Parallel Numerical Algorithms

Chapter 5 – Eigenvalue Problems
Section 5.2 – Eigenvalue Computation

Michael T. Heath and Edgar Solomonik

Department of Computer Science
University of Illinois at Urbana-Champaign

CS 554 / CSE 512

Outline

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Eigenvalues and Eigenvectors

- Given $n \times n$ matrix A, find scalar λ and nonzero vector x such that

$$Ax = \lambda x$$

- λ is *eigenvalue* and x is corresponding *eigenvector*

- A always has n eigenvalues, but they may be neither real nor distinct

- May need to compute only one or few eigenvalues, or all n eigenvalues

- May or may not need corresponding eigenvectors

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Problem Transformations

- *Shift*: for scalar σ, eigenvalues of $A - \sigma I$ are eigenvalues of A shifted by σ, $\lambda_i - \sigma$

- *Inversion*: for nonsingular A, eigenvalues of A^{-1} are reciprocals of eigenvalues of A, $1/\lambda_i$

- *Powers*: for integer $k > 0$, eigenvalues of A^k are kth powers of eigenvalues of A, λ_i^k

- *Polynomial*: for polynomial $p(t)$, eigenvalues of $p(A)$ are values of p evaluated at eigenvalues of A, $p(\lambda_i)$

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Similarity Transformations

- B is *similar* to A if there is nonsingular T such that

$$B = T^{-1}AT$$

- Then

$$By = \lambda y \;\Rightarrow\; T^{-1}ATy = \lambda y \;\Rightarrow\; A(Ty) = \lambda(Ty)$$

  so A and B have same eigenvalues, and if y is
  eigenvector of B, then $x = Ty$ is eigenvector of A

- Similarity transformations preserve eigenvalues, and
  eigenvectors are easily recovered

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Similarity Transformations

Forms attainable by similarity transformation

A	T	B
distinct eigenvalues	nonsingular	diagonal
real symmetric	orthogonal	real diagonal
complex Hermitian	unitary	real diagonal
normal	unitary	diagonal
arbitrary real	orthogonal	real upper Hessenberg
arbitrary	unitary	upper triangular (Schur)
arbitrary	nonsingular	almost diagonal (Jordan)

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Preliminary Reduction

- Eigenvalues easier to compute if matrix first reduced to simpler form by similarity transformation

- Diagonal or triangular most desirable, but cannot always be reached in finite number of steps

- Preliminary reduction usually to tridiagonal form (for symmetric matrix) or Hessenberg form (for nonsymmetric matrix)

- Preliminary reduction usually done by orthogonal transformations, using algorithms similar to QR factorization, but transformations must be applied from both sides to maintain similarity

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Definitions
Transformations
Parallel Algorithms

Parallel Algorithms for Eigenvalues

- Algorithms for computing eigenvalues and eigenvectors employ basic operations such as
  - vector updates (saxpy)
  - inner products
  - matrix-vector and matrix-matrix multiplication
  - solution of triangular systems
  - orthogonal (QR) factorization

- In many cases, parallel implementations will be based on parallel algorithms we have already seen for these basic operations, although there will sometimes be new sources of parallelism

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Power Iteration

- Simplest method for computing one eigenvalue-eigenvector pair is *power iteration*, which in effect takes successively higher powers of matrix times initial starting vector

- $x_0 =$ arbitrary nonzero vector
  for $k = 1, 2, \ldots$
    $y_k = A x_{k-1}$
    $x_k = y_k / \|y_k\|_\infty$
  end

- If A has unique eigenvalue λ_1 of maximum modulus, then power iteration converges to eigenvector corresponding to dominant eigenvalue

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Power Iteration

- Convergence rate of power iteration depends on ratio $|\lambda_2/\lambda_1|$, where λ_2 is eigenvalue having second largest modulus

- It may be possible to choose shift, $A - \sigma I$, so that ratio is more favorable and yields more rapid convergence

- Shift must then be added to result to obtain eigenvalue of original matrix

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Parallel Power Iteration

- Power iteration requires repeated matrix-vector products, which are easily implemented in parallel for dense or sparse matrix, as we have seen

- Additional communication may be required for normalization, shifts, convergence test, etc.

