Jacobi’s Ideas on Eigenvalue Computation in a modern context

Henk van der Vorst
vorst@math.uu.nl

Mathematical Institute
Utrecht University
General remarks

\[ Ax = \lambda x \]

**Nonlinear** problem:

for \( n > 4 \) no explicit solution

Essentially iterative methods
General remarks

\[ Ax = \lambda x \]

**Nonlinear problem:**

for \( n > 4 \) no explicit solution

Essentially iterative methods

Oldest methods:

Leverrier 1840

Jacobi 1845-1846

No matrix notation in that time
\[ Ax = \lambda x \]

**Nonlinear problem:**
For \( n > 4 \) no explicit solution

Essentially iterative methods

**Oldest methods:**
- Leverrier 1840
- Jacobi 1845-1846

No matrix notation in that time

Masterthesis of Anjet de Boer, 1991, Utrecht
Early paper by Leverrier (1811-1877)

*Sur les Variations sèculaires des Éléments elliptiques des sept Planètes principales: Mercure, Vénus, la Terre, Mars, Jupiter, Saturne et Uranus, 1840*

based on Laplace’s work (1789)
Early paper by Leverrier (1811-1877)

Sur les Variations sèculaires des Éléments elliptiques
des sept Planètes principales: Mercure, Vénus, la Terre, Mars, Jupiter, Saturne et Uranus, 1840

based on Laplace’s work (1789)

Perturbations to the orbits of planets caused by the
presence of other planets

linear eigensystem from system of 7 diff. equations
Early paper by Leverrier (1811-1877)

*Sur les Variations sèculaires des Eléments elliptiques des sept Planètes principales: Mercure, Vénus, la Terre, Mars, Jupiter, Saturne et Uranus, 1840*

based on Laplace’s work (1789)

**Perturbations to the orbits of planets caused by the presence of other planets**

linear eigensystem from system of 7 diff. equations

coefficients of characteristic polynomial

He neglected some small elements: factors of degree 3 and 4
Papers by Jacobi (1804-1851)

Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommende lineare Gleichungen, 1845
Papers by Jacobi (1804-1851)

Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommende lineare Gleichungen, 1845

Inspired by work of Gauss (1823)

New method for solution of sym. linear systems;

Jacobi-rotations as ”preconditioner” for G-Jacobi method
Papers by Jacobi (1804-1851)

Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommende lineare Gleichungen, 1845

Inspired by work of Gauss (1823)

New method for solution of sym. linear systems;

Jacobi-rotations as ”preconditioner” for G-Jacobi method

He announces the application for eigenproblems
Über ein leichtes Verfahren, die in der Theory der Säculärstörungen vorkommenden Gleichungen numerisch aufzulösen, 1846

based on earlier work of Lagrange (1778) and Cauchy (1829)
Second paper by Jacobi

Über ein leichtes Verfahren, die in der Theory der Säculärstörungen vorkommenden Gleichungen numerisch aufzulösen, 1846

based on earlier work of Lagrange (1778) and Cauchy (1829)

He applied his 1845-method to the system studied by Leverrier

Claim: easier and more accurate method (unsupported)

refers to Leverriers work
Second paper by Jacobi

Über ein leichtes Verfahren, die in der Theory der Säculärstörungen vorkommenden Gleichungen numerisch aufzulösen, 1846

based on earlier work of Lagrange (1778) and Cauchy (1829)

He applied his 1845-method to the system studied by Leverrier

Claim: easier and more accurate method (unsupported)

refers to Leverrier's work

Bodewig (1951): Jacobi knew his methods before 1840

(inconclusive) evidence: letter of Schumacher to Gauss (1842)
The Jacobi-rotations appear to be forgotten until \( \approx 1950 \)
The Jacobi-rotations appear to be forgotten until $\approx 1950$

Whittaker (1924) described G-Jacobi

Von Mises (1929) described the method without reference to J.
The Jacobi-rotations appear to be forgotten until $\approx 1950$

Whittaker (1924) described G-Jacobi

Von Mises (1929) described the method without reference to J.

