University of Maryland College Park

Institute for Advanced Computer Studies TR–99–64
Department of Computer Science TR–4071

# On Orthogonalization in the Inverse Power Method*

G. W. Stewart†

September 1999

ABSTRACT

When the inverse power method is used to compute eigenvectors of a symmetric matrix corresponding to close eigenvalues, the computed eigenvectors may not be orthogonal. The cure for the problem is to orthogonalize the vectors using the Gram–Schmidt algorithm. In this note it is shown that the orthogonalization process does not cause the quality of the eigenvectors to deteriorate.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports` or on the web at `http://www.cs.umd.edu/~stewart/`.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. Work supported by the National Science Foundation under Grant No. 970909-8562

# On Orthogonalization in the Inverse Power Method

G. W. Stewart

ABSTRACT

When the inverse power method is used to compute eigenvectors of a symmetric matrix corresponding to close eigenvalues, the computed eigenvectors may not be orthogonal. The cure for the problem is to orthogonalize the vectors using the Gram–Schmidt algorithm. In this note it is shown that the orthogonalization process does not cause the quality of the eigenvectors to deteriorate.

The inverse power method is widely used to produce approximate eigenvectors corresponding to a set of precomputed eigenvalues of a symmetric matrix $A$. Since the method involves solutions of systems of the form $(A - \mu I)v = u$, it is especially well suited for band matrices of narrow bandwidth, which can be easily factored. The method has been implemented for tridiagonal matrices in several packages, including the Handbook codes [7], EISPACK [3], and LAPACK [1].

When the inverse power method is applied to find several eigenvectors, the vectors corresponding to close eigenvalues may not be orthogonal. The cure for this problem is to enforce orthogonality using the Gram–Schmidt algorithm. In practice this procedure has been found to work well. However, no analysis of the process has been given, and in fact a superficial analysis suggests that the orthogonalization can cause a deterioration in the quality of the solution. It is the purpose of this note to show why such a deterioration does not occur.

In the following informal analysis, we will ignore the effects of rounding error, except in the solution of the system in (1) below. We will let $\|\cdot\|$ denote the Euclidean vector norm and the spectral matrix norm. To simplify the analysis we will assume that $\|A\| = 1$.

Let $\lambda$ be an eigenvalue of $A$ with normalized right and left eigenvectors $x$ and $y$. Let $\mu$ be an approximation to $\lambda$. To compute an approximation $z$ to $x$ one chooses a vector $u$ with $\|u\| = 1$ and performs the following calculations.

$$\begin{array}{ll} 1. & \text{Solve the system } (A - \mu I)v = u. \\ 2. & z = v/\|v\|. \end{array} \qquad (1)$$

This method has been analyzed in many places (especially [6]). Specifically, is is easy to see that

$$\|v\| \geq \frac{|y^{\mathrm{T}}u|}{|\lambda - \mu|}.$$

Consequently, if $|y^{\mathrm{T}}u|$ is not pathologically small and $|\mu|$ is a good approximation to $\lambda$, then $v$ is large, and the residual

$$e = (A - \mu I)z = \frac{u}{\|v\|}$$

has norm $\|v\|^{-1}$, which is small. This has the implication that $(\mu, z)$ is an exact eigenvector of $A + E$, where $\|E\| \leq 3/\|v\|$ (see [5, p. 135]). Thus if $u$ contains a strong component along $y$ and $\lambda$ differs from $\mu$ by a few units in the last place, then the computed eigenvector is the eigenvector of a matrix differing from $A$ by terms of order of the rounding unit. A particularly satisfying aspect of this analysis, is that the decision of whether to accept $z$ depends only on the size of the vector $v$, which is known.

The analysis does not guarantee the accuracy of the approximation $z$. The standard rounding error analysis for the solution of linear systems shows that the computed vector $v$ satisfies

$$(A + F - \mu I)v = u, \tag{2}$$

where $\|F\|$ is of the order of the rounding unit. If $\hat{\lambda}$ is the eigenvalue nearest $\lambda$, then standard perturbation theory [5] shows that that the the eigenvector $x$ will be perturbed by an amount bounded by $\|E\|/|\lambda - \hat{\lambda}|$, and hence we cannot expect any greater accuracy in $z$. There is not much to be done about this loss of accuracy, since in general comparable errors would have been introduced when $A$ was rounded to fit the floating-point arithmetic of the computer in question. However, it does have an undesirable consequence.

Suppose that we use the inverse power method to compute an approximation $\hat{z}$ to the eigenvector corresponding to $\hat{\lambda}$. If $\hat{\lambda}$ is close to $\lambda$, the vector $\hat{z}$ will be inaccurate. Since there is no correlation between the errors in $z$ and $\hat{z}$, the two vectors will not be orthogonal, and the loss of orthogonality will be proportional to $|\lambda - \hat{\lambda}|^{-1}$. This is troublesome, because many numerical procedures involving symmetric matrices assume a set of orthonormal vectors.

The usual cure for this problem is to reorthogonalize the current vector against the previously computed vectors. Since the problem does not arise with a well separated eigenvalue, one only orthogonalizes with respect to the nearby eigenvalues. Specifically, suppose that we have computed approximate eigenpairs

$$(\mu_i, z_i), \qquad i = 1, 2, \ldots, k,$$

where the $\mu_i$ are close according to some tolerance and the $z_i$ are orthonormal to working accuracy. We will assume that the residuals

$$e_i = (A - \mu_i I)z_i$$

are satisfactorily small.

