Iterative projection methods for general nonlinear eigenproblems

Heinrich Voss

voss@tu-harburg.de

Hamburg University of Technology
Institute of Mathematics

TUHH
Technische Universität Hamburg-Harburg

Outline

- Nonlinear eigenvalue problems

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods
  - Arnoldi method

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods
  - Arnoldi method
  - Jacobi–Davidson method

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods
  - Arnoldi method
  - Jacobi–Davidson method
  - Variants of Jacobi–Davidson

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods
  - Arnoldi method
  - Jacobi–Davidson method
  - Variants of Jacobi–Davidson
- Minmax characterization for nonlinear eigenproblems

Outline

- Nonlinear eigenvalue problems
- Iterative projection methods
  - Arnoldi method
  - Jacobi–Davidson method
  - Variants of Jacobi–Davidson
- Minmax characterization for nonlinear eigenproblems
- Numerical examples

Nonlinear eigenvalue problems

Let $D \subset \mathbb{C}$ be an open set (maybe unbounded), and let $T(\lambda) \in \mathbb{C}^{n\times n}$, $\lambda \in D$ be a family of matrices.

Find $\lambda \in D$ and $x \neq 0$ such that

$$T(\lambda)x = 0.$$

Then λ is called an eigenvalue of $T(\cdot)$, and x a corresponding eigenvector.

Nonlinear eigenvalue problems

Let $D \subset \mathbb{C}$ be an open set (maybe unbounded), and let $T(\lambda) \in \mathbb{C}^{n\times n}$, $\lambda \in D$ be a family of matrices.

Find $\lambda \in D$ and $x \neq 0$ such that

$$T(\lambda)x = 0.$$

Then λ is called an eigenvalue of $T(\cdot)$, and x a corresponding eigenvector.

Problems of this type arise in vibrations of conservative gyroscopic systems, damped vibrations of structures, problems with retarded argument, lateral buckling problems, fluid-solid vibrations, quantum dot heterostructures, and sandwich plates, e.g.

Example 1: Quantum dots

Semiconductor nanostructures have attracted tremendous interest in the past few years because of their special physical properties and their potential for applications in micro– and optoelectronic devices.

Example 1: Quantum dots

Semiconductor nanostructures have attracted tremendous interest in the past few years because of their special physical properties and their potential for applications in micro– and optoelectronic devices.

In such nanostructures, the free carriers are confined to a small region of space by potential barriers, and if the size of this region is less than the electron wavelength, the electronic states become quantized at discrete energy levels.

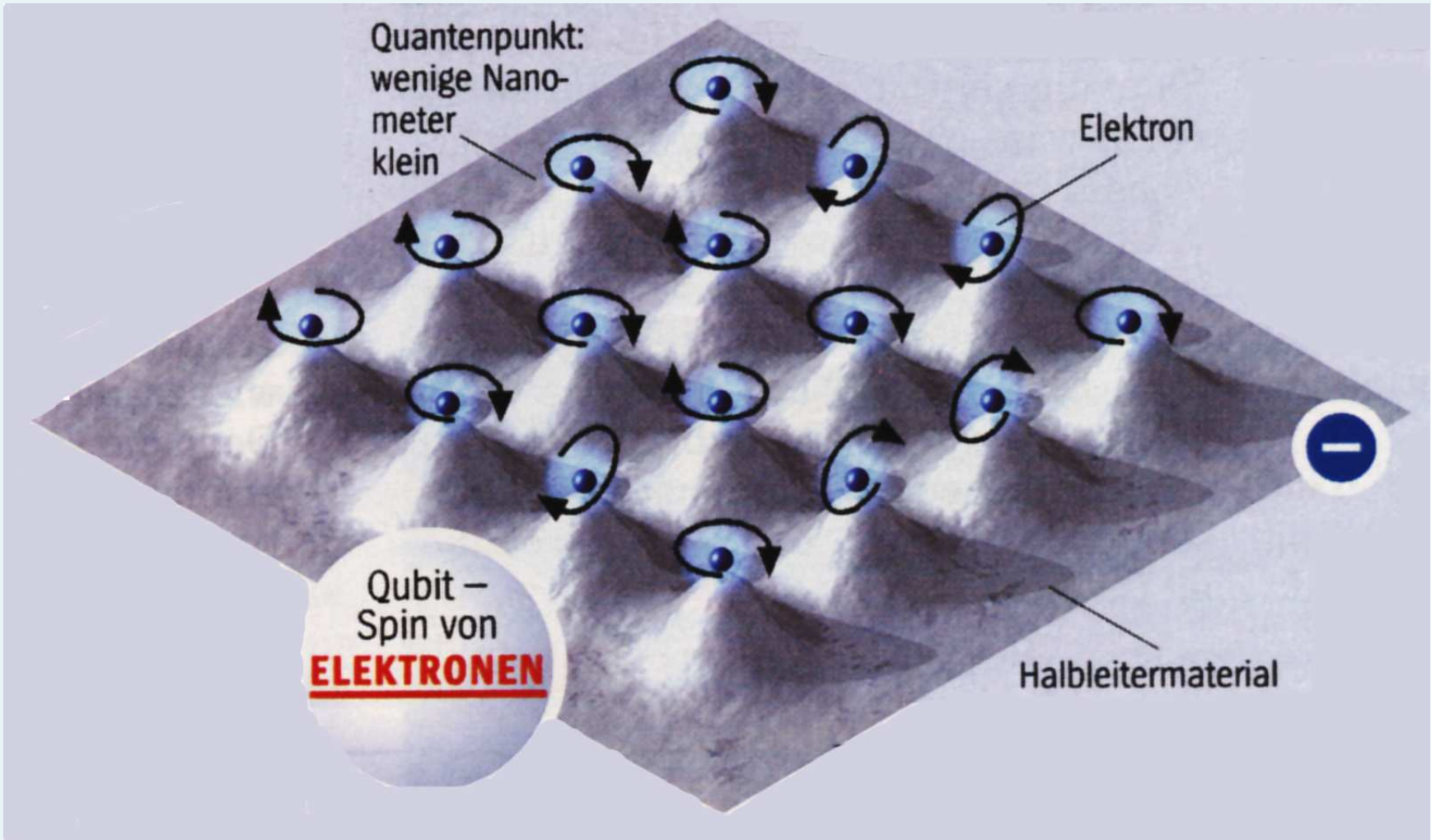
Example 1: Quantum dots

Semiconductor nanostructures have attracted tremendous interest in the past few years because of their special physical properties and their potential for applications in micro– and optoelectronic devices.

In such nanostructures, the free carriers are confined to a small region of space by potential barriers, and if the size of this region is less than the electron wavelength, the electronic states become quantized at discrete energy levels.

The ultimate limit of low dimensional structures is the quantum dot, in which the carriers are confined in all three directions, thus reducing the degrees of freedom to zero.
Therefore, a quantum dot can be thought of as an artificial atom.

