An Explanation of the MRRR algorithm to compute eigenvectors of symmetric tridiagonal matrices

Beresford N. Parlett
Departments of Mathematics
and
Computer Science Division, EECS dept.
University of California, Berkeley

Inderjit S. Dhillon
Department of Computer Sciences
University of Texas, Austin

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 :

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

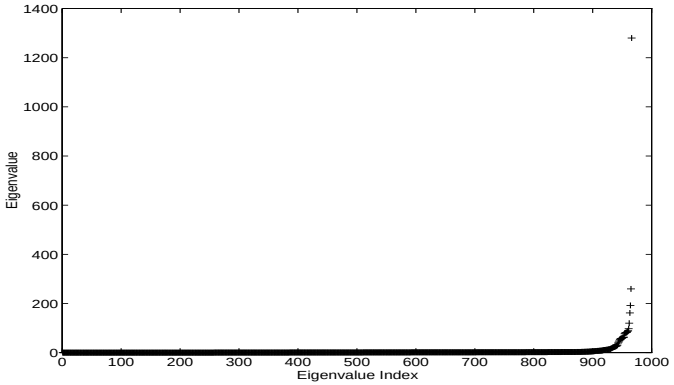
Assume all off-diagonals not negligible, so $\mathsf{Gap}(\hat{\lambda})$ not zero, but can be small:

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

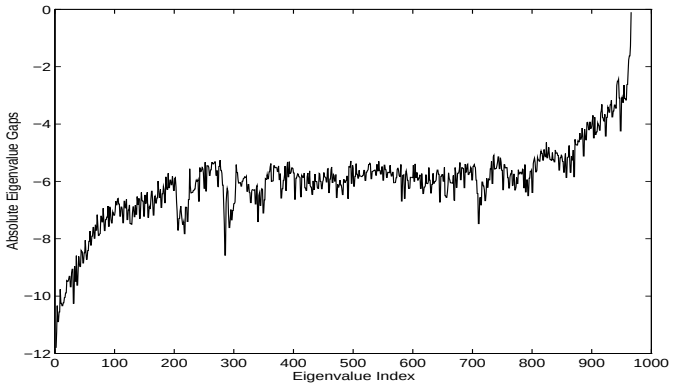
When eigenvalues are close, independently computed eigenvectors **WILL NOT** be mutually orthogonal.

Eigenvalues of Biphenyl Matrix

- Plot eigenvalues :



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



- LAPACK — one big "cluster" $\lambda_1, \lambda_2, \ldots, \lambda_{939}$.
- Tridiagonal solution takes 80% of Total Time.

Ideal Solution

- Fundamental Limitation:

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

- Get smallest possible residual norm.

  – Compute eigenvalue to greatest accuracy possible :

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

  – Compute eigenvector to high relative accuracy :

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

- Gap Theorem implies :

$$\sin \angle(v, \hat{v}) \;=\; \frac{O(\varepsilon|\lambda|)}{\mathsf{Gap}(\hat{\lambda})} \;=\; \frac{O(\varepsilon)}{\mathsf{Relgap}(\hat{\lambda})}.$$

- **Can we achieve the above ?**

Key Idea 1. Discard Tridiagonals, Embrace Bidiagonals

Factored Forms yield Better Representations

- Tridiagonals DO NOT determine their eigenvalues to high relative accuracy.
- Bidiagonals determine their singular values to high relative accuracy.

$$T + \mu I \quad = \quad L \quad L^T$$

$$\begin{bmatrix} x & x & & & \\ x & x & . & & \\ & . & . & . & \\ & & . & . & . \\ & & & . & x & x \\ & & & & x & x \end{bmatrix} = \begin{bmatrix} x & & & & \\ x & x & & & \\ & x & . & & \\ & & . & . & \\ & & & . & x \\ & & & & x & x \end{bmatrix} \begin{bmatrix} x & x & & & \\ & x & x & & \\ & & . & . & \\ & & & . & . \\ & & & & x & x \\ & & & & & x \end{bmatrix}$$

- Bidiagonal Factors are "better" since they allow us to
  - compute eigenvalues to high relative accuracy,
  - compute eigenvectors to high relative accuracy.
- High accuracy $\Rightarrow$ **Orthogonality**.
- For interior eigenvalues, extends to indefinite factorization LDL^T.