Suppose we now wish to add an approximation to the eigenvector corresponding to $\mu$. The procedure is to compute $z$ as in (1) and then orthogonalize $z$ as follows.

$$\begin{array}{ll} 1. & r_i = z_i^{\mathrm{T}} z, \quad i = 1, 2, \ldots, k. \\ 2. & w = z - r_1 z_1 - r_2 z_2 - \cdots - r_k z_k. \\ 3. & \tilde{x} = w / \|w\|. \end{array} \tag{3}$$

It is well known [2, 4] that if there is no severe cancellation in the computation of $w$ — which is to say that $z$ is reasonably independent of the $z_k$ — then $w$ will be orthogonal to the $z_k$ working accuracy. In what follows we will assume that $z$ is independent of the $z_k$, an assumption which is observed to be true in practice.[1]

As a practical matter, this orthogonalization procedure works quite well. However, that very fact requires an explanation. The reason is that in order to preserve reasonable orthogonality between vectors in different clusters, a fairly loose clustering criterion must be used. For example, the software mentioned above regards an eigenvalues as part of a cluster if they it is separated from its companion by $10^{-3}$. But if $\mu_1$, say, and $\mu$ differ by $10^{-3}$ then $(A - \mu I)z_1$ must be of order $10^{-3}$. It follows that the formula 2 in (3) expresses $x$ as a linear combination of vectors that may not have small residuals with respect to $\mu$. Why then does the residual of $x$ remain small?

To answer this question, it is convenient to cast the algorithm in terms of matrices. Let

$$Z = (z_1 \ \cdots \ z_k), \quad E = (e_1 \ \cdots \ e_k), \quad \text{and} \quad M = \mathrm{diag}(\mu_1, \ldots, \mu_k),$$

so that

$$AZ = ZM + E. \tag{4}$$

Then our orthogonalization algorithm can be written as follows.

$$\begin{array}{ll} 1. & r = Z^{\mathrm{T}} z. \\ 2. & w = z - Zr. \\ 3. & \tilde{z} = w / \|w\|. \end{array}$$

[1] The reader may have noticed that the method we have given here is the classical Gram–Schmidt algorithm, which is inferior to the modified Gram–Schmidt algorithm. However, for our application, in which we do not expect cancellation, the two methods give essentially the same results.

It is now easy to calculate the residual corresponding to $w$:

$$\begin{aligned}(A-\mu I)w &= (A-\mu I)(I-ZZ^{\mathrm{T}})z \\ &= (A-\mu I)z+(A-\mu I)Zr \\ &= (A-\mu I)z+(A-\mu I)Zr \\ &= (A-\mu I)z+(E+ZM-\mu Z)r \qquad \text{by (4)} \\ &= (A-\mu I)z+Er+Z(M-\mu I)r.\end{aligned} \tag{5}$$

The last expression in (5) quantifies the problem raised above. The first term $(A-\mu I)z$ is small provided $v$ in (1) is large. Since $\|r\| \le 1$ and the vectors $z_i$ are assumed to have small residuals, the term $Er$ is also small. However, the matrix $(M-\mu I)$ can be as large as the clustering tolerance. Thus this formula alone does not guarantee a small residual.

The cure for the problem is to investigate the structure of the vector $r$. Since

$$(A+F-\mu I)v = u$$

[see (2)], it follows that

$$\begin{aligned}\frac{Z^{\mathrm{T}}u}{\|v\|} &= Z^{\mathrm{T}}Az+Z^{\mathrm{T}}Fz-\mu Z^{\mathrm{T}}z \\ &= E^{\mathrm{T}}z+Z^{\mathrm{T}}Fz+(M-\mu I)Z^{\mathrm{T}}z.\end{aligned}$$

Hence

$$r = Z^{\mathrm{T}}z = (M-\mu I)^{-1}\left(\frac{Z^{\mathrm{T}}u}{\|v\|}-E^{\mathrm{T}}z-Z^{\mathrm{T}}Fz\right). \tag{6}$$

If this expression is substituted in the last expression in (5), we get

$$(A-\mu I)w = (A-\mu I)z+Er+Z\left(\frac{Z^{\mathrm{T}}u}{\|v\|}-E^{\mathrm{T}}z-Z^{\mathrm{T}}Fz\right). \tag{7}$$

This expression is wholly satisfactory. As we have noted, the first two terms in the right-hand side are small. Then quantity $Z^{\mathrm{T}}u/\|v\|$ is small because $v$ is large, and the other terms are small because $E$ and $F$ are small. Thus the residual for $w$ is small. Since we are assuming no significant cancellation in the computation of $w$, its normalization does not magnify the residual significantly.

The result depends on the multiplicative cancellation of $M-\mu I$ and $(M-\mu I)^{-1}$ when (5) and (6) are combined to give (7). The common sense of this cancellation is as follows. By (6) the lack of orthogonality between, say, $z_1$ and $z$ is inversely proportional to $|\mu_1-\mu|$. If this quantity is large (e.g., $10^{-3}$), then $z_1$ and $z$ are nearly orthogonal, and $r_1 = z_1^{\mathrm{T}}z$ must be small. Thus the $z_i$ with large residuals with respect to $\mu$ in (5) have small multipliers $r_i$.

**References**

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, second edition, 1995.

[2] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 1996.

[3] B. S. Garbow, J. M. Boyle, J. J. Dongara, and C. B. Moler. Matrix eigensystem routines-eispack guide extension. In *Lecture Notes in Computer Science*. Springer-Verlag, New York, 1977.

[4] G. W. Stewart. *Matrix Algorithms I: Basic Decompositions*. SIAM, Philadelphia, 1998.

[5] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, New York, 1990.

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

[7] J. H. Wilkinson and C. Reinsch. *Handbook for Automatic Computation. Vol. II Linear Algebra*. Springer-Verlag, New York, 1971.