The Jacobi (rotation) method was forgotten, but J. described the two methods as one single algorithm
In 1951 Goldstine presented the rotation method
joint work with Murray and Von Neumann
Jacobi method (2)

In 1951 Goldstine presented the rotation method
joint work with Murray and Von Neumann

Ostrowski pointed out that they had reinvented J.’s method
Jacobi method (2)

In 1951 Goldstine presented the rotation method
target work with Murray and Von Neumann

Ostrowski pointed out that they had reinvented J.’s method

also Runge, Hessenberg, Krylov, Magnier, Bodewig knew the method
In 1951 Goldstine presented the rotation method
joint work with Murray and Von Neumann

Ostrowski pointed out that they had reinvented J.’s method

also Runge, Hessenberg, Krylov, Magnier, Bodewig knew the method

Bodewig (1950, 1951) described the full J-method
He claimed the rediscovery
Jacobi in Matrix Notation

(1) First plane rotations to make $A$ diagonal dominant. Suppose that $a_{11}$; $a_{11}$ is largest element then $a_{11}$ and $x = e_1$ ($Ax = x$)

(2) Consider orthogonal complement of $e_1$: $A_1 w = a_{11}$; $a_{11} c^T c F_1 w = a_{11} w$. $w = a_{11} w + c^T w (F^T I w) = c w$
(1) First plane rotations to make $A$ diag. dom.
(1) First plane rotations to make $A$ diag. dom.

suppose that $a_{1,1}$ is largest element

then $\lambda \approx a_{1,1}$ and $x \approx e_1$ ($Ax = \lambda x$)
(1) First **plane rotations** to make $A$ diag. dom.

suppose that $a_{1,1}$ is largest element

then $\lambda \approx a_{1,1}$ and $x \approx e_1$ ($Ax = \lambda x$)

(2) Consider orthogonal complement of $e_1$:

$$A \begin{pmatrix} 1 \\ w \end{pmatrix} = \begin{pmatrix} a_{1,1} & c^T \\ c & F \end{pmatrix} \begin{pmatrix} 1 \\ w \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ w \end{pmatrix}$$
(1) First **plane rotations** to make $A$ diag. dom.

suppose that $a_{1,1}$ is largest element

then $\lambda \approx a_{1,1}$ and $x \approx e_1$ ($Ax = \lambda x$)

(2) **Consider orthogonal complement of** $e_1$:

$$A \begin{pmatrix} 1 \\ w \end{pmatrix} = \begin{pmatrix} a_{1,1} & c^T \\ c & F \end{pmatrix} \begin{pmatrix} 1 \\ w \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ w \end{pmatrix}$$

leads to

$$\lambda = a_{1,1} + c^T w$$

$$(F - \lambda I)w = -c$$
Jacobi (2)

start with \( w = 0, \theta = a_{1,1} \)

Solve \( w \) from \( (F - \theta I)w = -c \) with G-J iterations
start with $w = 0, \theta = a_{1,1}$

Solve $w$ from $(F - \theta I)w = -c$ with G-J iterations

J. applied 10 rotations before switch to G-J
start with \( w = 0, \theta = a_{1,1} \)

Solve \( w \) from \((F - \theta I)w = -c\) with G-J iterations

J. applied 10 rotations before switch to G-J

He applied 2 G-J steps before updating \( \theta \)

Both decisions without further comment
Jacobi (2)

start with $w = 0, \theta = a_{1,1}$

Solve $w$ from $(F - \theta I)w = -c$ with G-J iterations

J. applied $10$ rotations before switch to G-J

He applied $2$ G-J steps before updating $\theta$

Both decisions without further comment

Bodewig (1959) advocated this method (without success?)

