Orthogonal Eigenvectors and Gram-Schmidt

Inderjit S. Dhillon

The University of Texas at Austin

Beresford N. Parlett

The University of California at Berkeley

Joint GAMM-SIAM Conference on Applied Linear Algebra

University of Düsseldorf, Germany

July 25, 2006

- One of FOUR main papers on this work
- Algorithm MR^3 or MRRR
  - Acronym for Multiple Relatively Robust Representations
  - Accurate but turgid title
  - Jim Demmel has a more catchy title...
- Guiding Principle: No Gram-Schmidt

- Excerpt from the Introduction

interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- Excerpt from the Introduction

> interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- What this talk will not do

- Excerpt from the Introduction

> interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- What this talk will not do
  - Roundoff Error Analysis

- Excerpt from the Introduction

> interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- What this talk will not do
  - Roundoff Error Analysis
  - Theorems, Proofs

- Excerpt from the Introduction

> interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- What this talk will not do
  - Roundoff Error Analysis
  - Theorems, Proofs
  - Performance Numbers

- Excerpt from the Introduction

> interest very close to 0. The middle part of this paper presents the relevant error analysis. Although essential for our results, this analysis will be indigestible for most readers, but it tells us that changes of only 3 or 4 units in the last digit of each entry of the input L, D and the output twisted factors suffice to give the exact relation.

- What this talk will not do
  - Roundoff Error Analysis
  - Theorems, Proofs
  - Performance Numbers
  - Stick closely to the paper

Diagonal of the Inverse

- Let J be a tridiagonal that is irreducible and invertible
- Not necessarily symmetric

Diagonal of the Inverse

- Let J be a tridiagonal that is irreducible and invertible
- Not necessarily symmetric
- Perform triangular factorization “down” and “up” (no pivoting)

$$J = L_+ D_+ U_+ = U_- D_- L_-$$

- D_+ are “forward” pivots while D_- are “backward” pivots

Diagonal of the Inverse

- Let J be a tridiagonal that is irreducible and invertible
- Not necessarily symmetric
- Perform triangular factorization “down” and “up” (no pivoting)

$$J = L_+ D_+ U_+ = U_- D_- L_-$$

- D_+ are “forward” pivots while D_- are “backward” pivots
- Is there a relation between D_+ and D_-?

Diagonal of the Inverse

- Let J be a tridiagonal that is irreducible and invertible
- Not necessarily symmetric
- Perform triangular factorization “down” and “up” (no pivoting)

$$J = L_+ D_+ U_+ = U_- D_- L_-$$

- D_+ are “forward” pivots while D_- are “backward” pivots
- Is there a relation between D_+ and D_-?
- Beautiful identity:

$$D_+ + D_- = \mathrm{diag}\,(J) + \mathrm{diag}\,(J^{-1})^{-1}$$

- FOUR New Ideas

- FOUR New Ideas
  - Replace the tridiagonal with a bidiagonal

- FOUR New Ideas
  - Replace the tridiagonal with a bidiagonal
  - Shift close to clusters, but with differential transforms

- FOUR New Ideas
  - Replace the tridiagonal with a bidiagonal
  - Shift close to clusters, but with differential transforms
  - Twist, again with differential transforms

- FOUR New Ideas
  - Replace the tridiagonal with a bidiagonal
  - Shift close to clusters, but with differential transforms
  - Twist, again with differential transforms
  - Analyze with a Representation Tree

Difficulties

- All eigenvalues of T are easily computed in $O(n^2)$ time
- Given $\hat{\lambda}$, inverse iteration computes the eigenvector:

$$(T - \hat{\lambda} I) x_{i+1} \;=\; x_i, \quad i = 0, 1, 2, \ldots$$

  - Costs $O(n)$ per iteration
  - Typically, 1-3 iterations are enough
- **BUT**, inverse iteration only guarantees

$$\|T\hat{v} - \hat{\lambda}\hat{v}\| \;=\; O(\varepsilon \|T\|)$$

Fundamental Limitations

Gap Theorem :

$$\sin \angle(v, \hat{v}) \;\leq\; \frac{\|T\hat{v} - \hat{\lambda}\hat{v}\|}{\mathrm{Gap}(\hat{\lambda})}$$

