# GENERALIZATIONS OF THE RAYLEIGH QUOTIENT ITERATION FOR THE ITERATIVE REFINEMENT OF THE EIGENVECTORS OF REAL SYMMETRIC MATRICES

*Maziar Nikpour, Knut Hüper, Jonathan H. Manton*

Centre for Systems Engineering and Applied Mechanics
Department of Mathematical Engineering
Catholic University of Louvain
1348 Louvain-la-Neuve
Belgium,
nikpour@inma.ucl.ac.be

National ICT Australia
Canberra Research Laboratory
Systems Engineering and Complex Systems Program
Locked Bag 8001, Canberra ACT 2601
Australia
Knut.Hueper@nicta.com.au

Department of Electrical and Electronic Engineering
The University of Melbourne
Victoria 3010
Australia
j.manton@ee.mu.oz.au

## ABSTRACT

We present new algorithms for refining the estimates of the eigenvectors of a real symmetric matrix. Using techniques from calculus, we show that the algorithms converge locally cubically fast. By this we mean that locally *all* eigenvalue-eigenvector pairs *simultaneously* converge at a cubic rate. This is in contrast to well known shifted QR algorithms, which depending on the shifting strategy employed, have only one (or at most a small subset) of the eigenvalue-eigenvector pairs converging cubically at any one time.

The algorithms are well suited to the situation where one needs to compute the eigenvectors of a perturbed matrix $A + E$ based on a good estimate of the eigenvectors of a matrix $A$. Such a situation frequently appears in tracking applications.

M. Nikpour acknowledges the Belgian Programme on Inter-University Poles of Attraction initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence Program.

## 1. INTRODUCTION

The well known Rayleigh Quotient Iteration (RQI) is an iterative algorithm by which a single eigenvector-eigenvalue pair of a symmetric matrix $A$ can be computed with a locally cubic rate of convergence. It is also well understood that the RQI is implicitly performed in the shifted QR algorithm for computing the full set of eigenvectors of $A$ [2, 3].

The shifted QR algorithm is not suited to the problem of computing the eigenvectors of a perturbed symmetric matrix $A + E$ given a good estimate of the eigenvectors of $A$. There are two reasons for this. One is that efficient implementation of the shifted QR algorithm requires a preprocessing step of transforming the matrix into tridiagonal form which needs to be recomputed from scratch if the matrix changes. The second reason is that the shifted QR algorithm employs deflation. Without deflation, the algorithm would converge significantly more slowly. A pervasive myth is that the shifted QR algorithm converges cubically, when in fact it is only one eigenvector that is converging cubically at a given moment in the algorithm's operation. In [4] it was shown that there are no scalar valued shift strategies to ensure even quadratic convergence of the shifted QR algo-

rithm. This indicates that either deflation or a matrix-valued shift strategy is required for the shifted QR algorithm to be efficient. For the adaptive eigenvector computation problem deflation is not desirable and a matrix-valued shift strategy might be computationally too complex.

Recently and independently the authors looked at developing algorithms in which all eigenvalue-eigenvector pairs simultaneously converge cubically [1, 2, 4, 5, 6]. These algorithms took the RQI as a starting point and then developed generalisations to the full matrix case. These algorithms achieved simultaneous local cubic convergence of all eigenvector-eigenvalue pairs without any deflation. A drawback of these algorithms was that they required one or more orthogonalisations of the full set of vectors which is computationally inefficient. In this work we show that local cubic convergence of all eigenvectors can be achieved with more simple operations than full orthogonalisation. This comes at the price of losing possible good global convergence properties in the algorithms of [1, 2, 6], but we maintain fast local convergence which makes the algorithms useful for adaptive applications.

The algorithms we present can be thought of as generalisations of the RQI. Unlike other generalisations of the RQI for computing invariant subspaces and eigenvectors that are defined on Graßmannians or flag manifolds, our iterations are defined on the product of spheres.

*Notation*: Let upper case letters denote matrices. Let the lower case notation $x^{(j)}$ denote the $j$-th column of the matrix $X$ with entries $x_{ij}$. The symbol $\|X\| = \sqrt{\sum_{i,j=1}^{n} x_{ij}^2}$ denotes the Frobenius norm. By $X^\top$ we denote the transpose of $X$. The set $S^{n-1}$ denotes the set of vectors in $\mathbb{R}^n$ with unit norm. $I$ denotes the identity matrix.

