Reading course on
Numerical Solution of Eigenvalue Problems

Lecture 5

The Jacobi-Davidson method

Melina Freitag

May 24, 2005

1 Some basics

We have the following standard eigenvalue problem:

$$Ax = \lambda x.$$

We use a so called **Galerkin approach** to solve it:

- Suppose we have a subspace $\mathcal{S}$ which generates an orthogonal basis $q_1, \ldots, q_k$
- With $Q_k := [q_1, \ldots, q_k] \in \mathbb{R}^{n,k}$, a matrix with orthogonal columns: Galerkin condition: find s and θ such that

  $$r := A \underbrace{Q_k s}_{y} - \theta \underbrace{Q_k s}_{y} \perp \{q_1, \ldots, q_k\} = \mathcal{S},\ y \in \mathcal{S}$$

  or

  $$Q_k^H A Q_k s = \theta s,$$

  H_k is the orthogonal projection of A onto $\mathcal{S}$

Definition 1.1. $(\theta, Q_k s) = (\theta, y)$ *is called a* Ritz pair *associated with the subspace (search space)* $\mathcal{S} = span\{q_1, \ldots, q_k\}$. (θ, y) *with* Ritz residual r *approximates the eigenpair* (λ, x) *of* A.

In practice we want $k << n$!

- Generate an orthonormal system $q_1, \ldots, q_k$ and we wish to add q_{k+1}
- find an expansion vector v for the subspace Q_k
- expand the subspace by orthogonalisation of v against $q_1, \ldots, q_k$ (modified Gram-Schmidt)
- solve the slightly bigger projected problem

  $$Q_{k+1}^H A Q_{k+1} s = \theta s,$$

Remark 1.1. *Modified Gram-Schmidt and repeated Gram-Schmidt (for reorthogonalisation) are used in practice.*

There are different choices for the expansion vector v, assuming we have the initial subspace span$\{q_1\}$.

- *Arnoldi's method*: $v = Aq_k$, then $H := Q_k^H A Q_k$ is upper Hessenberg

Algorithm 1 Subspace iteration

choose initial subspace Q_1
for $j = 1, 2$ **do**
  $W_j \to AQ_j$, $H_j \to Q_j^H W_j$
  Compute desired eigenpair (θ, s) of H_j, with $\|s\| = 1$
  $y \to Q_j s$
  $r \to Ay - \theta y$
  Stop if satisfied
  Compute an expansion vector v
  Expand subspace $Q_{j+1} \to ModGS[Q_j, v]$
end for

- *Lanczos' method*: (for $A = A^H$) $v = Aq_k$, then $H := Q_k^H A Q_k$ is tridiagonal
- for both these methods the search space span$\{Q_k\}$ = span$\{q_1, Aq_1, \ldots, A^{k-1}q_1\}$ is a so-called *Krylov subspace*
- Arnoldi and Lanczos favour extremal eigenvalues:
- Shift-and-Invert Arnoldi: $v = (A - \tau I)^{-1} q_k$: approach favours eigenvalues close to τ
- for large problems $(A - \tau I)v = q_k$ is expensive and has to be done accurately!!

2 Davidson's method

- expand the search space span$\{q_1, \ldots, q_k\}$ in the direction

  $$v = (D_A - \theta I)^{-1} r$$

  where D_A is the diagonal of A.
- q_{k+1} is obtained by orthogonalisation of v w.r.t. span$\{Q_k\}$.
- used for strongly diagonal dominant matrices
- problem: for diagonal matrices

  $$v = (D_A - \theta I)^{-1} r = y \in \text{span}\{Q_j\}$$

  does not lead to the expansion of the search space span$\{Q_j\}$.

