

A Spectrum Algorithm and Optimization Notes

Jonathan Barzilai

Dalhousie University
Halifax, Nova Scotia
Barzilai@ScientificMetrics.com

© Jonathan Barzilai 2017

Abstract

We construct an eigenvector-based symmetric spectrum algorithm. At each step of this algorithm an eigenvector and its associated eigenvalue are computed without using previously computed eigenvalues. The limitations of optimization algorithms, in this context and in general, are noted.

Keywords

Symmetric Spectrum, Eigenvector, Eigenvalue, Rayleigh Quotient, Conjugate Gradients, Optimization Limitations.

1 Introduction

While the structure of the eigenvalues of a real symmetric matrix is arbitrary, a spectrum algorithm can take advantage of the structure of its eigenvectors. Generalizing the Rayleigh quotient of a vector with respect to a matrix, we construct an efficient eigenvector-based symmetric spectrum algorithm that takes advantage of the Rayleigh Quotient Iteration, the Conjugate Gradient algorithm, the fact that a symmetric matrix has a full set of orthogonal eigenvectors, and sequential problem-size reduction. For detailed treatments of the symmetric spectrum problem see [2–5]. In addition, the limitations of optimization algorithms and theory in this context and beyond are noted.

2 A Symmetric Spectrum Algorithm

2.1 A Generalized “Ratio”

Define the generalized “quotient” of the $n \times k$ matrix X with respect to the $n \times n$ matrix A as the $k \times k$ matrix $(X^T A X) \cdot (X^T X)^{-1}$ (in the complex case the Hermitian inner product). If the columns of X are independent, the inverse (rather than generalized inverse)

of X^tX exists. If the columns of X are orthonormal, i.e. $X^tX = I$ where I denotes an identity matrix, the ratio is X^tAX . If the columns of X are orthonormal eigenvectors of A , X^tAX is diagonal, displaying the corresponding eigenvalues of A .

2.2 Reduction

In what follows A is assumed to be a real, symmetric matrix of order n . Assume that T is an orthonormal basis of the null space of a proper subset of independent eigenvectors of A that have been computed. Then $T^tT = I$ and the size of $B = T^tAT$ – the generalized quotient of T with respect to A – is smaller than the size of A . If a vector v in the space spanned by T is an eigenvector of A , then $v = Tm$ for some m and $Av = \lambda v$ where λ is a scalar. Since $ATm = \lambda Tm$, $T^tATm = Bm = \lambda T^tTm = \lambda m$ and $Bm = \lambda m$. Finally, since T is a one-to-one mapping,

$$Av = \lambda v \Leftrightarrow Bm = \lambda m \quad (1)$$

and $(m, \lambda) \rightarrow (v, \lambda)$ is a one-to-one correspondence.

2.3 Algorithm

In this section we treat the real symmetric case, combining the reduction with the Rayleigh Quotient Iteration:

- *Initialization:* Start with an empty set L of computed A -eigenvectors, an identity matrix as an orthonormal basis T for the null space of L , and $B = A$.
- *Reduction step:* Compute an eigenvector of B using the Rayleigh Quotient Iteration (RQI, Algorithm 27.3 in [4]). Denote this B -eigenvector by m . Then Tm is an eigenvector of A . Update the set L of computed A -eigenvectors (and eigenvalues). Update the orthonormal basis T for the null space of L and the B matrix.
- *Repeat* the reduction step n times where n is the size of A .

2.4 Notes

To solve RQI's linear system use the Conjugate Gradient algorithm (CG, Algorithm 38.1 in [4]). The properties of the RQI and CG algorithms are well known. We now show that when the size of B is $k \times k$ (and T 's size is $n \times k$), the orthonormal basis T and the B matrix can be updated by $k-1$ two-dimensional rotations. Therefore, this update requires $O(n^2)$ multiplications.

Given $v = Tm$, $v \notin L$, compute a rotation R_{12} of the first two columns of T such that v is orthogonal on the first column of $T_{12} = TR_{12}$ (if a and b are the inner products of v with columns (1, 2) of T respectively, $x = b/\sqrt{a^2 + b^2}$, $y = a/\sqrt{a^2 + b^2}$, and