## 2. ALGORITHMS

The central computational step in our algorithms is the Rayleigh Quotient Iteration (RQI). The RQI is defined by the mapping

$$r : S^{n-1} \to S^{n-1},\ x \mapsto \frac{(A - x^\top A x I)^{-1} x}{\|(A - x^\top A x I)^{-1} x\|}. \qquad (1)$$

It is well known that for an initial condition outside a set of measure zero, the RQI converges to an eigenvector of $A = A^\top$ with an asymptotic cubic rate of convergence.

The RQI may converge to any eigenvector of $A$, even one which is not the closest to the current estimate. A key issue then, when extending the RQI to computing more than one eigenvector is to try and prevent convergence to the same eigenvector. This can be achieved, at least locally, by exploiting the orthogonality property of the set of eigenvectors of $A$. The key idea is to transmit the information obtained by the update of one eigenvector estimate to the others. In the algorithm of [4, 5] this was achieved by a full orthogonalisation of the set of vectors after performing RQI on each in parallel. In an algorithm in [1, 6] it was done by orthogonalisation of the full set after each RQI update of *one* vector. Another closely related algorithm in [1, 6] also achieved this by replacing the least recently updated vector by a vector orthogonal to all the others before performing the RQI update on it. In [2] it was discussed how these algorithms relate to the well known shifted QR algorithm. These algorithms all achieved local cubic convergence and some of them even displayed good global convergence properties.

The algorithms proposed in this paper achieve local cubic convergence even without a full orthogonalisation or the computation of the orthogonal complement of the span of a set of vectors.

Algorithm 2.1 performs an RQI on one column (eigenvector estimate) and then subtracts those components of the other columns which lie in the direction of the updated one. This is done to each column once per iteration of the algorithm.

**Algorithm 2.1** *1. Initialise* $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(n)}] \in \mathbb{R}^{n\times n}$ *such that* $X_0^\top X_0 = I$. *Let* $k = 0$.

*2. For* $i = 1, 2, \ldots, n$ *do:*

*(a) Evaluate* $x_{k+1}^{(i)} = r(x_k^{(i)})$.

*(b) For* $j \neq i$, *compute* $y = (I - x_{k+1}^{(i)} x_{k+1}^{(i)\top}) x_k^{(j)}$ *and set* $x_{k+1}^{(j)} = y/\|y\|$.

*3. Set* $X_{k+1} = [x_{k+1}^{(1)}, x_{k+1}^{(2)}, \ldots, x_{k+1}^{(n)}]$.

*4. If all* $x_k^{(i)}$ *have converged to eigenvectors then stop, otherwise goto 2.*

Step 2.(b) of Algorithm 2.1 which projects the other vectors onto the orthogonal complement of the vector most recently updated, requires less computation than a full orthogonalisation and is more simple to implement since it requires less memory storage.

Let

$$\mathcal{S} := S^{n-1} \times \ldots \times S^{n-1}. \qquad (2)$$

Let $A = A^\top$ be given with distinct eigenvalues. To remove the discontinuities at the fixed points of $r$ defined by (1) we define another map, which we again call RQI. Consider the map

$$\widehat{r} : S^{n-1} \to S^{n-1}, \quad x \mapsto \frac{\operatorname{adj}(A - x^\top A x I_n) x}{\|\operatorname{adj}(A - x^\top A x I_n) x\|}. \qquad (3)$$

Here the operator adj denotes the classical adjoint, i.e., $\mathrm{adj}(X) = X^{-1} \cdot \det(X)$. The differentiability properties of (3) are well understood, see [4, 5]. We consider Algorithm 2.1 as the self map

$$\sigma : \mathcal{S} \to \mathcal{S} \tag{4}$$

where

$$\sigma := \sigma_n \circ \cdots \circ \sigma_1 \tag{5}$$

and for all $i = 1, \ldots, n$

$$\begin{aligned} &\sigma_i : \mathcal{S} \to \mathcal{S}, \\ &\quad X \mapsto \left[y^{(1)}, \ldots, y^{(i-1)}, \widehat{r}(x^{(i)}), y^{(i+1)}, \ldots, y^{(n)}\right], \end{aligned} \tag{6}$$

where for all $j = 1, \ldots, n$ with $j \neq i$

$$y^{(j)} := \frac{\left(I - \widehat{r}(x^{(i)})\widehat{r}(x^{(i)})^\top\right) x^{(j)}}{\| \left(I - \widehat{r}(x^{(i)})\widehat{r}(x^{(i)})^\top\right) x^{(j)} \|}. \tag{7}$$

The following theorem proves that local cubic convergence can be achieved by Algorithm 2.1. The proof uses the same calculus approach utilised in [4] to prove local cubic convergence of matrix eigenvector algorithms.

