INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 1, Number 2, Pages 147–156

NEUTRALLY STABLE FIXED POINTS OF THE QR ALGORITHM

DAVID M. DAY AND ANDREW D. HWANG

(Communicated by Pavel Bochev)

Abstract. Practical QR algorithm for the real unsymmetric algebraic eigenvalue problem is considered. The global convergence of shifted QR algorithm in finite precision arithmetic is addressed based on a model of the dynamics of QR algorithm in a neighborhood of an unreduced Hessenberg fixed point. The QR algorithm fails at a "stable" unreduced fixed point. Prior analyses have either determined some unstable unreduced Hessenberg fixed points or have addressed stability to perturbations of the reduced Hessenberg fixed points. The model states that sufficient criteria for stability (e.g. failure) in finite precision arithmetic are that a fixed point be neutrally stable both with respect to perturbations that are constrained to the orthogonal similarity class and to general perturbations from the full matrix space. The theoretical analysis presented herein shows that at an arbitrary unreduced fixed point "most" of the eigenvalues of the Jacobian(s) are of unit modulus. A framework for the analysis of special cases is developed that also sheds some light on the robustness of the QR algorithm.

Key Words. Unsymmetric eigenvalue problem, QR algorithm, unreduced fixed point

1. Background

QR iteration is the standard method for computing the eigenvalues of an unsymmetric matrix [13, 16, 9, 2]. The global convergence properties of *un*shifted QR iteration are well established [14, 5]. For the shifted QR iteration there is no proof of convergence, and yet in practice failure is extremely rare.

A brief review of QR iteration follows. Please see [9] for a comprehensive discussion. The matrix $A = [a_{i,j}]$ has lower bandwidth k if $i > j + k$ implies that $a_{i,j} = 0$. A matrix with unit lower bandwidth is called an (*upper*) *Hessenberg* matrix. Any square matrix is orthogonally similar to a Hessenberg matrix.

In the eigenvalue problem, a given matrix is first reduced by orthogonal similarity transformations to Hessenberg form. A Hessenberg matrix is unreduced if no entry on the first subdiagonal vanishes. QR iteration is applied to the irreducible diagonal blocks consecutively.

Shifted QR iteration from H_0 with shift function $p(\cdot)$ is defined by

$$p(B_m) = Q_m R_m \qquad \text{and} \quad p(B_{m+1}) = R_m Q_m.$$

Received by the editors January 10, 2004.

2000 *Mathematics Subject Classification.* 65F15 (Primary), 65P40 (Secondary).

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

The matrices Q_m and R_m are orthogonal and upper triangular respectively. In the unshifted iteration, $p(x) = x$. QR iteration preserves the lower bandwidth of B.

The goal of QR iteration is to reduce a matrix by a sequence of orthogonal similarity transformations to a block upper triangular matrix, with one by one and two by two diagonal blocks. The decomposition is called a real Schur form.

QR iteration defines a matrix valued function whose singularities are the reduced Hessenberg matrices. In [3, 1]. the unshifted QR algorithm is viewed as a fixed-point iteration on the flag manifold, and the stability properties of the *reduced* fixed points are studied. Unreduced Hessenberg fixed points are always degenerate critical points (not obvious, see for example Theorem E), and are not structurally stable.

Though QR iteration has unreduced fixed points, in the cases considered for example in [4], convergence at an unreduced fixed point is nice in floating point arithmetic because the unreduced fixed points are strongly repelling. The observed robustness of the shifted iteration in finite precision arithmetic is due in part to the scarcity of "stable" fixed points.

Consider for example $f(x) = x + x^2 - x^3$. The fixed point zero is not strongly repelling and $f^k(x) = x + kx^2 + O(x^3)$. If x is slightly larger than the square root of the machine precision, then the number of iterations required for convergence (to $x = 1$) is inversely proportional to the machine precision.

QR iteration fails if the number of iterations to decouple a given matrix exceeds a maximum value. Herein the dynamics of QR iteration near *un*reduced Hessenberg fixed points in finite precision arithmetic are studied. The use of finite precision arithmetic introduces perturbations. At "unstable" fixed points, iteration amplifies the perturbations and the iterates escape the fixed point.

The convergence properties of the shifted QR iteration depend on the shift strategy (the map from the Hessenberg matrix and the iteration number to the shift polynomial). The implementations [13, 16, 2] all have evolved subtly different shift strategies in the attempt to enhance the convergence properties. The present study develops general results that apply to any shift strategy.

