

A SYMPLECTIC METHOD FOR APPROXIMATING
ALL THE EIGENVALUES OF A HAMILTONIAN MATRIX*

C. Van Loan

TR 82-494
May 1982

Department of Computer Science
Cornell University
Ithaca, New York 14853

*This report is also available through the Department of Mathematics,
University of Manchester, Manchester, England. (Numerical Analysis
Report #71.)

A Symplectic Method for Approximating
All the Eigenvalues of a Hamiltonian Matrix

C. Van Loan

Department of Computer Science
Cornell University
Ithaca, New York, 14853.

Abstract

A fast method for computing all the eigenvalues of a Hamiltonian matrix M is given. The method relies on orthogonal symplectic similarity transformations which preserve structure and have desirable numerical properties. The algorithm is about four times faster than the standard Q-R algorithm. The computed eigenvalues are shown to be the exact eigenvalues of a matrix $M+E$ where $\|E\|$ depends on the square root of the machine precision. The accuracy of a computed eigenvalue depends both on its condition and its magnitude, larger eigenvalues typically being more accurate.

1. Motivation

A real $2n$ -by- $2n$ matrix of the form

$$(1.1) \quad M = \begin{bmatrix} A^T & G \\ F & -A \end{bmatrix}$$

is called a Hamiltonian matrix if $A \in \mathbb{R}^{n \times n}$, $G^T = G \in \mathbb{R}^{n \times n}$, and $F^T = F \in \mathbb{R}^{n \times n}$. It is easy to verify that if

$$J = \begin{bmatrix} 0_n & I_n \\ -I_n & 0_n \end{bmatrix},$$

then

$$\mathcal{H} = \{M \in \mathbb{R}^{2n \times 2n} \mid J^T M J = -M^T\}$$

is precisely the set of all $2n$ -by- $2n$ real Hamiltonian matrices.

Here, I_n is the n -by- n identity.

For any matrix T let $\lambda(T)$ denote the set of its eigenvalues.

It follows that

$$(1.2) \quad M \in \mathcal{H}, \lambda \in \lambda(M) \Rightarrow -\lambda, \bar{\lambda}, -\bar{\lambda} \in \lambda(M).$$

This is because (a) M and $-M^T$ are similar and (b) the complex eigenvalues of a real matrix occur in conjugate pairs.

Hamiltonian matrices arise in several areas including control theory. For example, in the simplest kind of linear regulator problem we are asked to minimize

$$J(u) = \int_0^{\infty} [y^T y + u^T u] dt$$

subject to

$$\begin{aligned} \dot{x} &= Ax + Bu & A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times p} \\ y &= Cx & C \in \mathbb{R}^{m \times n} \end{aligned}$$

where $x(0) = x_0$ is given.

It is widely known [3] that if this system is stabilizable and detectable and we set $F = BB^T$ and $G = C^T C$ in (1.1), then

Unfortunately, M's rich Hamiltonian structure is ignored throughout MQR. This wastes computer time and storage. A way round these practical problems is to make use of symplectic matrices. $S \in \mathbb{R}^{2n \times 2n}$ is symplectic if

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}^{-1} = \begin{bmatrix} S_{22}^T & -S_{12}^T \\ -S_{21}^T & S_{11}^T \end{bmatrix} \quad S_{ij} \in \mathbb{R}^{n \times n},$$

or, equivalently, if S belongs to the set

$$\mathcal{S} = \{S \in \mathbb{R}^{2n \times 2n} \mid J^T S J = S^{-T}\}.$$

Symplectic similarity transformations preserve Hamiltonian structure:

$$(1.2) \quad S \in \mathcal{S}, M \in \mathcal{H} \Rightarrow S^{-1} M S \in \mathcal{H}:$$

This follows because

$$J^T (S^{-1} M S) J = (J^T S J)^{-1} (J^T M J) (J^T S J) = - (S^{-1} M S)^T.$$

In this paper we present an algorithm for computing the eigenvalues of a Hamiltonian matrix that relies on orthogonal symplectic transformations. The new technique is about four times faster than MQR and requires half the storage. However, these positive attributes are partially offset by a somewhat less favourable error analysis.

