An Invariant Deflation for Lower Banded Matrices

Chen Dao-qi
Department of Applied Mathematics
Zhejiang University, P.R.C.
and
Department of Applied Mathematics Statistics
State University of New York at Stony Brook
Stony Brook, New York 11794

Submitted by Richard A. Brualdi

ABSTRACT

Using an eigenvector of a complex matrix A, a unitary matrix U is constructed, so that U^HAU deflates A, and this deflation preserves some special structural properties of A, e.g., the Hessenberg form, the lower banded structure, and the symmetry (in case A is real).

In practice we often need to solve the eigenvalue (or eigensystem) problem for a given matrix. Usually we first reduce the given matrix to a special condensed form [1], such as the Hessenberg form, the tridiagonal form, or a general band form, by means of similarity transformations. After finding some eigenvalues and the corresponding eigenvectors, we use a method to deflate the condensed matrix to a lower order one. Usually the classical deflations destroy the structural properties of the matrix. Lajos László [2] gave a method which deflates real symmetric matrices and preserves the band structure. In this paper we present a generalization of this method which preserves several structural properties of the given matrix and furthermore can be applied to more general matrices.

We will make use the following basic definitions in this paper later. Let A be a given $n \times n$ complex matrix.

DEFINITION 1. A matrix A is called lower band structure with lower bandwidth l if

$$a_{ij} = 0 \qquad \text{whenever} \quad i > j + l, \quad 1 \leqslant i, j \leqslant n.$$

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/010237-09$3.00

DEFINITION 2. A matrix A is called (upper) Hessenberg if

$$a_{ij} = 0 \qquad \text{whenever} \quad i > j+1, \quad 1 \leqslant i, j \leqslant n.$$

DEFINITION 3. A matrix A is said to have a band structure with bandwidth l if

$$a_{ij} = 0 \qquad \text{when} \quad |i-j| > l.$$

We start with our first observation.

LEMMA 1. *Let $c_1, c_2, \dots, c_k$, $k > 1$, $c_k \neq 0$, be given complex numbers. Then there exists a unique negative number d and a unique complex number λ such that*

$$d\bar{c}_1 + \lambda \sum_{i=2}^{k} |c_i|^2 = 0, \tag{1}$$

$$d^2 + |\lambda|^2 \sum_{i=2}^{k} |c_i|^2 = 1. \tag{2}$$

Proof. It is obvious that Equations (1) and (2) hold for

$$d := -\left(\frac{\sum_{i=2}^{k} |c_i|^2}{\sum_{j=1}^{k} |c_j|^2} \right)^{1/2}$$

and for

$$\lambda := \frac{\bar{c}_1}{\left(\sum_{i=2}^{k} |c_i|^2 \cdot \sum_{j=1}^{k} |c_j|^2 \right)^{1/2}}.$$

The uniqueness follows from the fact that Equations (1) and (2) only have two pairs of solutions

$$d = \pm \left(\frac{\sum_{i=2}^{k} |c_i|^2}{\sum_{j=1}^{k} |c_j|^2} \right)^{1/2},$$

$$\lambda = - \frac{d\bar{c}_1}{\left(\sum_{i=2}^{k} |c_i|^2 \right)^{1/2}}$$

and only one pair of which satisfies $d < 0$. ∎

THEOREM 1. *For any unit complex k-dimensional column vector*

$$u = (u_{11}, \dots, u_{k1})^T, \qquad u_{k1} \neq 0, \quad k > 1,$$

there exists a lower Hessenberg unitary matrix $U = (u_1, \dots, u_k)$ *of order k with negative superdiagonal elements such that*

(a) $u = u_1$ *is the first column of* U,
(b) *and there are complex numbers* $\mu_2, \dots, \mu_k$ *with the property that*

$$u_{ij} = \mu_j u_{i1}, \qquad i = j, j+1, \dots, k, \quad 2 \leqslant j \leqslant k.$$

Proof. Since $u_{k1} \neq 0$, applying Lemma 1, we can find negative numbers d_i and complex numbers μ_i which satisfy

$$d_i \bar{u}_{i-1} + \mu_i \sum_{j=i}^{k} |u_{j1}|^2 = 0, \qquad i = 2, \dots, k. \tag{1'}$$

$$d_i^2 + |\mu_i|^2 \sum_{j=i}^{k} |u_{j1}|^2 = 1, \qquad i = 2, \dots, k. \tag{2'}$$

Let us define $u_1 = u$ and

$$u_i = d_i e_{i-1} + \mu_i (0, \dots, u_{i1}, \dots, u_{k1})^T, \qquad i = 2, 3, \dots, k,$$

where e_i is the ith column vector of the identity matrix I_k.

