# A new eigenvalue algorithm for unitary Hessenberg matrices via quasiseparable representations

**Yuli Eidelman**
*Tel-Aviv University*
*Joint work with I. Haimovici*

Kalamata, July, 2018

## Unitary upper Hessenberg matrices.

The $N \times N$ matrix $U$ is called *upper Hessenberg* if its entries below the first subdiagonal are zeros, i.e. $U_{ij} = 0$ for $i > j+1$. We consider here upper Hessenberg matrices which are also unitary. Also, we suppose that $U$ is unreduced, which means that the subdiagonal entries could not be $0$.

If the subdiagonal entries of the matrix $U$ are nonnegative the matrix $U$ has the representation

$$\begin{bmatrix} -\rho_1\rho_0^* & -\rho_2\mu_1\rho_0^* & -\rho_3\mu_2\mu_1\rho_0^* & \cdots & -\rho_{N-1}\mu_{N-2}\cdots\mu_1\rho_0^* & -\rho_N\mu_{N-1}\cdots\mu_1\rho_0^* \\ \mu_1 & -\rho_2\rho_1^* & -\rho_3\mu_2\rho_1^* & \cdots & -\rho_{N-1}\mu_{N-2}\cdots\mu_2\rho_1^* & -\rho_N\mu_{N-1}\cdots\mu_2\rho_1^* \\ 0 & \mu_2 & -\rho_3\rho_2^* & \cdots & -\rho_{N-1}\mu_{N-2}\cdots\mu_3\rho_2^* & -\rho_N\mu_{N-1}\cdots\mu_3\rho_2^* \\ \vdots & \ddots & \mu_3 & & \vdots & \vdots \\ \vdots & & \ddots & \ddots & -\rho_{N-1}\rho_{N-2}^* & -\rho_N\mu_{N-1}\rho_{N-2}^* \\ 0 & \cdots & \cdots & 0 & \mu_{N-1} & -\rho_N\rho_{N-1}^* \end{bmatrix},$$

where $\mu_k > 0$, $|\rho_k|^2 + \mu_k^2 = 1$ $(k = 1, \ldots, N-1)$, $\rho_0 = -1$, $|\rho_N| = 1$.

## The quasiseparable representation.

For an $N \times N$ matrix:

Quasiseparable representation:

$$A_{ij} = \begin{cases} p_i a_{i-1} \cdots a_{j+1} q_j, & 1 \le j < i \le N, \\ d_i, & 1 \le i = j \le N, \\ g_i b_{i+1} \cdots b_{j-1} h_j, & 1 \le i < j \le N. \end{cases} .$$

Here

$p_i : 1 \times r^L_{i-1}$, $q_j : r^L_j \times 1$, $a_k : r^L_k \times r^L_{k-1}$

$d_i : 1 \times 1$

$g_i : 1 \times r^U_i$, $h_j : r^U_{j-1} \times 1$, $b_k : r^U_{k-1} \times r^U_k$

are called quasiseparable generators of the matrix $A$ with orders $r^L_k, r^U_k$.

## The Hermitian matrices.

For a Hermitian matrix the diagonal entries $d(k)$ are real and the upper quasiseparable generators could be obtained from the lower ones by taking

$$g_k = q_k^*,\ h_k = p_k^*,\ b_k = a_k^* \quad k = 2, \ldots, N-1,$$

$$g_1 = q_1^*,\ h_N = p_N^*.$$

The eigenvalue computation methods for Hermitian matrices with quasiseparable representations were developed. We use essentially the results by

E., I. Haimovici, In Operator Theory: Advances and Applications (OTAA) volume in honour of Rien Kaashoek, accepted

## The basic approach.

For a unitary matrix $U$ of size $N \times N$, all its eigenvalues are on the unit circle, i.e. they are of the form

$$\lambda_k = \cos\theta_k + i \cdot \sin\theta_k, \quad \theta_k \in [0, 2\pi), \quad k = 1, \ldots, N,$$

The cosine part and respectively the sine part are the eigenvalues of the Hermitian matrices

$$A = \frac{1}{2}(U + U^*), \quad B = \frac{i}{2}(U^* - U).$$

So that if we find the eigenvalues of $A$, we readily know the eigenvalues of $U$ up to the sign of the imaginary part:

$$\lambda_k = \cos\theta_k \pm i \cdot \sqrt{1 - \cos^2\theta_k}, \quad \theta_k \in [0, 2\pi), \quad k = 1, \ldots, N.$$