The organization of the paper is as follows. In Section 2 we develop Householder symplectic and Givens symplectic transformations. We then show how these computational tools can be used to zero the (2,1) block of the matrix M^2 . Why this is a constructive course of action to take is motivated in Section 3. The main algorithm and its computer implementation are detailed in Sections 4 and 5. In the final section we examine the numerical properties of the new method with a mixture of examples and analysis.

Readers already familiar with Householder and Givens transformations may find our exposition a bit lengthy, particularly if they have read [5]. However, we feel it is important to acquaint

$J(k, \theta)$ is an "ordinary" $2n$ -by- $2n$ Givens rotation that rotates in planes k and $k+n$. See [8].

As we mentioned, Householder and Givens symplectic transformations can be used to zero prescribed entries in a vector. In the Householder case we have

Algorithm H

Given k ($1 \leq k < n$) and $y, z \in \mathbb{R}^n$, the following algorithm determines $w = (w_k, \dots, w_n)^T$ such that if

$$H(k, w) \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} v \\ x \end{pmatrix}$$

then $x_i = 0$ for $i = k+1, \dots, n$.

$$\sigma := (z_k^2 + \dots + z_n^2)^{\frac{1}{2}}$$

$$w_k := z_k + \text{sign}(z_k)\sigma$$

$$w_i := z_i \quad \text{for } i = k+1, \dots, n$$

end

We point out that by interchanging the roles of y and z , Algorithm H can be used to determine $H(k, w)$ such that $v_i = 0$ for $i = k+1, \dots, n$.

While Householder symplectics can be used to zero large portions of a vector, Givens symplectics can be used to zero single entries;

Algorithm J

Given k ($1 \leq k \leq n$) and $y, z \in \mathbb{R}^n$, the following algorithm determines $c = \cos(\theta)$ and $s = \sin(\theta)$ such that if

$$J(k, \theta) \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} v \\ x \end{pmatrix}$$

then $x_k = 0$.

$$\sigma := (y_k^2 + z_k^2)^{\frac{1}{2}}$$

If $\sigma = 0$

then $c := 1$ and $s := 0$

else $c := y_k/\sigma$ and $s := z_k/\sigma$

end

This follows from (1.2) and the implication

$$N = M^2 \Rightarrow S^{-1}NS = (S^{-1}MS)^2.$$

Second, if $M \in \mathcal{H}$ and

$$\lambda(M) = \{\lambda_1, -\lambda_1, \dots, \lambda_n, -\lambda_n\},$$

then

$$\lambda(M^2) = \{\lambda_1^2, \lambda_1^2, \dots, \lambda_n^2, \lambda_n^2\}.$$

Thus, $\lambda(M)$ can be readily deduced once we compute $\mu_i = \lambda_i^2$ for $i=1, \dots, n$.

These observations form the basis of the following algorithm for computing the eigenvalues $\lambda_1, \dots, \lambda_{2n}$ of $M \in \mathcal{H}$:

Step 1.

$$\text{Form } N = \begin{bmatrix} A^T & G \\ F & -A \end{bmatrix}^2$$

Step 2.

Compute an orthogonal symplectic Q such that

$$Q^T N Q = \begin{bmatrix} H & R \\ 0 & H^T \end{bmatrix}$$

where H is upper Hessenberg ($h_{ij} = 0, i > j + 1$).

Step 3.

Use the QR algorithm to compute $\lambda(H) = \{\mu_1, \dots, \mu_n\}$.

Step 4.

For $i=1, \dots, n$ compute $\lambda_i = \sqrt{\mu_i}$ taking the square root located in the left half plane. Set $\lambda_{n+i} = -\lambda_i$ for $i=1, \dots, n$.

Step 2 is the only step in this process that requires immediate clarification. Details are given in the next section.