**Theorem 2.1** *Let $A = A^\top$ be given with distinct eigenvalues. Assume that Algorithm 2.1 is initialised by an orthogonal matrix which is sufficiently close to an orthogonal matrix consisting of eigenvectors of $A$. Using $\widehat{r}$ instead of $r$, Algorithm 2.1 converges locally cubically fast.*

PROOF (SKETCH). Exploiting the fact that the mapping $\sigma$ is smooth in an open neighborhood of the identity matrix $X_* = I \in \mathcal{S}$ we can show that the first and second derivatives of $\sigma$ at $X_*$ vanish. It is easily seen that for arbitrary tangent elements $H \in T_{X_*}\mathcal{S}$ the derivative

$$\mathrm{D}\,\sigma(X_*)H = \mathrm{D}\,\sigma_n(X_*) \circ \cdots \circ \mathrm{D}\,\sigma_1(X_*)H = 0. \tag{8}$$

Eq. (8) holds true (i) by the chain rule, (ii) by the quadratic convergence of RQI induced by $\widehat{r}$, and (iii) by the fixed point property of each $\sigma_i$, i.e., $\sigma_i(X_*) = X_*$, see [4, 5] for details. Realising that asymptotically $\sigma$ decouples into $n$ individual RQI iterations on the $n$ columns of $X_*$ gives the result, because each individual RQI is cubically convergent in itself, [4, 5]. □

**Remark 2.1** *Note that the iterations induced by $r$ and $\widehat{r}$ roughly have the same dynamics. The difference is just in*

$$\operatorname{sign} \det(A - x^\top A x I). \tag{9}$$

*Another way to look at these iterations would certainly be as iterations on projective space $\mathbb{RP}^{n-1}$ instead of $S^{n-1}$. See also the comments in [7], but we will not go into details here, we refer to the forthcoming paper by U. Helmke, K. Hüper and J. Trumpf where RQI and its relation to different Newton-type methods is revisited.*

Algorithm 2.1 represents a reduction in complexity over previous algorithms. Further simplifications are possible if instead of projecting all columns onto the space orthogonal to the one just updated, we project only the next column. This modification gives:

**Algorithm 2.2** *1. Initialise $X_0 = [x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(n)}] \in \mathbb{R}^{n \times n}$ such that $X_0^\top X_0 = I$. Let $k = 0$.*

*2. For $i = 1, 2, \ldots, n$ do:*

*(a) Evaluate $x_{k+1}^{(i)} = r(x_k^{(i)})$.*

*(b) If $i \neq n$, compute $y = (I - x_{k+1}^{(i)} x_{k+1}^{(i)\top}) x_k^{(i+1)}$ and set $x_{k+1}^{(i+1)} = y/\|y\|$.*
*If $i = n$, compute $y = (I - x_{k+1}^{(i)} x_{k+1}^{(i)\top}) x_k^{(1)}$ and set $x_{k+1}^{(1)} = y/\|y\|$.*

*3. Set $X_{k+1} = [x_{k+1}^{(1)}, x_{k+1}^{(2)}, \ldots, x_{k+1}^{(n)}]$.*

*4. If all $x_k^{(i)}$ have converged to eigenvectors then stop, otherwise goto 2.*

Using the same technique as for the proof of Theorem 2.1, we have:

**Theorem 2.2** *Let $A = A^\top$ be given with distinct eigenvalues. Assume that Algorithm 2.2 is initialised by an orthogonal matrix which is sufficiently close to an orthogonal matrix consisting of eigenvectors of $A$. Then Algorithm 2.2 converges locally cubically fast.*

## 3. NUMERICAL EXAMPLE

To illustrate the utility of our algorithms in non-stationary environments, we consider a classical signal processing example. Suppose we have a symmetric matrix $A$ whose elements evolve according to an AR(1) process;

$$A_{k+1} = \alpha A_k + E_k, \quad E_k = E_k^\top. \tag{10}$$

At each time instant $k$, we run each algorithm for just one iteration (i.e. one run of step 2) and compare the eigenvalue estimates of our algorithms to MATLAB's `eig` routine which employs a variation of the single shift QR algorithm. We initialise $A_0$ to be the diagonal matrix

$$A_0 = \mathrm{diag}\{1, 2, 9, 10, 11\}. \tag{11}$$

The symmetric perturbation matrices $E_k$ we have used have elements drawn from a zero mean, variance 0.1 normal distribution, and we set $\alpha = 0.99$.