Algorithm Outline

1. Choose μ such that $T + \mu I$ is positive definite.

2. Compute the factorization :
$$T + \mu I \;=\; LDL^T.$$

3. Compute eigenvalues of LDL^T to high relative accuracy (by *dqds* or *bisection*).

4. Given eigenvalues, compute accurate eigenvectors of LDL^T.

   — **HOW?**

Key Idea 2. Shift with Differential QD

How do we get an eigenvector such that

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

Differential Transformations

- Inverse iteration — Solve for z :

$$LDL^T - \hat{\lambda} I \;=\; L_+ D_+ L_+^T.$$
$$L_+ D_+ L_+^T \, z \;=\; \text{random vector.}$$

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 (dqds) :

$$s_1 \;:=\; -\hat{\lambda}$$
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
$$D_+(n) \;:=\; s_n + d_n$$

Computing an Eigenvector

- Compute the appropriate Twisted Factorization :

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

where D_r is diagonal, $N_r \;=$

$$\begin{bmatrix}
\mathsf{x} & & & & & & \\
\mathsf{x} & \mathsf{x} & & & & & \\
 & \cdot & \cdot & & & & \\
 & & \mathsf{x} & \textcircled{\gamma_r} & \mathsf{x} & & \\
 & & & & \cdot & \cdot & \\
 & & & & & \mathsf{x} & \mathsf{x} \\
 & & & & & & \mathsf{x}
\end{bmatrix}$$

and r is chosen to minimize $|\gamma_r|$ (it will be $O(\lambda - \hat{\lambda})$).

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

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

- Solves an open problem posed by Wilkinson (1965).

Main Theorem

THEOREM. [Dhillon & Parlett, 2003] Eigenvectors computed by twisted factorization are numerically orthogonal if eigenvalues of LDL^T have large relative gaps. In particular,

$$(\hat{v}_i, \hat{v}_j) \;=\; \frac{O(\varepsilon)}{\mathsf{Relsep}(\lambda_i, \lambda_j)},$$

where

$$\mathsf{Relsep}(\lambda_i, \lambda_j) \;=\; \frac{|\lambda_i - \lambda_j|}{\max(|\lambda_i|, |\lambda_j|)}.$$

- Example of Large Relsep :

$$\lambda_1 = 10^{-16},\ \lambda_2 = 10^{-15} \quad \Rightarrow \quad \mathsf{Relsep}(\lambda_1, \lambda_2) \approx 1$$

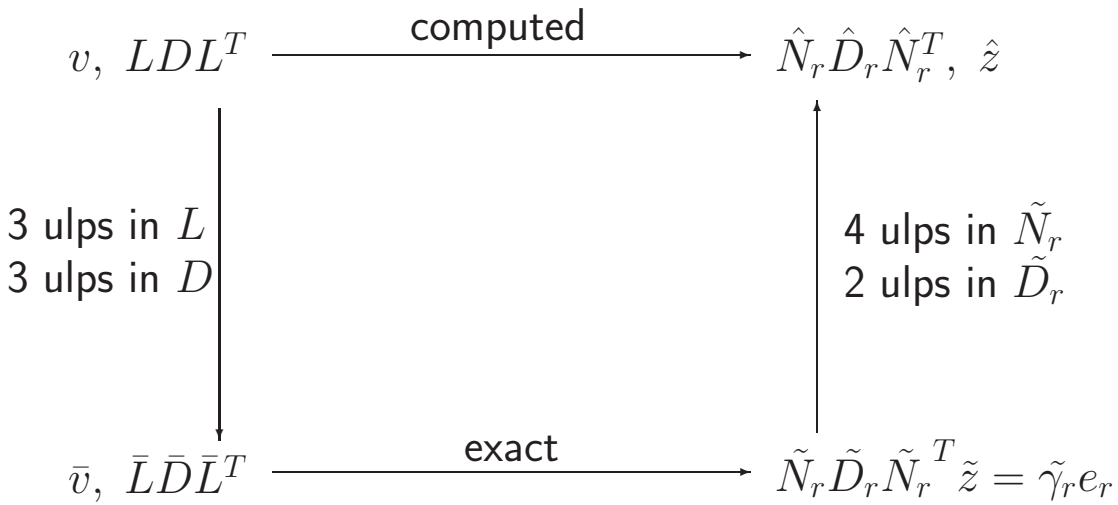
Above Theorem $\Rightarrow$ Automatic Orthogonality.

- Example of Small Relsep :

$$\begin{aligned} \lambda_1 &= 1.000000000000000\mathbf{1}, \\ \lambda_2 &= 1.000000000000000\mathbf{2}. \end{aligned}$$

Proof of Correctness

- Desired Relationship: $LDL^T - \hat{\lambda} I = N_r D_r N_r^T$, and $N_r D_r N_r^T z = \gamma_r e_r$.



- Exact Mathematical relationship holds : $\bar{L}\bar{D}\bar{L}^T - \hat{\lambda} I \ = \ \tilde{N}_r \tilde{D}_r \tilde{N}_r{}^T$.

- **Key step in proof** is to relate $\hat{z}$ to v in 3 steps :
  1. $\hat{z}$ is close to $\tilde{z}$, (only multiplications),
  2. $\sin \angle(\bar{v}, \tilde{z}) = O(\varepsilon|\bar{\lambda}|)/\mathsf{gap}(\hat{\lambda})$, $(|\tilde{\gamma}_r| = O(\varepsilon|\bar{\lambda}|))$,
  3. $\sin \angle(\bar{v}, v) = O(\varepsilon)/\mathsf{relgap}(\hat{\lambda})$ (relative perturbation theory).

$$\Rightarrow \quad \sin \angle(\hat{z}, v) \ = \ \frac{O(\varepsilon)}{\mathsf{Relgap}(\hat{\lambda})}.$$

Key Idea 3. Shift for Separation, again differentially

Algorithm MR^3 (Multiple RRRs)

1. Choose μ such that $T + \mu I$ is positive definite.

2. Compute the factorization :
$$T + \mu I \;=\; LDL^T.$$

3. Compute eigenvalues of LDL^T to high relative accuracy (by *dqds* or *bisection*).

4. Group eigenvalues according to their **Relative Gaps** :

   a) isolated (agree in $<$ 3 digits). Compute eigenvector using a twisted factorization.

   b) clustered (agree in $>$ 3 digits).