4. The Reduction of a Squared Hamiltonian

We now show how an orthogonal symplectic $Q \in \mathbb{R}^{2n \times 2n}$ can be determined such that

The next step is to zero v_{21} using a Givens symplectic similarity transformation. This can be achieved by applying Algorithm J with $k = 2$, $y = De_1$ and $z = Ve_1$. Let $J_1 = J(2,\theta)$ be the resulting transformation and notice that only the second row and column of D, U , and V are affected by the update $J_1 N J_1^T$. This implies that

$$N = \begin{bmatrix} D & U \\ V & D^T \end{bmatrix} := J_1 N J_1^T = \begin{matrix} & & & & x & x & x & x & & o & x & x & x \\ & & & & x & x & x & x & & x & o & x & x \\ & & & & x & x & x & x & & x & x & o & x \\ & & & & x & x & x & x & & x & x & x & o \\ & & & & o & o & o & o & & x & x & x & x \\ & & & & o & o & x & x & & x & x & x & x \\ & & & & o & x & o & x & & x & x & x & x \\ & & & & o & x & x & o & & x & x & x & x \end{matrix}$$

Next, we compute a Householder symplectic to zero d_{31} and d_{41} . In particular, by applying Algorithm H with $k = 2$, $y = Ve_1$, and $z = De_1$ we obtain $G_1 = H(2,w)$ with the property that

$$N = \begin{bmatrix} D & U \\ V & D^T \end{bmatrix} := G_1 N G_1^T = \begin{matrix} & & & & x & x & x & x & & o & x & x & x \\ & & & & x & x & x & x & & x & o & x & x \\ & & & & o & x & x & x & & x & x & o & x \\ & & & & o & x & x & x & & x & x & x & o \\ & & & & o & o & o & o & & x & x & o & o \\ & & & & o & o & x & x & & x & x & x & x \\ & & & & o & x & o & x & & x & x & x & x \\ & & & & o & x & x & o & & x & x & x & x \end{matrix}$$

Note that only rows and columns 2,3, and 4 of D, U , and V are altered by this similarity transformation.

This completes the zeroing in the first column of N . The computations we have illustrated, however, are typical of the general k -th step. Overall we have the following "square-reduction" procedure:

Algorithm SR

Given $N = \begin{bmatrix} D & U \\ V & D^T \end{bmatrix} \in \mathcal{N}^2$, the

following algorithm overwrites D with an upper Hessenberg matrix H having the property that $\lambda(N) = \lambda(H) \cup \lambda(H)$.

The first point to discuss is the squaring of M . Normally, this matrix would be represented in two n -by- n arrays -- one for A and one that is "shared" by the symmetric matrices F and G . (Of course, an n -vector is required to store the diagonal of either F or G .) An equal amount of storage is necessary for N -- an array for $A^2 + FG$ and an array for the skew-symmetric matrices $FA^T - AF$ and $A^T G - GA$. Unfortunately, a $3n^2/2$ workspace is required for the computation of N .

A way round this annoying problem is to work only implicitly with the matrix N in Algorithm SR. Instead of applying the H_k, J_k , and G_k to N , we apply them to M . Of course, the construction of these orthogonal symplectic matrices requires access to certain components of Ne_k , the k -th column of the current N . However, at any instant we can calculate these quantities from the current M via the formula $Ne_k = M(Me_k)$.

The resulting "implicit SR" algorithm is mathematically equivalent to the "explicit" version detailed in the previous section. In particular, it overwrites the original M with the Hamiltonian

$$(5.1) \quad M_0 = \begin{bmatrix} A_0^T & G_0 \\ F_0 & -A_0 \end{bmatrix} = Q^T M Q$$

where Q is defined by (4.1). It follows that

$$A_0^2{}^T + G_0 F_0 = H$$

is upper Hessenberg and that

$$F_0 A_0^T - A_0 F_0 = 0.$$

Hamiltonian matrices with the property that their square has a zero (2,1) block are called "square-reduced".