The current work relies on the first author's study of the 4×4 Hamiltonian fixed points of shifted QR iteration as implemented in [16] or [2]. The connection between integrable flows and the QR algorithm is well known (see [12], page 59 or [17, 7]). The elementary explanations in [8] are inherently lengthy and provide only a relatively limited insight into the problem at hand.

The basis for our analysis is an **empirical model** of the dynamics of the QR algorithm implemented in finite precision arithmetic and applied to matrices near to a fixed point. The dynamics of the QR algorithm in exact arithmetic is much more complicated. For instance, the stability on center manifolds must be addressed. The empirical model was derived in [8], and is the simplest model whose predictions coincide with all the observations known to the first author.

In finite precision arithmetic, QR iteration is backward stable [9] but forward unstable [11]. Consecutive QR iterates are nearly orthogonally similar, to within machine precision, but the computed iterate is not necessarily near to the iterate determined in exact arithmetic. The term orthogonal similarity class refers to the set of Hessenberg matrices orthogonally similar to a given matrix. The derivative along the orthogonal similarity class (which is low dimensional and tractable) predicts the dynamics in many situations, but not near to the unreduced fixed points [8]. Suppose that QR iteration is applied to a n by n matrix. There the difference between consecutive iterates is transverse to the orthogonal similarity class, and the derivative over matrix space, a map from R^{n^2} to R^{n^2}, also influences the dynamics near to unreduced fixed points.

Our empirical criteria is the following. In QR iteration applied to an $n \times n$ matrix, if the derivative along the orthogonal similarity has no eigenvalues outside the closed unit disk, and furthermore the derivative over matrix space, a map from R^{n^2} to R^{n^2}, also has no eigenvalues outside the closed unit disk, then QR iteration in finite precision arithmetic will fail for matrices sufficiently near to the exact fixed point. Matrices satisfying the stringent criteria for the implementations [16, 2] are described in [8].

1.1. Summary. The theoretical analysis is presented for use in selecting shift strategies the enhance the convergence properties of the shifted QR iteration. some general results with short proofs about the spectrum of the Jacobian at a fixed point. The results make tractable the problem of checking whether the empirical criteria are satisfied both for the map within the orthogonal similarity class in §2 and for the map over matrix space in §6. Low dimensional problems are tractable. The orthogonal similarity class is low dimensional. In general maps over matrices are much less tractable. Here we prove a result that the degenerate manifold has some special structure. Theorem E proves that over matrix space, the Jacobian has at least $\frac{1}{2}n(n+1)$ eigenvalues of unit modulus. The complimentary space is low dimensional, making the problem tractable.

Prerequisite results on unshifted QR iteration are also established. QR iteration restricted to the orthogonal similarity class is considered in §2, and Theorem A shows that for unshifted QR, the derivative along the orthogonal similarity class at any unreduced fixed point is an orthogonal matrix. In §4, the derivative of the unshifted QR iteration as a transformation of matrices is analyzed. Theorem B shows that over matrix space, all of the eigenvalues of the Jacobian are of unit modulus. In other words, the stringent empirical criteria for a fixed point to cause the QR iteration to fail in finite precision arithmetic are automatically satisfied by *all* fixed points of unshifted QR iteration.

1.2. Notation and Preliminaries. The space of $n \times n$ real matrices is denoted $\mathbf{M}_n$, and may be identified with $\mathbf{R}^{n^2}$ under a fixed (but arbitrary) indexing of entries. The Frobenius inner product on $\mathbf{M}_n$, defined by $\langle X, Y \rangle = \mathrm{tr}(X^\top Y)$, corresponds to the Euclidean inner product on $\mathbf{R}^{n^2}$. Denote by $\mathbf{Up}(n)$, $\mathbf{Sym}(n)$, and $\mathbf{Skew}(n)$ the spaces of $n \times n$ upper triangular, symmetric, and skew-symmetric real matrices. Each of these subspaces is a closed submanifold of $\mathbf{M}_n$, and each is canonically identified with its own tangent space at the zero matrix.

The set of invertible $n \times n$ matrices with real entries is denoted $GL(n)$, and the orthogonal group $\mathbf{O}_n$ is the set of elements of $GL(n)$ that preserve the Euclidean inner product on $\mathbf{R}^n$. As is well-known, $\mathbf{O}_n$ is a compact Lie subgroup of $GL(n)$ (itself an open subset of $\mathbf{M}_n$), whose tangent space at I_n is $\mathbf{Skew}(n)$. If $A \in \mathbf{M}_n$, then the *adjoint orbit* of A is the set $\{Q^\top A Q \mid Q \in \mathbf{O}_n\}$ of conjugates of A by orthogonal matrices. The remarks above imply that every adjoint orbit is a compact subset of $\mathbf{M}_n$.