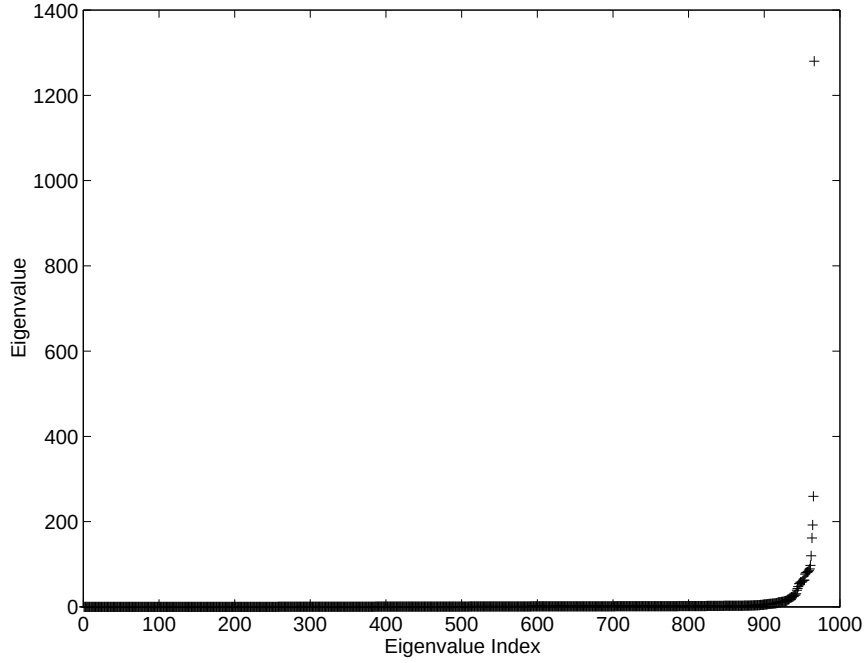
$\mathrm{Gap}(\hat{\lambda})$ can be small :

$$\begin{bmatrix} 1 & \varepsilon_1 & & \\ \varepsilon_1 & 1 & \varepsilon_2 & \\ & \varepsilon_2 & 1 & \varepsilon_3 \\ & & \varepsilon_3 & 1 \end{bmatrix}$$

When eigenvalues are close, independently computed eigenvectors **WILL NOT** be mutually orthogonal

Example from Quantum Chemistry

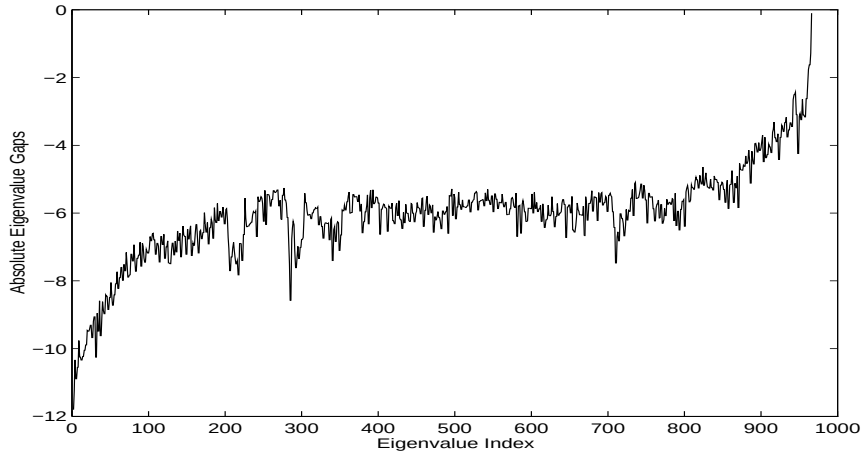
- Symmetric positive definite eigenproblem, $n = 966$
- Occurs in Møller-Plesset theory in the modeling of biphenyl
- Eigenvalue Distribution:



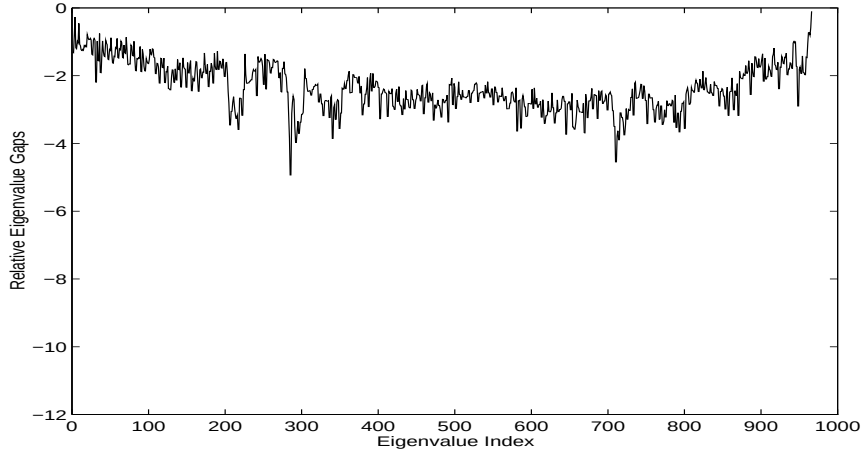
- LAPACK "clusters" — numbered $1, \ldots, 939$ and smaller ones

Example from Quantum Chemistry

Plot of $\mathrm{Absgap}(i) = \log_{10}(\min(\lambda_{i+1} - \lambda_i, \lambda_i - \lambda_{i-1})/\|T\|)$ versus i :



Plot of $\mathrm{Relgap}(i) = \log_{10}(\min(\lambda_{i+1} - \lambda_i, \lambda_i - \lambda_{i-1})/|\lambda_i|)$ versus i :