Algorithm SR (Implicit)

$$\text{Given } M = \begin{bmatrix} A^T & G \\ F & -A \end{bmatrix} \in \mathcal{H}, \text{ the}$$

$$M = \begin{bmatrix} A_{11}^T & A_{21}^T & G_{11} & G_{12} & k \\ A_{12}^T & A_{22}^T & G_{12}^T & G_{22} & n-k \\ F_{11} & F_{12} & -A_{11} & -A_{12} & k \\ F_{12}^T & F_{22} & -A_{21} & -A_{22} & n-k \\ k & n-k & k & n-k & \end{bmatrix}$$

then

$$H_k M H_k^T = \begin{bmatrix} A_{11}^T & A_{21}^T P & G_{11} & G_{12}^T P \\ PA_{12} & PA_{22}^T P & PG_{12}^T & PG_{22}^T P \\ F_{11} & F_{12} P & -A_{11} & -A_{12} P \\ PF_{12}^T & PF_{22} P & -PA_{21} & -PA_{22} P \end{bmatrix}$$

Thus, the update of M amounts to a collection of "ordinary" Householder updates. Computational details may be found in [8].

Update by a Givens symplectic similarity $J_k = J(k, \theta)$ is equally simple. Set $c = \cos(\theta)$, $s = \sin(\theta)$, $w = A_{1k}^T e_k$, $x = A_{2k} e_k$, $y = G_{1k} e_k$, and $z = F_{2k} e_k$. Since $J_k M J_k^T$ affects just the k -th row and column of A , F , and G , we need only perform the following calculations:

For $i = 1, \dots, k-1$

$$\left[\begin{array}{l} a_{ik} := cx_i + sz_i \\ a_{ki} := cw_i + sy_i \\ f_{ki} := -sx_i + cz_i \\ g_{ik} := cy_i - sw_i \end{array} \right.$$

$$a_{kk} := (c^2 - s^2)w_k + cs(y_k + z_k)$$

$$f_{kk} := c^2 z_k - s^2 y_k - 2csw_k$$

$$g_{kk} := c^2 y_k - s^2 z_k - 2csw_k$$

Essentially, the two algorithms involve the same amount of work, about 25% of that required by MQR. It should be stressed, however, that the above style of quantifying work is only approximate.

6. Numerical Properties and Examples

Suppose $\hat{\lambda}_{\text{MQR}}$ is a computed eigenvalue of M obtained by using Algorithm MQR. If t -digit base b floating point arithmetic is used, then it can be shown that

$$(6.1) \quad \hat{\lambda}_{\text{MQR}} \in \lambda(M+E)$$

where $E_{\text{MQR}} \in \mathbb{R}^{2n \times 2n}$ satisfies

$$(6.2) \quad \|E_{\text{MQR}}\|_2 \approx b^{-t} \|M\|_2.$$

This says that $\hat{\lambda}_{\text{MQR}}$ is an exact eigenvalue of a matrix relatively "near" to M , an optimum result in that the mere storage of M results in rounding errors of order $b^{-t} \|M\|_2$.

In general, if $\hat{\lambda}_{\text{MQR}}$ is the computed analog of $\lambda \in \lambda(M)$ and λ is a simple eigenvalue, then

$$(6.3) \quad |\lambda - \hat{\lambda}_{\text{MQR}}| \approx \frac{b^{-t} \|M\|_2}{s(\lambda)}$$

where the quantity $1/s(\lambda)$ is the condition of λ . (The denominator $s(\lambda)$ is the cosine of the angle between the left and right eigenvectors associated with λ , a number that can be very small.) Results (6.1) - (6.3) follow immediately from the classical analysis in [8].