## The quasiseparable generators for unitary Hessenbeg matrices.

For a unitary upper Hessenberg matrix $U$ we use the generators

$$p_k = \mu_{k-1},\ h_k = -\rho_k, \quad k = 2, \ldots, N,$$

$$g_k = \mu_k\overline{\rho_{k-1}},\ q_k = 1, \quad k = 1, \ldots, N-1,$$

$$a_k = 0,\ b_k = \mu_k, \quad k = 2, \ldots, N-1,$$

$$d_1 = \rho_1,\ d_k = -\rho_k\overline{\rho_{k-1}}, \quad k = 2, \ldots, N.$$

Similarly for $U^*$.

## The quasiseparable generators for Hermitian matrices.

For the Hermitian matrix $A = \frac{1}{2}(U + U^*)$ we can use the lower quasiseparable generators

$$p_k = \begin{bmatrix} \frac{1}{2} & -\frac{\overline{\rho_k}}{2} \end{bmatrix}, \quad k = 2, \ldots, N,$$

$$q_k = \begin{bmatrix} \mu_k & \mu_k \rho_{k-1} \end{bmatrix}^T, \quad k = 1, \ldots, N-1,$$

$$a_k = \begin{bmatrix} 0 & 0 \\ 0 & \mu_k \end{bmatrix}, \quad k = 2, \ldots, N-1,$$

$$d_1 = \Re(\rho_1), \; d_k = -\Re(\rho_{k-1}\overline{\rho_k}), \quad k = 2, \ldots, N.$$

Similarly for the Hermitian matrix $B$.

## The bisection method.

Let $\gamma_k(\lambda) = \det(A(1:k, 1:k) - \lambda I)$ be characteristic polynomials of the principal submatrices of the quasiseparable matrix $A$.

The number of sign changes for the sequence of polynomials $\gamma_k(\lambda)$ equals the number of negative elements in the sequence $D_k(\lambda)$ for the ratio of characteristic polynomials

$$D_k(\lambda) = \frac{\gamma_k(\lambda)}{\gamma_{k-1}(\lambda)}.$$

## The quasiseparable recursive relations for Hermitian matrices.

1. $D_1(\lambda) = d_1 - \lambda, \quad f_1(\lambda) = \frac{q_1 q_1^*}{D_1(\lambda)}$

2. For $k = 2, \ldots, N-1$:

$$D_k(\lambda) = d_k - \lambda - p_k f_{k-1}(\lambda) p_k^*,$$

$$z_k(\lambda) = \frac{1}{D_k(\lambda)}[q_k^* - p_k f_{k-1}(\lambda) a_k^*],$$

$$f_k(\lambda) = a_k f_{k-1}(\lambda) a_k^* + [q_k - a_k f_{k-1}(\lambda) p_k^*] z_k(\lambda).$$

3. $D_N(\lambda) = d_N - \lambda - p_N f_{N-1}(\lambda) p_N^*$. Here $f_k(\lambda)$ are auxiliary matrix valued rational functions.

## The bisection step for the cosines.

The present algorithm receives the Schur parameters $\rho(k), \mu_k,\ k = 0, \ldots, N$ of the $N \times N$ unitary upper Hessenberg matrix $U$ and a real number $\lambda$ and it computes the number $\nu$ of the eigenvalues of the matrix $A = \frac{1}{2}(U + U^*)$ which are less than $\lambda$.

1. Compute $D_1 = \Re(\rho_1) - \lambda,\ z_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},\ f_1 = z_1 z_1^* \mu_1^2 / D_1$. Set $\nu = 1$ if $D_1 < 0$ and $\nu = 0$ otherwise.

