1. Then for $k = 1, 2, \ldots$, do

(i) Form $A(c^k)$ by (1).

(ii) Compute all the eigenvalues $\lambda_i(c^k)$ and normalized eigenvectors $q_i(c^k)$ of $A(c^k)$.

(iii) Stop if $\max_{i=1,\ldots,n} |\lambda_i(c^k) - \lambda_i^*|$ is small enough. Otherwise, continue.

(iv) Form $J(c^k)$ by (8).

(v) Compute the next iterate c^{k+1} by solving (11).

We note that in step (ii), the exact eigenvalues $\{\lambda_i(c^k)\}_{i=1}^n$ and eigenvectors $\{q_i(c^k)\}_{i=1}^n$ of $A(c^k)$ are computed. For a general matrix, it will require approximately $5n^3$ operations. One way to minimize the cost is to approximate the eigenvalues and eigenvectors of $A(c^k)$ instead of computing them exactly. The following quasi-Newton method given in [1] is based on using the inverse power method to find the approximate eigenvectors q_i^k to $q_i(c^k)$. In the following, we will denote the diagonal matrix $\mathrm{diag}(\lambda_1^*, \ldots, \lambda_n^*)$ by Λ^*.

Method II
Choose a starting vector c^1. Then form $A(c^1)$ by (1) and compute its exact eigenvalues λ_i^1 and the normalized eigenvectors q_i^1, $1 \le i \le n$. Then for $k = 1, 2, \cdots$, do

(i) Form $Q = [q_1^k, \ldots, q_n^k]$, the matrix with the ith column given by q_i^k.

(ii) Stop if $\|Q^T A(c^k) Q - \Lambda^*\|_F$ is small enough. Otherwise, continue.

(iii) Form J_k (cf. (8)) where

$$[J_k]_{i,j} = (q_i^k)^T A_j q_i^k, \qquad 1 \le i, j \le n. \tag{12}$$

(iv) Compute the next iterate c^{k+1} by solving (cf. (11))

$$J_k c^{k+1} = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)^T. \tag{13}$$

(v) Form $A(c^{k+1})$ by (1).

(vi) For each $i = 1, \ldots, n$, solve v_i^k in

$$(A(c^{k+1}) - \lambda_i^* I) v_i^k = q_i^k.$$

Here I is the identity matrix.

(vii) Normalize v_i^k, $i = 1, \ldots, n$, to get the next approximate eigenvectors q_i^{k+1}:

$$q_i^{k+1} = \frac{v_i^k}{\|v_i^k\|}.$$

We note that the main cost per iteration of Method II is at step (vi) where n linear systems are to be solved. However, we can first find the LU decomposition of $A(c^{k+1})$ and use it in the solution of all v_i^k, $i = 1, \ldots, n$, see [1]. In doing so, the cost of step (vi) can be reduced to approximately $3n^3$ operations. Since the eigenvalues and eigenvectors of $A(c^1)$ are computed exactly, we see that the iterates c^2 generated by Methods I and II are the same.

Here, we remark that the new iterates in both methods actually do not depend on the signs of the eigenvectors or the approximate eigenvectors, because the Jacobians do not change as the signs of the eigenvectors and approximate eigenvectors are changed, see (8) and (12). As in [1], we will ignore the choice of sign of the eigenvectors and the approximate eigenvectors in this paper too.

3 The Convergence Rate

The convergence rate of both Methods I and II has been studied in [1]. For Method I, the convergence rate is quadratic. It was also proved that the convergence rate of Method II is quadratic.