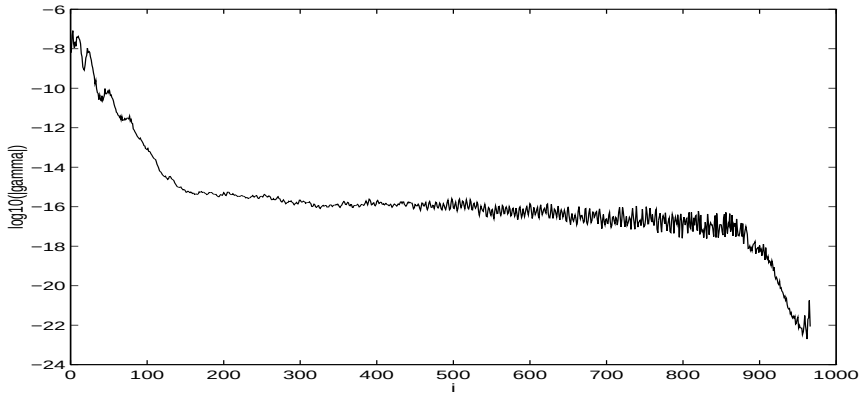


First Steps

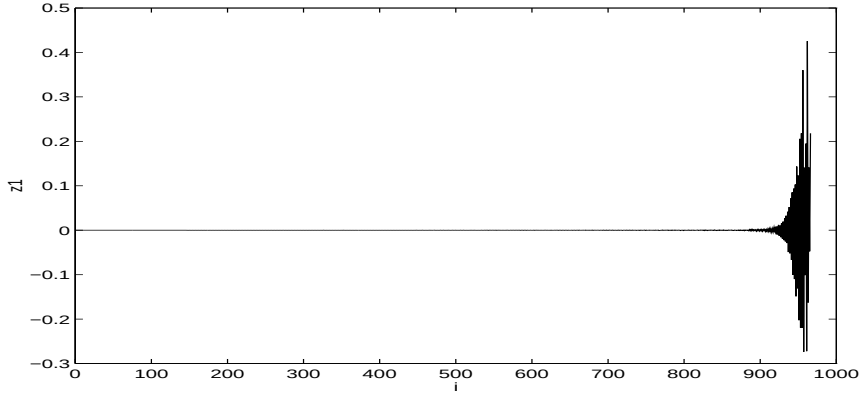
- Factor $T = LDL^T$
- Compute eigenvalues of LDL^T — by dqds or bisection
- For each eigenvalue λ, compute eigenvector by inverse iteration

Computing Eigenvector #1

- $\hat{\lambda}_1 = 4.1338 \times 10^{-08}$, $\hat{\lambda}_2 = 4.3417 \times 10^{-08}$, $\hat{\lambda}_3 = 4.5\ldots \times 10^{-08}$
- Factor $T_1 = LDL^T - \hat{\lambda}_1 I = L_+ D_+ L_+^T = U_- D_- U_-^T$ (up & down)
- Compute $\gamma(i) = D_+(i) + D_-(i) - T_1(i,i)$

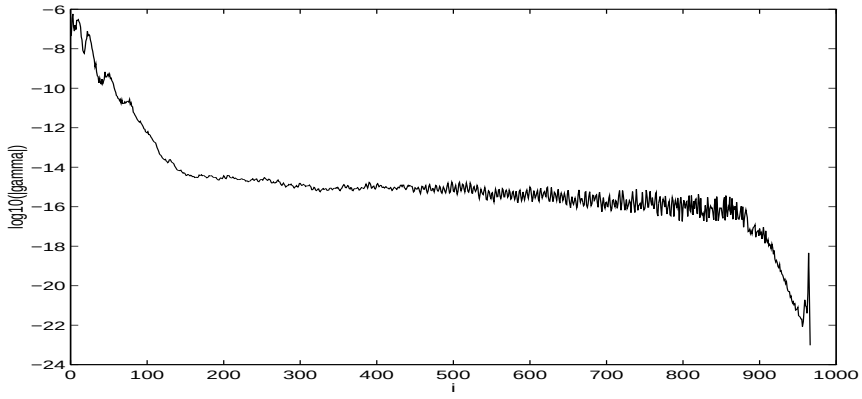


- Solve $T_1 z_1 = \gamma_r e_r$, where $\gamma_r = \min_k |\gamma_k|$

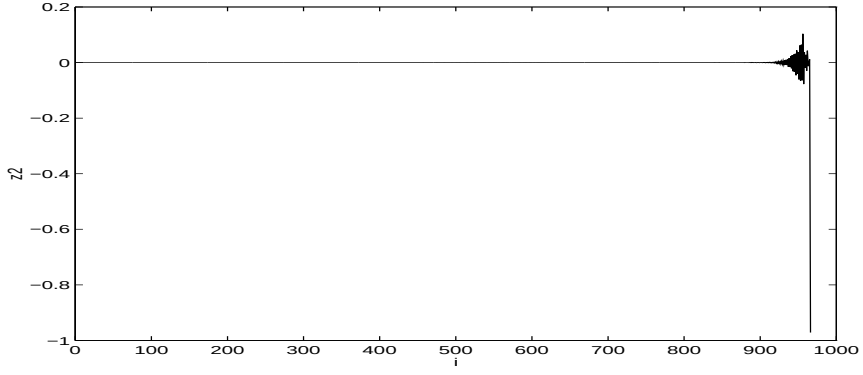


Computing Eigenvector #2

- $\hat{\lambda}_1 = 4.1338 \times 10^{-08}$, $\hat{\lambda}_2 = 4.3417 \times 10^{-08}$, $\hat{\lambda}_3 = 4.5\ldots \times 10^{-08}$
- Factor $T_2 = LDL^T - \hat{\lambda}_2 I = L_+ D_+ L_+^T = U_- D_- U_-^T$ (up & down)
- Compute $\gamma(i) = D_+(i) + D_-(i) - T_2(i,i)$