Generally, Q denotes an orthogonal matrix, R and U denote upper triangular matrices (R has non-negative diagonal entries, U is general), and S denotes a skew-symmetric matrix.

2. An Equivalent Vector Iteration

QR iteration is applied to unreduced Hessenberg matrices. The unreduced Hessenberg fixed points of unshifted QR iteration in $Gl(n)$ for $n > 2$ are multiples of orthogonal matrices [15].

Questions about the dynamics of QR iteration may be addressed through an equivalent vector iteration (see [5]).

QR iteration is equivalent to a vector iteration, $f(\cdot)$, (the power method) [6]. Next the derivative Df of the QR iteration map along the orthogonal similarity class is reviewed. Such expressions are known [3]. We evaluate Df at an unreduced fixed point with unit lower bandwidth, and study the sensitivity of the equivalence between the vector and matrix iterations.

For a given $n \times n$ matrix C, one defines the function $\Gamma(v, C)$ from a unit n-vector v as follows: Let $P(v)$ denote the n by n matrix whose ith column is $C^{i-1}v$:

$$P(v) = [v, Cv, \ldots, C^{n-1}v] \tag{2.1}$$

Next decompose $P(v) = VR$ as a product of an orthogonal matrix and an upper triangular matrix with nonnegative diagonal entries. Finally

$$\Gamma(v, C) = V^{\mathsf{T}}CV \tag{2.2}$$

The function $\Gamma(\cdot, C)$ is well-defined and by [6] maps onto the set of matrices with unit lower bandwidth that are orthogonally similar to C. If R is singular, then $\Gamma(v, C)$ is reduced,

Unshifted QR iteration is semiconjugate to a power method [5]. For

$$f(v) = Cv/\|Cv\|, \tag{2.3}$$

if $\Gamma(v, C) = B = QR$, then $\Gamma(f(v), C) = Q^{\mathsf{T}}BQ$ and the following diagram commutes:

$$\begin{array}{ccc} B & \xrightarrow{\text{QR}} & Q^{\mathsf{T}}BQ \\ \uparrow & & \uparrow \\ \{v\} & \xrightarrow{Cq} & \{f(v)\} \end{array}$$

In exact arithmetic, $D_v f$ governs the dynamics of QR iteration.

Theorem A. *For unshifted QR iteration the vector iteration $f(\cdot)$ at a point x that corresponds to an unreduced fixed point $C = T\gamma$ of the unshifted QR algorithm, has derivative $D_v f$, that is orthogonal $D_v f = \pm T|_{x^\perp}$.*

Proof. One may show that $D_x f(x) = (I - ff^T)C/\|Cx\|$. If B is a fixed point, then by [15], $C = T\gamma$ is a multiple, γ, of an orthogonal matrix T. The domain of $f(.)$ is $n-1$ dimensional. The tangent space to the domain at the unit vector x is $x^\perp = \{h : x^h = 0\}$. At first glance,

$$D_v f = \pm(I - ff^T)T|_{x^\perp},$$

but that $x^T h = 0$ implies $f(x)^T Tx = 0$. □

Remark 2.1 An unreduced fixed point of QR iteration does not correspond to a unreduced fixed point of an equivalent vector iteration. To interpret the derivative of the equivalent vector iteration at a unreduced fixed point of QR iteration, the tangent space of the vector iteration must be lifted up to the tangent space of QR iteration.

3. The Adjoint Action and (Skew-)Symmetric Matrices

This section is devoted to a proof of the following more-or-less standard fact:

Proposition 3.1. *Let* $Q \in \mathbf{O}_n$*. The linear transformation* $\mathrm{Ad}_Q : \mathbf{M}_n \to \mathbf{M}_n$ *defined by* $\mathrm{Ad}_Q X = Q^\mathsf{T} X Q$ *is orthogonal with respect to the Frobenius inner product, and preserves the spaces* $\mathbf{Skew}(n)$ *and* $\mathbf{Sym}(n)$*.*

Note that while $Q \in \mathbf{M}_n$, Ad_Q acts on a space of dimension n^2. In words, conjugation by an orthogonal matrix is an orthogonal transformation on the *space of matrices*. The proof is broken into three lemmas.