Quadratic convergence of J-rotations already fast enough?
start with $w = 0, \theta = a_{1,1}$

Solve $w$ from $(F - \theta I)w = -c$ with G-J iterations

J. applied 10 rotations before switch to G-J

He applied 2 G-J steps before updating $\theta$

Both decisions without further comment

Bodewig (1959) advocated this method (without success?)

Quadratic convergence of J-rotations already fast enough?

Goldstine suggested J’s rotations only for proving real eigenvalues
Krylov subspaces (1)

Krylov suggested in 1931 the subspace:

\[ K_m(A; x) = \text{span}\{x, Ax, \ldots, A^{m-1}x\} \]

for some convenient starting vector for construction of characteristic polynomial. But in his case: \( m = 6 \).

How to make things work for large \( m \)?
Krylov suggested in 1931 the subspace:

$$K_m(A; x) = \text{span}\{x, Ax, \ldots, A^{m-1}x\}$$

for some convenient starting vector $x$

for construction of characteristic polynomial
Krylov suggested in 1931 the subspace:

$$K_m(A; x) = \text{span}\{x, Ax, \ldots, A^{m-1}x\}$$

for some convenient starting vector $x$

for construction of characteristic polynomial

illconditioned basis, but in his case: $m = 6$

How to make things work for large $m$?
In the early 1950s: orthogonal basis
Krylov subspaces (2)

In the early 1950s: **orthogonal basis**

It does not help to build basis first
Krylov subspaces (2)

In the early 1950s: **orthogonal basis**

It does not help to build basis first

Start with $v_1 = x / \|x\|$  

Form $Av_1$ and orthogonalize w.r.t. $v_1$

Normalize: $v_2$ (so far nothing new!)
Krylov subspaces (2)

In the early 1950s: **orthogonal basis**

It does not help to build basis first

**Start with** \( v_1 = \frac{x}{\|x\|} \)

Form \( Av_1 \) and orthogonalize w.r.t. \( v_1 \)

Normalize: \( v_2 \) (so far nothing new!)

**Instead of** \( A^2 v_1 \), compute \( Av_2 \)
Krylov subspaces (2)

In the early 1950s: **orthogonal basis**

It does not help to build basis first

**Start with** $v_1 = x/\|x\|$

**Form** $Av_1$ and orthogonalize w.r.t. $v_1$

**Normalize:** $v_2$ (so far nothing new!)

**Instead of** $A^2v_1$, **compute** $Av_2$

**Orthogonalize w.r.t** $v_1$, $v_2$ and **normalize:** $v_3$
Krylov subspaces (3)

The general step is:

Form $AV_i$, Orth. w.r.t $v_1, \ldots, v_i$, normalize: $v_{i+1}$
Krylov subspaces (3)

The general step is:

Form $A v_i$, Orth. w.r.t $v_1, \ldots, v_i$, normalize: $v_{i+1}$

(modified) Gram-Schmidt orthogonalization
The general step is:

Form $A v_i$, Orth. w.r.t $v_1, \ldots, v_i$, normalize: $v_{i+1}$

(modified) Gram-Schmidt orthogonalization

Results in well-conditioned basis (Stewart, SIAM books)
Krylov methods (3)

Write $V_m = [v_1|v_2|\ldots|v_m]$
Krylov methods (3)

write \( V_m = [v_1 \mid v_2 \mid \ldots \mid v_m] \)

then G-S in matrix notation: \( AV_m = V_m H_m + c_m v_{m+1} e_m^T \)
Krylov methods (3)

write \( V_m = [v_1|v_2|\ldots|v_m] \)

then G-S in matrix notation: \( AV_m = V_m H_m + c_m v_{m+1} e_m^T \)