- Solve $T_2 z_2 = \gamma_r e_r$, where $\gamma_r = \min_k |\gamma_k|$



Accuracy of the vectors

- $|z_1^T z_2| < 2\varepsilon$
- Residual norms are $< \varepsilon \|T\|$

In general,

- $\min_k |\gamma_k| \approx \varepsilon |\hat{\lambda}|$
- Dot Product between two vectors $\propto \frac{\varepsilon |\hat{\lambda}|}{\mathrm{Gap}}$
- If two eigenvalues share d leading digits, then dot product $\approx 10^d \varepsilon$

Smaller Relative Gaps

- $\hat{\lambda}_{280}$ to $\hat{\lambda}_{303}$ are :

$$0.\mathbf{28}52950617$$
$$0.\mathbf{28}77656004$$
$$\vdots$$
$$0.\mathbf{2901}068244$$
$$0.\mathbf{29015}14166$$
$$0.\mathbf{29014}80490$$

- Some eigenvalues agree in 5 digits

Smaller Relative Gaps

- $\hat{\lambda}_{280}$ to $\hat{\lambda}_{303}$ are :

$$\begin{array}{c} 0.\mathbf{28}52950617 \\ 0.\mathbf{28}77656004 \\ \vdots \\ 0.\mathbf{2901}068244 \\ 0.\mathbf{29015}14166 \\ 0.\mathbf{29014}80490 \end{array}$$

- Some eigenvalues agree in 5 digits
- Compute *new* shifted representations:

$$LDL^T - 0.2852950617I = L_1 D_1 L_1^T$$

$$\begin{array}{c} 0.60738244647 \times 10^{-15} \\ 0.24705386756 \times 10^{-02} \\ \vdots \\ 0.\mathbf{48}11762683 \times 10^{-02} \\ 0.\mathbf{485}2987244 \times 10^{-02} \\ 0.\mathbf{485}5354851 \times 10^{-02} \end{array}$$

Smaller Relative Gaps

- $\hat{\lambda}_{280}$ to $\hat{\lambda}_{303}$ are :

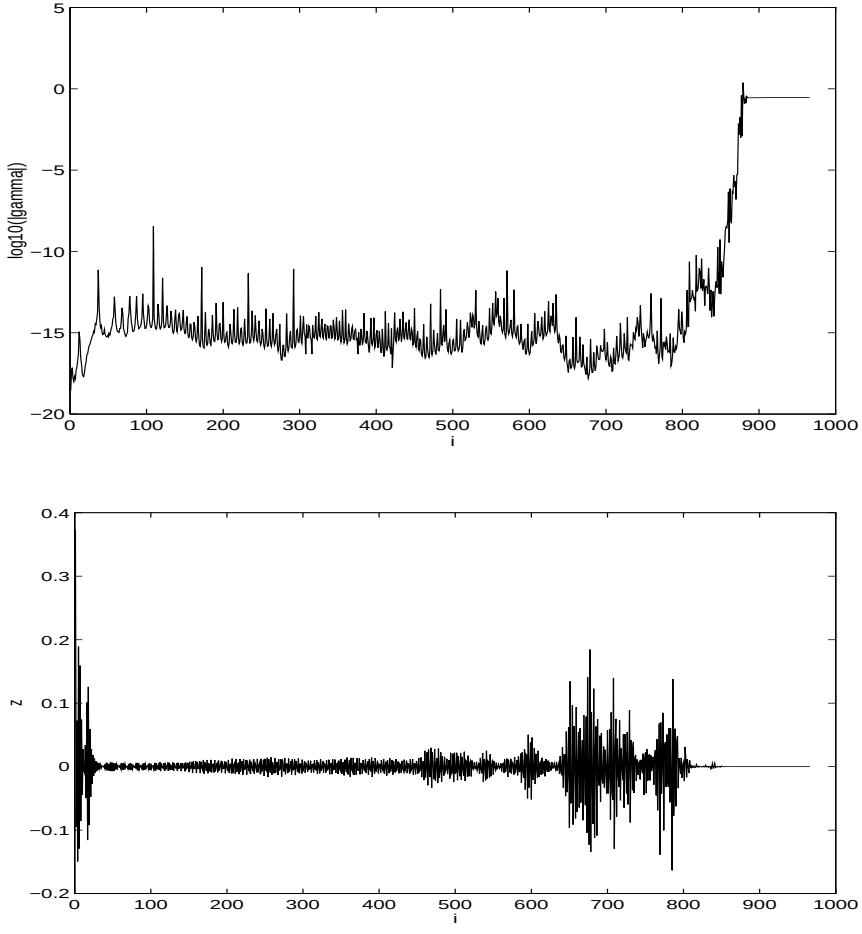
$$0.\mathbf{28}52950617$$
$$0.\mathbf{28}77656004$$
$$\vdots$$
$$0.\mathbf{2901}068244$$
$$0.\mathbf{29015}14166$$
$$0.\mathbf{29014}80490$$