Algorithm 2 Davidson's method

choose q_1 with $\|q_1\| = 1$, $Q_1 = [q_1]$
for $j = 1, 2$ **do**
  $w_j = Aq_j$
  for $k = 1$ to $j - 1$ **do**
    $b_{kj} = q_k^H w_j$
    $b_{jk} = q_j^H w_k$
  end for
  $b_{jj} = q_j^H w_j$
  Compute largest eigenpair (θ, s) of B, with $\|s\| = 1$
  $y = Q_j s$
  $r = Ay - \theta y$
  $v = (D_A - \theta I)^{-1} r$
  $v = v - Q_j Q_j^H v$
  $q_{j+1} = \frac{v}{\|v\|}$
  $Q_{j+1} = [Q_j, q_{j+1}]$
end for

3 Jacobi's method

- Let A be diagonal dominant and $\alpha = a_{11}$ the largest diagonal element. The α is an approximation of the largest eigenvalue λ and e_1 is an approximation for the corresponding eigenvector q.

- Hence the problem

$$A \begin{bmatrix} 1 \\ z \end{bmatrix} = \begin{bmatrix} \alpha & c^T \\ b & F \end{bmatrix} \begin{bmatrix} 1 \\ z \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ z \end{bmatrix}$$

- interested in eigenvalue that is close to α and in corresponding eigenvector $q = [1, z^T]^T$ with z the component of q orthogonal to e_1

- equivalent is

$$\begin{aligned} \lambda &= \alpha + c^T z \\ (F - \lambda I)z &= -b \end{aligned}$$

- apply Jacobi iteration with $z_1 = 0$

$$\begin{aligned} \theta_k &= \alpha + c^T z_k \\ (D - \theta_k I) z_{k+1} &= (D - F) z_k - b \end{aligned}$$

  where D is the diagonal entry of F.

- at all stages we look for the orthogonal complement to the initial approximation $q_1 = e_1$, not taking into account that better approximations $q_k = [1, z_k^T]^T$ become available at each state (it may be more efficient to calculate $q - (q^T q_k) q_k$).

4 The Jacobi-Davidson method

We try to find an *optimal expansion* of the subspace!

- Jacobi and Davidson attempt to find corrections to some initially given eigenvector approximation , they both use fixed operators
- Jacobi-Davidson: find the orthogonal complement for current approximation y_k with respect to the desired eigenvector A
- let y_k be an approximation to the eigenvector of A and θ_k the corresponding Ritz value, i.e. $y_k \in \mathcal{S}$, where $\mathcal{S}$ is a low k-dimensional subspace (the search space)
- interested in seeing what happens in the subspace $y_k^\perp$
- orthogonal projection of A onto that subspace is given by (with $\|y_k\| = 1$)

  $$B = (I - y_k y_k^H) A (I - y_k y_k^H).$$

  B is the restriction of A to the subspace orthogonal to y_k. (Note that for $y_k = e_1$ we have that F is the restriction of B with respect to $e_1^\perp$).
- with $\theta_k = y_k^H A y_k$ it follows that

  $$A = B + A y_k y_k^H + y_k y_k^H A - \theta_k y_k y_k^H$$

- Look for an eigenvalue λ of A close to θ_k, we want to have the correction $v \perp y_k$ (orthogonal correction) such that

  $$A(y_k + v) = \lambda(y_k + v),$$

  or with $B y_k = 0$

  $$(B - \lambda I)v = -r + (\lambda - \theta_k - y_k^H A v) y_k$$

- Since LHS and r are orthogonal to y_k the last term must be zero and hence the correction satisfies
$$(B - \lambda I)v = -r$$
or
$$(I - y_k y_k^H)(A - \lambda I)(I - y_k y_k^H)v = -r.$$

- replace unknown λ by known θ_k if approximate eigenvalue is already good enough or by some target τ

- *JD Correction equation:*
$$\underbrace{(I - y_k y_k^H)(A - \theta_k I)(I - y_k y_k^H)}_{B-\theta_k I} v = -r. \tag{1}$$
i.e. $A - \theta_k I$ is restricted to the orthogonal complement of y_k

- expand the subspace by v (using GS or modification) and compute new Ritz pair in expanded subspace

- Combination of the Jacobi approach of looking for the orthogonal complement of a given eigenvector approximation and Davidson's algorithm for expanding the subspace in which the eigenvector approximations are constructed

Remark 4.1. *Modified Gram-Schmidt and repeated Gram-Schmidt (for reorthogonalisation) are used in practice.*