Note that \( H_m = V_m^T A V_m \)
**Krylov methods (3)**

Write $V_m = [v_1|v_2|\ldots|v_m]$

Then G-S in matrix notation: $AV_m = V_m H_m + c_m v_{m+1} e_m^T$

Note that $H_m = V_m^T A V_m$

The eigenvalues $\theta$ of $H_m$:

approximations for eigenvalues of $A$
Krylov methods (3)

write \( V_m = [v_1|v_2|\ldots|v_m] \)

then G-S in matrix notation: \( AV_m = V_m H_m + c_m v_{m+1} e_m^T \)

Note that \( H_m = V_m^T A V_m \)

The eigenvalues \( \theta \) of \( H_m \):

approximations for eigenvalues of \( A \)

\( H_m y = \theta y, z = V_m y \) is approximation for eigenvector of \( A \)
write $V_m = [v_1 | v_2 | \ldots | v_m]$

then G-S in matrix notation: $AV_m = V_m H_m + c_m v_{m+1} e_m^T$

Note that $H_m = V_m^T A V_m$

The eigenvalues $\theta$ of $H_m$:

approximations for eigenvalues of $A$

$H_m y = \theta y, z = V_m y$ is approximation for eigenvector of $A$

$A$ symmetric: LANCZOS METHOD (1952)
Krylov methods (3)

write \( V_m = [v_1 | v_2 | \ldots | v_m] \)

then G-S in matrix notation: \( AV_m = V_m H_m + c_m v_{m+1} e_m^T \)

Note that \( H_m = V_m^T AV_m \)

The eigenvalues \( \theta \) of \( H_m \):

approximations for eigenvalues of \( A \)

\( H_m y = \theta y, z = V_m y \) is approximation for eigenvector of \( A \)

\( A \) symmetric: LANCZOS METHOD (1952)

\( A \) unsymmetric: ARNOLDI METHOD (1952)
Davidson’s subspace

Compute residual $r = Az$.

Precondition $r$ (inverse iteration):

$$t = (DAI)^{-1}r$$

orthonormalize $t$ and expand subspace.

Claim: Newton method (Arnoldi? Davidson opens ways for other subspaces.)
Davidson’s subspace

Krylov subspaces popular after 1976 (Paige)

Compute residual $r = Az$

Preconditioning using inverse iteration:

$t = (D - A)^{-1}r$

Orthonormalize $t$ and expand subspace.

Claim: Newton method (?) preconditioned Arnoldi? Davidson opens ways for other subspaces.
Davidson’s subspace

Krylov subspaces popular after 1976 (Paige)

Davidson (1975) suggested other subspace:

Compute residual \( r = Az - \theta z \)
Davidson’s subspace

Krylov subspaces popular after 1976 (Paige)

Davidson (1975) suggested other subspace:

Compute residual $r = Az - \theta z$

Precondition $r(\approx \text{inverse iteration})$:

$t = (D_A - \theta I)^{-1}r$
Davidson’s subspace

Krylov subspaces popular after 1976 (Paige)

Davidson (1975) suggested other subspace:

Compute residual \( r = Az - \theta z \)

Precondition \( r \approx \text{inverse iteration} \):
\[
t = (D_A - \theta I)^{-1}r
\]

orthonormalize \( t \) and expand subspace
Krylov subspaces popular after 1976 (Paige)

**Davidson** (1975) suggested other subspace:

Compute residual $r = Az - \theta z$

Precondition $r(\approx \text{inverse iteration})$:

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

orthonormalize $t$ and expand subspace

claim: Newton method (?)

preconditioned Arnoldi?
Davidson’s subspace

Krylov subspaces popular after 1976 (Paige)

Davidson (1975) suggested other subspace:

Compute residual $r = Az - \theta z$

Precondition $r(\sim \text{inverse iteration})$:

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

orthonormalize $t$ and expand subspace

claim: Newton method (?)

preconditioned Arnoldi?