- Some eigenvalues agree in 5 digits
- Compute *new* shifted representations:

$LDL^T - 0.2852950617I = L_1 D_1 L_1^T$	$L_1 D_1 L_1^T - 0.4855354851 \times 10^{-02} I = L_2 D_2 L_2^T$
$0.60738244647 \times 10^{-15}$	
$0.24705386756 \times 10^{-02}$	
$\vdots$	$-0.4459216793 \times 10^{-04}$
$0.\mathbf{48}11762683 \times 10^{-02}$	$-0.3367607591 \times 10^{-05}$
$0.\mathbf{485}2987244 \times 10^{-02}$	$-0.1463151420 \times 10^{-17}$
$0.\mathbf{485}5354851 \times 10^{-02}$	

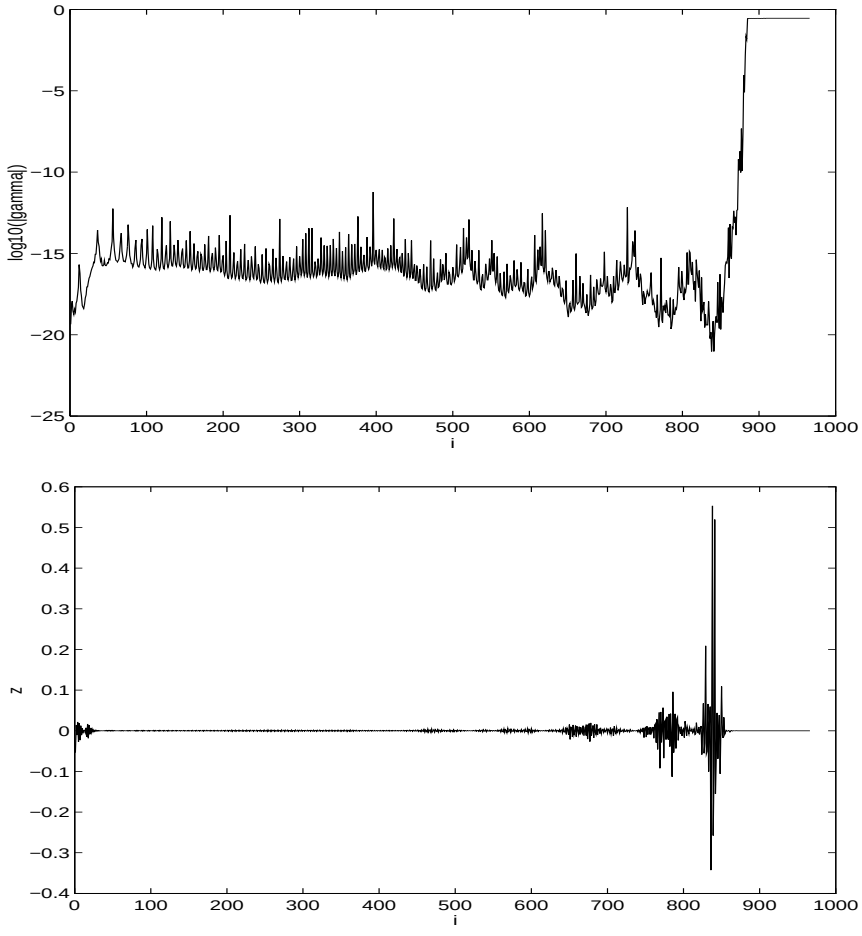
The computed vectors

- $\hat{\lambda} = -0.4459216793 \times 10^{-04}$



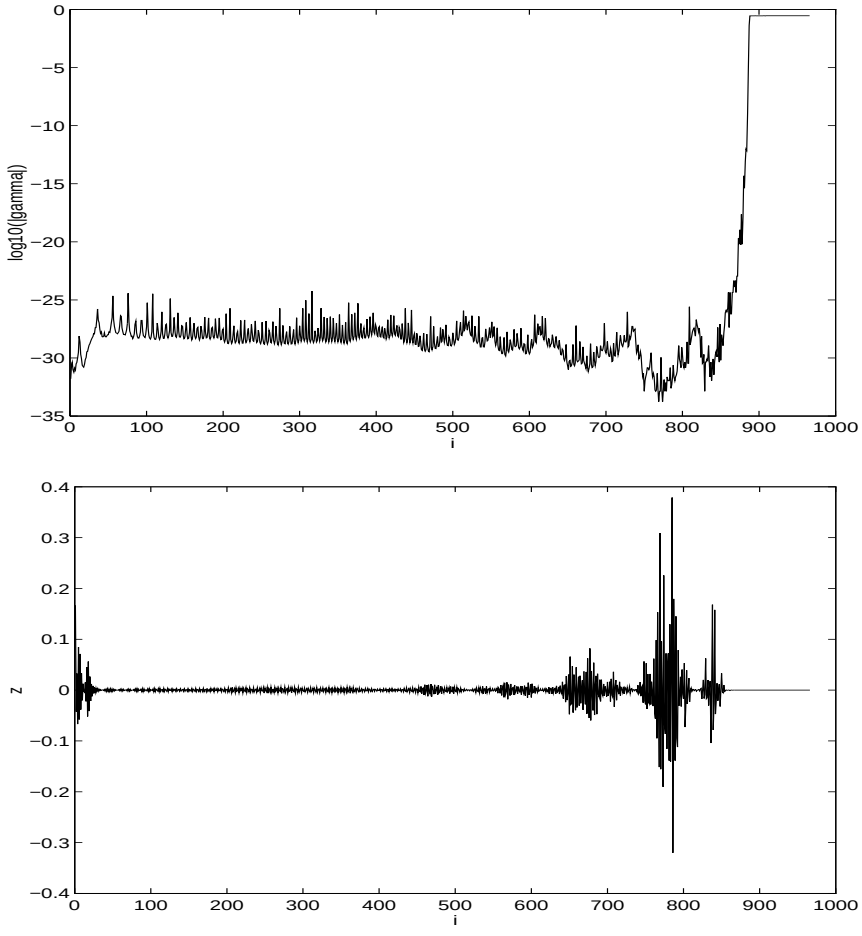
The computed vectors

- $\hat{\lambda} = -0.3367607591 \times 10^{-05}$



The computed vectors

- $\hat{\lambda} = -0.1463151420 \times 10^{-17}$



- Maximum Dot Product $< 3\varepsilon$.

Key Idea 1 — Replace the tridiagonal with a bidiagonal

- Bidiagonal Factorization

$$T \longrightarrow LDL^T$$

Key Idea 1 — Replace the tridiagonal with a bidiagonal

- Bidiagonal Factorization

$$T \longrightarrow LDL^T$$

- Relative condition number

$$\text{relcond}(\lambda) = \frac{v^T L|D|L^T v}{v^T LDL^T v}$$

Key Idea 1 — Replace the tridiagonal with a bidiagonal

- Bidiagonal Factorization

$$T \longrightarrow LDL^T$$

- Relative condition number

$$\text{relcond}(\lambda) = \frac{v^T L|D|L^T v}{v^T LDL^T v}$$

- Seminal 1991 Demmel-Kahan paper (SIAG/LA Prize)
  - Small relative changes in the entries of a bidiagonal cause small relative changes in all its singular values

Key Idea 1 — Replace the tridiagonal with a bidiagonal

- Bidiagonal Factorization

$$T \longrightarrow LDL^T$$

- Relative condition number

$$\text{relcond}(\lambda) = \frac{v^T L|D|L^T v}{v^T LDL^T v}$$

- Seminal 1991 Demmel-Kahan paper (SIAG/LA Prize)
  - Small relative changes in the entries of a bidiagonal cause small relative changes in all its singular values