5 Jacobi-Davidson as Newton-method/Rayleigh quotient iteration

If we solve the correction equation (1) *exactly*, then, since $v \perp y_k$ we have $(I - y_k y_k^H)v = v$ and
$$(A - \theta_k I)v = -r + \alpha y_k \tag{2}$$
where $\alpha \in \mathbb{C}$ s.t. $v \perp y_k$. Then
$$v = -(A - \theta_k I)^{-1} r + \alpha (A - \theta_k I)^{-1} y_k = -y_k + \alpha (A - \theta_k I)^{-1} y_k.$$
The solution v is used to expand the search space.

- y_k is already in the search space the expansion vector is effectively $(A - \theta_k I)^{-1} y_k$, which is the same as for *inverse iteration* for fixed θ_k and the same as for *RQI* for $\theta_k = y_k^H A y_k$ (from Ritz residual $r = A y_k - \theta_k y_k$.

Algorithm 3 Jacobi-Davidson method

choose q_1 with $\|q_1\| = 1$, $Q_1 = [q_1]$
for $j = 1, 2$ **do**
  $w_j = Aq_j$
  for $k = 1$ to $j - 1$ **do**
    $b_{kj} = q_k^H w_j$
    $b_{jk} = q_j^H w_k$
  end for
  $b_{jj} = q_j^H w_j$
  Compute largest eigenpair (θ, s) of B, with $\|s\| = 1$
  $y = Q_j s$
  $r = Ay - \theta y$
  if $\|r\| \leq tol$ **then**
    $\lambda = \theta$, $x = y$, STOP
  end if
  Solve (approximately)

$$(I - yy^H)(A - \theta I)(I - yy^H)v = -r.$$

  $v = v - Q_j Q_j^H v$
  $q_{j+1} = \frac{v}{\|v\|}$
  $Q_{j+1} = [Q_j, q_{j+1}]$
end for

- JD where we solve correction equation exactly is a *subspace accelerated* inverse iteration or RQI
- subspace acceleration: $(A-\theta_k I)^{-1}y_k$ does not directly give the next approximate eigenvector, a hopefully even better approximation is sought in the subspace formed by $\mathcal{S}$ expanded by this new vector.
- convergence of RQI: quadratic or cubic for Hermitian matrices
- JD can also be viewed as a Newton's method: equation (2):

$$(A-\theta_k I)v = -r + \alpha y_k, \quad v \perp y_k$$

can be written as

$$\begin{bmatrix} A-\theta_k I & y_k \\ y_k^H & 0 \end{bmatrix} \begin{bmatrix} v \\ -\alpha \end{bmatrix} = \begin{bmatrix} -r \\ 0 \end{bmatrix}$$

- generally quadratic convergence (for Hermitian problems even cubic)

Now, the key idea of JD is to solve the correction equation (1) only *inexactly* by an iterative method.

- JD combined with an iterative solver: accelerated *inexact* Newton's method or accelerated *inexact* Inverse Iteration/RQI (quadr. convergence..)
- numerical observation: even for approximate solution of the correction equation (1) we get quite fast convergence

6 Solution of the correction equation

Iterative solvers are used to solve the large linear system:

- CG (Hermitian positive definit), MINRES (Hermitian), GMRES
- use preconditioner K for $A-\theta_k I$, i.e. $K^{-1}(A-\theta_k I) \approx I$, but usually fixed
- have to restrict K to the same subspace

$$\tilde{K} = (I - y_k y_k^H)K(I - y_k y_k^H)$$

- for a Krylov solver, in each step we have to find a vector $z = \tilde{K}^{-1}\tilde{A}w$, with

$$\tilde{A} = (I - y_k y_k^H)(A - \theta_k I)(I - y_k y_k^H)$$

- First
$$\tilde{A}w = (I - y_k y_k^H)(A - \theta_k I)(I - y_k y_k^H)w = (I - y_k y_k^H)g$$
with $g = (A - \theta_k I)w$, since $y_k^H w = 0$.