Lemma 3.2. *If* Q *is orthogonal, then* Ad_Q *preserves the Frobenius inner product.*

Proof. If X and Y are $n \times n$ matrices, then

$$\begin{aligned}\langle \mathrm{Ad}_Q X\,, \mathrm{Ad}_Q Y\rangle &= \mathrm{tr}\big((Q^\mathsf{T} X Q)^\mathsf{T} Q^\mathsf{T} Y Q\big)\\ &= \mathrm{tr}\big((Q^\mathsf{T} X^\mathsf{T})(YQ)\big) = \mathrm{tr}\big((YQ)(Q^\mathsf{T} X^\mathsf{T})\big)\\ &= \mathrm{tr}(X^\mathsf{T} Y) = \langle X\,, Y\rangle.\end{aligned}$$

□

The maps skew and sym, projectors from $\mathbf{M}_n$ to the space of (skew-)symmetric matrices, are given by the polarization identities

$$\mathsf{sym}\, A = \tfrac{1}{2}(A + A^\mathsf{T}) \qquad \text{and} \quad \mathsf{skew}\, A = \tfrac{1}{2}(A - A^\mathsf{T}).$$

The image of each map is the kernel of the other, and a short calculation shows that the spaces of symmetric and skew-symmetric matrices are perpendicular under the Frobenius norm. Further, conjugation by an orthogonal matrix commutes with sym and skew:

Lemma 3.3. *If* Q *is orthogonal, then* $\mathrm{Ad}_Q(\mathsf{sym} X) = \mathsf{sym}(\mathrm{Ad}_Q X)$ *for all* X*, and similarly for* skew*.*

Proof. If $X \in \mathbf{M}_n$, then

$$\mathrm{Ad}_Q(X^\mathsf{T}) = Q^\mathsf{T} X^\mathsf{T} Q = (Q^\mathsf{T} X Q)^\mathsf{T} = (\mathrm{Ad}_Q X)^\mathsf{T}.$$

Since Ad_Q commutes with transposition, it commutes with the (skew-)symmetric projector. □

Lemma 3.4. *If* Q *is orthogonal, then* Ad_Q *preserves the spaces of symmetric and skew-symmetric matrices, and therefore defines an orthogonal transformation on each.*

Proof. Clearly, a matrix X is symmetric iff $\mathsf{skew} X = 0$. Lemma 3.3 implies that $\mathsf{skew}(\mathrm{Ad}_Q X) = \mathrm{Ad}_Q(\mathsf{skew} X)$, so X is symmetric iff $\mathrm{Ad}_Q X$ is symmetric. The same argument works if X is skew-symmetric. □

This completes the proof of Proposition 3.1.

4. Eigenvalues of the Derivative of Unshifted QR Iteration

Next we will show that *any* fixed point of the unshifted QR iteration automatically satisfies the criteria for stability in finite precision arithmetic (see Section 1).

Unshifted QR iteration is the mapping $\mathsf{F} : Gl(n) \to Gl(n)$ defined by

$$\mathsf{F}(QR) = RQ = Q^\mathsf{T}(QR)Q = \mathrm{Ad}_Q(QR).$$

A *fixed point* of F is a matrix H such that $F(H) = H$. A fixed point that is also Hessenberg must be a scalar multiple of an orthogonal matrix: $R = \gamma I_n$.

First we establish the differentiability of F. The singular value decomposition of an invertible matrix H gives a lower bound on the distance from H to the set of singular matrices: The distance is the smallest absolute value of a singular value. In a sufficiently small ball about H, the Gram-Schmidt factorization is given by real-analytic functions of the entries of H—rational functions, and square roots with radicand bounded away from 0. It follows that the QR iteration map F is real-analytic in the standard coordinates on $\mathbf{M}_n$.