- Though easily parallelized, power iteration is often too slow to be useful in this form

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Inverse Iteration

- *Inverse iteration* is power iteration applied to A^{-1}, which converges to eigenvector corresponding to dominant eigenvalue of A^{-1}, which is reciprocal of *smallest* eigenvalue of A

- Inverse of A is not computed explicitly, but only factorization of A (and only once) to solve system of linear equations at each iteration

- $x_0 = $ arbitrary nonzero vector
  for $k = 1, 2, \ldots$
    Solve $A y_k = x_{k-1}$ for y_k
    $x_k = y_k / \| y_k \|_\infty$
  end

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Inverse Iteration

- Shifting strategy can greatly accelerate convergence

- Inverse iteration is especially useful for computing eigenvector corresponding to approximate eigenvalue, since it converges rapidly when approximate eigenvalue is used as shift

- Inverse iteration also useful for computing eigenvalue closest to any given value β, since if β is used as shift, then desired eigenvalue corresponds to smallest eigenvalue of shifted matrix

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Parallel Inverse Iteration

- Inverse iteration requires initial factorization of matrix A and solution of triangular systems at each iteration, so it appears to be much less amenable to efficient parallel implementation than power iteration

- However, inverse iteration is often used to compute eigenvector in situations where
  - approximate eigenvalue is already known, so using it as shift yields very rapid convergence
  - matrix has previously been reduced to simpler form (e.g., tridiagonal) for which linear system is easy to solve

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Power Iteration
Inverse Iteration
Simultaneous Iteration

Simultaneous Iteration

- To compute many eigenvalue-eigenvector pairs, could apply power iteration to several starting vectors simultaneously, giving *simultaneous iteration*

  $X_0 =$ arbitrary $n \times q$ matrix of rank q
  for $k = 1, 2, \ldots$
  $\qquad X_k = A X_{k-1}$
  end

- span(X_k) converges to invariant subspace determined by q largest eigenvalues of A, provided $|\lambda_q| > |\lambda_{q+1}|$

- Normalization is needed at each iteration, and columns of X_k become increasingly ill-conditioned basis for span(X_k)

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

Orthogonal Iteration

- Both issues addressed by computing QR factorization at each iteration, giving *orthogonal iteration*

$X_0 =$ arbitrary $n \times q$ matrix of rank q
for $k = 1, 2, \ldots$
   Compute reduced QR factorization
   $\hat{Q}_k R_k = X_{k-1}$
   $X_k = A\hat{Q}_k$
end

- Converges to block triangular form, and blocks are triangular where moduli of consecutive eigenvalues are distinct

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

Parallel Orthogonal Iteration

- Orthogonal iteration requires matrix-matrix multiplication and QR factorization at each iteration, both of which we know how to implement in parallel with reasonable efficiency

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

QR Iteration

- If we take $X_0 = I$, then orthogonal iteration should produce *all* eigenvalues and eigenvectors of A
- Orthogonal iteration can be reorganized to avoid explicit formation and factorization of matrices X_k
- Instead, sequence of unitarily similar matrices is generated by computing QR factorization at each iteration and then forming reverse product, giving *QR iteration*

$$A_0 = A$$
for $k = 1, 2, \ldots$
    Compute QR factorization
       $Q_k R_k = A_{k-1}$
   $A_k = R_k Q_k$
end

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

QR Iteration

- In simple form just given, each iteration of QR method requires $\Theta(n^3)$ work

- Work per iteration is reduced to $\Theta(n^2)$ if matrix is in Hessenberg form, or $\Theta(n)$ if symmetric matrix is in tridiagonal form

- Preliminary reduction is usually done by Householder or Givens transformations

- In addition, number of iterations required is reduced by preliminary reduction of matrix

- Convergence rate also enhanced by judicious choice of shifts

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

Parallel QR Iteration

- Preliminary reduction can be implemented efficiently in parallel, using algorithms analogous to parallel QR factorization for dense matrix