Davidson opens ways for other subspaces
Davidson - num. analysis

\[ r = (A I)^{-1} z \]
\[ t = (D A I)^{-1} \]
\[ r = z \]

With preconditioner

\[ (A I)^{-1} \]

no expansion of subspace

Insightful paper by Crouzeix, Philippe, Sadkane (1994)

Analysis for

\[ t = M^1_k \]

\[ k r \]

Suspect!

But successful for Chemistry problems

Idea: apply preconditioner instead of Jacobi rotations and use Jacobi's idea for new update of \[ z \]
\[ r = (A - \theta I)z \]
Davidson - num. analysis

\[ r = (A - \theta I)z \]

\[ t = (D_A - \theta I)^{-1}r \approx (A - \theta I)^{-1}r = z \]
Davidson - num. analysis

\[ r = (A - \theta I)z \]
\[ t = (D_A - \theta I)^{-1}r \approx (A - \theta I)^{-1}r = z \]

With preconditioner \((A - \theta I)^{-1}\) no expansion of subspace
Davidson - num. analysis

\[ r = (A - \theta I)z \]
\[ t = (D_k - \theta I)^{-1}r \approx (A - \theta I)^{-1}r = z \]

With preconditioner \((A - \theta I)^{-1}\) no expansion of subspace

Insightful paper by Crouzeix, Philippe, Sadkane (1994)

Analysis for \( t = M_k^{-1}r \)
Davidson - num. analysis

\[ r = (A - \theta I)z \]
\[ t = (D_A - \theta I)^{-1}r \approx (A - \theta I)^{-1}r = z \]

With preconditioner \((A - \theta I)^{-1}\) no expansion of subspace

Insightful paper by Crouzeix, Philippe, Sadkane (1994)

Analysis for \( t = M_k^{-1}r \)

\( M_k \) should not be close to \((A - \theta I)\)

Suspect! But successful for Chemistry problems
\[ r = (A - \theta I)z \]
\[ t = (D_A - \theta I)^{-1}r \approx (A - \theta I)^{-1}r = z \]

With preconditioner \((A - \theta I)^{-1}\) no expansion of subspace

Insightful paper by Crouzeix, Philippe, Sadkane (1994)

Analysis for \(t = M_k^{-1}r\)

\(M_k\) should not be close to \(A - \theta I\)

Suspect! But successful for Chemistry problems

Idea: apply preconditioner instead of Jacobi rotations and

use Jacobi’s idea for new update of \(z\)
Jacobi-Davidson

In Jacobi’s case:

\[ e_1 \] is the approximation for \( x \).

In subspace method we have approximation \( z \).

J. computes update in subspace \( e \)?

Sleijpen en VDV (1996): compute update in \( z \)? \((A I)\) restricted to \( z \)? is given by

\[ B = (I z z)(A I)(I z z) \] is given by

Expand subspace with (approx.) solution of \( B^t = r \).

Jacobi-Davidson method, SIMAX 1996

Newton method for RQ
Jacobi-Davidson

In Jacobi’s case: $e_1$ is the approximation for $x$

In subspace method we have approximation $z$
In Jacobi’s case: $e_1$ is the approximation for $x$

In subspace method we have approximation $z$

J. computes update in subspace $e_1$

Sleijpen en VDV (1996): compute update in $z^\perp$
In Jacobi's case: $e_1$ is the approximation for $x$

In subspace method we have approximation $z$

J. computes update in subspace $e_1^\perp$

Sleijpen en VDV (1996): compute update in $z^\perp$

$(A - \theta I)$ restricted to $z^\perp$ is given by

\[ B = (I - zz^*)(A - \theta I)(I - zz^*) \]
Jacobi-Davidson

In Jacobi’s case: \( e_1 \) is the approximation for \( x \)

In subspace method we have approximation \( z \)

J. computes update in subspace \( e_1^\perp \)

Sleijpen en VDV (1996): compute update in \( z^\perp \)

\((A - \theta I)\) restricted to \( z^\perp \) is given by

\[ B = (I - zz^*) (A - \theta I) (I - zz^*) \]