To compute the derivative of F at $H_0 = Q_0 R_0$, consider a path

$$H(t) = H_0 + \dot{H}_0 t + O(t^2) = Q_0\big(I_n + St + O(t^2)\big)R_0\big(I_n + Ut + O(t^2)\big)$$

through H_0 with velocity $\dot{H}_0$, and differentiate:

$$D\mathsf{F}(H_0)(\dot{H}_0) = \frac{d}{dt}\Big|_{t=0}\mathsf{F}\big(H(t)\big) = \frac{d}{dt}\Big|_{t=0}R(t)Q(t).$$

Short calculations show that

$$\text{(4.1)} \qquad \dot{H}_0 = Q_0(SR_0 + R_0 U), \qquad D\mathsf{F}(H_0)(\dot{H}_0) = R_0(Q_0 S + UQ_0).$$

Recall that at a fixed point, $R_0 = \gamma I_n$; assume without loss of generality that $\gamma = 1$. With this notation, the derivative at a fixed point becomes

$$\text{(4.2)} \qquad D\mathsf{F}(Q_0)\big(Q_0(S+U)\big) = Q_0 S + UQ_0 = Q_0(S + \mathrm{Ad}_{Q_0} U).$$

The derivative mapping (which is not orthogonal) bears a strong formal resemblance to the mappings considered in Section 3. In detail, we have a direct sum decomposition $\mathbf{M}_n = \mathbf{Skew}(n) \oplus \mathbf{Up}(n)$ and a map that acts by $\mathrm{Id} + \mathrm{Ad}_Q$. The technical snag is that the summand $\mathrm{Ad}_Q U$ is not upper triangular; the map does not respect the decomposition. However, the same formal definition, regarded as a function of a skew-symmetric matrix S and a *symmetric* matrix X, is both orthogonal and similar to the map of interest, as we now show.

Theorem B. *Let Q be an $n \times n$ orthogonal matrix. The linear transformation $f_Q(S+U) = S + Q^{\mathsf{T}}UQ = S + \mathrm{Ad}_Q U$ is similar to an orthogonal transformation.*

Proof. Lemmas 3.2 and 3.4 imply that if S is skew-symmetric and X is symmetric, then the mapping $A : S + X \mapsto S + \mathrm{Ad}_Q X$ is an orthogonal transformation on $\mathbf{M}_n = \mathbf{Skew}(n) \oplus \mathbf{Sym}(n)$. Now, let U be upper triangular, and consider the isomorphism

$$P : S + U \in \mathbf{Skew}(n) \oplus \mathbf{Up}(n) \mapsto S + \mathsf{sym}U \in \mathbf{Skew}(n) \oplus \mathbf{Sym}(n)$$

that symmetrizes the triangular summand. Since sym commutes with Ad_Q by Lemma 3.3, $P^{-1}AP(S+U) = S + \mathrm{Ad}_Q U$. Thus f_Q is therefore similar to the orthogonal transformation A, see Figure 4.1. □

5. Shifted QR Iteration

The practical QR iteration is more complicated than the unshifted iteration. Selected computational details are reviewed here, and some theoretical results are strengthened.

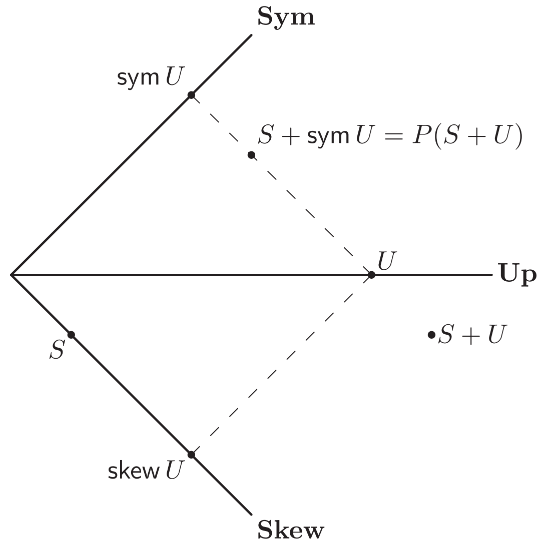

FIGURE 4.1. The (skew-)symmetrization projectors, and the isomorphism P.

Schur form. A matrix is unitarily similar to at least one complex upper triangular matrix. A real matrix is orthogonally similar to a real quasi-upper triangular matrix with 1×1 blocks corresponding to real eigenvalues and 2×2 blocks corresponding to complex conjugate pairs of eigenvalues. The matrices are also called Schur forms.

Definition 5.1 A matrix $A \in \mathbf{M}_n$ is *derogatory* if at least one eigenvalue of A has geometric multiplicity greater than one. Otherwise, A is *nonderogatory.*

Remark 5.2 An unreduced Hessenberg matrix is nonderogatory.

Lemma 5.3. *$A\mathbf{M}_n$ is nonderogatory if and only if every matrix that commutes with A is a polynomial in A.*

Proof. See [10], Corollary 4.4.18, pages 275–276. □

Definition 5.4 The *centralizer* of $A \in \mathbf{M}_n$ is the set $Z(A) \subset \mathbf{M}_n$ of matrices that commute with A.