- L and D almost always define the small eigenvalues of LDL^T to high relative accuracy
  - Even in the face of element growth!

Key Idea 2 — Shift close to clusters

- Connection between residuals and orthogonality. Let

$$r = Ax - x\hat{\lambda} \neq 0, \quad s = Ay - y\hat{\mu} \neq 0,$$

Key Idea 2 — Shift close to clusters

- Connection between residuals and orthogonality. Let

$$r = Ax - x\hat{\lambda} \neq 0, \quad s = Ay - y\hat{\mu} \neq 0,$$

- Then

$$x^T y(\hat{\mu} - \hat{\lambda}) = y^T r - x^T s$$

Key Idea 2 — Shift close to clusters

- Connection between residuals and orthogonality. Let

$$r = Ax - x\hat{\lambda} \neq 0, \quad s = Ay - y\hat{\mu} \neq 0,$$

- Then

$$x^T y(\hat{\mu} - \hat{\lambda}) = y^T r - x^T s$$

- Suppose, by some miracle, $\|r\| \leq \varepsilon_1 |\hat{\lambda}|$ and $\|s\| \leq \varepsilon_2 |\hat{\mu}|$, then

Key Idea 2 — Shift close to clusters

- Connection between residuals and orthogonality. Let

$$r = Ax - x\hat{\lambda} \neq 0, \quad s = Ay - y\hat{\mu} \neq 0,$$

- Then

$$x^T y(\hat{\mu} - \hat{\lambda}) = y^T r - x^T s$$

- Suppose, by some miracle, $\|r\| \leq \varepsilon_1 |\hat{\lambda}|$ and $\|s\| \leq \varepsilon_2 |\hat{\mu}|$, then

$$|x^T y| \leq \max(\varepsilon_1, \varepsilon_2) \Big/ \frac{|\hat{\lambda} - \hat{\mu}|}{|\hat{\lambda}| + |\hat{\mu}|}$$

- Thus, orthogonality depends on the relative separation

Key Idea 2 — Shift close to clusters

- Connection between residuals and orthogonality. Let

$$r = Ax - x\hat{\lambda} \neq 0, \quad s = Ay - y\hat{\mu} \neq 0,$$

- Then

$$x^T y(\hat{\mu} - \hat{\lambda}) = y^T r - x^T s$$

- Suppose, by some miracle, $\|r\| \leq \varepsilon_1 |\hat{\lambda}|$ and $\|s\| \leq \varepsilon_2 |\hat{\mu}|$, then

$$|x^T y| \leq \max(\varepsilon_1, \varepsilon_2) \Big/ \frac{|\hat{\lambda} - \hat{\mu}|}{|\hat{\lambda}| + |\hat{\mu}|}$$

