Lanczos Method for Eigensystems

Aiichiro Nakano

Collaboratory for Advanced Computing & Simulations
Department of Computer Science
Department of Physics & Astronomy
Department of Chemical Engineering & Materials Science
Department of Biological Sciences
University of Southern California

Email: anakano@usc.edu

B. N. Parlett
The Symmetric Eigenvalue Problem
(Prentice-Hall, ’80) Secs. 11-13
History’s Top 10 Algorithms Again

In putting together this issue of Computing in Science & Engineering, we knew three things: it would be difficult to list just 10 algorithms; it would be fun to assemble the authors and read their papers; and, whatever we came up with in the end, it would be controversial. We tried to assemble the 10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century. Following is our list (here, the list is in chronological order; however, the articles appear in no particular order):

- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues
- Quicksort Algorithm for Sorting
- Fast Fourier Transform
- Integer Relation Detection
- Fast Multipole Method

IEEE Comput. Sci. Eng. 2(1), 22 ('00)
Rayleigh Quotient

Theorem
Let A be an \( n \times n \) real symmetric matrix, \( \lambda_1[A] \leq \ldots \leq \lambda_n[A] \) its eigenvalues in ascending order, \( x \in \mathbb{R}^n \), & the Rayleigh quotient

\[
\rho(x; A) = \frac{x^TAx}{x^Tx}
\]

then

\[
\begin{align*}
\lambda_1[A] &= \min_{x \in \mathbb{R}^n} \rho(x; A) \\
\lambda_n[A] &= \max_{x \in \mathbb{R}^n} \rho(x; A)
\end{align*}
\]

Proof
Let \( q^{(k)} \) be the \( k \)-th orthonormalized eigenvector of \( A \), \( Aq_k = \lambda_k q_k \), & orthogonal transformation matrix, \( Q = [q_1 q_2 \cdots q_n] \), then

\[
Q^T A Q = \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_n
\end{bmatrix}
\]

Let \( x = Qz \) (note \( Q^T Q = I \)), then

\[
\rho(x; A) = \frac{z^T Q^T A Q z}{z^T Q Q z} = \frac{z_1^2 \lambda_1 + \cdots + z_n^2 \lambda_n}{z_1^2 + \cdots + z_n^2}
\]

which is a weighted average of \( \lambda_1, \ldots, \lambda_n \), & the minimum is when \( z^T = (1,0,\ldots,0) = e_1 \) & \( x = Qe_1 = q_1 \).
Rayleigh-Ritz Procedure

**Theorem**

Let \( \{q_1, \ldots, q_m\} \) be an orthonormal set that spans \( \mathbb{R}^m \) (\( m < n \)) \( \subset \mathbb{R}^n \), so that any vector \( x \in \mathbb{R}^m \) is expressed as a linear combination of \( q_1, \ldots, q_m \):

\[
x = z_1 q_1 + \cdots + z_m q_m \quad \text{or} \quad
n \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = m \begin{bmatrix} q_1 & \cdots & q_m \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} = Qz
\]

then the best approximations for \( \lambda_1[A] \) & \( \lambda_n[A] \) are obtained by diagonalizing

\[
H = Q^T A Q
\]
as \( \lambda_1[H] \) & \( \lambda_m[H] \).

**Proof**

Note \( \left(Q^T Q\right)_{ij} = \sum_{k=1}^{n} Q_{ki} Q_{kj} = q_i \cdot q_j = \delta_{ij} \) \( 1 \leq i, j \leq m \)
then

\[
\rho(x; A) = \frac{z^T Q^T A Q z}{z^T Q^T Q z} = \frac{z^T H z}{z^T z} = \frac{z_1^2 \lambda_1(H) + \cdots + z_m^2 \lambda_m(H)}{z_1^2 + \cdots + z_m^2}
\]

the minimum of which is \( \lambda_1[H] \).
Orthogonalization by QR Decomposition

• **Gram-Schmidt orthonormalization:** The orthonormal set $Q$ required for the Rayleigh-Ritz procedure is obtained starting from an arbitrary set of $m$ vectors, $S = [s_1...s_m] \ (s_j \in \mathbb{R}^n)$ as:

$$q_1 = s_1 / |s_1|$$

for $i = 2$ to $m$

$$q'_i = s_i - \sum_{j=1}^{i-1} q_j (q_j \cdot s_i)$$

Projection!