Adjoint orbits. The set of unreduced Hessenberg matrices orthogonally similar to a given matrix is a smooth manifold. There is a smooth map $G(,)$ from real projective space onto the orthogonal similarity class. Define a map G from $(\mathbf{RP}^{n-1}, \mathbf{M}_n)$ to $\mathbf{M}_n$ by $[x, Sx, S^2x, \ldots] = QR$, Q in $\mathbf{O}_n$, R in $\mathbf{Up}(n)$,

$$SQ = QH, \tag{5.1}$$

and $G(x, S) = H$. The following facts are well known in numerical analysis: $G(\cdot, S)$ is onto the orthogonal similarity class (this is called the Implicit Q theorem), H is Hessenberg, and furthermore R is nonsingular if and only if H is unreduced.

Next we show that $G(\cdot, S)$ is one to one modulo $Z(S) \cap \mathbf{O}_n$.

Theorem C. *Suppose S is nonderogatory and $G(x, S)$ is unreduced. If $G(x, S) = G(y, S)$, then there exists a polynomial $p()$ of degree $< n$ such that $p(S)x = y$ and $p(S) \in \mathbf{O}_n$.*

Proof. To fix notation, let $[y, Sy, ...] = PU$, P in $\mathbf{O}_n$, U in $\mathbf{Up}(n)$, and $SP = PH$ or equivalently

$$P^{\mathsf{T}} S = H P^{\mathsf{T}} \tag{5.2}$$

Substitute equations (1) and (2) to find that QP^{T} is in the centralizer of S: $SQP^{\mathsf{T}} = QHP^{\mathsf{T}} = QP^{\mathsf{T}}S$. By Lemma 5.3 there is a polynomial $p()$ of degree $< n$ so that $p(S) = QP^{\mathsf{T}}$. □

Shifts. The fundamental theorem of algebra states that a polynomial factors into a product of translations or shifts. In QR iteration a sequence of shift polynomials is equivalent to one shift by the product of the shift polynomials.

Convergence. A sequence of QR iterates determined by a shift strategy has at least one convergent subsequence, by compactness. The subsequence corresponds to QR iteration from the same matrix with an induced shift strategy that allows shifts of variable degree, and the limit point is a fixed point. For matrices with distinct eigenvalues, QR iteration is quadratically convergent, see [18].

If the shift strategy detects that the limit point is reduced, then the limit point is an attracting fixed point. We will work hard to show that the unreduced fixed points are never attracting.

Shifts revisited. QR iteration shift strategies, the reasonable ones at least, have two invariance properties. For any matrix, there is a sequence of shifts. If the matrix is translated, then the shifts are also translated. If the matrix is multiplied by a scalar, then the shifts are also multiplied by a scalar.

Next a result from [15] is strengthened.

For $n > 2$, an irreducible matrix with unit lower bandwidth is nonderogatory (eigenvalues have unit geometric multiplicity). The matrix $B = QR$ is invariant under QR iteration if $QB = BQ$. For $n > 2$, an irreducible $n \times n$ matrix B with unit lower bandwidth is invariant under unshifted QR iteration if and only if $B = Qr_{1,1}$ where $R = [r_{i,j}]$, see [15].

Theorem D. *If an unreduced Hessenberg H in $\mathbf{M}_n$ is a cyclic point under QR with some shift strategy, i.e. there exists a shift polynomial such that $p(H) = QR$ and $[H, Q] = 0$, then R must be a multiple of the identity.*

Proof. $[H, R] = 0$ and the Lemma 5.3 imply that there exists a polynomial $q()$ of degree less than n such that $q(H) = R$. The only possibility is for $q()$ to be constant. □

Remark 5.5 A shift polynomial that differs from the characteristic polynomial only in the constant term always generates a fixed point.

Fixed Points. The preimage of $\mathbf{O}_n$ under polynomials has the following characterization with respect to the Schur form.