Theorem 1 *Suppose that the inverse eigenvalue problem (2) has a solution c^* and that the Jacobian matrix $J(c^*)$ is nonsingular. Then there exist scalars $\epsilon, \rho > 0$ such that if $\|c^1 - c^*\| < \epsilon$, then the iterates c^k of Method II converge quadratically to c^*, i.e.*

$$\|c^{k+1} - c^*\| \le \rho \|c^k - c^*\|^2, \qquad k = 1, 2, \ldots.$$

In the paper, the theorem was proved as follows: Let $Q = [q_1^k, \cdots, q_n^k]$ and $P = [q_1(c^*), \cdots, q_n(c^*)]$. Define X by

$$e^X = Q^T P. \tag{14}$$

Then it was claimed that X is a skew-symmetric matrix. Hence, by Corollary 3.1 in [1],

$$\|X\| \le \sigma \|Q - P\|, \tag{15}$$

where σ is a constant independent of k. Since P is a matrix of eigenvectors of $A(c^*)$ and X is skew-symmetric, then

$$e^X \Lambda^* e^{-X} = e^X \Lambda^* (e^X)^T = Q^T P \Lambda^* P^T Q = Q^T A(c^*) Q. \tag{16}$$

By expanding (16), we get

$$\Lambda^* + X\Lambda^* - \Lambda^* X = Q^T A(c^*) Q + O(\|X\|^2). \tag{17}$$

By comparing the diagonal entries of the matrices in (17), we see that

$$\lambda_i^* = (q_i^k)^T A(c^*) q_i^k + O(\|X\|^2).$$

From (12), we know that

$$[J_k c^*]_i = (q_i^k)^T A(c^*) q_i^k, \quad i = 1, \ldots n, \tag{18}$$

(cf. (9)). Hence

$$\lambda_i^* = [J_k c^*]_i + O(\|X\|^2), \quad i = 1, \ldots n.$$

By substracting it from the iteration formula (13), we thus have

$$J_k (c^{k+1} - c^*) = O(\|X\|^2). \tag{19}$$

Then by the nonsingularity assumption on $J(c^*)$ and (15), the quadratic convergence follows.

We note that the above deduction is incorrect since X is assumed to be a skew-symmetric matrix, which is not true. The reason is as follows. Since the matrix Q in Method II is computed by the one step inverse power method, it is not guaranteed to be orthogonal. Therefore $Q^T P$ in general is not an orthogonal matrix. Hence X defined by (14) may not exist. Even if it exists, it will not be a skew-symmetric matrix. Therefore, Corollary 3.1 of [1] cannot be used to derive (15). Moreover, $(e^X)^{-1} \neq (e^X)^T$ in general and therefore (16) may be incorrect. Thus we cannot obtain the expansion (17) and (19). In particular, we cannot use (19) and (15) to get the required quadratic convergence.

In the rest of the paper, we will give a proof of this quadratic convergence. We will follow the line of proof given in [1] and use the mathematical induction to prove that if c^1 is sufficiently close to c^*, then the following two inequalities hold for $k = 1, 2, \ldots$:

$$\|q_i^k - q_i(c^*)\| \le \gamma \|c^k - c^*\|, \quad i = 1, 2, \cdots, n, \tag{20}$$

and

$$\|c^{k+1} - c^*\| \le \rho \|c^k - c^*\|^2. \tag{21}$$

Here γ and ρ are constants independent of k. It is clear that (21) implies Theorem 1.

4 The Mathematical Induction

As remarked in §2, the second iterates c^2 for Methods I and II are the same, since the exact eigenvalues and eigenvectors are computed in the first iteration in both methods. Therefore (20) and (21) hold for $k = 1$.

We assume that (20) and (21) are true for the case $k-1$. We now prove that they are true for the case k. In [1, (3.57)], it has already been shown that under the induction hypothesis, (20) holds for the case k. Therefore, we only consider (21) for the case k.