2. For $k = 2, \ldots, N-1$ set $p_k = \begin{bmatrix} \frac{1}{2} & -\frac{\overline{\rho_k}}{2} \end{bmatrix}$ and compute

$$D_k = -\Re(\rho_{k-1}\overline{\rho_k}) - \lambda - p_k f_{k-1} p_k^*,$$

$$z_k = \begin{pmatrix} 1 \\ \rho_{k-1} - \frac{1}{2}(f_{k-1}(2,1) - f_{k-1}(2,2)\rho_k) \end{pmatrix}$$

and

$$\phi_k = z_k z_k^* / D_k,$$

$$f_k = \mu_k^2 \begin{pmatrix} \phi_k(1,1) & \phi_k(1,2) \\ \phi_k(2,1) & \phi_k(2,2) + f_{k-1}(2,2) \end{pmatrix}.$$

Set $\nu = \nu + 1$ if $D_k < 0$.

3. $p_N = \begin{bmatrix} \frac{1}{2} & -\frac{\overline{\rho_N}}{2} \end{bmatrix}$ and compute

$$D_N = -\Re(\rho_{N-1}\overline{\rho_N}) - \lambda - p_N f_{N-1} p_N^*.$$

Set $\nu = \nu + 1$ if $D_N < 0$.

## The complexity.

When completely optimized, without $2 \times 2$ and $2 \times 1$ small matrices or vectors and when all what is possible is pre-computed, the complexity is $c_\nu = 16.5N$ arithmetic operations on scalars. The 0.5 comes from the fact that only if $D_k < 0$, i.e. in half of the cases, one computes $\nu = \nu + 1$.

The complexity is only 3.3 times the one for symmetric tridiagonal matrices (Wilkinson).

## The bisection procedure.

1. For $k = 1, \ldots, N$ set $L_0(k) = -1, R_0(k) = 1$. Set $U = 1$.

2. For $k = N, \ldots, 2, 1$ perform the following steps.

2.1 Set $L = \max_{1 \le j \le k} L_0(j)$ and $U = \min\{U, R_0(k)\}$.

2.2. While $U - L$ is larger than the machine precision, perform the following. 2.2.1. Set $\lambda = (L + U)/2$.

2.2.2. Use the algorithm to obtain the number $\nu$ of eigenvalues of the matrix $A$ which are less than $\lambda$.

2.2.3. If $\nu \ge k$ set $U = \lambda$, else set $L = \lambda, L_0(\nu + 1) = L$ and $R_0(\nu) = L$ if $\nu \ne 0$ and $R_0(\nu) > L$.

2.3. Set $R_0(k) = (U + L)/2$ to contain the $k^{\text{th}}$ eigenvalue.

## Finding the sign of sines.

Assume that $\alpha$ is an eigenvalue of $A$ and $\beta = \sqrt{1-\alpha^2}$.

If the multiplicity of $\alpha$ equals 2 then the numbers $\alpha \pm i\beta$ are eigenvalues of the unitary matrix $U$. This happens always for real orthogonal matrices.

If the multiplicity of $\alpha$ equals 1 then one of the numbers $\alpha + i\beta$ and $\alpha - i\beta$ is an eigenvalue of $U$.

If $\beta$ is an eigenvalue of the matrix $B$ then $\alpha + i\beta$ is an eigenvalue of $U$. If $-\beta$ is an eigenvalue of the matrix $B$ then $\alpha - i\beta$ is an eigenvalue of $U$.

It is sufficient to find the difference of the numbers $\nu$ just before and just after non-negative values of $\beta$ to find out which of $\pm\beta$ gives an eigenvalue.

## The cases.

However, we do not want to call find $\nu$ for the matrix $B$ $2N$ times and we want to determine precisely what means "just before", so that we perform first some un-expensive linear time steps and then we call it only $0.72N$ times at most.

In the rare case when both $\pm\alpha$ are simple eigenvalues of $A$ and both $\pm\beta$ are simple eigenvalues of $B$, then one of the pairs $\alpha + i\beta, -\alpha - i\beta$ or $\alpha - i\beta, -\alpha + i\beta$ are the eigenvalues of $U$. To recognize which is indeed, some eigenvector computation is used.

We find the eigenvector $v$ such that $Bv = \beta v$, we find out for which index $k_M$ we have $|v(k_M)| = \max_{1 \le k \le N} |v(k)|$ and we compute $\lambda = U(k_M, 1 : N)v/v(k_M)$.

Finding the sign of all sines takes only $1.7\%$ of overall time.

## The unitary Hessenberg eigenvector algorithm.

Fast but nonstable.