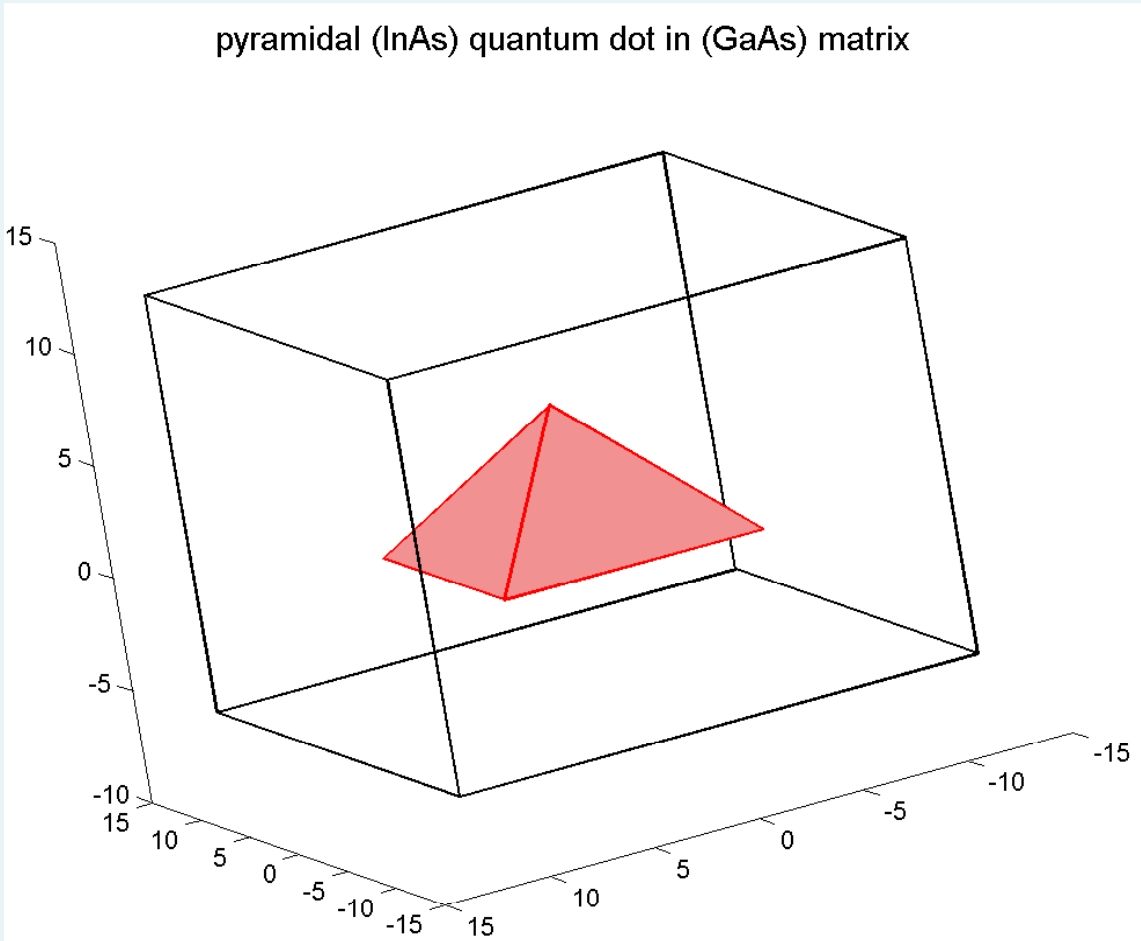
Quantum dots ct.



DER SPIEGEL, March 14, 2005

Quantum dots ct.

Determine relevant energy states (i.e. eigenvalues) and corresponding wave functions (i.e. eigenfunctions) of a three-dimensional quantum dot embedded in a matrix.



Schrödinger equation

$$-\nabla\cdot\left(\frac{\hbar^2}{2m(x,\lambda)}\nabla u\right)+V(x)u=\lambda u,\quad x\in\Omega_1\cup\Omega_2$$

where $\hbar$ is the reduced Planck constant,

$$m(x,\lambda)=\begin{cases}\frac{1}{m_1(\lambda)}, & x\in\Omega_1\\ \frac{1}{m_2(\lambda)}, & x\in\Omega_2\end{cases}\qquad V(x)=\begin{cases}0, & x\in\Omega_1\\ V_2, & x\in\Omega_2\end{cases}$$

$$\frac{1}{m_j(\lambda)}=\frac{P_j^2}{\hbar^2}\left(\frac{2}{\lambda+g_j-V_j}+\frac{1}{\lambda+g_j-V_j+\delta_j}\right)$$

where m_j is the electron effective mass, V_j the confinement potential, P_j the momentum, g_j the main energy gap, and δ_j the spin-orbit splitting in the jth region.

Quantum dot ct.

Boundary and interface conditions

$$u = 0 \quad \text{on outer boundary of matrix } \Omega_1$$

BenDaniel–Duke condition

$$\left.\frac{1}{m_2}\frac{\partial u}{\partial n}\right|_{\partial\Omega_2} = \left.\frac{1}{m_1}\frac{\partial u}{\partial n}\right|_{\partial\Omega_1} \quad \text{on interface}$$

Quantum dot ct.

Boundary and interface conditions

$$u = 0 \quad \text{on outer boundary of matrix } \Omega_1$$

BenDaniel–Duke condition

$$\left.\frac{1}{m_2}\frac{\partial u}{\partial n}\right|_{\partial\Omega_2} = \left.\frac{1}{m_1}\frac{\partial u}{\partial n}\right|_{\partial\Omega_1} \quad \text{on interface}$$

Variational form

Find $\lambda \in \mathbb{R}$ *and* $u \in H_0^1(\overline{\Omega}_1 \cup \Omega_2)$, $u \neq 0$, *such that*

$$\begin{aligned} a(u,v;\lambda) &:= \frac{1}{m_1(\lambda)}\int_{\Omega_1} \nabla u \nabla v\, dx + \frac{1}{m_2(\lambda)}\int_{\Omega_2} \nabla u \nabla v\, dx + V_2 \int_{\Omega_2} uv\, dx \\ &= \lambda \int_{\Omega_1 \cup \Omega_2} uv\, dx =: \lambda b(u,v) \quad \text{for every } v \in H_0^1(\overline{\Omega}_1 \cup \Omega_2) \end{aligned}$$

Quantum dot ct.

Boundary and interface conditions

$$u = 0 \quad \text{on outer boundary of matrix } \Omega_1$$

BenDaniel–Duke condition
$$\frac{1}{m_2}\frac{\partial u}{\partial n}\bigg|_{\partial\Omega_2} = \frac{1}{m_1}\frac{\partial u}{\partial n}\bigg|_{\partial\Omega_1} \quad \text{on interface}$$

Variational form

Find $\lambda \in \mathbb{R}$ *and* $u \in H_0^1(\overline{\Omega}_1 \cup \Omega_2)$, $u \neq 0$, *such that*

$$\begin{aligned} a(u,v;\lambda) &:= \frac{1}{m_1(\lambda)}\int_{\Omega_1} \nabla u \nabla v\, dx + \frac{1}{m_2(\lambda)}\int_{\Omega_2} \nabla u \nabla v\, dx + V_2 \int_{\Omega_2} uv\, dx \\ &= \lambda \int_{\Omega_1 \cup \Omega_2} uv\, dx =: \lambda b(u,v) \quad \textit{for every } v \in H_0^1(\overline{\Omega}_1 \cup \Omega_2) \end{aligned}$$

Discretization (FEM, FVM) $\rightarrow$ rational matrix eigenvalue problem

Iterative projection methods

For linear sparse eigenproblems

$$T(\lambda) = \lambda B - A$$

very efficient methods are iterative projection methods (Lanczos, Arnoldi, Jacobi–Davidson method, e.g.), where approximations to the wanted eigenvalues and eigenvectors are obtained from projections of the eigenproblem to subspaces of small dimension which are expanded in the course of the algorithm.