Expand subspace with (approx.) solution of \( Bt = r \)

Jacobi-Davidson method, SIMAX 1996

Newton method for RQ
Numerical example

\[ n = 100, \ A = \text{tridiag}(1, 2.4, 1) \]
\[ x = (1, 1, \ldots, 1)^T \]
Numerical example

\[ n = 100, \ A = \text{tridiag}(1, 2.4, 1) \]

\[ x = (1, 1, \ldots, 1)^T \]

**Davidson**: \( M_k = A - \theta_k I \): stagnation
Numerical example

\[ n = 100, \quad A = \text{tridiag}(1, 2.4, 1) \]
\[ x = (1, 1, \ldots, 1)^T \]

**Davidson:** \( M_k = A - \theta_k I: \) stagnation

**Jacobi-Davidson:** \( M_k = A - \theta_k I: \) it's
Numerical example

\[ n = 100, \ A = \text{tridiag}(1, 2.4, 1) \]
\[ x = (1, 1, \ldots, 1)^T \]

**Davidson:** \[ M_k = A - \theta_k I: \text{stagnation} \]

**Jacobi-Davidson:** \[ M_k = A - \theta_k I: \text{it's} \]

Davidson, prec. with GMRES(5) for \( (A - \theta_k I) \tilde{t} = r \):

**slow convergence** (since \( \theta_k \approx \lambda \))
Numerical example

\[ n = 100, \ A = \text{tridiag}(1, 2.4, 1) \]
\[ x = (1, 1, \ldots, 1)^T \]

**Davidson:** \[ M_k = A - \theta_k I: \text{stagnation} \]

**Jacobi-Davidson:** \[ M_k = A - \theta_k I: 5 \text{ it’s} \]

Davidson, prec. with GMRES(5) for \( (A - \theta_k I)\tilde{t} = r: \)

slow convergence (since \( \theta_k \approx \lambda \))

Jac.Dav., GMRES(5) for \( F\tilde{t} = r \) with

\[ F = (I - zz^T)(A - \theta_k I)(I - zz^T): 13 \text{ it’s} \]

Note that \( F \) has no small eigenvalues
More practical example

Acoustics, attachment line:

\[ Ax + \lambda Bx + \lambda^2 Cx = 0 \]
More practical example

Acoustics, attachment line:
\[ Ax + \lambda Bx + \lambda^2 Cx = 0 \]

For problem coming from **acoustics**:
\[ A, \ C \text{ 19-diagonal, } B \text{ complex, } n = 136161 \]
More practical example

Acoustics, attachment line:

\[ Ax + \lambda Bx + \lambda^2 Cx = 0 \]

For problem coming from **acoustics**: 

- \( A \), 19-diagonal, \( B \) complex, \( n = 136161 \)

Results for interior isolated eigenvalue (resonance) on a **Cray T3D**

<table>
<thead>
<tr>
<th>Processors</th>
<th>Elapsed time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>206.4</td>
</tr>
<tr>
<td>32</td>
<td>101.3</td>
</tr>
<tr>
<td>64</td>
<td>52.1</td>
</tr>
</tbody>
</table>

June 3, 2006, Michel Crouzeix – p.18/18
More practical example

Acoustics, attachment line:

\[ Ax + \lambda Bx + \lambda^2 Cx = 0 \]

For problem coming from acoustics:

\( A, C \) 19-diagonal, \( B \) complex, \( n = 136161 \)

Results for interior isolated eigenvalue (resonance)
on a **CRAY T3D**

<table>
<thead>
<tr>
<th>Processors</th>
<th>Elapsed time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>206.4</td>
</tr>
<tr>
<td>32</td>
<td>101.3</td>
</tr>
<tr>
<td>64</td>
<td>52.1</td>
</tr>
</tbody>
</table>

For \( n = 274625 \), on 64 processors: **93.3 seconds**

1 invert step \( \approx 3 \) hours