Proving Rapid Global Convergence for the Shifted QR Algorithm

Jess Banks, Jorge Garza-Vargas, Nikhil Srivastava

Abstract

The design of efficient and reliable algorithms for computing the eigenvalues and eigenvectors of a matrix is of unquestionable importance in both science and engineering. However, despite significant advancements in various practical aspects, fundamental theoretical questions about the eigenvalue problem remain poorly understood. In this talk I will discuss work [BGVSa, BGVSb, BGVSc] that provides nearly optimal rigorous guarantees, on all inputs, for the shifted QR algorithm. Similar results were established by Wilkinson in [Wil68] and Dekker and Traub in [DT71] for Hermitian inputs; however, despite sustained interest and several attempts, the non-Hermitian case remained elusive for the last five decades.

The QR iteration. The QR algorithm, which originated in the works of Francis [Fra61, Fra62] and Kublanovskaya [Kub62] (see [GU09] for some history), has been listed as one of the top ten most influential algorithms of the 20th century [DS00] and is the preferred method for computing the full eigendecomposition of an arbitrary input matrix.

In its simpler form, the QR algorithm starts by putting the input matrix $A \in \mathbb{C}^{n\times n}$ into Hessenberg form, that is, it computes a unitary matrix U such that $H = U^*AU$ is an upper Hessenberg matrix.[1] Then, it computes a sequence of Hessenberg matrices $H_0 = H, H_1, H_2 \dots$ via the iteration:

$$\begin{aligned} [Q_t, R_t] &= \mathrm{qr}(H_t), \\ H_{t+1} &= Q_t^* H_t Q_t. \end{aligned} \tag{1}$$

Where $Q_tR_t = H_t$ is the QR decomposition of H_t, and from $H_{t+1} = Q_t^*H_tQ_t$ we see that

$$A = U_t H_t U_t^* \quad \text{for} \quad U_t = UQ_0 \cdots Q_t.$$

This iteration has the fascinating property (see [Wat82]) that for generic[2] inputs A, as t goes to infinity, the H_t converge to an upper triangular matrix, say, T. In such situation, we can set $V = \lim_{t\to\infty} U_t$, so that

$$A = VTV^*,$$

therefore obtaining the Schur decomposition of A.[3] The appeal of this method resides on the simplicity of the iteration described in (1). The drawback is that the convergence $H_t \to T$ happens at a prohibitively slow rate for most inputs, ultimately turning it into an impractical algorithm.

The shifted QR algorithm. In practice the QR iteration is endowed with "shifts" that seek to accelerate convergence. Concretely, at each time t, a polynomial $p_t(z)$ is computed as a function of H_t (see Wilkinson's shift below for an example) and the iteration now is given by:

$$\begin{aligned} [Q_t, R_t] &= \mathrm{qr}(p_t(H_t)), \\ H_{t+1} &= Q_t^* H_t Q_t. \end{aligned} \tag{2}$$

[1] This can be done by applying a sequence of $n-1$ suitably chosen Householder transformations. This procedure is numerically stable and can be executed in $O(n^3)$ arithmetic operations, see [Wat08] for details.

[2] That is, all but a set of Lebesgue measure zero.

[3] Recall that one can read the eigenvalues of A from the diagonal entries of T, and if desired, easily compute the eigenvectors of A from the columns of V.

Intuitively, one should think of the roots of $p_t(z)$ as "guesses" for the eigenvalues of H_t (which by unitary equivalence are the same as the eigenvalues of A) and, the better the guesses the more progress towards convergence one will make while going from H_t to H_{t+1}. Moreover, the closer H_t is to an upper triangular matrix, the more its eigenvalues have been "revealed", which allows one to make better guesses, all together yielding a virtuous cycle that is in part responsible for the undefeated performance of the shifted QR algorithm. This intuition can be made rigorous by understanding the connection between the shifted QR algorithm and shifted inverse iteration [Wat82, Wat08], where the aforementioned "virtuous cycle" can be established via a *local* analysis of convergence, e.g. see [Par74] or [Par98, §4.7].