Iterative projection methods

For linear sparse eigenproblems

$$T(\lambda) = \lambda B - A$$

very efficient methods are iterative projection methods (Lanczos, Arnoldi, Jacobi–Davidson method, e.g.), where approximations to the wanted eigenvalues and eigenvectors are obtained from projections of the eigenproblem to subspaces of small dimension which are expanded in the course of the algorithm.

Essentially two types of methods are in use:

- methods which project the problem to a sequence of Krylov spaces like Lanczos or Arnoldi, and

Iterative projection methods

For linear sparse eigenproblems

$$T(\lambda) = \lambda B - A$$

very efficient methods are iterative projection methods (Lanczos, Arnoldi, Jacobi–Davidson method, e.g.), where approximations to the wanted eigenvalues and eigenvectors are obtained from projections of the eigenproblem to subspaces of small dimension which are expanded in the course of the algorithm.

Essentially two types of methods are in use:

- methods which project the problem to a sequence of Krylov spaces like Lanczos or Arnoldi, and
- methods which aim at a specific eigenpair like the Jacobi–Davidson method.

Generalizations to nonlinear sparse eigenproblems

- nonlinear rational Krylov: Ruhe (2000,2005), Hager (2001)
  Regula falsi for the nonlinear problem and Arnoldi for the linear problem are knit together to form a sequence of subspaces $V_k \in \mathbb{C}^n$ and corresponding Hessenberg matrices H_k which approximate the projection of $T(\sigma)^{-1}T(\lambda_k)$ on V_k for some shift σ close to the wanted eigenvalues.

  Can be interpreted as projection method for $T(\sigma)^{-1}T(\lambda)$ (Jarlebring, V. (2003)) .

Generalizations to nonlinear sparse eigenproblems

- nonlinear rational Krylov: Ruhe (2000,2005), Hager (2001)
  Regula falsi for the nonlinear problem and Arnoldi for the linear problem are knit together to form a sequence of subspaces $V_k \in \mathbb{C}^n$ and corresponding Hessenberg matrices H_k which approximate the projection of $T(\sigma)^{-1}T(\lambda_k)$ on V_k for some shift σ close to the wanted eigenvalues.

  Can be interpreted as projection method for $T(\sigma)^{-1}T(\lambda)$ (Jarlebring, V. (2003)) .

- Arnoldi method: V. (2003, 2004); quadratic probl.: Meerbergen (2001)

Generalizations to nonlinear sparse eigenproblems

- nonlinear rational Krylov: Ruhe (2000,2005), Hager (2001)
  Regula falsi for the nonlinear problem and Arnoldi for the linear problem are knit together to form a sequence of subspaces $V_k \in \mathbb{C}^n$ and corresponding Hessenberg matrices H_k which approximate the projection of $T(\sigma)^{-1}T(\lambda_k)$ on V_k for some shift σ close to the wanted eigenvalues.

  Can be interpreted as projection method for $T(\sigma)^{-1}T(\lambda)$ (Jarlebring, V. (2003)) .

- Arnoldi method: V. (2003, 2004); quadratic probl.: Meerbergen (2001)

- Jacobi-Davidson: Betcke, V. (2002), Schwetlick (2005)
  polynomial problems: Sleijpen, Boten, Fokkema, van der Vorst (1996), Hwang, Lin, Wang, Wang (2003,2004)

Expanding the subspace

Given subspace $V \subset \mathbb{C}^n$. Expand V by a direction with high approximation potential for the next wanted eigenvector.

Expanding the subspace

Given subspace $V \subset \mathbb{C}^n$. Expand V by a direction with high approximation potential for the next wanted eigenvector.

Let θ be an eigenvalue of the projected problem

$$V^H T(\lambda) V y = 0$$

and $x = Vy$ corresponding Ritz vector, then inverse iteration yields suitable candidate

$$v := T(\theta)^{-1} T'(\theta) x$$

Expanding the subspace

Given subspace $V \subset \mathbb{C}^n$. Expand V by a direction with high approximation potential for the next wanted eigenvector.

Let θ be an eigenvalue of the projected problem

$$V^H T(\lambda) V y = 0$$

and $x = Vy$ corresponding Ritz vector, then inverse iteration yields suitable candidate

$$v := T(\theta)^{-1} T'(\theta) x$$

BUT: In each step have to solve large linear system with varying matrix

Way out: Residual inverse It.

start with an approximation x_1 to an eigenvector of $T(\lambda)x = 0$
for $\ell = 1, 2, \ldots$ until convergence **do**
    solve $e^H T(\sigma)^{-1} T(\mu_{\ell+1}) x_\ell = 0$ for $\mu_{\ell+1}$
    compute the residual $r_\ell = T(\mu_{\ell+1}) x_\ell$
    solve $T(\sigma) d_\ell = r_\ell$
    set $x_{\ell+1} = x_\ell - d_\ell$, $x_{\ell+1} = x_{\ell+1} / \|x_{\ell+1}\|$
end for

Way out: Residual inverse It.

1: start with an approximation x_1 to an eigenvector of $T(\lambda)x = 0$
2: **for** $\ell = 1, 2, \ldots$ until convergence **do**
3: solve $e^H T(\sigma)^{-1} T(\mu_{\ell+1}) x_\ell = 0$ for $\mu_{\ell+1}$
4: compute the residual $r_\ell = T(\mu_{\ell+1})x_\ell$
5: solve $T(\sigma)d_\ell = r_\ell$
6: set $x_{\ell+1} = x_\ell - d_\ell$, $x_{\ell+1} = x_{\ell+1}/\|x_{\ell+1}\|$
7: **end for**

Theorem (Neumaier 1985)

Let $T(\lambda)$ be twice continously differentiable. Assume that $\hat{\lambda}$ is a simple eigenvalue of $T(\lambda)x = 0$, and let $\hat{x}$ be a corresponding eigenvector normalized by $\|\hat{x}\| = 1$. Then the residual inverse iteration converges for all σ sufficently close to $\hat{\lambda}$, and it holds

$$\frac{\|x_{\ell+1} - \hat{x}\|}{\|x_\ell - \hat{x}\|} = \mathcal{O}(|\sigma - \hat{\lambda}|), \quad \text{and} \quad |\lambda_{\ell+1} - \hat{\lambda}| = \mathcal{O}(\|x_\ell - \hat{x}\|).$$