- Thus, orthogonality depends on the relative separation
- Relative Gaps can be made larger by shifting

$$LDL^T - \xi I = \tilde{L}\tilde{D}\tilde{L}^T$$

- Different representations for different clusters!
  - Essential to use differential transforms

Differential Transforms

$$LDL^T - \hat{\lambda} I \quad = \quad L_+ D_+ L_+^T$$

Simple qd :

$D_+(1) \;:=\; d_1 - \hat{\lambda}$

for $i = 1, n-1$

$L_+(i) \;:=\; (d_i l_i)/D_+(i)$

$D_+(i+1) \;:=\; d_i l_i^2 + d_{i+1} - L_+(i) d_i l_i - \hat{\lambda}$

end for

Differential qd :

$s_1 \;:=\; -\hat{\lambda}$

for $i = 1, n-1$

$D_+(i) \;:=\; s_i + d_i$

$L_+(i) \;:=\; (d_i l_i)/D_+(i)$

$s_{i+1} \;:=\; L_+(i) l_i s_i - \hat{\lambda}$

end for

$D_+(n) \;:=\; s_n + d_n$

Key Idea 3 — Twist, again with differential transformations

- Godunov et al. [1985], Fernando [1995]
- Compute the appropriate Twisted Factorization :

$$LDL^T - \hat{\lambda} I \;=\; N_r D_r N_r^T,$$

where $N_r \;=$

$$\begin{bmatrix}
\mathrm{x} & & & & & & & \\
\mathrm{x} & \mathrm{x} & & & & & & \\
 & \cdot & \cdot & & & & & \\
 & & \mathrm{x} & \textcircled{\mathrm{x}} & \mathrm{x} & & & \\
 & & & & \cdot & \cdot & & \\
 & & & & & \mathrm{x} & \mathrm{x} & \\
 & & & & & & \mathrm{x} & \mathrm{x} \\
 & & & & & & & \mathrm{x}
\end{bmatrix}$$

- Solve for z, $N_r D_r N_r^T z = \gamma_r e_r$ ($\Rightarrow N_r^T z = e_r$) :

$$z(i) \;=\; \begin{cases} 1, & i = r, \\ -L_+(i) \cdot z(i+1), & i = r-1, \ldots, 1, \\ -U_-(i-1) \cdot z(i-1), & i = r+1, \ldots, n. \end{cases}$$

Key Idea 4 — Representation Tree

- Eigenvalues: ε, $1+\sqrt{\varepsilon}$, $1+2\sqrt{\varepsilon}$, 2

Key Idea 4 — Representation Tree

- Eigenvalues: ε, $1+\sqrt{\varepsilon}$, $1+2\sqrt{\varepsilon}$, 2
- Extra representation needed at $\sigma = 1$:

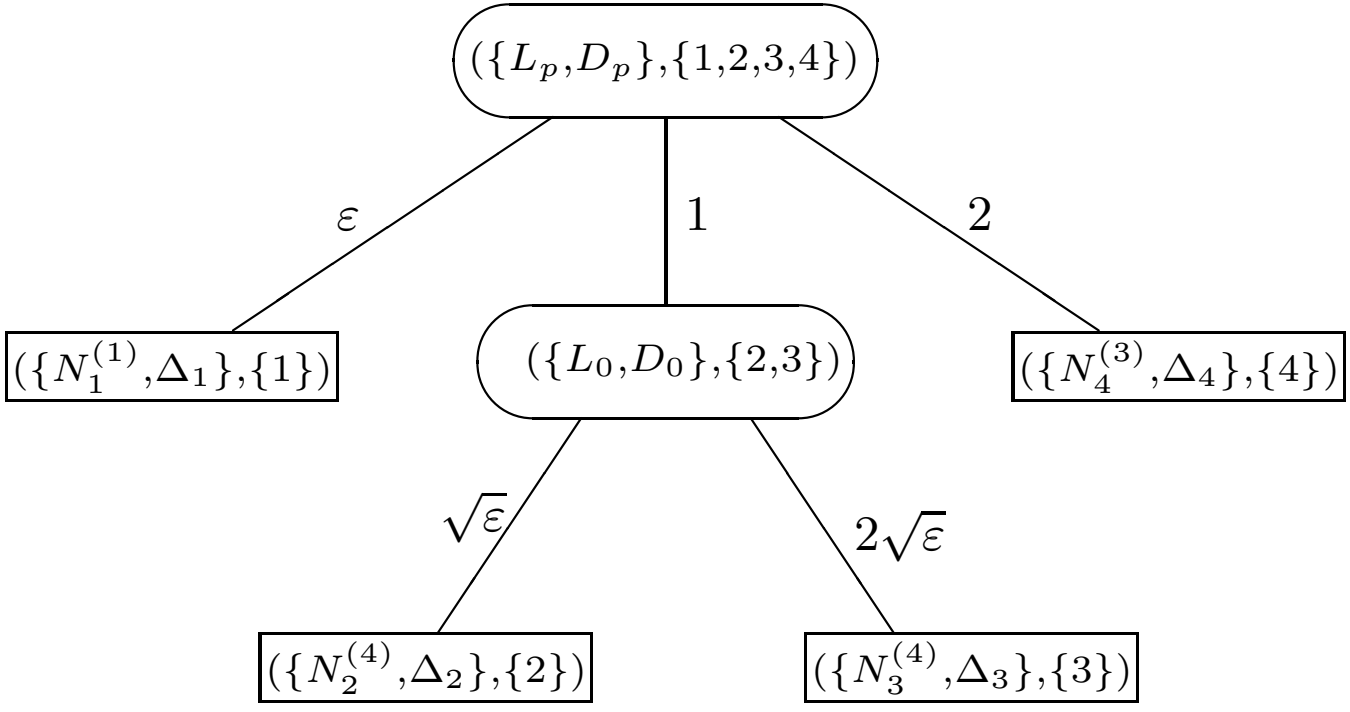
$$L_p D_p L_p^T - I \;=\; L_0 D_0 L_0^T$$