Although Algorithm MQR is "perfectly stable", it has one disconcerting numerical feature. Because it ignores Hamiltonian structure the computed eigenvalues will not come in plus-minus pairs. Indeed, we have constructed examples where $n+1$ of the computed eigenvalues are situated in the open left half plane. This complicates the identification of the closed loop eigenvalues in the regulator

Example 2

$M = Q\text{diag}(D, -D)Q^T$ where $D = \text{diag}(1, 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8})$ and Q is a randomly generated orthogonal symplectic matrix. Because M is symmetric, the eigenvalues are perfectly well-conditioned, i.e., for each eigenvalue $s(\lambda) = 1$. For stable eigenvalues we have

λ	$ \hat{\lambda}_{SR} - \lambda $	$ \hat{\lambda}_{MQR} - \lambda $
-1	10^{-15}	10^{-15}
-10^{-2}	10^{-15}	10^{-16}
-10^{-4}	10^{-13}	10^{-17}
-10^{-6}	10^{-12}	10^{-16}
-10^{-8}	10^{-9}	10^{-17}

As predicted by (6.6), the accuracy of $\hat{\lambda}_{SR}$ diminishes with its magnitude.

Example 3

$M = Q\text{diag}(H, -H)Q^T$ where Q is a randomly generated orthogonal symplectic matrix and H is the 12-by-12 Frank matrix. (See [7].) M has some very ill-conditioned eigenvalues. For the four worst conditioned stable eigenvalues we have

$\approx \lambda$	$\approx s(\lambda)$	$ \hat{\lambda}_{SR} - \hat{\lambda}_{MQR} $
-0.1436	10^{-7}	10^{-7}
-0.0812	10^{-8}	10^{-6}
-0.0495	10^{-8}	10^{-6}
-0.0310	10^{-8}	10^{-7}

Thus, even if an eigenvalue is not particularly small, $\hat{\lambda}_{SR}$ can differ significantly from $\hat{\lambda}_{MQR}$ if λ is ill-conditioned.

In the remainder of this section we justify (6.4) - (6.6). To do this we must establish the following theorem concerned with the singular values of a shifted Hamiltonian. (Readers unfamiliar

$$0 = (M^2 - \hat{\lambda}_{SR}^2 I + E)x = (M - \hat{\lambda}_{SR} I)(M + \hat{\lambda}_{SR} I)x - Ex$$

Thus,

$$(6.9) \quad \|E\|_2 \geq \|Ex\|_2 = \|(M - \hat{\lambda}_{SR} I)(M + \hat{\lambda}_{SR} I)x\|_2$$

Let $\sigma_{\min}(W)$ denote the smallest singular value of a square matrix W . It is well-known that

$$\|Wx\|_2 \geq \sigma_{\min}(W) \|x\|_2$$

for any vector x . Moreover,

$$\sigma_{\min}(W) = \min_{\det(W+E)=0} \|E\|_2 .$$

Consequently, using (6.9) and the theorem we have

$$\begin{aligned} \|E\|_2 &\geq \sigma_{\min}(M - \hat{\lambda}_{SR} I) \sigma_{\min}(M + \hat{\lambda}_{SR} I) \\ &= [\sigma_{\min}(M - \hat{\lambda}_{SR} I)]^2 \end{aligned}$$

It follows that there is a matrix E_{SR} (possibly complex) satisfying

$$\|E_{SR}\|_2 = \sigma_{\min}(M - \hat{\lambda}_{SR} I) \leq \sqrt{\|E\|_2} \approx \sqrt{b^{-t}} \|M\|$$

such that $M - \hat{\lambda}_{SR} I + E_{SR}$ is singular. This establishes (6.4) and (6.5).

If $\hat{\lambda}_{SR}$ is the computed analog of λ , then from standard eigenvalue perturbation theory we have

$$(6.10) \quad |\lambda - \hat{\lambda}_{SR}| \approx \frac{\sqrt{b^{-t}} \|M\|}{s(\lambda)}$$

where $1/s(\lambda)$ is the condition of λ . On the other hand, it follows from (6.8) that

Acknowledgements

I wish to thank Ralph Byers for performing the numerical experiments and for rigorously establishing equation (6.8). I am also grateful to Ian Gladwell for reading the original manuscript and suggesting improvements.

This paper was produced while I was a visitor in the Department of Mathematics at the University of Manchester, England. Financial support for my visit was provided by the Science Research Council (Grant number GR/B/95257) and the National Science Foundation (Grant MCS80-04106).