Arnoldi method

If θ is an eigenvalue of the projected problem $V^H T(\lambda) V y = 0$ and $\tilde{x} = V\tilde{y}$ is a corresponding Ritz vector, then expand V by new direction

$$v = \tilde{x} - T(\sigma)^{-1} T(\theta)\tilde{x}$$

Arnoldi method

If θ is an eigenvalue of the projected problem $V^H T(\lambda) V y = 0$ and $\tilde{x} = V\tilde{y}$ is a corresponding Ritz vector, then expand V by new direction

$$v = \tilde{x} - T(\sigma)^{-1} T(\theta)\tilde{x}$$

In projection methods the new direction is orthonormalized against the previous ansatz vectors. Since the Ritz vector $\tilde{x}$ is contained in span V we expand V by

$$v = T(\sigma)^{-1} T(\theta)\tilde{x}.$$

Arnoldi method

If θ is an eigenvalue of the projected problem $V^H T(\lambda)Vy = 0$ and $\tilde{x} = V\tilde{y}$ is a corresponding Ritz vector, then expand V by new direction

$$v = \tilde{x} - T(\sigma)^{-1}T(\theta)\tilde{x}$$

In projection methods the new direction is orthonormalized against the previous ansatz vectors. Since the Ritz vector $\tilde{x}$ is contained in span V we expand V by

$$v = T(\sigma)^{-1}T(\theta)\tilde{x}.$$

For the linear problem $T(\lambda) = \lambda I - A$ this is exactly the Cayley transformation or shifted-and-inverted Arnoldi method.

Arnoldi method

If θ is an eigenvalue of the projected problem $V^H T(\lambda)Vy = 0$ and $\tilde{x} = V\tilde{y}$ is a corresponding Ritz vector, then expand V by new direction

$$v = \tilde{x} - T(\sigma)^{-1}T(\theta)\tilde{x}$$

In projection methods the new direction is orthonormalized against the previous ansatz vectors. Since the Ritz vector $\tilde{x}$ is contained in span V we expand V by

$$v = T(\sigma)^{-1}T(\theta)\tilde{x}.$$

For the linear problem $T(\lambda) = \lambda I - A$ this is exactly the Cayley transformation or shifted-and-inverted Arnoldi method.

If the linear system $T(\sigma)v = T(\theta)\tilde{x}$ is too expensive to solve for v we may choose as new direction $v = MT(\lambda)\tilde{x}$ with $M \approx T(\sigma)^{-1}$, and for the linear problem we obtain the preconditioned Arnoldi method.

Therefore the resulting iterative projection method is called nonlinear Arnoldi method

Nonlinear Arnoldi Method

start with initial basis V, $V^H V = I$; set $k = m = 1$
determine preconditioner $M \approx T(\sigma)^{-1}$, σ close to first wanted eigenvalue
while $m \leq$ number of wanted eigenvalues **do**
  solve $V^H T(\mu) V y = 0$ for (μ, y) and set $u = Vy$, $r_k = T(\mu)u$
  if $\|r_k\|/\|u\| < \epsilon$ **then**
    Accept eigenpair $\lambda_m = \mu$, $x_m = u$,
    if $m ==$ number of wanted eigenvalues **then** STOP **end if**
    $m = m + 1$
    if $(k > 1)$ & $(\|r_{k-1}\|/\|r_k\| >$ tol) **then**
      choose new pole σ, determine preconditioner $M \approx T(\sigma)^{-1}$
    end if
    restart if necessary
    Choose approximations μ and u to next eigenvalue and eigenvector
    determine $r = T(\mu)u$ and set $k = 0$
  end if
  $v = Mr$, $k = k + 1$
  $v = v - VV^H v$,$\tilde{v} = v/\|v\|$, $V = [V, \tilde{v}]$ and reorthogonalize if necessary
end while

Second Way out: Jacobi–Davidson

Solve (approxemately) the correction equation

$$(I - \frac{T'(\theta)xx^T}{x^T T'(\theta)x})T(\theta)(I - xx^T)t = -r, \qquad t \perp x$$

where (x, θ) is the current Ritz pair and $r = T(\theta)x$.

Second Way out: Jacobi–Davidson

Solve (approxemately) the correction equation

$$(I - \frac{T'(\theta)xx^T}{x^T T'(\theta)x})T(\theta)(I - xx^T)t = -r, \qquad t \perp x$$

where (x, θ) is the current Ritz pair and $r = T(\theta)x$.

If the correction equation is solved exactly, then

$$T(\theta)t - \alpha T'(\theta)x = -r$$

where α is chosen such that $t \perp x$. Solving for t yields

$$t = -x + \alpha T(\theta)^{-1}T'(\theta)x.$$

Since $x = Vy$ for some y, and t is orthogonalized against V the expansion by t is equivalent to the expansion by inverse iteration.

Solving correction equation

The correction equation is solved approximately by a few steps of an iterative solver (GMRES or BiCGStab).

Solving correction equation

The correction equation is solved approximately by a few steps of an iterative solver (GMRES or BiCGStab).

The operator $T(\sigma)$ is restricted to map the subspace $x^{\perp}$ into itself. Hence, if $M \approx T(\sigma)$ is a preconditioner of $T(\sigma)$, $\sigma \approx \theta$, then a preconditioner for an iterativ solver of the correction equation should be modified correspondingly to

$$\tilde{M} := (I - \frac{T'(\theta)uu^H}{x^H T(\theta)u})M(I - \frac{uu^H}{u^H u}).$$

Solving correction equation

The correction equation is solved approximately by a few steps of an iterative solver (GMRES or BiCGStab).

The operator $T(\sigma)$ is restricted to map the subspace $x^\perp$ into itself. Hence, if $M \approx T(\sigma)$ is a preconditioner of $T(\sigma)$, $\sigma \approx \theta$, then a preconditioner for an iterativ solver of the correction equation should be modified correspondingly to

$$\tilde{M} := (I - \frac{T'(\theta)uu^H}{x^H T(\theta)u})M(I - \frac{uu^H}{u^H u}).$$

Taking into account the projectors in the preconditioner, i.e. using $\tilde{M}$ instead of M, raises the cost of the preconditioned Krylov solver only slightly (cf. Sleijpen, van der Vorst). Only one additional linear solve with system matrix M is required.

Variant of Hwang, Lin, Wang, Wang (2004)

Solve

$$\Big(I - \frac{T'(\theta)uu^H}{u^H T'(\theta)u}\Big)T(\theta)\Big(I - \frac{uu^H}{u^H u}\Big)t = -r, \quad t \perp u$$

approximately

Variant of Hwang, Lin, Wang, Wang (2004)

Solve

$$\Big(I - \frac{T'(\theta)uu^H}{u^H T'(\theta)u}\Big)\cancel{T(\theta)}^{M}\Big(I - \frac{uu^H}{u^H u}\Big)t = -r, \quad t \perp u$$

approximately by computing

$$t = M^{-1}r + \alpha M^{-1}T'(\theta)u \quad \text{with } \alpha := \frac{u^H M^{-1} r}{u^H M^{-1} T'(\theta) u}$$

where M is a preconditioner of $T(\theta)$.