- But subsequent QR iteration for reduced matrix is inherently serial, and permits little parallel speedup for this portion of algorithm

- This may not be of great concern if iterative phase is relatively small portion of total time, but it does limit efficiency and scalability

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

Reduction to Hessenberg/Tridiagonal

- To reduce to Hessenberg can execute QR on each column

- If H_i is the Householder transformation used to annihilate ith subcolumn, perform similarity transformation
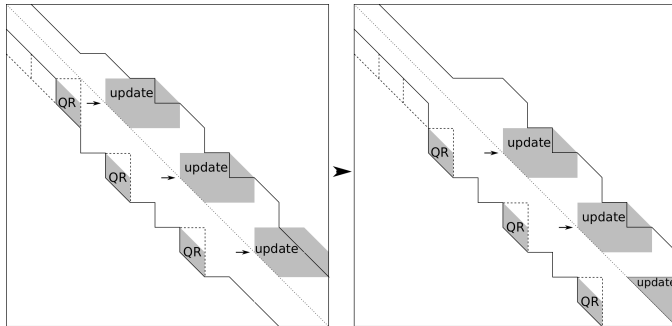
$$A^{(i+1)} = H_i A^{(i)} H_i^T$$

- More generally, run QR on b subcolumns of A to reduce to band-width b

$$B^{(i+1)} = Q_i B^{(i)} Q_i^T$$

- To avoid fill Q_i^T must not operate on the b columns which Q_i reduces

- Once reduction completed to a band-width subsequent eliminations introduce fill (bulges) but can be done

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Orthogonal Iteration
QR Iteration

Reduction from Banded to Tridiagonal

- Two-sided successive orthogonal reductions generate bulges of fill when applied to a banded matrix
- Bulges can be chased with pipelined parallelism

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Krylov Subspace Methods

- *Krylov subspace* methods reduce matrix to Hessenberg (or tridiagonal) form using only matrix-vector multiplication

- For arbitrary starting vector x_0, if

$$\boldsymbol{K}_k = \begin{bmatrix} \boldsymbol{x}_0 & \boldsymbol{A}\boldsymbol{x}_0 & \cdots & \boldsymbol{A}^{k-1}\boldsymbol{x}_0 \end{bmatrix}$$

  then

$$\boldsymbol{K}_n^{-1}\boldsymbol{A}\boldsymbol{K}_n = \boldsymbol{C}_n$$

  where $\boldsymbol{C}_n$ is upper Hessenberg (in fact, companion matrix)

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Krylov Subspace Methods

- To obtain better conditioned basis for span(K_n), compute QR factorization

$$Q_n R_n = K_n$$

so that

$$Q_n^H A Q_n = R_n C_n R_n^{-1} \equiv H$$

with H upper Hessenberg

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Krylov Subspace Methods

- Equating kth columns on each side of equation $AQ_n = Q_nH$ yields recurrence

$$Aq_k = h_{1k}q_1 + \cdots + h_{kk}q_k + h_{k+1,k}q_{k+1}$$

  relating q_{k+1} to preceding vectors $q_1, \ldots, q_k$

- Premultiplying by q_j^H and using orthonormality,

$$h_{jk} = q_j^H A q_k, \quad j = 1, \ldots, k$$

- These relationships yield *Arnoldi iteration*, which produces upper Hessenberg matrix column by column using only matrix-vector multiplication by A and inner products of vectors

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Arnoldi Iteration

x_0 = arbitrary nonzero starting vector
$q_1 = x_0/\|x_0\|_2$
for $k = 1, 2, \ldots$
    $u_k = A q_k$
    for $j = 1$ **to** k
        $h_{jk} = q_j^H u_k$
        $u_k = u_k - h_{jk} q_j$
    end
    $h_{k+1,k} = \|u_k\|_2$
    if $h_{k+1,k} = 0$ **then** stop
    $q_{k+1} = u_k/h_{k+1,k}$
end