$$q_i = q'_i / |q'_i|$$

endfor

• The Gram-Schmidt amounts to QR decomposition, $S = QR$, where $R$ is an $m\times m$ right-triangle matrix:

$$\begin{bmatrix}
  n & m \\
  s_1 & s_2 & s_3 & s_4 \\
\end{bmatrix} = \begin{bmatrix}
  n & m \\
  q_1 & q_2 & q_3 & q_4 \\
\end{bmatrix} \begin{bmatrix}
  m \\
  |q'_1| & q_1 \cdot s_2 & q_1 \cdot s_3 & q_1 \cdot s_4 \\
  0 & |q'_2| & q_2 \cdot s_3 & q_2 \cdot s_4 \\
  0 & 0 & |q'_3| & q_3 \cdot s_4 \\
  0 & 0 & 0 & |q'_4| \\
\end{bmatrix}$$

$$\therefore s_i = |q'_i| q_i + \sum_{j=1}^{i-1} q_j (q_j \cdot s_i)$$
1. Start from \( S = [s_1 \ldots s_m] \) (\( s_j \in \mathbb{R}^n \)) & do Gram-Schmidt orthonormalization, \( S = QR \), to obtain an orthonormal set \( Q = [q_1 \ldots q_m] \)

2. Form \( H = Q^T AQ \)

3. Diagonalize \( H \) to get \( \lambda_1[H], \ldots, \lambda_m[H] \): \( Hg_k = \lambda_k[H]g_k \) \( (k = 1, \ldots, m) \)

4. The approximations of \( \lambda_1[A] \) & \( \lambda_n[A] \) are given by \( \lambda_1[H] \) & \( \lambda_m[H] \) with the corresponding eigenvectors, \( y_k = Qg_k \) \( (k = 1 \& m) \).
Krylov Subspace

- **Krylov subspace** $S_m$ is spanned by a Krylov matrix, $K^m(f) = [f \, Af \, \ldots \, A^{m-1}f]$ ($f \in \mathbb{R}^n$)

**Theorem**

Let $Q_m$ be the orthonormal basis obtained by QR factorization, $K_m(f) = Q_mR$, then $T_m = Q_m^TAQ_m$ is a tridiagonal matrix

**Proof**

For $i > j+1$, $q_i^T(Aq_j) = 0$, since $Aq_j \subset S_{j+1}$ by construction & $q_i \perp S_{j+1}$ by Gram-Schmidt orthonormalization for $i > j+1$. By the symmetry of $A$, $q_i^T(Aq_j) = q_j^T(A^Tq_i) = q_j^T(Aq_i) = 0$ for $j > i+1$ or $i < j-1$.

$$T_m = \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \alpha_2 & \beta_2 \\ \vdots & \ddots & \ddots \\ \beta_{m-2} & \alpha_{m-1} & \beta_{m-1} \\ \beta_{m-1} & \alpha_m \end{bmatrix}$$

\[
\begin{align*}
\alpha_j &= q_j^T Aq_j \\
\beta_j &= q_{j+1}^T Aq_j \\
\end{align*}
\]

\[
\begin{align*}
\alpha_j &= q_j^T Aq_j & j = 1, \ldots, m \\
\beta_j &= q_{j+1}^T Aq_j & j = 1, \ldots, m - 1 \\
\end{align*}
\]

- **Tridiagonal matrix can be diagonalized in $O(N)$ time**

Alexei Krylov with daughter Anna, later Anna Kapitsa, wife of Pyotr Kapitsa (1904)
Recursion Formula

- Due to the tridiagonality, $Aq_i$ is a linear combination of $q_{i-1}$, $q_i$ & $q_{i+1}$:
  \[ Aq_i = \beta_{i-1}q_{i-1} + \alpha_i q_i + \beta_i q_{i+1} \quad (2 \leq i \leq m-1) \]

  If we define $q_0 = 0$, the above equation is valid for $i = 1$ as well. Let $r_i = \beta_i q_{i+1}$ ($r_i$ is a component of $Aq_i$ orthogonal to $q_j$ for $j \leq i$), then
  \[ r_i = Aq_i - \beta_{i-1}q_{i-1} - \alpha_i q_i \quad (1 \leq i \leq m-1) \]

- Lanczos algorithm:

  Given $r_0, \beta_0 = \|r_0\|$ ($q_0 = 0$)
  
  for $i = 1, \ldots, m$
  
  \[ q_i \leftarrow \frac{r_{i-1}}{\beta_{i-1}} \]
  \[ r_i \leftarrow Aq_i - \beta_{i-1}q_{i-1} \]
  \[ \alpha_i \leftarrow q_i^T r_i \quad \therefore q_i^T (Aq_i - \beta_{i-1}q_{i-1}) = q_i^T Aq_i = \alpha_i \quad \text{(orthogonality)} \]
  \[ r_i \leftarrow r_i - \alpha_i q_i \]
  \[ \beta_i = \|r_i\| \quad \text{(only when} \quad i \leq m-1) \]
  
  endfor

Keep increasing $m$ until $\lambda_1[T_m]$ converges
An Application of Rayleigh-Ritz/Lanczos

- Search for transition states (with a negative eigenvalue of the Hessian matrix, $\frac{\partial^2 E}{\partial r_i \partial r_j}$, by following the eigenvector with the smallest eigenvalue
  — Rayleigh-Ritz: Kumeda, Wales & Munro, Chem. Phys. Lett. 341, 185 (’01)
  — Lanczos: Mousseau et al., J. Mol. Graph. Model. 19, 78 (’01)

Figure from Prof. H. B. Schlegel; http://chem.wayne.edu/schlegel
**Algorithm Lanczos**

**Input:**
- \( \mathbf{R} \in \mathbb{R}^{3N} \): a state
- logical \textit{initialize}: TRUE for the first call in each event generation; FALSE otherwise

**Output:**
- \( \lambda_1 \): the minimum eigenvalue of the Hessian matrix, \( \mathbf{H}(\mathbf{R}) = \partial^2 \mathcal{V} / \partial \mathbf{R}^2 \)
- \( \mathbf{V}^1 \in \mathbb{R}^{3N} \): the Hessian eigenvector corresponding to \( \lambda_1 \)

**Steps:**
- \textbf{if} \textit{initialize}
  - randomize \( \Delta \in \mathbb{R}^{3N} \), such that it contains no translational motion
  - \( s \leftarrow 0 \)
  - \( \beta^s \leftarrow \| \Delta \| \)
  - \( \mathbf{Q}^s \in \mathbb{R}^{3N} \leftarrow 0 \)
  - \textbf{do}
    - \( s \leftarrow s + 1 \)
    - \( \mathbf{Q}^s \leftarrow \Delta / \beta^{s-1} \)
    - \( c_{id} \leftarrow \max_{l,t} \{ |q_{il}^t|, |q_{il}^t| \} \) \( i = 1, \ldots, N; \alpha = x, y, z \) / \( \delta_{id} \)
    - \( \Delta \leftarrow c_{id}[ -F(\mathbf{R} + \mathbf{Q}^s / c_{id}) + F(\mathbf{R}) ] - \beta^{s-1} \mathbf{Q}^{s-1} \)
    - \( \alpha^s \leftarrow \mathbf{Q}^{s-1} \Delta \)
    - \( \Delta \leftarrow \Delta - \alpha^s \mathbf{Q}^s \)
    - \( \beta^s \leftarrow \| \Delta \| \)
  - \textbf{end do}
- \textbf{diagonalize} \( \mathbf{T}_s \)
  - \[ \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \alpha_2 & \beta_2 \\ \beta_2 & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \beta_{s-2} & \alpha_{s-1} & \beta_{s-1} \\ \beta_{s-1} & \alpha_s \end{bmatrix} \]
  - so that \( \mathbf{Q}_s^1 \mathbf{T}_s \mathbf{Q}_s = \text{diag}(\bar{\lambda}_1^s, \ldots, \bar{\lambda}_s^s) \)
- \textbf{while} \( |(\bar{\lambda}_1^s - \bar{\lambda}_1^1) / \bar{\lambda}_1^1| > \Delta_{\text{eigen}} \)
  - \( \lambda_1 \leftarrow \bar{\lambda}_1^s \)
  - \( \mathbf{V}^1 \leftarrow \sum_{s=1}^{\infty} \mathbf{Q}^s \mathbf{\hat{q}}_k^1 \)
  - \( \mathbf{V}^1 \leftarrow \mathbf{V}^1 / \| \mathbf{V}^1 \| \)

\* \( \text{diag}(\bar{\lambda}_1^1, \ldots, \bar{\lambda}_s^1) \) is an \( s \) by \( s \) diagonal matrix, with its diagonal elements given by \( \bar{\lambda}_1^1, \ldots, \bar{\lambda}_s^1 \). \( \mathbf{\hat{q}}_k^1 \) is an \( s \) by \( s \) orthogonal matrix, with \( \mathbf{\hat{q}}_m^1 \in \mathbb{R}^s \) is the \( m \)th eigenvector of \( \mathbf{T}_s \).
Sample Run of Lanczos Program
Electronic Energy Bands of GaAs