Variant of Hwang, Lin, Wang, Wang (2004)

Solve

$$\Big(I - \frac{T'(\theta)uu^H}{u^H T'(\theta)u}\Big)\cancel{T(\theta)}^{M}\Big(I - \frac{uu^H}{u^H u}\Big)t = -r, \quad t \perp u$$

approximately by computing

$$t = M^{-1}r + \alpha M^{-1}T'(\theta)u \quad \text{with } \alpha := \frac{u^H M^{-1} r}{u^H M^{-1} T'(\theta) u}$$

where M is a preconditioner of $T(\theta)$.

Method combines preconditioned Arnoldi method ($M^{-1}r$) and simplified inverse iteration ($M^{-1}T'(\theta)u$)

Primal-Dual JD Update: Schwetlick (2005)

Given: triplet (u, v, θ), s.t. $\|u\| = \|v\| = 1$, $v^H T(\theta) u = 0$

Primal-Dual JD Update: Schwetlick (2005)

Given: triplet (u, v, θ), s.t. $\|u\| = \|v\| = 1$, $v^H T(\theta) u = 0$

Step: $(u, v, \theta) \mapsto (u_+, v_+, \theta_+)$

- solve $(I - vv^H) T(\theta)(I - uu^H)s = -T(\theta)u,\ u^H s = 0$
- solve $(I - uu^H) T(\theta)^H (I - vv^H)t = -T(\theta)^H v,\ v^H t = 0$
- set $u_+ = (u + s)/\|u + s\|$, $v_+ = (v + t)/\|v + t\|$
- determine θ_+ from $v_+^H T(\theta_+) u_+ = 0$

Primal-Dual JD Update: Schwetlick (2005)

Given: triplet (u, v, θ), s.t. $\|u\| = \|v\| = 1$, $v^H T(\theta) u = 0$

Step: $(u, v, \theta) \mapsto (u_+, v_+, \theta_+)$

- solve $(I - vv^H)T(\theta)(I - uu^H)s = -T(\theta)u,\ u^H s = 0$
- solve $(I - uu^H)T(\theta)^H(I - vv^H)t = -T(\theta)^H v,\ v^H t = 0$
- set $u_+ = (u + s)/\|u + s\|$, $v_+ = (v + t)/\|v + t\|$
- determine θ_+ from $v_+^H T(\theta_+)u_+ = 0$

Experience of Schwetlick: If Standard JD works, then it is faster than P-D JD; if $v^H T'(\theta)u$ is small then P-D JD is more robust than Standard JD

Primal-Dual JD

Given orthogonal bases V and U of left and right search spaces, and left and right approximate eigenvectors $v \in \mathrm{span}(V)$ and $u \in \mathrm{span}(U)$, and approximate eigenvalue θ s.t. $v^H T(\theta) u = 0$

Primal-Dual JD

Given orthogonal bases V and U of left and right search spaces, and left and right approximate eigenvectors $v \in \mathrm{span}(V)$ and $u \in \mathrm{span}(U)$, and approximate eigenvalue θ s.t. $v^H T(\theta)u = 0$

Solve (approximately)

$$(I - vv^H)T(\theta)(I - uu^H)s = -T(\theta)u, \quad s \perp u$$

$$(I - uu^H)T(\theta)^H(I - vv^H)t = -T(\theta)^H v, \quad t \perp v$$

Primal-Dual JD

Given orthogonal bases V and U of left and right search spaces, and left and right approximate eigenvectors $v \in \mathrm{span}(V)$ and $u \in \mathrm{span}(U)$, and approximate eigenvalue θ s.t. $v^H T(\theta)u = 0$

Solve (approximately)

$$(I - vv^H)T(\theta)(I - uu^H)s = -T(\theta)u, \quad s \perp u$$

$$(I - uu^H)T(\theta)^H(I - vv^H)t = -T(\theta)^H v, \quad t \perp v$$

Expand search spaces $V \leftarrow [V, t]$, $U \leftarrow [U, s]$

Primal-Dual JD

Given orthogonal bases V and U of left and right search spaces, and left and right approximate eigenvectors $v \in \mathrm{span}(V)$ and $u \in \mathrm{span}(U)$, and approximate eigenvalue θ s.t. $v^H T(\theta)u = 0$

Solve (approximately)

$$(I - vv^H)T(\theta)(I - uu^H)s = -T(\theta)u, \quad s \perp u$$

$$(I - uu^H)T(\theta)^H(I - vv^H)t = -T(\theta)^H v, \quad t \perp v$$

Expand search spaces $V \leftarrow [V, t]$, $U \leftarrow [U, s]$

Solve

$$V^H T(\theta)Us = 0 \quad \text{and} \quad U^H T(\theta)^H Vt = 0 \quad \text{for } (s, t, \theta)$$

and set

$$u = Us \quad \text{and} \quad v = Vt.$$

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

- linear problems: (incomplete) Schur decomposition

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

- linear problems: (incomplete) Schur decomposition
- quadratic problems: Meerbergen (2001) based on linearization and Schur form of linearized problem (lock 2 vectors in each step)

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

- linear problems: (incomplete) Schur decomposition
- quadratic problems: Meerbergen (2001) based on linearization and Schur form of linearized problem (lock 2 vectors in each step)
- cubic problem: Hwang, Lin, Liu, Wang (2005) based on linearization

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

- linear problems: (incomplete) Schur decomposition
- quadratic problems: Meerbergen (2001) based on linearization and Schur form of linearized problem (lock 2 vectors in each step)
- cubic problem: Hwang, Lin, Liu, Wang (2005) based on linearization
- Approach of Hwang et al. can be generalized directly to general problems if all eigenvalues of projected problems are determined

Locking of converged eigenvectors

A major problem with iterative projection methods for nonlinear eigenproblems when approximating more than one eigenvalue is to inhibit the method from converging to an eigenpair which was detected already previously.

- linear problems: (incomplete) Schur decomposition
- quadratic problems: Meerbergen (2001) based on linearization and Schur form of linearized problem (lock 2 vectors in each step)
- cubic problem: Hwang, Lin, Liu, Wang (2005) based on linearization
- Approach of Hwang et al. can be generalized directly to general problems if all eigenvalues of projected problems are determined
- for Hermitean problems one can often take advantage of a variational characterization of eigenvalues

Nonlinear minmax theory

Let $T(\lambda) \in \mathbb{C}^{n\times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval. Assume that

$$f : \begin{cases} J \times \mathbb{C}^n & \to & \mathbb{R} \\ (\lambda, x) & \mapsto & x^H T(\lambda) x \end{cases}$$

is continuously differentiable, and that for every fixed $x \in \mathbb{C}^n$, $x \neq 0$ the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Nonlinear minmax theory

Let $T(\lambda) \in \mathbb{C}^{n\times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval. Assume that

$$f : \begin{cases} J \times \mathbb{C}^n & \to & \mathbb{R} \\ (\lambda, x) & \mapsto & x^H T(\lambda) x \end{cases}$$

is continuously differentiable, and that for every fixed $x \in \mathbb{C}^n$, $x \neq 0$ the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Then equation $f(\lambda, x) = 0$ implicitly defines a functional p on some subset D of $\mathbb{C}^n$ which we call the Rayleigh functional. If $D = \mathbb{C}^n \setminus \{0\}$ then $T(\lambda)x = 0$ is called overdamped.