It follows from Equation (2′) that

$$\|u_j\|_2^2 = d_j^2 + |\mu_j|^2 \sum_{i=j}^{k} |u_{i1}|^2 = 1, \qquad j = 2,\dots,k,$$

and in view of Equation (1′) we get

$$u_1^H u_j = \bar{u}_{j-1\,1} d_j + \mu_j \sum_{i=j}^{k} |u_{i1}|^2 = 0 \qquad \text{for} \quad j = 2,3,\dots,k.$$

When $1 < j < s \leqslant k$, we have

$$u_j^H u_s = \bar{\mu}_j u_1^H u_s = 0.$$

Hence $U^H U = I$, or $U^H = U^{-1}$, and it is evident that U is lower Hessenberg. ■

In Theorem 1, the lower Hessenberg unitary U is constructed from the given complex column vector

$$u = (u_{11}, u_{21}, \dots, u_{k1})^T, \quad k > 1, \qquad u_{k1} \neq 0;$$

therefore we denote it by U_u or U_{u_1}.

Since in Lemma 1, d and λ are uniquely determined, the lower Hessenberg unitary matrix U of Theorem 1 is unique.

Let $0 \leqslant \varphi_1, \varphi_2, \dots, \varphi_{n-1} \leqslant \pi$, $c_i = \cos\varphi_i$, $s_i = \sin\varphi_i$, $i = 1,2,\dots,n-1$. We can easily prove that the real lower Hessenberg matrix

$$U(\varphi_1, \varphi_2, \dots, \varphi_{n-1}) = (u_{ij}),$$

with

$$u_{ij} = \begin{cases} 0 & n \geqslant j > i+1, \\ -s_j, & n \geqslant j = i+1, \\ c_{j-1} s_j \cdots s_{i-1} c_i, & n \geqslant i \geqslant j \geqslant 1, \quad c_0 = 1, \quad c_n = 1, \end{cases}$$

is a unitary matrix.

For a given real unit column vector $u=(u_{11},\dots,u_{k1})^T$, $u_{k1}>0$, we compute the c_i's and s_i's as follows:

$$\sigma_k^2 := u_{k1}^2$$

$$\sigma_s^2 := \sigma_{s+1}^2 + u_{s1}^2, \qquad s=k-1,\dots,1,$$

$$c_i := u_{i1}/\sigma_i, \qquad \sigma_i>0, \quad i=1,2,\dots,k-1,$$

$$s_i := \sigma_{i+1}/\sigma_i, \qquad i=1,2,\dots,k-1.$$

It is obvious that the real Hessenberg unitary matrix $U(\varphi_1,\dots,\varphi_{k-1})$ has negative superdiagonal elements and has u as its first column.

Note that the number of arithmetical operations for calculating the c_i's and the s_i's from the given real column vector $u=(u_{11},\dots,u_{k1})^T$, $u_{k1}>0$, is no more than $3k-2$ multiplications (including squares and divisions), $k-1$ additions, and $k-1$ square roots.

We would like to point out that for any complex lower Hessenberg unitary matrix $U_{n\times n}$ we can find two diagonal unitary matrices

$$D_1=\operatorname{diag}\left(e^{i\alpha_1},\dots,e^{i\alpha_n}\right)$$

and

$$D_2=\operatorname{diag}\left(e^{i\beta_1},\dots,e^{i\beta_n}\right)$$

such that $D_1UD_2=U(\varphi_1,\varphi_2,\dots,\varphi_{n-1})$ for some $0\leqslant\varphi_1,\dots,\varphi_{n-1}\leqslant\pi$. We omit the proof here, since it is not needed for the purpose of this paper.

LEMMA 2. *Let $u=(u_{11},\dots,u_{k1})^T$, $k>1$, $u_{k1}\neq 0$, be a unit column vector. If $U=U_u$ is the lower Hessenberg unitary matrix constructed from u, and $v=(0,\dots,0,v_s,\dots,v_k)$, $v_s\neq 0$, is the row vector which satisfies*

$$vu_1=\sum_{i=s}^{k} v_iu_{i1}=0,$$

then the first s components of the row vector vU_u are zeros.

Proof. Let

$$vU_u=(r_1,\dots,r_k);$$

then we have

$$r_j = \sum_{i=s}^{k} v_i u_{ij}.$$

When $1 \leqslant j \leqslant s$, then from Theorem 1 it follows that

$$r_j = \mu_j v u_1 = 0, \qquad j = 1, 2, \dots, s.$$

■

Using Theorem 1 and Lemma 2 we can prove the main result of our paper.

THEOREM 2. *If a given $n \times n$ matrix A has a lower band structure with lower bandwidth l, if $u = (u_{11}, \dots, u_{k1}, 0, \dots, 0)^T$, $k > 1, u_{k1} \neq 0$, is a unit eigenvector of A, and if*

$$U = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix},$$

where U_k is the lower Hessenberg unitary matrix of order k constructed from $(u_{11}, \dots, u_{k1})^T$, then

$$U^H A U = \begin{pmatrix} \lambda_1 & x & \cdots & x \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix},$$

where λ_1 is the eigenvalue of A corresponding to the eigenvector u, and $A^{(1)}$, which is of order $n-1$, has a lower band structure with lower bandwidth $\leqslant l$.

