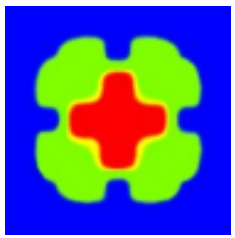
Approximation of Eigenvalue Problems II
Solving the discrete eigenvalue problem

Guido Kanschat

IWR, Universität Heidelberg



IWR School 2017: Mathematical Methods in Quantum Mechanics

1. Vector iterations
2. Krylov space methods
3. The Jacobi-Davidson Method

Literature: algebraic EVP

- Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, H. van der Vorst, eds. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide.* SIAM. 2000
- Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* SIAM. 2011

The standard EVP

$$Au = \lambda u$$

- Eigenvalues: λ_k
- Eigenvectors: u_k
- Eigenspaces U_k

The generalized EVP

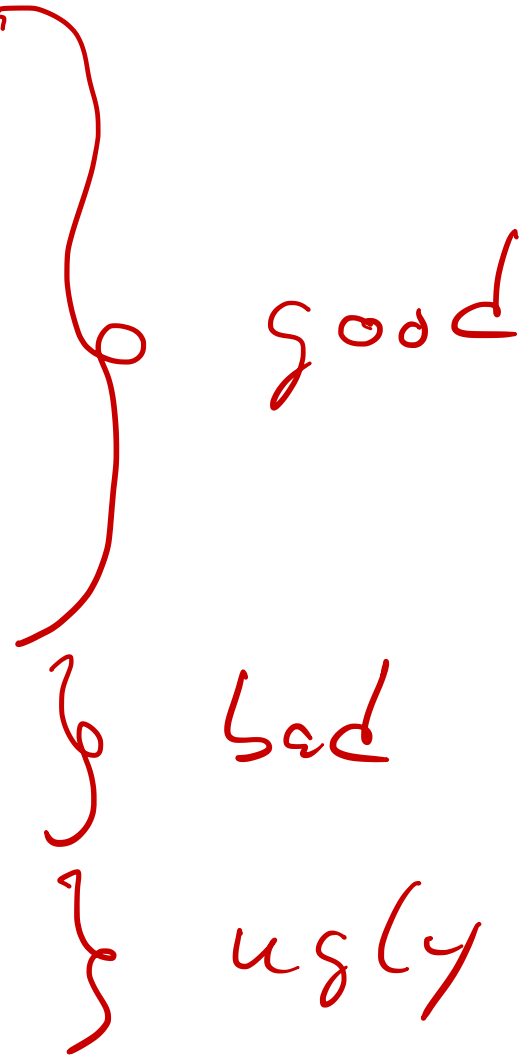
$$Au = \lambda Mu$$

- Generally, $A, M \in C^{n\times n}$
- Variational EVP always lead to the second form
  - M symmetric (Hermitian), positive definite
  - semi-definite possible

IWR

Properties of A

- Hermitian (symmetric)
  - $\mathbb{C}^n = \text{span}(u_1, \ldots, u_n)$
  - $(u_i, u_j) = \delta_{ij}$
  - λ_k real
- normal
  - $\mathbb{C}^n = \text{span}(u_1, \ldots, u_n)$
  - $(u_i, u_j) = \delta_{ij}$
  - λ_k complex
- diagonalizable
  - $\mathbb{C}^n = \text{span}(u_1, \ldots, u_n)$
  - λ_k complex
- other
  - less than n eigenvectors
  - $\mathbb{C}^n \neq \text{span}(u_1, \ldots, u_n)$



Restriction to real, symmetric matrices

- Real eigenvalues and eigenvectors
- Orthogonality: for any $v \in \mathbb{R}^n$ holds

Parseval

$$v = \sum_{k=1}^{n} \alpha_k u_k, \qquad \|v\|^2 = \sum_{k=1}^{n} \alpha_k^2$$

- Even if we do not know the eigenvectors in advance, knowing their existence helps

Power iteration (von Mises)

Initial vector v
for $i := 1, \ldots$ **do**
$v := Av$
end for
$\lambda := \frac{a(v,v)}{(v,v)}$

Normalization

- Vectors may grow or shrink beyond numerical range
- Normalization in every step or every few steps

Power iteration (von Mises)

Initial vector v
for $i := 1, \ldots$ **do**
    $v := Av$
    $v /= \|v\|$
end for
$\lambda := a(v, v)$