The chosen algorithm for computing the $p_t(z)$ as a function of the H_t is refered to as the *shifting strategy*, and the main purpose of any shifting strategy is to guarantee rapid *global* convergence, that is, rapid convergence to an upper triangular matrix regardless of the starting condition H_0. Although local convergence is intuitive (as explained above) and typically easy to establish, devising a shifting strategy that ensures rapid global convergence remained an important open problem throughout the years [Par74, Mol78, Dem97, Sma97, HDG+15].

Exploiting the Hessenberg structure. Working with Hessenberg matrices has several computational advantages that ultimately permit obtaining the full eigendecomposition of the input in nearly n^3 operations, which is the initial cost of putting the input matrix into Hessenberg form.

Easy deflation. In practice, one can only hope to compute an approximate Schur form (resp. approximate eigedecomposition) for the input matrix. In turn, when seeking to solve the an approximate version of the eigenvalue problem, one can exploit the Hessenberg structure to accelerate the algorithm as follows. We will say that an upper Hessenberg matrix H is δ-decoupled if one of its subdiagonals satisfies $|H(i, i-1)| \leq \delta \|H\|$. So, in the iteration (2), once one of the matrices H_t is δ-*decoupled* for some δ small enough, one can zero out the small subdiagonal:

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & \text{small} & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix} \longrightarrow \left(\begin{array}{cc:ccc} * & * & * & * & * \\ * & * & * & * & * \\ \hdashline 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{array}\right).$$

This procedure, called *deflation*, incurs a small error in the computation but has the advantage that the resulting matrix is now block upper triangular, and therefore the spectrum of the big matrix is the union of the spectra of each of the smaller block diagonal parts, which happen to again have a Hessenberg structure. Moreover, the eigenvectors of the big matrix can be related in a similar way to the eigenvectors of the smaller diagonal blocks. With this, the eigenvalue problem has been reduced to two subproblems of smaller dimension, on which one can again call the QR algorithm.

Implicit shifts. Another advantage of the Hessenberg structure is that in the iteration (2) one can compute H_{t+1} from H_t without having to explicitly compute $p_t(H_t)$. Concretely, if $p_t(z)$ is of degree k and H_t is of dimension n, one can compute H_{t+1} from H_t in $O(kn^2)$ operations using a procedure commonly known as *chasing the bulge* (see [Tis96] or [Wat08]). Moreover, when the input matrix is Hermitian, the iterates H_t are tridiagonal, and in this case H_{t+1} can be computed from H_t in $O(kn)$ operations.

Meaningful corners. If H is a normal upper Hessenberg matrix then the lower-right corners of H can be related to the orthogonal polynomials associated to a natural probability measure supported on the spectrum of H, and, there is a natural potential theory interpretation of the subdiagonals

of such corners. In the general non-normal case these interpretations are no longer valid, but still provide great intuition for the dynamics of the shifted QR algorithm. In part, this is the reason why many of the shifting strategies use small lower-right corners of the H_t to compute $p_t(z)$.

Previous theoretical guarantees. When the input $A \in \mathbb{C}^{n\times n}$ is Hermitian, and therefore all the iterates H_t are too, Wilkinson introduced a shifting strategy that guarantees rapid global convergence. At time t, Wilkinson's shift computes the two eigenvalues of the lower-right 2×2 matrix of H_t and takes the one (call it w_t) that is closest to $H_t(n,n)$ to then set $p_t(z) = z - w_t$. In [Wil68] Wilkinson proved that for any initial Hermitian H_0 , if one runs the iteration (2) using his shifting strategy, it holds that $\lim_{t\to\infty} H_t(n, n-1) = 0$, which in particular implies that for any $\delta > 0$, the matrix H_t is δ-decoupled once t is large enough. This was then revisited by Dekker and Traub [DT71] who obtained a rate of convergence for Wilkinson's shift by showing that

$$|H_{t+1}(n, n-1)^2 H_{t+1}(n-1, n-2)| \leq \frac{|H_t(n, n-1)^2 H_t(n-1, n-2)|}{\sqrt{2}}, \quad \text{for all } t \geq 0. \tag{3}$$