Let $Q = [q_1^k, \cdots, q_n^k]$ and $P = [q_1(c^*), \cdots, q_n(c^*)]$. Instead of (14), we define

$$I + V \equiv Q^T P. \tag{22}$$

Then

$$Q^T A(c^*) Q = Q^T P \Lambda^* P^T Q = (I+V)\Lambda^*(I+V)^T = \Lambda^* + \Lambda^* V^T + V\Lambda^* + V\Lambda^* V^T.$$

Comparing the diagonal entries of the matrices in the above equation, one gets

$$(q_i^k)^T A(c^*) q_i^k = \lambda_i^* + 2\lambda_i^* [V]_{i,i} + \sum_{j=1}^n \lambda_j^* [V]_{i,j}^2, \quad 1 \le i \le n.$$

Using (18), we have

$$J_k c^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)^T + w, \tag{23}$$

where $w = \text{diag}(\Lambda^* V^T + V\Lambda^* + V\Lambda^* V^T)$, i.e.

$$[w]_i = 2\lambda_i^* [V]_{i,i} + \sum_{j=1}^n \lambda_j^* [V]_{i,j}^2, \quad 1 \le i \le n. \tag{24}$$

By taking the difference of (23) with the iteration formula (13), we get

$$J_k (c^* - c^{k+1}) = w. \tag{25}$$

Thus we only need to estimate $\|w\|$. For this, we first note that by the definition of V in (22) and the fact that P is orthogonal, we have

$$I + V + V^T + VV^T = (I+V)(I+V)^T = Q^T P P^T Q = Q^T Q.$$

Since $\{q_j^k\}_{j=1}^n$ are unit vectors, we see that the main diagonal entries of $Q^T Q$ are 1. Hence the main diagonal entries of $V + V^T + VV^T$ are zeros. Therefore, we get

$$[V]_{i,i} = -\frac{1}{2} \sum_{j=1}^n [V]_{i,j}^2, \quad 1 \le i \le n. \tag{26}$$

Putting this back into (24), we then have

$$\begin{aligned}\sum_{i=1}^{n}[w]_i^2 &\leq 2\left\{\sum_{i=1}^{n}4(\lambda_i^*)^2[V]_{i,i}^2+\sum_{i=1}^{n}(\sum_{j=1}^{n}\lambda_j^*[V]_{i,j}^2)^2\right\}\leq 4\max_{1\leq i\leq n}|\lambda_i^*|^2\sum_{i=1}^{n}(\sum_{j=1}^{n}[V]_{i,j}^2)^2\\ &\leq 4\max_{1\leq i\leq n}|\lambda_i^*|^2(\sum_{i=1}^{n}\sum_{j=1}^{n}[V]_{i,j}^2)^2\leq 4\max_{1\leq i\leq n}|\lambda_i^*|^2\|V\|_F^4.\end{aligned}$$

However, since (20) holds for k, we have

$$\begin{aligned}\|V\|_F &= \|V^T\|_F=\|P^TQ-I\|_F=\|Q-P\|_F\\ &= \left(\sum_{i=1}^{n}\|q_i^k-q_i(c^*)\|^2\right)^{1/2}\leq\gamma\sqrt{n}\|c^k-c^*\|. \end{aligned}\tag{27}$$

Thus, we get

$$\|w\|\leq 2\gamma^2 n\max_{1\leq i\leq n}|\lambda_i^*|\|c^k-c^*\|^2.$$

Therefore by the nonsingular assumption on J^* and (25), (21) for the case k follows. Hence Theorem 1 is proved.

We conclude that the convergence rate of Method II in [1] is still quadratic, even though the method is a quasi-Newton type method. Numerical experiments in [1] have already confirmed this. For Method III in [1], the matrix Q is orthogonal because it is the product of the previous iterate and the Cayley transform which is exactly orthogonal. Therefore the proof in [1] is correct.

Acknowledgments: We would like to thank Prof. Overton for the helpful discussions in the preparation of this paper.

References

[1] 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), 634–667.

[2] J. G. Sun, *Eigenvalues and Eigenvectors of a Matrix Dependent on Several Parameters*, J. Comput. Math., 3 (1985), 351–364.

[3] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.