Power iteration with stopping criterion

Initial vector v with $\|v\| = 1$

repeat

$w := Av$

$\theta := (w, v)$

$r = w - \theta v$

$v = \frac{1}{\|w\|} w$

until $\|r\| \leq \varepsilon$

$\lambda := \theta$

$u := v$

Convergence estimate

- Use the fact that A is orthogonally diagonalizable

$$v = \sum_{k=1}^{n} \alpha_k u_k \qquad Av = \lambda_n \left(\alpha_n u_n + \sum_{k=1}^{n-1} \frac{\lambda_k}{\lambda_n} \alpha_k u_k \right)$$

- Renormalize for simplicity $\alpha_n = 1$
- Result of step j

$$v^{(j)} = u_n + \mathcal{O}\left(\left(\frac{\lambda_{n-1}}{\lambda_n} \right)^j \right)$$

$$\theta^{(j)} = \lambda_n \left(1 + \mathcal{O}\left(\left(\frac{\lambda_{n-1}}{\lambda_n} \right)^{2j+1} \right) \right)$$

$\alpha_n = 0$ cured by floating point arithmetic

$|\lambda_n| \neq |\lambda_{n-1}|$

- Residual

$$w^{(j)} - \theta^{(j)} v^{(j)} = \lambda_n \mathcal{O}\left(\left(\frac{\lambda_{n-1}}{\lambda_n} \right)^{j+1} \right)$$

IWR

Spectral transformations

- Power iteration can only approximate the largest eigenvalue

Spectral transformations

- Power iteration can only approximate the largest eigenvalue
- Smallest eigenvalue: apply iteration to A^{-1}

Spectral transformations

- Power iteration can only approximate the largest eigenvalue
- Smallest eigenvalue: apply iteration to A^{-1}
- Convergence speed depends on ratio of the two largest eigenvalues
- Shift and invert: choose σ close to the desired eigenvalue λ_* and apply iteration to

$$(A - \sigma I)^{-1} \qquad \mu_* = \frac{1}{\lambda_* - \sigma} \ll \frac{1}{\lambda_k - \sigma} = \mu_k$$

Inverse iteration (Wieland)

Initial vector v with $\|v\| = 1$
repeat
  Solve: $(A - \sigma I)w := v$
  $\theta := (w, v)$
  $r = w - \theta v$
  $v := \frac{1}{\|w\|} w$
until $\|r\| \leq \varepsilon$
$\lambda := \sigma + \frac{1}{\theta}$
$u := v$

Adapt shift in every step

Rayleigh quotient iteration

Initial vector v with $\|v\| = 1$
Initial shift σ
repeat
  Solve: $(A - \sigma I)w := v$
  if singular **then stop**
  $\theta := \|w\|$
  $\sigma += \frac{(w,v)}{\theta^2}$
  $v := \frac{1}{\theta} w$
until $\theta \geq \varepsilon^{-1/2}$
$\lambda := \sigma$
$u := v$

Eigenvalues converge third order!

How to compute the second eigenvalue?

How to compute the second eigenvalue?

- Iterate two vectors at a time: $V = (v_1, v_2) \in \mathbb{R}^{n\times 2}$

$\tilde{V} = (Av_1, Av_2) \in \mathbb{R}^{n\times 2}$ $\quad \tilde{V} = AV$

How to compute the second eigenvalue?

- Iterate two vectors at a time: $V = (v_1, v_2) \in \mathbb{R}^{n\times 2}$

$$\tilde{V} = AV$$

- Orthogonalize and normalize (Gram-Schmidt)

$$v_1 = \frac{1}{\|w_1\|} w_1 \qquad w_1 = \tilde{v}_1$$

$$v_2 = \frac{1}{\|w_2\|} w_2 \qquad w_2 = \tilde{v}_2 - (\tilde{v}_2, v_1)\, v_1$$

Gram-Schmidt revisited

- Simultaneous iteration of k vectors

$$\begin{aligned}
\gamma_1 v_1 &= \tilde{v}_1 \\
\gamma_2 v_2 + \beta_{21} v_1 &= \tilde{v}_2 \\
&\vdots \\
\gamma_k v_k + \beta_{k1} v_1 + \cdots + \beta_{k,k-1} v_{k-1} &= \tilde{v}_k
\end{aligned}$$

- Matrix form

$$VR = \tilde{V}, \qquad R = \begin{pmatrix} \gamma_1 & \beta_{21} & \cdots & \beta_{k1} \\ & \gamma_2 & \cdots & \beta_{k2} \\ & & \ddots & \\ & & & \gamma_k \end{pmatrix}$$

Simultaneous vector iteration (truncated QR method)

Initial matrix $V \in \mathbb{R}^{n\times k}$ with orthogonal columns
while true **do**
    $W := AV$
    $H := V^T W$
    if $\|W - VH\| \leq \varepsilon$ **then stop**
    QR-factorize: $VR = W$
end while
$\lambda_i = h_{ii}$
$u_i = V_{\cdot,i}$ or QR-factorize $VR = W$