## Numerical tests.

All the numerical experiments have been performed on a computer with an i7-5820 microprocessor, 31.9 gigabytes installed memory (RAM) at 3.30GHz and another 4GB in the video card GTX, which is exploited by Matlab as well. The operating system is Windows 10, 64 bits and the machine precision is 2.2204e-16, as given by the Matlab command eps. We wrote the program using Matlab version $R2016B$.

## The errors estimate.

The formula used for plotting the errors is the average error over the $k = 1, \ldots, N$ eigenvalues of each of the $n = 20$ considered matrices of each size $N \times N$, $N$ being a power of 2 starting with $2^2$ and finishing with $2^{13} = 8,192$.

If we denote by $\lambda_T^j(s)$ the true known eigenvalue $s$ of one of the $j = 1, \ldots, 20$ considered matrices and by $\lambda^j(k)$ the eigenvalue obtained by bisection, then the formula is

$$\frac{1}{20} \sum_{j=1}^{20} \left( \frac{1}{N} \sum_{k=1}^{N} \min_{s \in S} |\lambda^j(k) - \lambda_T^j(s)| \right),$$

where the set of indexes $S$ contains those $s \in \{1, \ldots, N\}$ that have not been previously used by another $\lambda^j(k)$.

## Random data.

We build generators for random unitary Hessenberg matrices, which in principle rarely have multiple eigenvalues, since this is not even possible if the matrix is unreduced and the latter takes place whenever none of the Schur parameters $\rho_k,\ k = 1, \ldots, N-1$ is on the unit circle. We build the (sub) unit complex numbers $\rho_k,\ k = 0, \ldots, N$, such that $\rho_0 = -1$ and $|\rho_N| = 1$, and

$$\rho_k = exp(2\pi \cdot i \cdot rand) \cdot rand, \quad k = 1, \ldots, N-1,$$

where $rand$ gives a random number between 0 and 1. The subdiagonal entries of $U$ are

$$\mu_k^2 = 1 - |\rho_k|^2, \quad k = 1, \ldots, N-1.$$

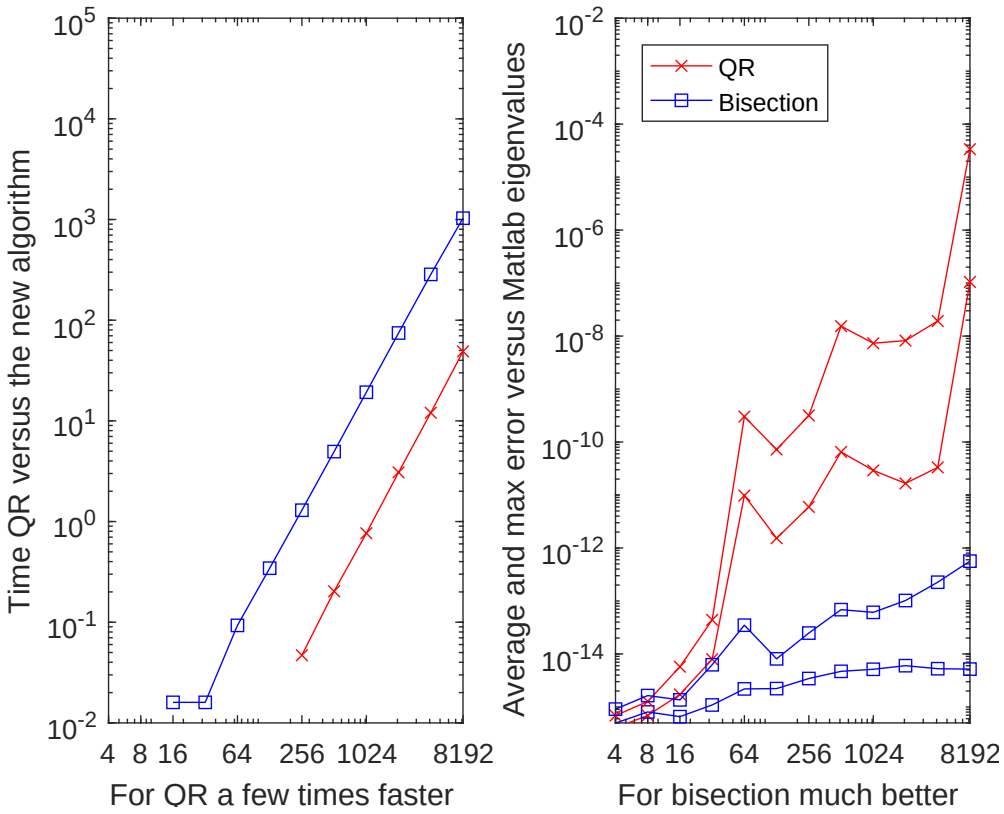
In this way any possible unitary upper Hessenberg unreduced matrix could be obtained.