Lemma 5.6. *Given H in $Gl(n)$ and a polynomial $p()$ such that $p(H)$ is in $\mathbf{O}_n$. Suppose that the spectrum of $p(H)$ consists of q distinct complex conjugate pairs. Then H admits a real Schur form $diag(H_1, ..., H_q)$ such that the spectrum of each H_k is the preimage under $p()$ of the kth pair of eigenvalues of $p(H)$.*

Proof. There exists an orthogonal similarity transform with respect to which $p(H)$ becomes quasi-diagonal. The orthogonal similarity transforms H to the desired form. □

Fixed points satisfy a geometric constraint and a combinatoric constraint. The geometric constraint is that the spectrum of the Hessenberg matrix must be on a lemniscate with foci at the shifts (the roots of the shift polynomial). The combinatoric constraint is that each group of non-orthogonal eigenvectors must be mapped by the shift polynomial to the same value.

Remark 5.7 If the shift polynomial is of degree one, then it is a bijection, and the fixed points are the linear combinations of an orthogonal matrix and the identity. Similarly some computational scientists have reported that the complex QR iteration (with a linear shift polynomial) is more robust than the real QR iteration (with a quadratic shift polynomial).

Remark 5.8 The dimension of the orthogonal similarity class is generically the maximum number of coefficients that one may hope to match in a monic shift polynomial. The dimension of the orthogonality similarity class is at most $n-1$.

6. The Derivative of Shifted QR Iteration

A QR iteration on the space of Hessenberg matrices is defined as follows. the matrix H_i determines a shift polynomial p_i, and the iteration is defined by

$$p_i(H_i) = Q_i R_i, \qquad Hi+1 = \mathsf{F}(H_i) = Q_i^{\mathsf{T}} H_i Q_i.$$

For example, H_0 determines p_0, next $p_0(H_0) = Q_0 R_0$, and $\mathsf{F}(H_0) = Q_0^{\mathsf{T}} H_0 Q_0$. Recall that by Theorem D at a fixed point, $R_0 = \gamma I$ here.

To differentiate F at a fixed point, write $H(t) = H_0 + t\dot{H}_0$, and by abuse of notation let p_t be the shift polynomial associated to $H(t)$. Then

$$p_t\big(H(t)\big) = Q(t)R(t), \qquad Q(t) = Q_0(I_n + tS + \ldots), \quad R(t) = \gamma(I_n + tU + \cdots).$$

As in the unshifted case,

$$D\mathsf{F}(H_0)(\dot{H}_0) = Q_0^{\mathsf{T}} \dot{H}_0 Q_0 + H_0 S - S H_0 = Q_0^{\mathsf{T}} \dot{H}_0 Q_0 + [H_0, S].$$

Theorem E. *The derivative $DF(H_0)$ of the shifted QR iteration map $\mathsf{F} : Gl(n) \to Gl(n)$ at an unreduced Hessenberg fixed point H_0 has at least $\frac{1}{2}n(n+1)$ eigenvalues of unit modulus.*

Proof. $DF(H_0)$ is the sum of the isometry Ad_{Q_0} and the bracket $[H_0, S]$. As in the unshifted case, $DF(H_0)$ acts orthogonally on the space of tangent vectors for which "$S = 0$". There may be tangent vectors for which $S \neq 0$ and yet $[H_0, S] = 0$, but this only increases the dimension of the space on which $DF(H_0)$ acts orthogonally. It suffices to establish the claim that in the shifted case, the space of such tangent vectors has dimension at least $\frac{1}{2}n(n+1)$.

The justification of the claim depends on a more precise representation of the derivative. In the shifted case, the shift polynomial contains a derivative in the direction $\dot{H}_0$, or in other words

$$p_t = p_0 + tDp_0(\dot{H}_0) + \cdots .$$

Define S and U by $Dp_0(H_0)(\dot{H}_0) = \gamma Q_0(S + U)$. Fix H_0 and let $Dp_0(H_0) : \mathbf{M}_n \to \mathbf{M}_n$ be the derivative of the shift map at H_0.

If $\operatorname{rank} Dp_0(H_0) = k$, then $\operatorname{null} Dp_0(H_0) = n^2 - k$. Decompose the *image* of $Dp_0(H_0)$ using $\mathbf{M}_n \simeq Q_0 \cdot \big(\mathbf{Skew}(n) \oplus \mathbf{Up}(n)\big)$, and let k_1 and k_2 be the dimensions of the respective spaces, so that $k_1 + k_2 = k$. The first space is contained in $Q_0 \cdot \mathbf{Skew}(n)$, so $k_1 \leq \dim \mathbf{Skew}(n) = \frac{1}{2}n(n-1)$.