Nonlinear minmax theory

Let $T(\lambda) \in \mathbb{C}^{n\times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval. Assume that

$$f : \begin{cases} J \times \mathbb{C}^n & \to & \mathbb{R} \\ (\lambda, x) & \mapsto & x^H T(\lambda) x \end{cases}$$

is continuously differentiable, and that for every fixed $x \in \mathbb{C}^n$, $x \neq 0$ the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Then equation $f(\lambda, x) = 0$ implicitly defines a functional p on some subset D of $\mathbb{C}^n$ which we call the Rayleigh functional. If $D = \mathbb{C}^n \setminus \{0\}$ then $T(\lambda)x = 0$ is called overdamped.

Suppose that

$$x^H T'(p(x))x > 0 \quad \text{for every } x \in D.$$

which generalizes the definiteness condition for linear eigenproblems.

Enumeration of eigenvalues

If $\lambda \in J$ is an eigenvalue of $T(\cdot)$ then $\theta = 0$ is an eigenvalue of the linear problem $T(\lambda)y = \theta y$, and therefore there exists $\ell \in \mathbb{N}$ such that

$$0 = \max_{V \in H_\ell} \min_{v \in V \setminus \{0\}} \frac{v^H T(\lambda) v}{\|v\|^2}$$

where H_ℓ denotes the set of all ℓ–dimensional subspaces of $\mathbb{R}^n$.

Enumeration of eigenvalues

If $\lambda \in J$ is an eigenvalue of $T(\cdot)$ then $\theta = 0$ is an eigenvalue of the linear problem $T(\lambda)y = \theta y$, and therefore there exists $\ell \in \mathbb{N}$ such that

$$0 = \max_{V \in H_\ell} \min_{v \in V \setminus \{0\}} \frac{v^H T(\lambda) v}{\|v\|^2}$$

where H_ℓ denotes the set of all ℓ–dimensional subspaces of $\mathbb{R}^n$.

In this case λ is called an ℓ-th eigenvalue of $T(\cdot)$.

Minmax characterization

Under the conditions given above it holds:

(i) For every $\ell \in \mathbb{N}$ there is at most one ℓ-th eigenvalue of $T(\cdot)$ which can be characterized by

$$\lambda_\ell = \min_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v). \qquad (*)$$

The set of eigenvalues of $T(\cdot)$ in J is at most countable.

Minmax characterization

Under the conditions given above it holds:

(i) For every $\ell \in \mathbb{N}$ there is at most one ℓ-th eigenvalue of $T(\cdot)$ which can be characterized by

$$\lambda_\ell = \min_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v). \qquad (*)$$

The set of eigenvalues of $T(\cdot)$ in J is at most countable.

(ii) If

$$\lambda_\ell := \inf_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v) \in J$$

for some $\ell \in \mathbb{N}$ then λ_ℓ is the ℓ-th eigenvalue of $T(\cdot)$ in J, and (*) holds.

Minmax characterization

Under the conditions given above it holds:

(i) For every $\ell \in \mathbb{N}$ there is at most one ℓ-th eigenvalue of $T(\cdot)$ which can be characterized by

$$\lambda_\ell = \min_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v). \qquad (*)$$

The set of eigenvalues of $T(\cdot)$ in J is at most countable.

(ii) If

$$\lambda_\ell := \inf_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v) \in J$$

for some $\ell \in \mathbb{N}$ then λ_ℓ is the ℓ-th eigenvalue of $T(\cdot)$ in J, and (*) holds.

(iii) If for $\mu < \nu$ there exist a μ-th and a ν-th eigenvalue of $T(\cdot)$ in J, then for $\ell = \mu, \mu+1, \ldots, \nu$ there is an ℓ-th eigenvalue $\lambda_\ell \in J$, and $\lambda_\mu \leq \lambda_{\mu+1} \leq \cdots \leq \lambda_\nu$.

Safeguarded iteration

The minimum in

$$\lambda_\ell = \min_{\substack{V \in H_\ell \\ V \cap D \neq \emptyset}} \sup_{v \in V \cap D} p(v)$$

is attained by the invariant subspace of $T(\lambda_\ell)$ corresponding to the ℓ largest eigenvalues, and the maximum by every eigenvector corresponding to the eigenvalue 0. This suggests

Safeguarded iteration

The minimum in

$$\lambda_\ell = \min_{\substack{V\in H_\ell \\ V\cap D\neq\emptyset}} \sup_{v\in V\cap D} p(v)$$

is attained by the invariant subspace of $T(\lambda_\ell)$ corresponding to the ℓ largest eigenvalues, and the maximum by every eigenvector corresponding to the eigenvalue 0. This suggests

Safeguarded iteration

1: Start with an approximation μ_1 to the ℓ-th eigenvalue of $T(\lambda)x = 0$
2: **for** $k = 1, 2, \ldots$ until convergence **do**
3: determine eigenvector u corresponding to the ℓ-largest eigenvalue of $T(\mu_k)$
4: evaluate $\mu_{k+1} = p(u)$
5: **end for**

Safeguarded iteration

The minimum in

$$\lambda_\ell = \min_{\substack{V\in H_\ell \\ V\cap D\neq\emptyset}} \sup_{v\in V\cap D} p(v)$$

is attained by the invariant subspace of $T(\lambda_\ell)$ corresponding to the ℓ largest eigenvalues, and the maximum by every eigenvector corresponding to the eigenvalue 0. This suggests

Safeguarded iteration

1: Start with an approximation μ_1 to the ℓ-th eigenvalue of $T(\lambda)x = 0$
2: **for** $k = 1, 2, \ldots$ until convergence **do**
3: determine eigenvector u corresponding to the ℓ-largest eigenvalue of $T(\mu_k)$
4: evaluate $\mu_{k+1} = p(u)$
5: **end for**

Convergence quadratic; global to the first eigenvalue, local to further eigenvalues.

Computing several eigenvalues

If $T(\lambda)$ is a family of Hermitean matrices allowing a minmax characterization of its eigenvalues in an open interval J, and if the columns of $V \in \mathbb{C}^{n\times k}$ form a basis of the current search space $\mathcal{V}$, then the projeceted problem

$$T_V(\lambda)y := V^H T(\lambda)Vy = 0$$

inherits this property.

Computing several eigenvalues

If $T(\lambda)$ is a family of Hermitean matrices allowing a minmax characterization of its eigenvalues in an open interval J, and if the columns of $V \in \mathbb{C}^{n\times k}$ form a basis of the current search space $\mathcal{V}$, then the projeceted problem

$$T_V(\lambda)y := V^H T(\lambda)Vy = 0$$

inherits this property.

If J contains a first eigenvalue $\lambda_1 = \min_{x\in D} p(x)$ and $\mathcal{V} \cap D \neq \emptyset$ then the safeguarded iterations converges globally to the first eigenvalue of the projected problem.

Computing several eigenvalues

If $T(\lambda)$ is a family of Hermitean matrices allowing a minmax characterization of its eigenvalues in an open interval J, and if the columns of $V \in \mathbb{C}^{n\times k}$ form a basis of the current search space $\mathcal{V}$, then the projeceted problem

$$T_V(\lambda)y := V^H T(\lambda)Vy = 0$$

inherits this property.