• 8-band $k\cdot p$ model

\[
H = \begin{pmatrix}
A & 0 & V^* & 0 & \sqrt{3} V & -\sqrt{2} U & -U & \sqrt{2} V^* \\
0 & A & -\sqrt{2} U & -\sqrt{3} V^* & 0 & -V & \sqrt{2} V & U \\
V & -\sqrt{2} U & -P + Q & -S^* & R & 0 & \sqrt{1/3} S & -\sqrt{2} Q \\
0 & -\sqrt{3} V & -S & -P - Q & 0 & R & -\sqrt{2} R & \frac{1}{\sqrt{2}} S \\
\sqrt{3} V^* & 0 & R^* & 0 & -P - Q & S^* & \frac{1}{\sqrt{2}} S^* & \sqrt{2} R^* \\
-\sqrt{2} U & -V^* & 0 & R^* & S & -P + Q & \sqrt{2} Q & \sqrt{1/3} S^* \\
-U & \sqrt{2} V^* & \sqrt{1/3} S^* & -\sqrt{2} R^* & \frac{1}{\sqrt{2}} S & \sqrt{2} Q & -P - \Delta & 0 \\
\sqrt{2} V & U & -\sqrt{2} Q & \frac{1}{\sqrt{2}} S & \sqrt{2} R & \sqrt{1/3} S & 0 & -P - \Delta
\end{pmatrix}
\]

\[
A = E_c - \frac{\hbar^2}{2m_0} (\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + \hat{\sigma}_z^2),
\]

\[
P = -E_v - \gamma_1 \frac{\hbar^2}{2m_0} (\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + \hat{\sigma}_z^2),
\]

\[
Q = -\gamma_2 \frac{\hbar^2}{2m_0} (\hat{\sigma}_x^2 + \hat{\sigma}_y^2 - 2\hat{\sigma}_z^2),
\]

\[
R = \sqrt{3} \frac{\hbar^2}{2m_0} [\gamma_2 (\hat{\sigma}_x^2 - \hat{\sigma}_y^2) - 2i \gamma_3 \hat{\sigma}_x \hat{\sigma}_y],
\]

\[
S = -\sqrt{3} \frac{\hbar^2}{m_0} \hat{\sigma}_z (\hat{\sigma}_x - i \hat{\sigma}_y),
\]

\[
U = -\frac{i}{\sqrt{3}} P_0 \hat{\sigma}_z,
\]

\[
V = -\frac{i}{\sqrt{6}} P_0 (\hat{\sigma}_x - i \hat{\sigma}_y).
\]

C. Pryor, Phys. Rev. B 57, 7190 ('98)
Lanczos Program in Fortran

do s = 1,NWF
    q(:,:,,:,s) = v/bet(s-1)
    call hamiltonian_op(q(:,:,,:,s),hv) ! Operates Hamiltonian H on Q(S)
    v = hv-bet(s-1)*q(:,:,,:,s-1)
    alp(s) = inner_product(q(:,:,,:,s),v)
    v = v-alp(s)*q(:,:,,:,s)
    bet(s) = sqrt(inner_product(v,v))
    call tridiag(eval,s) ! Diagonalize the S by S tridiagonal matrix
end do ! Lanczos iteration S

Given \( r_0, \beta_0 = \|r_0\| \) (\( q_0 = 0 \))

for \( i = 1, \ldots, m \)

\( q_i \leftarrow r_{i-1} / \beta_{i-1} \)
\( r_i \leftarrow Aq_i - \beta_{i-1}q_{i-1} \)
\( \alpha_i \leftarrow q_i^T r_i \)
\( r_i \leftarrow r_i - \alpha_i q_i \)
\( \beta_i = \|r_i\| \) (only when \( i \leq m-1 \))

endfor
Band-edge Wave Functions

- Band-edge states in an array of GaN quantum dots in AlN matrix

Valence-band top

0.802 eV 3.405 eV 3.438 eV 3.438 eV

3.519 eV 3.553 eV 3.580 eV 3.612 eV

3.622 eV 3.657 eV 3.692 eV 3.717 eV

Conduction-band states

S. Sburlan, Ph.D. dissertation, USC ('13)