- Then solve
$$\tilde{K}z = (I - y_k y_k^H)g,$$
and since $z \perp y_k$ we have
$$Kz = g - \beta y_k, \quad z = K^{-1}g - \beta K^{-1} y_k$$
and with $z \perp y_k$ we get
$$\beta = \frac{y_k^H K^{-1} g}{y_k K^{-1} y_k},$$
so in each step of the Krylov solver the system $K\tilde{g} = g$ has to be solved plus $K\tilde{y} = y_k$ at the beginning.

7 Interior eigenvalues - Harmonic Ritz values

Definition 7.1 (Harmonic Ritz value)**.** *Let τ be a complex (target) value that is not an eigenvalue of A. Then μ is a* Harmonic Ritz value *of A with target τ w.r.t. the space $\mathcal{S}$ if $(\mu - \tau)^{-1}$ is an ordinary Ritz value of $(A - \tau I)^{-1}$.*

- Revising the theory for Ritz values μ satisfies this property if and only if $(\mu - \tau)^{-1}$ is an eigenvalue of $Q^H(A - \tau I)^{-1}Q$, where span$\{Q\}$ is a basis of $\mathcal{S}$

- difficult to evaluate if A is large, because it involves solving a system

- obtain harmonic Ritz value w.r.t. another subspace $\mathcal{U} = (A - \tau I)\mathcal{S}$

- μ is a harmonic Ritz value of A with target τ with respect to the space $\mathcal{U}$ if and only if there is a $u \in \mathcal{U}$ s.t.
$$(A - \tau I)^{-1}u - (\mu - \tau)^{-1}u \perp \mathcal{U}$$

- with $u = (A - \tau I)Qs$, where s is uniquely determined we get
$$(A - \mu I)Qs \perp \mathcal{U}$$

- Let $v = Qs$. Then v is called the *harmonic Ritz vector* associated with the Ritz value μ

- finally, with $Y = (A - \tau I)Q \in \mathbb{C}^{n,k}$ we get

$$Y^H Y s = (\mu - \tau) Y^H Q s$$

which is a small $(k \times k)$ generalized eigenproblem, which has to be solved

- get μ, the harmonic Ritz value and $v = Qs$ the harmonic Ritz vector

8 Remarks

- **Restarts**: dispose the less promising vectors to reduce amount of storage, suppose we have $m = k + j$ orthonormal vectors, the colums of the matrix $Q \in \mathbb{C}^{n,m}$ and we want to discard j colums and keep k-dimensional subspace, let $H = Q^H A Q$ and compute Schur decomposition B $= UTU^H$, U unitary, T upper triangular with Ritz values on the diagonal, order them in a way s.t.

$$T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

where T_{11} contains the k most promising Ritz values, let $\hat{Q} = QU$ and partition $\hat{Q} = [\hat{Q}_1, \hat{Q}_2]$, where $\hat{Q}_1 \in \mathcal{C}^{n,k}$ is the vector we keep (*purging*)

- **Deflation**: used if an eigenpair (λ, x) is detected and we would like to find other pairs, use only subspaces that are spanned by the remaining vector, orthogonal deflation replaces A by $(1 - xx^H)A(1 - xx^H)$ after finding x, project out the converged subspaces, natural in the Jacobi-Davidson context, more general: replace A by $(1 - ZZ^H)A(1 - ZZ^H)$, where $AZ = ZS$ is a partial Schur decomposition of A with $Z \in \mathcal{C}^{n,k}$ orthonormal Schur vectors and $S \in \mathcal{C}^{k,k}$ upper triangular with eigenvalues of A on diagonal

- **Deflated Preconditioning**: is not much harder than calculating just one extremal eigenvalue, let (θ, y) be the current Ritz pair, then with $\tilde{U} = [Z, y]$ we have

$$\tilde{A} = (I - \tilde{U}\tilde{U}^H)(A - \theta I)(I - \tilde{U}\tilde{U}^H)$$

and for the preconditioner

$$\tilde{K} = (I - \tilde{U}\tilde{U}^H)K(I - \tilde{U}\tilde{U}^H)$$