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Arnoldi Iteration

- If
$$\boldsymbol{Q}_k = \begin{bmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_k \end{bmatrix},$$

  then

$$\boldsymbol{H}_k = \boldsymbol{Q}_k^H \boldsymbol{A} \boldsymbol{Q}_k$$

  is upper Hessenberg matrix

- Eigenvalues of $\boldsymbol{H}_k$, called *Ritz values*, are approximate eigenvalues of $\boldsymbol{A}$, and *Ritz vectors* given by $\boldsymbol{Q}_k \boldsymbol{y}$, where $\boldsymbol{y}$ is eigenvector of $\boldsymbol{H}_k$, are corresponding approximate eigenvectors of $\boldsymbol{A}$

- Eigenvalues of $\boldsymbol{H}_k$ must be computed by another method, such as QR iteration, but this is easier problem if $k \ll n$

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Arnoldi Iteration

- Arnoldi iteration expensive in work and storage because each new vector q_k must be orthogonalized against all previous columns of Q_k, which must be stored

- So Arnoldi process usually restarted periodically with carefully chosen starting vector

- Ritz values and vectors produced are often good approximations to eigenvalues and eigenvectors of A after relatively few iterations

- Work and storage costs drop dramatically if matrix symmetric or Hermitian, since recurrence then has only three terms and H_k is tridiagonal (so usually denoted T_k), yielding *Lanczos iteration*

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Lanczos Iteration

$q_0 = \mathbf{0}$

$\beta_0 = 0$

$x_0 = $ arbitrary nonzero starting vector

$q_1 = x_0 / \|x_0\|_2$

for $k = 1, 2, \ldots$

    $u_k = Aq_k$

    $\alpha_k = q_k^H u_k$

    $u_k = u_k - \beta_{k-1} q_{k-1} - \alpha_k q_k$

    $\beta_k = \|u_k\|_2$

    if $\beta_k = 0$ **then** stop

    $q_{k+1} = u_k / \beta_k$

end

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Lanczos Iteration

- α_k and β_k are diagonal and subdiagonal entries of symmetric tridiagonal matrix T_k

- As with Arnoldi, Lanczos iteration does not produce eigenvalues and eigenvectors directly, but only tridiagonal matrix T_k, whose eigenvalues and eigenvectors must be computed by another method to obtain Ritz values and vectors

- If $\beta_k = 0$, then algorithm appears to break down, but in that case invariant subspace has already been identified (i.e., Ritz values and vectors are already exact at that point)

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Lanczos Iteration

- In principle, if Lanczos algorithm is run until $k = n$, resulting tridiagonal matrix is orthogonally similar to A

- In practice, rounding error causes loss of orthogonality, invalidating this expectation

- Problem can be overcome by reorthogonalizing vectors as needed, but expense can be substantial

- Alternatively, can ignore problem, in which case algorithm still produces good eigenvalue approximations, but multiple copies of some eigenvalues may be generated

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Krylov Subspace Methods

- Virtue of Arnoldi and Lanczos iterations is ability to produce good approximations to extreme eigenvalues for $k \ll n$

- Moreover, they require only one matrix-vector multiplication by A per step and little auxiliary storage, so are ideally suited to large sparse matrices

- If eigenvalues are needed in middle of spectrum, say near σ, then algorithm can be applied to matrix $(A - \sigma I)^{-1}$, assuming it is practical to solve systems of form $(A - \sigma I)x = y$

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Krylov Subspaces
Arnoldi Iteration
Lanczos Iteration
Krylov Subspace Methods

Parallel Krylov Subspace Methods

- Krylov subspace methods composed of
  - vector updates (saxpy)
  - inner products
  - matrix-vector multiplication
  - computing eigenvalues/eigenvectors of tridiagonal matrices

- Parallel implementation requires implementing each of these in parallel as before

- For early iterations, Hessenberg or tridiagonal matrices generated are too small to benefit from parallel implementation, but Ritz values and vectors need not be computed until later

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Jacobi Method

- *Jacobi method* for symmetrix matrix starts with $A_0 = A$ and computes sequence

$$A_{k+1} = J_k^T A_k J_k$$

where J_k is plane rotation that annihilates symmetric pair of off-diagonal entries in A_k