- Can be applied with shift and invert
- Orthogonalization can be done every m steps

Polynomial acceleration

- Assumption: Some eigenvalues are less important
- Replace the product $A^m V$ by the polynomial $p_m(A)$
- Choose p_m such that it is
  - small where eigenvalues are not important
  - large where they are important
- Example: Chebyshev polynomials to suppress eigenvalues in $[\sigma - \rho, \sigma + \rho]$:

$$T_m\left(\frac{A - \sigma I}{\rho}\right)$$

Implicit shift

- In each step, QR-factorize $A - \sigma I$
- If $\sigma = \lambda$, then

$$RQ = \begin{pmatrix} \ddots & & & \\ \ddots & \ddots & & \\ & \ddots & \ddots & \\ 0 & \cdots & 0 & \lambda \end{pmatrix}$$

- The eigenvector gets reproduced exactly
- Find shifts close to λ_k to accelerate convergence

Remarks on the generalized EVP

$$Au = \lambda Mu$$

- Apply the vector iterations to

$$C = M^{-1}A$$

  - Not symmetric
- Cholesky decomposition $M = LL^T$ if M is s.p.d.

$$C = L^{-1}AL^{-T}$$

$M^{-1} = L^{-T}L^{-1}$

- Symmetric
- Stability if M is ill-conditioned?

IWR

Inverse iterations can be rewritten by changing inner product

Example: Inverse iteration for generalized EVP

Initial vector v with $\|v\|_M = 1$

repeat

Solve: $(A - \sigma M)w := Mv$

$\theta := (w, v)_M$

$r = w - \theta v$

$v := \frac{1}{\|w\|} w$

until $\|r\| \leq \varepsilon$

$\lambda := \sigma + \frac{1}{\theta}$

$u := v$

Why are we solving generalized EVP?

- We have to distinguish between the eigenvalues of the operator $A : H \to H^*$

$$a(u, v) = \lambda (u, v)$$

and the matrix A which is generated by applying the bilinear form to the basis functions

$$Ax = \lambda x$$

- The second depend on the basis, the first do not
- The first make sense in H, the second only discretely
- The Ritz isomorphism is missing in the second EVP

Matrix structures (FEM)

- Basis typically large, $n \approx 10^7$
- Applying matrix is cheap $O_m = n \ll n^2$

Optimal complexity

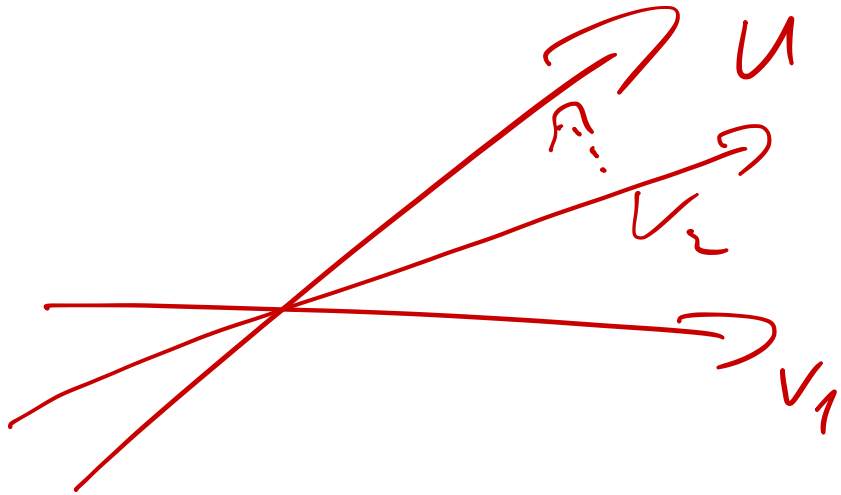
A single step of an iterative method is of optimal complexity, if its operation count is of the same order as O_m.