Key Idea 4 — Representation Tree

- Eigenvalues: ε, $1+\sqrt{\varepsilon}$, $1+2\sqrt{\varepsilon}$, 2
- Extra representation needed at $\sigma = 1$:

$$L_p D_p L_p^T - I \;=\; L_0 D_0 L_0^T$$

- Following Representation Tree captures the steps of the algorithm:



Caveats

- BLAS
  - Complexity is $O(nk)$ to compute k eigenpairs
  - However, all operations are BLAS 1
  - Closest competitor D&C — $O(n^3)$, but BLAS 3
- Very tight eigenvalue clusters
  - C. Vömel's torture tests

Wilkinson's Matrix W_{21}^+

- λ_{20} and λ_{21} are identical to working precision

Wilkinson's Matrix W_{21}^+

- λ_{20} and λ_{21} are identical to working precision
- Form new representation:

$$L_p D_p L_p^T - \hat{\lambda}_{21} I \;=\; L_0 D_0 L_0^T$$

Wilkinson's Matrix W_{21}^+

- λ_{20} and λ_{21} are identical to working precision
- Form new representation:

$$L_p D_p L_p^T - \hat{\lambda}_{21} I \;=\; L_0 D_0 L_0^T$$

- Roundoff to the rescue:

$$\begin{array}{ccc} \lambda_{20}(L_0 D_0 L_0^T) & \& & \lambda_{21}(L_0 D_0 L_0^T) \\ -7.28 \times 10^{-14} & \& & -1.22 \times 10^{-15} \end{array}$$

— no digits in common!

Wilkinson's Matrix W_{21}^+

- λ_{20} and λ_{21} are identical to working precision
- Form new representation:

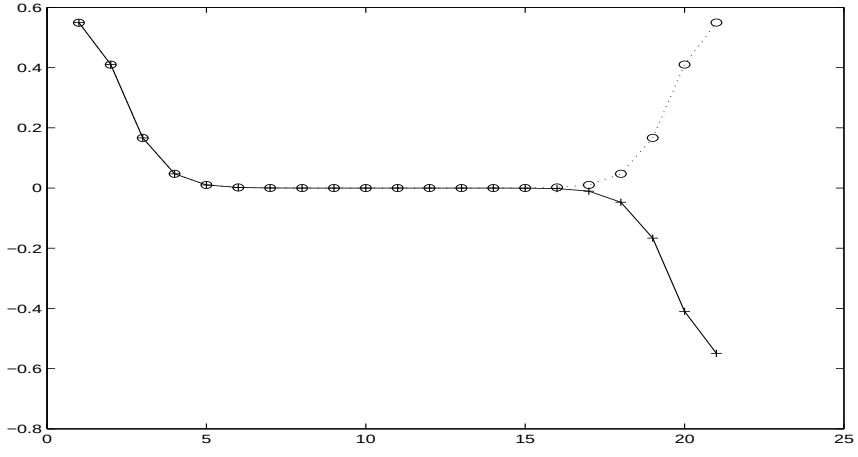
$$L_p D_p L_p^T - \hat{\lambda}_{21} I \;=\; L_0 D_0 L_0^T$$

- Roundoff to the rescue:

$$\begin{array}{ccc} \lambda_{20}(L_0 D_0 L_0^T) & \& & \lambda_{21}(L_0 D_0 L_0^T) \\ -7.28 \times 10^{-14} & \& & -1.22 \times 10^{-15} \end{array}$$

— no digits in common!

- Computed Eigenvectors $\hat{v}_{20}$ and $\hat{v}_{21}$ (inner product is 1.0×10^{-16}):



Caveats

- BLAS
  - Complexity is $O(nk)$ to compute k eigenpairs
  - However, all operations are BLAS 1
  - Closest competitor D&C — $O(n^3)$, but BLAS 3
- Very tight eigenvalue clusters
  - C. Vömel's torture tests — 5 copies of W_{101}^+ with glue of $\sqrt{\varepsilon}$
  - Required a tweak — Perturb base representation