If J contains a first eigenvalue $\lambda_1 = \min_{x\in D} p(x)$ and $\mathcal{V} \cap D \neq \emptyset$ then the safeguarded iterations converges globally to the first eigenvalue of the projected problem.

Usually, while computing the m-th eigenvalue λ_m the algorithm gathers enough information in the search space $\mathcal{V}$ about the next eigenvector to compute λ_{m+1} safely, and the eigenvalues in J can be computed one after the other without determining the same eigenvalue repeatedly.

Numerical example

pyramidal quantum dot:	baselength: 12.4 nm,	height: 6.2 nm
cubic matrix:	$24.8 \times 24.8 \times 18.6$ nm^3	

Numerical example

pyramidal quantum dot: baselength: 12.4 nm, height: 6.2 nm
cubic matrix: $24.8 \times 24.8 \times 18.6$ nm^3

Parameters (Hwang, Lin, Wang, Wang 2004)

$$P_1 = 0.8503, \quad g_1 = 0.42, \quad \delta_1 = 0.48, \quad V_1 = 0.0$$
$$P_2 = 0.8878, \quad g_2 = 1.52, \quad \delta_2 = 0.34, \quad V_1 = 0.7$$

Numerical example

pyramidal quantum dot: baselength: 12.4 nm, height: 6.2 nm
cubic matrix: $24.8 \times 24.8 \times 18.6$ nm^3

Parameters (Hwang, Lin, Wang, Wang 2004)

$$P_1 = 0.8503, \quad g_1 = 0.42, \quad \delta_1 = 0.48, \quad V_1 = 0.0$$
$$P_2 = 0.8878, \quad g_2 = 1.52, \quad \delta_2 = 0.34, \quad V_1 = 0.7$$

Discretization by FEM or FVM yields rational eigenproblem

$$T(\lambda)x = \lambda Mx - \frac{1}{m_1(\lambda)}A_1x - \frac{1}{m_2(\lambda)}A_2x - Bx = 0$$

where $T(\lambda)$ is symmetric and satisfies conditions of minmax characterization for $\lambda \geq 0$

Numerical example

pyramidal quantum dot: baselength: 12.4 nm, height: 6.2 nm
cubic matrix: $24.8 \times 24.8 \times 18.6$ nm^3

Parameters (Hwang, Lin, Wang, Wang 2004)

$$P_1 = 0.8503, \quad g_1 = 0.42, \quad \delta_1 = 0.48, \quad V_1 = 0.0$$
$$P_2 = 0.8878, \quad g_2 = 1.52, \quad \delta_2 = 0.34, \quad V_1 = 0.7$$

Discretization by FEM or FVM yields rational eigenproblem

$$T(\lambda)x = \lambda Mx - \frac{1}{m_1(\lambda)}A_1x - \frac{1}{m_2(\lambda)}A_2x - Bx = 0$$

where $T(\lambda)$ is symmetric and satisfies conditions of minmax characterization for $\lambda \geq 0$

All timings for MATLAB 7.0.4 on AMD Opteron Processor 248 $\times$ 860_64 with 2.2 GHz and 4 GB RAM

Numerical example ct.

FVM: Hwang, Lin, Wang, Wang (2004)

dim	λ_1	$\lambda_{2/3}$	λ_4	λ_5	CPU time
$2'475$	0.41195	0.58350	0.67945	0.70478	0.7 s
$22'103$	0.40166	0.57668	0.68418	0.69922	8.1 s
$186'543$	0.39878	0.57477	0.68516	0.69767	150.9 s
$1'532'255$	0.39804	0.57427	0.68539	0.69727	4017.7 s
$12'419'775$	0.39785	0.57415			overnight

Numerical example ct.

FVM: Hwang, Lin, Wang, Wang (2004)

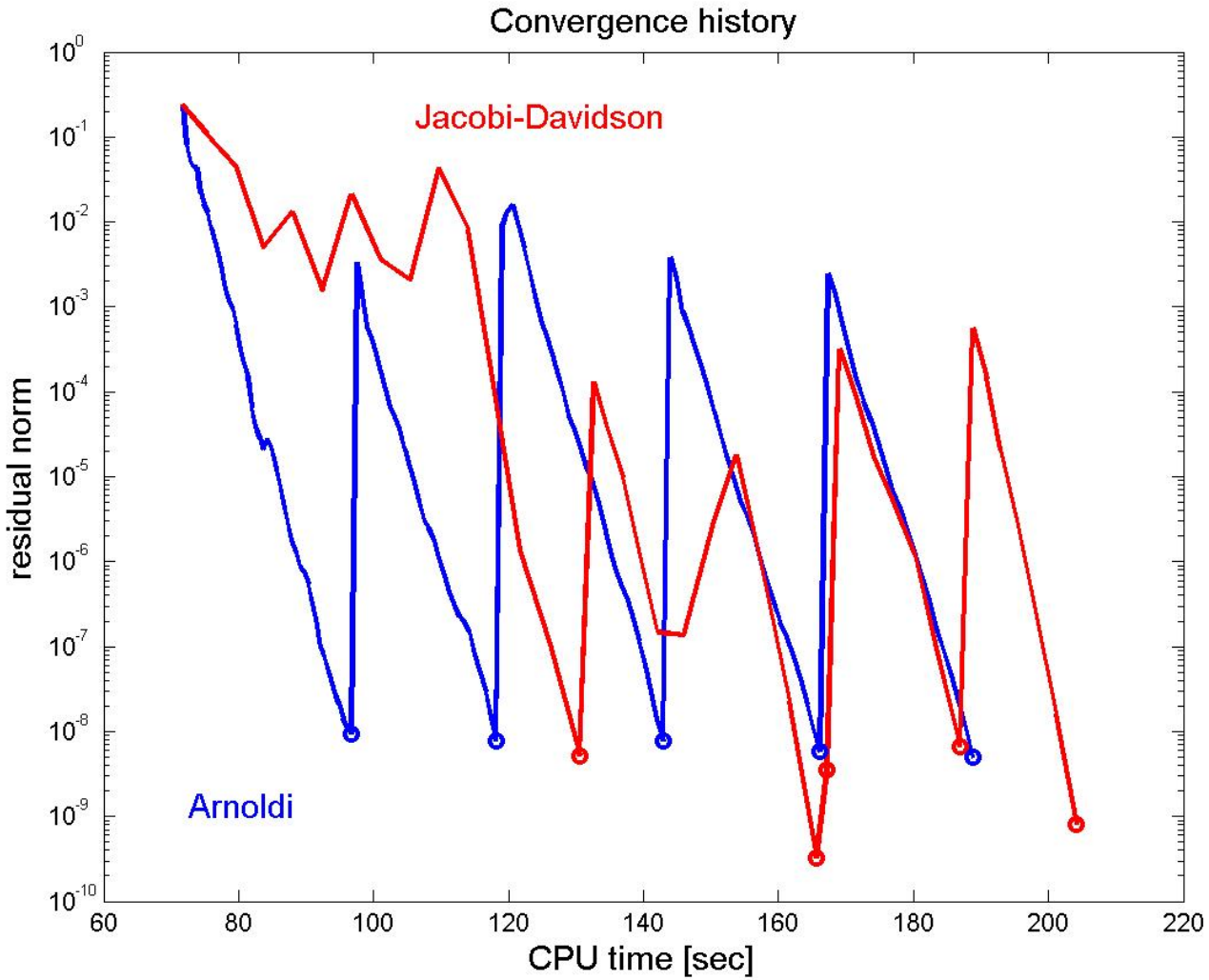
dim	λ_1	$\lambda_{2/3}$	λ_4	λ_5	CPU time
$2'475$	0.41195	0.58350	0.67945	0.70478	0.7 s
$22'103$	0.40166	0.57668	0.68418	0.69922	8.1 s
$186'543$	0.39878	0.57477	0.68516	0.69767	150.9 s
$1'532'255$	0.39804	0.57427	0.68539	0.69727	4017.7 s
$12'419'775$	0.39785	0.57415			overnight