Consider the preimages under $Dp_0(H_0)$ of the two spaces:

$$V_1 = Dp_0(H_0)^{-1}\big(Q_0 \cdot \mathbf{Skew}(n)\big), \qquad V_2 = Dp_0(H_0)^{-1}\big(Q_0 \cdot \mathbf{Up}(n)\big).$$

Thus $V_1 + V_2 = \mathbf{M}_n$ as vector spaces, and $V_1 \cap V_2 = \ker Dp_0(H_0)$. By the rank theorem,

$$\dim V_2 = (n^2 - k) + k_2 = n^2 - k_1 \geq n^2 - \tfrac{1}{2}n(n-1) = \tfrac{1}{2}n(n+1).$$

However, if $\dot{H}_0 \in V_2$, then $Dp(H_0)(\dot{H}_0) = \gamma Q_0 U$, so

$$D\mathsf{F}(H_0)(\dot{H}_0) = Q_0^{\mathsf{T}} \dot{H}_0 Q_0 = \mathrm{Ad}_{Q_0} \dot{H}_0.$$

Therefore $D\mathsf{F}(H_0)$ acts orthogonally on V_2, which establishes the theorem. □

Remark 6.1 Some of the eigenvalues of the derivative of the shifted QR map, $DF(H_0)$, may not have unit modulus.

References

[1] G.S. Ammar and C.F. Martin. The geometry of matrix eigenvalue methods. *Acta Appl. Mathematicae*, 5:239–278, 1986.

[2] 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 User's Guide.* SIAM, Philadelphia, PA, USA, 1994.

[3] S. Batterson. Structurally stable grassman transformations. *Trans. AMS*, 231:385–404, 1977.

[4] S. Batterson. Convergence of the shifted QR algorithm on 3×3 normal matrices. *Numer. Math.*, 58:341–352, 1990.

[5] S. Batterson. The dynamics of eigenvalue computation. In M. Hirsh, J. Marsden, and M. Shub, editors, *From Topology to Computation: Proceedings of the Smalefest*, pages 368–386. Springer Verlag, 1993.

[6] B.N.Parlett and W. Kahan. On the convergence of a practical QR algorithm. (with discussion). In A. Morrell, editor, *Information Processing 68 (Proc. IFIP Congress, Edinburgh, 1968), vol I*, pages 114–118. North-Holland, Amsterdam, 1969.

[7] M. Chu and L. Norris. Isospectral flows and abstract matrix factorizations. *SIAM J. Numer. Anal.*, 25(6):1383–1391, December 1988.

[8] D. M. Day. How the QR algorithm fails to converge and how to fix it. Technical Report SAND96-0913J, Sandia National Laboratory, PO Box 5800, Albuquerque, NM 87185-5800, 1996.

[9] G. Golub and C. Van Loan. *Matrix computations.* The Johns Hopkins University Press, 2nd edition, 1989.

[10] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis.* Cambridge University Press, 1st edition, 1995.

[11] R. B. Lehoucq. *Analysis and implementation of an implicitly restarted Arnoldi iteration.* PhD thesis, Rice University, Houston Texas, 1995.

[12] Jerrold Marsden and Tudor Ratiu. *Introduction to Mechanics and Symmetry.* Springer, 2nd edition, 1999.

[13] R. Martin, G. Peters, and J. Wilkinson. The QR algorithm for real Hessenberg matrices. *Numer. Math.*, 14:219–231, 1970.

[14] B. N. Parlett. Global convergence of the basic QR algorithm on Hessenberg matrices. *Math. Comp.*, 20:611—615, 1966.

[15] B. N. Parlett. Singular and invariant matrices under the QR transformation. *Math. Comp.*, 20:611—615, 1966.

[16] B. T. Smith, J.M. Boyle, J.J. Dongarra, B.S. Garbow, Y. Ikebe, V. Klema, and C. Moler. *Matrix Eigensystem Routines-Eispack Guide*, volume 6 of *Lecture Notes in Computer Science.* Springer-Verlag, 1974.

[17] D. Watkins. Isospectral flows. *SIAM Rev.*, 26(3):379–392, July 1984.

[18] D. Watkins. Convergence of algorithms of decomposition type for the eigenvalue problem. *Lin. Alg. Appl.*, 143:19–47, 1991.

David M. Day, Org. 9222/MS 1110, Sandia National Laboratories, P.O. Box 5800, Albuquerque, New Mexico, 87185
E-mail: `dmday@sandia.gov`

Andrew D. Hwang, Dept. of Math and CS, College of the Holy Cross, Worcester, MA 01610
E-mail: `ahwang@mathcs.holycross.edu`