the (1, 2) block of R_{12} is $\begin{bmatrix} x & y \\ -y & x \end{bmatrix}$. Then (i) v is orthogonal on the first column of T_{12} , (ii) since a rotation matrix is orthogonal, T_{12} is an orthonormal basis for the null space of L , (iii) the columns of T other than those rotated by R_{12} (and their inner products with v) are unchanged, and (iv) the product TR_{12} requires $O(n)$ multiplications. Rotate again with $T_{123} = T_{12}R_{23} = TR_{12}R_{23}$ such that v is orthogonal on the second column of T_{123} . Then T_{123} is an orthonormal basis for the null space of L ; the columns of T_{12} other than those rotated by R_{23} are unchanged in T_{123} ; and v is orthogonal on columns 1 and 2 of T_{123} . Continuing in this manner produces an orthonormal basis $T_{12\dots k}$ for the null space of L such that v is orthogonal on all its columns except the last one. Since v is also orthogonal on L , the last column of $T_{12\dots k}$ must be v . The first $k-1$ columns of $T_{12\dots k}$ are then the updated orthonormal basis of the null space of $L - \{v\}$, and computing this basis requires $O(n^2)$ multiplications. Since $B = T^t A T$, a rotation on T is reflected by two rotations on B : $(TR)^t A (TR) = R^t T^t A T R = R^t B R$. It follows that B can be updated simultaneously with T . The last row of $T_{12\dots k}^t A T_{12\dots k}$ is $[0 \dots 0 \lambda]$ where $A v = \lambda v$, $B m = \lambda m$ and, when the last column of $T_{12\dots k}$ is dropped, the last row and column of B are dropped. Updating B requires $O(n^2)$ multiplications.

3 Optimization Notes

If α is not an eigenvalue of A , then $A - \alpha I = 0$ is not singular and the only solution of the equation $(A - \alpha I)x = 0$ is $x = 0$ which is not an eigenvector since an eigenvector must be a non-zero vector. While the terms “not an eigenvalue” and “a non-zero vector” are meaningful as infinite precision concepts, the finite precision spectrum computation problem is closely related to singularity (if α is an eigenvalue of A) and ill-conditioning (if α is close to an eigenvalue of A) of $A - \alpha I$. A new algorithm (Algorithm B67 by this author) is stable, fast, and applicable on singular linear systems, but like all such iterative algorithms that do not terminate in a finite number of steps, its convergence rate is linear. Since an algorithm’s convergence is as slow as its slowest sub-algorithm, it is preferable to employ the CG algorithm, which converges in n steps where n is the size of A , as the RQI’s linear solver. In that case, the slowest part of our spectrum algorithm is the Rayleigh Quotient Iteration. The RQI’s asymptotic rate of convergence is cubic and Trefethen and Bau [4, p. 208] describe its convergence speed as “spectacular.” In practice, it converges in a very small number of steps which appear to be independent of the problem’s size, in which case the computation of the entire spectrum *practically* requires only $O(n^4)$ multiplications. Note that the behaviour of the CG algorithm if A is not a real symmetric matrix has not been studied and that the RQI algorithm may not converge (see [1]), or it may converge slowly in that case.

Despite the fact that optimization is a powerful tool, it has its theoretical and computational limitations. There are no efficient solutions for the general optimization (or solution of equations) problem of order 2, since some NP-complete problems are of this order. The solution of a linear system of equations is a sub-step of optimization algorithms (e.g. Newton's Method and RQI), yet the rate of convergence of all iterative algorithms for solving linear equations is linear, implying poor performance for ill-conditioned problems. Constructing a superlinearly convergent algorithm for solving the general linear equations problem (i.e. including the singular and ill-conditioned cases) or proving that such an algorithm does not exist is a difficult challenge.

Newton's Method (and its variants) for minimizing a general (non-convex) smooth function, even in the one-dimensional case, is not guaranteed to converge to a global minimum, to a local minimum, or to converge at all.

Establishing that an algorithm's convergence point satisfies optimality conditions is itself a difficult problem in the general case. There are no universal necessary *and* sufficient optimality conditions and the minimum of a smooth function on a real interval may be attained at its end-points where no optimality conditions are satisfied. In the case of constrained optimization in higher dimensions, even for a smooth function, the feasible region's boundary structure is generally complicated.

4 Summary

The proposed algorithm takes advantage of the Rayleigh Quotient Iteration and Conjugate Gradients algorithms, a new generalized matrix quotient, and a reduction that guarantees that the computed eigenvector/eigenvalue pairs constitute the entire spectrum. It yields high-precision results and it does not require preparatory matrix decompositions or eigenvalues estimates (cf. the QR algorithm's "shifts"). Implementing this algorithm does not require the very great amount of understanding that is needed to address all the subtleties that a state-of-the-art implementation of the QR algorithm requires (see Trefethen and Bau [4, p. 338]).

The use of the RQI and CG algorithms is motivated by considerations of rates of convergence of linear solvers near singularities. These considerations and related limitations of optimization algorithms are noted.

Acknowledgements

With thanks to my teacher, A. Ben-Tal, and his teacher, A. Ben-Israel.

References

- [1] Steve Batterson and John Smillie, "Rayleigh Quotient Iteration Fails for Nonsymmetric Matrices," *Applied Mathematics Letters*, Vol. 2, No. 1, pp. 19-20, 1989.

- [2] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins, 2013.
- [3] Beresford N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, 1998.
- [4] Lloyd Trefethen and David Bau, III, *Numerical Linear Algebra*, SIAM, 1997.
- [5] James H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, 1965.