FEM: Cubic Lagrangian elements on tetrahedal grid

dimension: 96'640 $((\mathrm{Dof}_{\mathrm{QD}}, \mathrm{Dof}_{\mathrm{mat}}, \mathrm{Dof}_{\mathrm{interf}}) = (43'615, 43'897, 9'128)$

dim	λ_1	λ_2	λ_3	λ_4	λ_5	CPU time
$96'640$	0.39779	0.57411	0.57411	0.68547	0.69714	
Arnoldi	44 it.	29 it.	29 it.	24 it.	21 it.	188.8 s
JD	15 it.	9 it.	1 it.	7 it.	7 it.	204.4 s
HLWW	45 it.	49 it.	5 it.	24 it.	21 it.	226.7 s

Convergence history



Preconditioning

incomplete LU with cut-off threshold τ

τ	JD	Arnoldi	HLWW	precond.
0.1	261.4	1084.1	1212.4	3.4
0.01	132.7	117.1	155.7	71.7
0.001	118.9	61.2	96.0	246.6
0.0001	155.6	46.6	71.1	665.6

Preconditioning

incomplete LU with cut-off threshold τ

τ	JD	Arnoldi	HLWW	precond.
0.1	261.4	1084.1	1212.4	3.4
0.01	132.7	117.1	155.7	71.7
0.001	118.9	61.2	96.0	246.6
0.0001	155.6	46.6	71.1	665.6

Eigenfunctions

Example 2: Damped vibrations of a solid

Using a viscoelastic constitutive relation to describe the material behaviour in the equations of motion yields a rational eigenvalue problem in the case of free vibrations.

Example 2: Damped vibrations of a solid

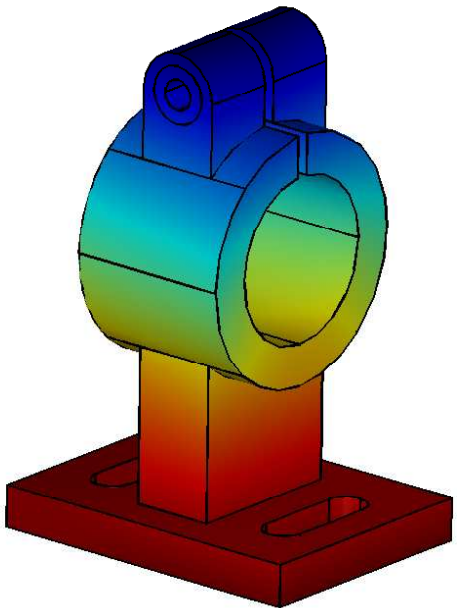
Using a viscoelastic constitutive relation to describe the material behaviour in the equations of motion yields a rational eigenvalue problem in the case of free vibrations.

Discretizing by finite elements yields

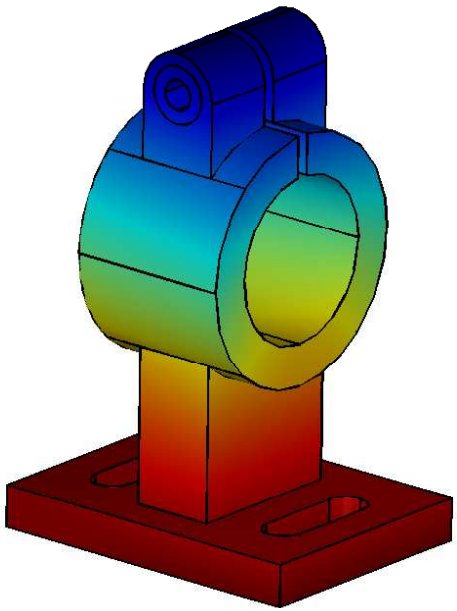
$$T(\omega)x := \left(\omega^2 M + K - \sum_{j=1}^{k} \frac{1}{1+b_j\omega}\Delta K_j\right)x = 0$$

where M is the consistent mass matrix, K is the stiffness matrix with the instantaneous elastic material parameters used in Hooke's law, and ΔK_j collects the contributions of damping from elements with relaxation parameter b_j.

Nonproportional damping



Nonproportional damping



FE model of feeder clamp

linear Lagrangean elements on tetrahedal grid

dimension	193′617
nnz(K)	7′670′533
nnz(M)	2′557′851

Determine 30 eigenvalues with maximal negative imaginary part.

Nonproportional damping ct.

Iterative projection methods

τ	JD	Arnoldi	HLWW	precond.
0.01	3276.6	no conv.	no conv.	49.0
0.001	2077.5	5240.4	5559.9	141.4
0.0001	2067.5	945.4	1452.7	356.3

Conclusions

- Iterative projection methods of Arnoldi or Jacobi–Davidson type are efficient for solving large nonlinear eigenproblems

Conclusions

- Iterative projection methods of Arnoldi or Jacobi–Davidson type are efficient for solving large nonlinear eigenproblems
- Particular care is necessary to prevent the method from converging to the same eigenvalue repeatedly

Conclusions

- Iterative projection methods of Arnoldi or Jacobi–Davidson type are efficient for solving large nonlinear eigenproblems
- Particular care is necessary to prevent the method from converging to the same eigenvalue repeatedly
- Taking advantage of symmetry properties and the minmax characterization of its eigenvalues the projected eigenproblems of small dimension can be solved efficiently by safeguarded iteration.

Conclusions

- Iterative projection methods of Arnoldi or Jacobi–Davidson type are efficient for solving large nonlinear eigenproblems
- Particular care is necessary to prevent the method from converging to the same eigenvalue repeatedly
- Taking advantage of symmetry properties and the minmax characterization of its eigenvalues the projected eigenproblems of small dimension can be solved efficiently by safeguarded iteration.
- The Arnoldi method seems to be faster than the Jacobi–Davidson method, if accurat preconditioners are available.

Conclusions

- Iterative projection methods of Arnoldi or Jacobi–Davidson type are efficient for solving large nonlinear eigenproblems
- Particular care is necessary to prevent the method from converging to the same eigenvalue repeatedly
- Taking advantage of symmetry properties and the minmax characterization of its eigenvalues the projected eigenproblems of small dimension can be solved efficiently by safeguarded iteration.
- The Arnoldi method seems to be faster than the Jacobi–Davidson method, if accurat preconditioners are available.
- The Jacobi–Davidson method seems to be more robust concerning only coarse preconditioners