In particular, this implies that for any $\delta > 0$, δ-decoupling occurs in $O(\log(1/\delta))$ iterations. Combining this with the deflation technique and the implicit shifts described above, one gets that any Hermitian matrix can be fully diagonalized to accuracy δ in $O(n^3 + \log(1/\delta)n^2)$ operations.

The case in which the input matrix $A \in \mathbb{C}^{n\times n}$ is unitary was later solved by Eberlein and Huang [EH75] and Wang and Gragg [WG02]. When H_0 is unitary the Wilkinson shift is no longer guaranteed to eventually produce decoupling. In fact, if the Wilkinson shift is used it can occur that $H_0 = H_1 = H_2 = \cdots$, and similarly many other natural shifting strategies have certain unitary matrices as fixed points (see [Par66]). The insight of Eberlein and Huang [EH75] was that these commonly used shifting strategies could be combined with an *exceptional shift* that avoids stagnation. In essence, their idea was to choose a *main shift* (e.g. one could choose the Wilkinson shift), and then exploit the knowledge that the input matrix is unitary to identify fixed points for the main shift, to then scape them by invoking the exceptional shift whenever necessary. Later, Gragg and Wang [WG02] revisited this idea and showed that, on unitary inputs, a mixed strategy that combines the Wilkinson shift and an exceptional shift satisfies a more complicated version of (3). Their analysis implies that this mixed strategy achieves δ-decoupling in $O(\log(1/\delta))$ iterations, ultimately implying that any unitary input can be diagonalized to accuracy δ in $O(\log(1/\delta)n^3)$ operations.

Beyond Hermitian and unitary matrices not much was known and proving rapid global convergence was open even in the normal case. We refer the reader to [BGVSa, §1.2] for a comprehensive literature review.

The main result. In the series [BGVSa, BGVSb, BGVSc] we introduced a shifting strategy that provably achieves global rapid convergence (in the space of all matrices). Hereon, if all the $p_t(z)$ in a shifting strategy are of degree k we will say that the shifting strategy is of degree k.

The condition number of the eigenvector matrix turned to be a fundamental quantity in our analysis. To be precise, if $A \in \mathbb{C}^{n\times n}$ is diagonalizable, define

$$\kappa_V(A) = \inf_{V:A=VDV^{-1}} \|V\|\|V^{-1}\|,$$

where $\|\cdot\|$ denotes the operator norm and the infimum runs over all diagonalizations of A. Note that when A is normal one has $\kappa_V(A) = 1$ and when A is non-diagonalizable the convention is

that $\kappa_V(A) = \infty$, so $\kappa_V(\cdot)$ can be viewed as a measure of non-normality. Fundamentally, [BGVSa] proves the following.[4]

Theorem 1. *For every positive integer k, there exists a shifting strategy of degree k that is ensured to achieve δ-decoupling in $\log(1/\delta)$ iterations provided that the starting matrix H_0 satisfies*

$$\log(1 + \kappa_V(H_0)) \cdot \log\big(1 + \log(1 + \kappa_V(H_0))\big) \leq ck, \tag{4}$$

where $c > 0$ is some absolute constant.