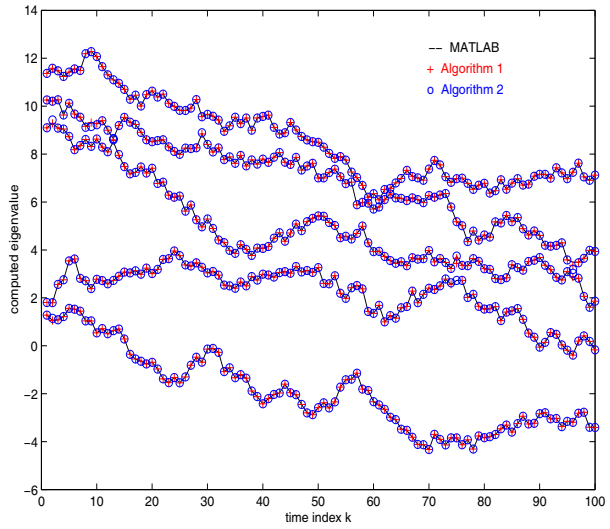


**Fig. 1**. Tracking of the eigenvalues of a time varying matrix. See section 3 for details.

Figure 1 illustrates the power of the local cubic convergence of our algorithms. The curves in the figure show the evolution of the eigenvalues of matrices $A_k$ as computed by MATLAB's `eig` routine. Also shown are the values of the eigenvalues as computed by Algorithms 2.1 and 2.2. The close proximity of the eigenvalues computed by our algorithms to that of MATLAB illustrates the good tracking capability of our algorithms. Whilst MATLAB performs the shifted QR algorithms from scratch at each $k$, our algorithms only require one iteration to keep track of the eigenvalues of $A$. In typical signal processing applications the variation of the eigenspace is also slower than the above example.

Although the algorithms are not meant for block computations, it is interesting to note that Algorithm 2.1 also displays good *global* convergence properties by converging to the correct set of eigenvectors of a given matrix from most initial conditions. Simulations show however, that it requires more iterations than the algorithms in [1, 2] and [4] which perform full orthogonalisation of the set of vectors in each iteration. Algorithm 2.2, which is even less complex than Algorithm 2.1 does not display good global convergence.

## 4. DISCUSSION

The results of this paper suggest that a full orthogonalisation of a set of eigenvector estimates is not required to prevent RQI iterates from converging to the same eigenvector provided that we are sufficiently close to the true eigenvector estimates. Furthermore, cubic convergence of all eigenvector estimates to the true eigenvectors is still achieved.

## 5. ACKNOWLEDGEMENT

This work was initiated when the second and third author met at MTNS 2002 in Notre Dame, USA. It was completed while the second author visited CESAME, Louvain, in July 2004. The hospitality of the Department of Mathematical Engineering, Catholic University of Louvain, is gratefully acknowledged. Special thanks to Paul Van Dooren, CESAME, for many fruitful discussions.

## 6. REFERENCES

[1] M. Nikpour, J.H. Manton, and R. Mahony, "Novel Newton Algorithms for the Hermitian Eigenvalue Problem," in *Proceedings of Information, Decision and Control*, Adelaide, Australia, February 2002.

[2] M. Nikpour, J.H. Manton, and I.M.Y. Mareels, "A novel approach to the shifted QR algorithm for the Hermitian eigenvalue problem," in *Proceedings of* $42^{nd}$ *IEEE Conference on Decision and Control*, Maui, USA, December 2003, vol. 3, pp. 2364 – 2367.

[3] D.S. Watkins, "Understanding the QR algorithm," *SIAM Review*, vol. 24, no. 4, pp. 427–440, 1982.

[4] K. Hüper, "A calculus approach to matrix eigenvalue algorithms," Habilitation Dissertation, July 2002, Mathematisches Institut, Universität Würzburg, Germany, http://www.rsise.anu.edu.au/~hueper/.

[5] K. Hüper, *Mathematical Systems Theory in Biology, Communications, Computation*, vol. 134 of *The IMA Volumes in Mathematics and its Applications*, chapter: A Dynamical System Approach to Matrix Eigenvalue Algorithms, pp. 257–274, Springer, New York, 2003.

[6] M. Nikpour, *Reduced Rank Signal Processing*, Ph.D. thesis, The University of Melbourne, Australia, 2002.

[7] M. Shub, "Some remarks on dynamical systems and numerical analysis," in *Dynamical systems and partial differential equations (Caracas, 1984)*, pp. 69–91. Univ. Simon Bolivar, Caracas, 1986.