- Plane rotations are applied repeatedly from both sides in systematic sweeps through matrix until magnitudes of all off-diagonal entries are reduced below tolerance

- Resulting diagonal matrix is orthogonally similar to original matrix, so diagonal entries are eigenvalues, and eigenvectors given by product of plane rotations

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Parallel Jacobi Method

- Jacobi method, though slower than QR iteration serially, parallelizes better

- Parallel implementation of Jacobi method performs many annihilations simultaneously, at locations chosen so that rotations do not interfere with each other (analogous to parallel Givens QR factorization)

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Divide-and-Conquer Method

- One method for computing eigenvalues and eigenvectors of real symmetric tridiagonal matrix is based on *divide-and-conquer*

- Express symmetric tridiagonal matrix T as

$$T = \begin{bmatrix} T_1 & O \\ O & T_2 \end{bmatrix} + \beta \, uu^T$$

- Can now compute eigenvalues and eigenvectors of smaller matrices T_1 and T_2

- To relate these back to eigenvalues and eigenvectors of original matrix requires solution of *secular equation*, which can be done reliably and efficiently

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Divide-and-Conquer Method

- Applying this approach recursively yields divide-and-conquer algorithm that is naturally parallel

- Parallelism in solving secular equations grows as parallelism in processing independent tridiagonal matrices shrinks, and vice versa

- Algorithm is complicated to implement and difficult questions of numerical stability, eigenvector orthogonality, and load balancing must be addressed

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Polar Decomposition

- The polar decomposition uses spectrum slicing to obtain the symmetric eigenvalue decomposition via $O(\log n)$ QR factorizations and matrix multiplications and $O(n^3)$ work

- The polar decomposition of a matrix A is composed of an orthogonal matrix U and symmetric positive semi-definite matrix H, $A = UH$

- It can be computed via the QR-based Dynamically Weighted Halley (QDWH) algorithm, an iterative scheme that starts with $X_0 = A/\alpha$ and computes

$$\begin{bmatrix} \sqrt{c_k} X_k \\ I \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R, \;\; X_{k+1} = \frac{b_k}{c_k} X_k + \frac{1}{\sqrt{c_k}} (a_k - b_k/c_k) Q_1 Q_2^T$$

- With appropriate choices of α, a_k, b_k, c_k, QDWH converges within 6 iterations for double precision

Basics
Power Iteration
QR Iteration
Krylov Methods
Other Methods

Jacobi Method
Divide-and-Conquer
Polar Decomposition

Eigenvalues via Polar Decomposition

- When A is symmetric, the orthogonal matrix U in the polar decomposition is given by the sign function,

$$U = \text{sign}(A) = A(A^2)^{-1/2}$$

- If the eigenvalue decomposition of A is $A = VDV^T$, then

$$U = V\text{sign}(D)V^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} I_k & \\ & -I_{n-k} \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

- Consequently, given an estimate σ of a median eigenvalue of A, the polar decomposition of $A - \sigma I$ allows one to partition the spectrum by finding V_1 from subspace iteration on $U + I = 2V_1 V_1^T$

References

- Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, 2000

- J. W. Demmel, M. T. Heath, and H. A. van der Vorst, Parallel numerical linear algebra, *Acta Numerica* 2:111-197, 1993

- P. J. Eberlein and H. Park, Efficient implementation of Jacobi algorithms and Jacobi sets on distributed memory architectures, *J. Parallel Distrib. Comput.*, 8:358-366, 1990

- G. W. Stewart, A Jacobi-like algorithm for computing the Schur decomposition of a non-Hermitian matrix, *SIAM J. Sci. Stat. Comput.* 6:853-864, 1985

- G. W. Stewart, A parallel implementation of the QR algorithm, *Parallel Comput.* 5:187-196, 1987

References

- Y. Nakatsukasa and N.J Higham, *Stable and efficient spectral divide and conquer algorithms for the symmetric eigenvalue decomposition and the SVD*, SIAM Journal on Scientific Computing, 2013.