In some sense, our analysis articulates that the complexity of shifted QR is tied to κ_V of the input. In particular, the above theorem implies that rapid global convergence on normal matrices is possible using a shifting strategy of degree $O(1)$, just as in the case of Hermitian and unitary matrices. In contrast, when the input is non-diagonalizable the strategy needed is "infinitely complex" and the theorem becomes vacuous. That said, the latter situation can be addressed using an idea from smoothed analysis [ST04] which in the context of the eigenvalue problem can be traced back to Davies [Dav08]. In short, to obtain guarantees for arbitrary inputs, instead of running the algorithm on the original input matrix $A \in \mathbb{C}^{n\times n}$ we run it on $A + \gamma G_n$, where γG_n is a tiny random perturbation of A. One can then invoke results from random matrix theory (e.g. from [ABB$^+$18, BKMS21, BGVKS24, JSS21]), which for example imply that if G_n is a normalized $n \times n$ Ginibre matrix[5], $\|A\| \leq 1$, and $\gamma > 0$, with high probability

$$\kappa_V(A + \gamma G_n) \leq \frac{n^4}{\gamma}. \tag{5}$$

Certainly, this *preprocessing* random perturbation incurs an error in the computation (just as the deflation step does), but if the scale of γ is chosen appropriately, it will not preclude one from being able to obtain an accurate approximate version of the eigenvalue problem. Then, putting (4) and (5) together, in [BGVSb] we were able to show that a randomized version of the QR algorithm can diagonalize any input matrix $A \in \mathbb{C}^{n\times n}$ with accuracy δ in $O(n^3 \log(n/\delta)^2 \log\log(n/\delta)^2)$ operations.

Our shifting strategy. As in [DT71] and other works that served as inspiration (e.g. [Bat94]), we used the lower subdiagonal entries of the iterates H_t to keep track of progress towards convergence. Specifically, to analyze the shifting strategy of degree k mentioned in Theorem 1, we used the potential function ψ_k which on a Hessenberg matrix H is defined as

$$\psi_k(H) = |H(n, n-1)H(n-1, n-2)\cdots H(n-k+1, n-k)|^{\frac{1}{k}}.$$

Then, as in [EH75, WG02], we used a mixed strategy consisting of a main shift and an exceptional shift. If at time t an iteration with the main shift *did not* satisfy that $\psi_k(H_{t+1}) \leq .8\psi_k(H_t)$ (i.e. if progress is not being made), then our shifting strategy recomputes H_{t+1}, this time using the exceptional shift, and in [BGVSa] we show that provided that κ_V of the input matrix satisfies the bound (5) the exceptional shift *does* succeed in guaranteeing $\psi_k(H_{t+1}) \leq .8\psi_k(H_t)$. Our mixed strategy then guarantees a geometric decrease of the quantity $\psi_k(H_t)$, which in turn implies that δ-decoupling will occur after $O(\log(1/\delta))$ iterations.

A final caveat. Our theoretical algorithm is not a prescription for practitioners and does not seek to replace the current very efficient LAPACK routines, which have been fine-tuned over the

[4]This theorem was not stated verbatim and strictly speaking only k's that are powers of 2 were treated in the paper, however, the ideas in [BGVSa] yield, with very little extra work, the theorem stated here.

[5]That is, and $n \times n$ matrix with i.i.d. complex Gaussian entries of variance $\frac{1}{n}$.

decades and for which several patches have been added to avoid convergence failures. We do warn the reader however, that such routines are by now quite sophisticated and do not come with theoretical guarantees. This does make one wonder if there is an algorithm that is as efficient as the existing implementations, but that is conceptually simple and for which one can give rigorous guarantees.

References

[ABB+18] Diego Armentano, Carlos Beltrán, Peter Bürgisser, Felipe Cucker, and Michael Shub. A stable, polynomial-time algorithm for the eigenpair problem. *Journal of the European Mathematical Society*, 20(6):1375–1437, 2018.

[Bat94] Steve Batterson. Convergence of the Francis shifted QR algorithm on normal matrices. *Linear algebra and its applications*, 207:181–195, 1994.

[BGVKS24] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Overlaps, eigenvalue gaps, and pseudospectrum under real ginibre and absolutely continuous perturbations. In *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques*, volume 60, pages 2736–2766. Institut Henri Poincaré, 2024.