Proof. It is obvious that U is unitary. We rewrite A as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11}, A_{22} are square lower banded matrices with lower bandwidth $\leqslant l$, and are of order k and $n-k$, respectively. Then

$$U^H A U = \begin{pmatrix} U_k^H & 0 \\ 0 & I_{n-k} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

$$= \begin{pmatrix} BU_k & U_k^H A_{12} \\ A_{12} U_k & A_{22} \end{pmatrix},$$

where $B = U_k^H A_{11}$.

Since U_k^H is an upper Hessenberg matrix and A_{11} is lower banded with lower bandwidth $\leqslant l$, therefore $U_k^H A_{11} = B$ is lower banded with lower bandwidth $\leqslant l+1$.

Because the first column of U is u, the unit eigenvector of A, we have

$$U^H A u = \lambda_1 U^H u = \lambda_1 e_1,$$

where λ_1 is the eigenvalue of A corresponding to u. Therefore, we have

$$B \begin{pmatrix} u_{11} \\ \vdots \\ u_{k1} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{k \times 1}, \tag{3}$$

and

$$A_{21} \begin{pmatrix} u_{11} \\ \vdots \\ u_{k1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{(n-k) \times 1}. \tag{4}$$

That implies that each nonzero row of the matrix B, except the first one, and each nonzero row vector of the matrix A_{21} satisfies the conditions of Lemma 2. It follows from Lemma 2 that (lower bandwidth of BU_k) $\leqslant$ (lower bandwidth of B) $-1 \leqslant (l+1)-1 = l$, and that before the nonzero component there are more zero components in each nonzero row vector of $A_{21}U_k$ than in the corresponding row vector of A_{21}. ∎

Notice that when $u = e_1$, then

$$A = \begin{pmatrix} \lambda_1 & x & \cdots & x \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix},$$

so we need only consider the cases for which $k > 1$.

It is easy to see that if A is real and the known eigenvector u is also real, then we can choose U to be real, and Theorem 2 still holds for the field of real numbers.

We can get following corollaries from Theorem 2:

COROLLARY 1. *If A is real symmetric and in a banded form, then the deflation preserves all these properties.*

COROLLARY 2. *The Hessenberg form is preserved in the deflation.*

In order to carry out our deflation, we suggest following algorithm: Let $a = (a_1, \dots, a_k)$ be a real row vector, and let U be a real lower Hessenberg unitary matrix with known parameters c_i, s_i, $i = 1, \dots, n-1$. We compute $a' = aU$ by the scheme

$$\begin{aligned} r_k &:= a_k, \\ a'_k &:= c_{k-1} r_k - s_{k-1} a_{k-1}, \\ r_i &:= c_i a_i + s_i r_{i+1}, & i &= k-1, \dots, 2, \\ a'_i &:= c_{i-1} r_i - s_{i-1} a_{i-1}, & i &= k-1, \dots, 2, \\ a'_1 &:= c_1 a_1 + s_1 r_2. \end{aligned}$$

The total number of arithmetical operations for computing a' is no more than $4k - 4$ multiplications and $2k - 1$ additions. So we see that, using our algorithm, the total number of arithmetical operations for deflating a given real matrix with a known real eigenvector $u = (u_{11}, \dots, u_{k1}, 0, \dots, 0)^T$, $u_{k1} > 0$, never exceeds $8nk - 8n + 3k$ multiplications, $4nk - 2n + k$ additions, and $k - 1$ square roots.

REMARK. Using a given eigenvector to generate a unitary U such that $U^H AU$ deflates A and preserves the lower band structure, we need to compute some square roots. If instead of using a unitary matrix U we use UD for some nonsingular diagonal matrix D, then we do not need to compute any square roots. This procedure still preserves the properties we need, namely, $D^{-1}U^H AUD$ still deflates A and this deflation preserves the lower band structure. But Corollary 1 is not true for a nonunitary deflation.

If, using nonunitary deflation, it is not necessary to calculate the square roots, then the total number of multiplications to deflate a matrix with a known eigenvector is approximately three-fourths that for unitary deflation. The details of description of the nonunitary deflation scheme we omit here. Nonunitary deflation may suffer numerical instability.

The author would like to express his appreciation to Professor R. P. Tewarson for his suggestions and encouragement, and Professor Y. M. Chen for his encouragement.

REFERENCES

1 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford U.P., 1965, pp. 345–410.

2 Lajos László, A bandstructure-invariant deflation for real symmetric matrices, *Linear Algebra Appl.* 33:9–13 (1980).

Received 5 January 1982; revised 19 July 1982