- Power iteration is of optimal complexity
- Inverse iteration is of optimal complexity if the linear solver is
- The truncated QR method is of optimal complexity
- The full QR method is not

1. Vector iterations

2. **Krylov space methods**
   - Basic idea
   - The Lanczos method
   - Implicitly restarted Lanczos

3. The Jacobi-Davidson Method



- Krylov space methods are projection methods
  - Approximate large EVP by exact solution to small EVP
  - Approximation space growing in each step
  - Ritz-Galerkin method

Krylov space

$$\mathcal{K}_m(A, v) = \text{span}\left(v, Av, \ldots, A^{m-1}v\right)$$

- Krylov spaces are generated by subsequent powers of A
- The vectors $A^j v$ become more and more parallel as j increases
- Methods must maintain an orthogonal basis for stability

Notation

- V_m is orthonormal basis of $\mathcal{K}_m(A, v)$
- The projected matrix

$$T_m = V_m^T A V_m$$

- (θ_k, s_k) eigenpairs of T_m
- Ritz vectors

$$u_k^{(m)} = V_m^T s_k$$

The Lanczos method

Initial vector r

for $j = 1, 2, \ldots$ **do**

$\beta_{j-1} = \|r\|_M$

$v_j = \frac{1}{\beta_{j-1}} r$

$r = Av_j$ or solve $(A - \sigma M)r = v_j$

$r -= \beta_{j-1} v_{j-1}$

$\alpha_j = (v_j, r)_M$

$r -= \alpha_j v_j$

Gram-Schmidt

$r \perp v_{j-1}$

$r \perp v_j$

Optional reorthogonalization

Compute T_j and diagonalize $T_j = S\Theta^{(j)}S^T$

if converged **then stop**

end for

- Variant with shift and invert
- Variant for generalized EVP
  - r is M-orthogonal to V_j

The projected system

$$T_m = \begin{pmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{m-1} \\ & & \beta_{m-1} & \alpha_m \end{pmatrix}$$

- Tridiagonal, symmetric
- QR factorization particularly simple
- Recursive construction

$$AV_j = V_j T_j + r_j e_j^T, \qquad V_j^T r_j = 0$$

Stopping criterion

- Approximation of Ritz pair $(\theta_k, v_k^{(m)} = V_m s_k)$

$$r_k^{(m)} = A v_k^{(m)} - \theta_k v_k^{(m)} = (AV_m - V_m T_m) s_k = v_{m+1} \beta_m s_{m,k}$$

- Estimate

$$\left\| r_k^{(m)} \right\| = |\beta_m s_{m,k}|$$

Reorthogonalization

- Not necessary for conjugate gradient method
  - Only last approximation counts
- Recursive orthogonality suffers as soon as an eigenvalue is well approximated
  - roundoff errors dominate
- Reorthogonalization of the whole basis V_m necessary

?

Multiple eigenvalues

- The Lanczos method works with multiple eigenvalues
  - Only one eigenvector in the eigenspace
  - Depending on initial vector
- Locking of eigenvectors
  - Flag eigenpairs as converged if residual is small
  - Do not iterate those anymore
  - Keep vectors in orthogonalization
  - Roundoff errors force additional eigenvectors

Restarting

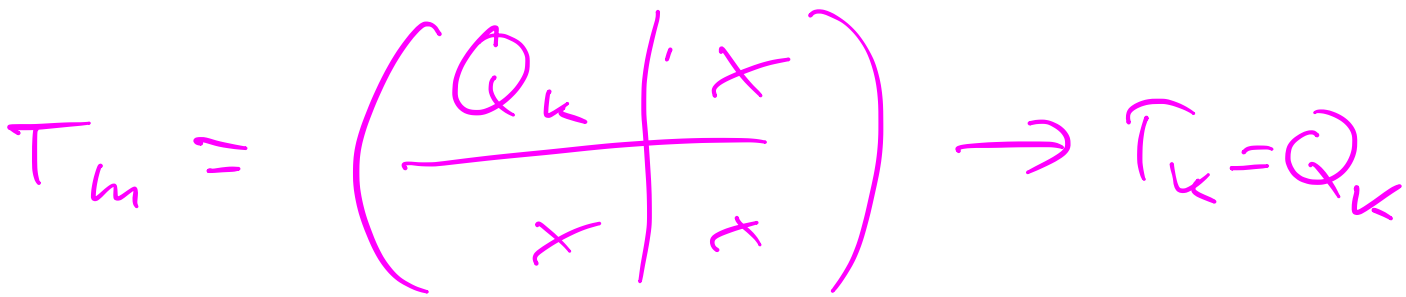
- The Lanczos basis may become very large before convergence
  - Reorthogonalization costly
- Solution: cutting off the basis (restart)
- Explicit restart
  - Delete basis after m steps
  - Begin with last iterate
- Implicit restart
  - Keep part of the basis and fill up again

Implicit restart

- Split Krylov space of size $m = k + p$ into
  - k desired Ritz vectors
  - p undesired Ritz vectors
- Selection criterion by selecting from $\theta_1, \ldots, \theta_m$
  - the largest Ritz values
  - other choices