## Comparison with implicit QR.

The implicit QR developed in Gragg, 1986 and Wang, Gragg, 2002 and presented in Vandebril, Van Barel, Mastronardi, Matrix computations and semiseparable matrices, Vol. II, Eigenvalue and singular value methods, 2008.

After we find the quasiseparable generators of the matrix $U$, we build the whole unitary upper Hessenberg matrix $U$ out of its generators and we use the Matlab function $eig(U)$ in order to find the Matlab eigenvalues, which we consider to be the golden, true ones.

# Comparison with implicit QR, the figure



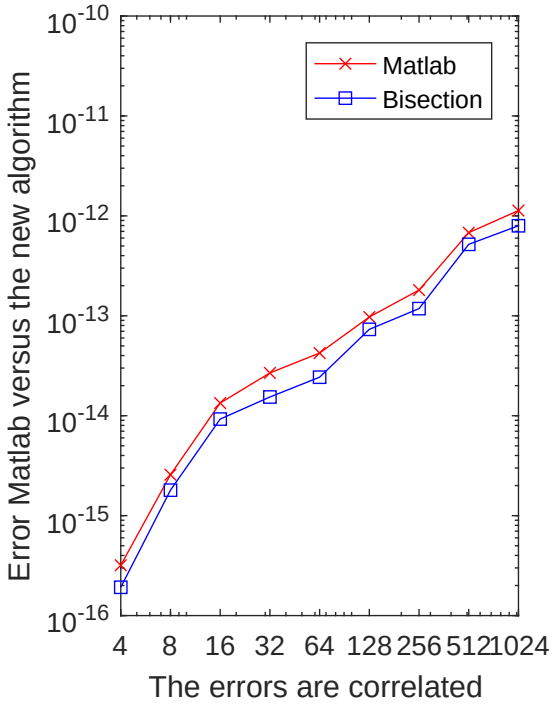
## Comparison with implicit QR.

Optimized implicit QR works much faster than the bisection algorithm presented here and it would be for sure faster, even without any optimization. However, surprisingly again, the bisection is much more exact. Among all the matrices that we checked, the largest error of the bisection, for the worst ever eigenvalue has been $4e-13$, while for implicit QR $5e-07$ and errors of $5e-08$ are very common. In general, for sizes 16 to 32 bisection is only 5 times more exact, while for 64 up to 256 it gives 3 correct decimal digits more than QR for the average error, and for larger sizes it is $10^4$ times more correct and in some experiments with matrices up to the size $2,048$, $10^5$ times. For one matrix of size $8,192$ QR gave an error of $3e-03$ as its worst error, while the worst error ever of bisection has been better by 8 decimal digits. Moreover, the error of QR does not increase smoothly in proportion with the size of the matrix.

## Eigenvalues known in advance.

In Ammar, Gragg, Reichel, 1991 the Schur parameters of a unitary upper Hessenberg unreduced matrix are built from eigenvalues which are given a priori. We implemented their method and we checked our algorithm and Matlab eig() function results when compared to the true eigenvalues, which are known in advance. Since the Schur parameters are built in $O(N^2)$ operations and for checking Matlab one must also build the whole matrix $U$ the results contained larger errors than when comparing our results to those of Matlab. What is important is that there exists a strong correlation between us and Matlab in all the eigenvalues.

# Eigenvalues known in advance.




Our implementation of the $O(N^2)$ function that builds a unitary Hessenberg matrix out of eigenvalues known in advance as in the paper of Ammar, Gragg and Reichel is not good enough and building the whole matrix for the application of Matlab function eig() also introduces errors. But we can see that bisection and Matlab find most of the time the same eigenvalues.