[BGVSa] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global convergence of Hessenberg shifted QR I: Dynamics. *To appear in Foundations of Computational Mathematics*.

[BGVSb] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global convergence of Hessenberg shifted QR II: Numerical stability. *To appear in SIAM Journal on Matrix Analysis and Applications*.

[BGVSc] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global convergence of Hessenberg shifted QR III: Approximate Ritz values via shifted inverse iteration. *To appear in SIAM Journal on Matrix Analysis and Applications*.

[BKMS21] Jess Banks, Archit Kulkarni, Satyaki Mukherjee, and Nikhil Srivastava. Gaussian regularization of the pseudospectrum and Davies' conjecture. *Communications on Pure and Applied Mathematics*, 74(10):2114–2131, 2021.

[Dav08] E. Brian Davies. Approximate diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1051–1064, 2008.

[Dem97] James W Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[DS00] Jack Dongarra and Francis Sullivan. Guest editors introduction to the top 10 algorithms. *Computing in Science & Engineering*, 2(01):22–23, 2000.

[DT71] Theodorus J Dekker and Joseph F Traub. The shifted QR algorithm for Hermitian matrices. *Linear Algebra Appl*, 4:137–154, 1971.

[EH75] Patricia J Eberlein and C. P. Huang. Global convergence of the QR algorithm for unitary matrices with some results for normal matrices. *SIAM Journal on Numerical Analysis*, 12(1):97–104, 1975.

[Fra61] John GF Francis. The QR transformation a unitary analogue to the LR transformation—part 1. *The Computer Journal*, 4(3):265–271, 1961.

[Fra62] John GF Francis. The QR transformation—part 2. *The Computer Journal*, 4(4):332–345, 1962.

[GU09] Gene Golub and Frank Uhlig. The QR algorithm: 50 years later its genesis by John Francis and Vera Kublanovskaya and subsequent developments. *IMA Journal of Numerical Analysis*, 29(3):467–485, 2009.

[HDG+15] Nicholas J Higham, Mark R Dennis, Paul Glendinning, Paul A Martin, Fadil Santosa, and Jared Tanner. *The Princeton companion to applied mathematics*. Princeton University Press Princeton, NJ, USA:, 2015.

[JSS21] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. On the real Davies' conjecture. *The Annals of Probability*, 49(6):3011–3031, 2021.

[Kub62] Vera N Kublanovskaya. On some algorithms for the solution of the complete eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, 1(3):637–657, 1962.

[Mol78] Cleve B Moler. Three research problems in numerical linear algebra. *Numerical Analysis*, 22:1–18, 1978.

[Par66] Beresford Parlett. Singular and invariant matrices under the QR transformation. *Mathematics of Computation*, 20(96):611–615, 1966.

[Par74] Beresford N Parlett. The Rayleigh quotient iteration and some generalizations for nonnormal matrices. *Mathematics of Computation*, 28(127):679–693, 1974.

[Par98] Beresford N Parlett. *The symmetric eigenvalue problem*. SIAM, 1998.

[Sma97] Steve Smale. Complexity theory and numerical analysis. *Acta numerica*, 6:523–551, 1997.

[ST04] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[Tis96] Francoise Tisseur. Backward stability of the QR algorithm. Technical report, 239, UMR 5585, Lyon Saint-Etienne, 1996.

[Wat82] David S Watkins. Understanding the QR algorithm. *SIAM review*, 24(4):427–440, 1982.

[Wat08] David S Watkins. The QR algorithm revisited. *SIAM review*, 50(1):133–145, 2008.

[WG02] Tai-Lin Wang and William Gragg. Convergence of the shifted QR algorithm for unitary Hessenberg matrices. *Mathematics of computation*, 71(240):1473–1496, 2002.

[Wil68] James Hardy Wilkinson. Global convergence of tridiagonal QR algorithm with origin shifts. *Linear Algebra and its Applications*, 1(3):409–420, 1968.