Implementation of implicit restart

1. Compute until Krylov space has dimension $m = k + p$
2. Apply implicitly shifted QR with the p undesired Ritz values
   - Eigenvalues of T_m with these eigenvalues will be in the end
   - First k vectors will be more desirable Ritz vectors



Implicitly restarted Lanczos method

Initial vector r_0
repeat
  Compute Lanczos basis V_m and projection T_m
  Compute spectrum of T_m
  Choose shifts $\theta_{j_1}, \ldots, \theta_{j_p}$
  $Q = I_m$
  for $i := 1, \ldots, p$ **do**
    QR-factorize: $Q_i R_i = T_m - \theta_{j_i} I$
    $T_m := Q_i^T T_m Q_i$
    $Q := Q Q_i$
  end for
  $r_k = t_{k+1,k} v_{k+1} + q_{mk} r_m$
  Truncate $T_m \longrightarrow T_k \in \mathbb{R}^{k \times k}$ and $Q \longrightarrow Q \in \mathbb{R}^{m \times k}$
  $V_k = V_m Q$
until T_k almost diagonal

Locking and implicit restart

1. Begin locking Ritz values θ whenever the residual is small
2. After k values have been locked and a new one converges
   1. If is more desirable than one of the previous
      - lock it and
      - release the least desirable locked one
   2. If it is not more desirable than any, replace v_{k+1} by a random initial vector and start from there
3. End if the random restart failed twice

randomized SVD ?

Implicitly restarted Arnoldi method

The same method for non-Hermitian EVP, yielding a projected matrix not in tridiagonal, but in Hessenberg form.

1. Vector iterations
2. Krylov space methods
3. **The Jacobi-Davidson Method**

Jacobi-Davidson for a single eigenvalue

Initial vector w
$v_0 = \frac{1}{\|w\|} w$
for $m := 1, \ldots$ **do**
  $v_m = \frac{1}{\|w\|} w$
  $T = V_m^T A V_m$
  Compute largest eigenpair (θ, s) of T
  $u = V_m s$
  $r = Au - \theta u$
  if $\|r\| < \varepsilon$ **then stop**
  Solve approximately $P_u^\perp (A - \theta I) P_u^\perp w = -r$
  Orthogonalize w with respect to V_m
end for

The projected system

- The Galerkin projection of A can be built recursively adding one row and column

$$T_m = V_m^T A V_m \left(\begin{array}{c|c} T_{m-1} & v_m^T A V_{m-1} \\ \hline V_{m-1}^T A v_m & v_m^T A v_m \end{array} \right)$$

- Since the vectors v_j are not obtained by short recursion, the matrix is full
- Full QR method needed

The new search direction

- Search orthogonal to the current Ritz vector u:

$$P_u^\perp = (I - uu^T)$$

- Solves the problem

$$(I - uu^T)(A - \vartheta I)(I - uu^T)\, w = -r$$

$$(A - \theta I) w = -r$$

  on the subspace orthogonal to u.
- $\|u\| = 1$ and the eigenvector has length one
  - Search tangential to the unit sphere
- Exact solution not necessary, since projected matrix is computed explicitly
  - Can be replaced by any other operator

Another view to the Davidson method

- The shift θ for inversion is adapted in each step
  - Projection version of the Rayleigh quotient iteration
- The matrix $A - \theta I$ becomes more and more singular as θ improves
  - Condition number grows
  - The projection improves the condition number

Jacobi-Davidson for several eigenvalues

Initial vector w

$v_0 = \frac{1}{\|w\|} w$

for $m := 1, \ldots$ **do**

$v_m = \frac{1}{\|w\|} w$

$T = V_m^T A V_m$

Compute largest eigenpairs (θ_k, s_k) of T

$u_k = V_m s_k, \quad k = 1, \ldots, \ell$

$r_k = A u_k - \theta u_k, \quad k = 1, \ldots, \ell$

if $\|R\| < \varepsilon$ **then stop**

Solve approximately $P_U^\perp (A - \theta I) P_U^\perp w = -r$

Orthogonalize w with respect to V_m

end for

Search direction

- The Ritz vectors form an orthogonal matrix

$$U = (u_1, \ldots, u_\ell) \in \mathbb{R}^{n \times \ell}$$

- The new orthogonal projection is

$$P_U^\perp = (I - UU^T)$$

- θ is still chosen as the largest Ritz value

Further improvements

- Locking of converged eigenvalues
- Implicit restart

Summary

- Vector iterations
  - Amplify desired eigenvectors by applying matrix powers
- Projection methods
  - Accelerate by solving projected problems exactly
  - Projected problems are of dimension greater one
  - Lanczos accelerating power and inverse iteration
  - Davidson accelerating Rayleigh quotient iteration

IWR