   - Pick μ near cluster to form $LDL^T - \mu I = L_1 D_1 L_1^T$ (by *dqds*).
   - "Refine" eigenvalues in cluster to high relative accuracy.
   - Set $L \leftarrow L_1$, $D \leftarrow D_1$. Repeat step 4 for eigenvalues in cluster.

Key Idea 4. Analyze the Representation Tree

Step 4 of the $\mathbf{MR}^3$ algorithm may be represented as a tree:

- At the root is the original factorization.
- At each internal node is a factorization for another shift μ.
- Each child of the node for μ corresponds to an isolated eigenvalue or a cluster.

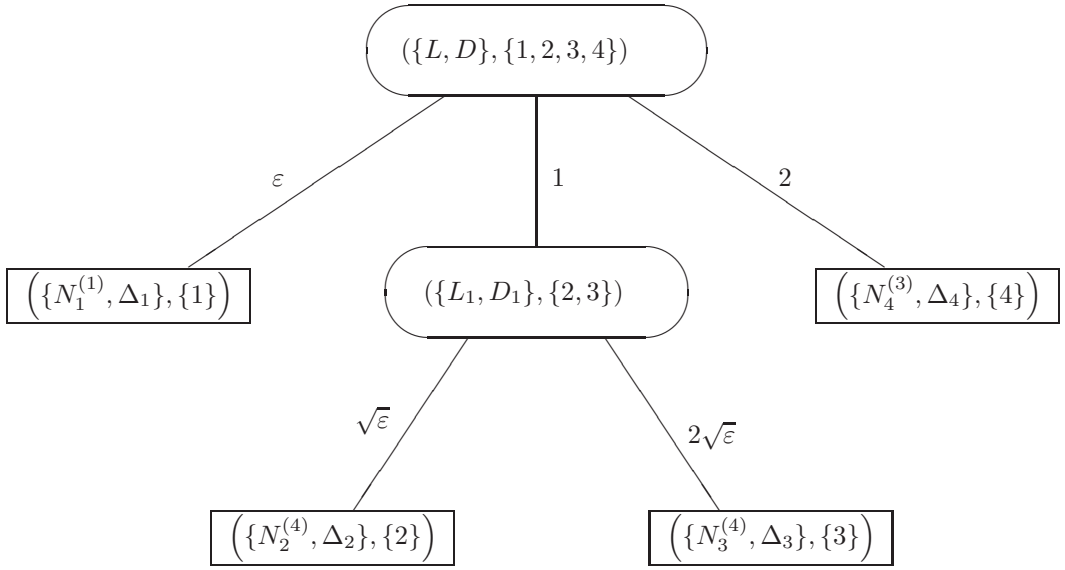
Small Example

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

- Extra representation needed at $\sigma = 1$:

$$LDL^T - I \;=\; L_1 D_1 L_1^T.$$

- The following **Representation Tree** captures the steps of the algorithm:



Wilkinson's Matrix

- W_{21}^{+} : 21×21 Wilkinson's matrix.
- λ_{20} and λ_{21} are identical to working precision.
- What happens in this case?

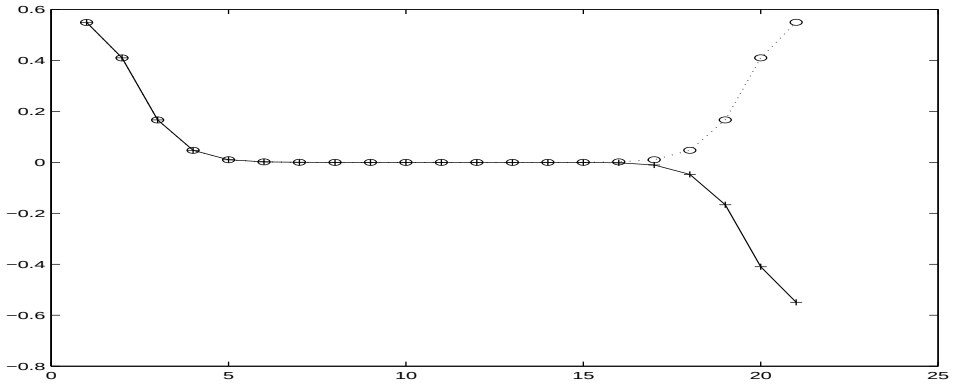
$$LDL^T - \hat{\lambda}_{21} I \;=\; L_1 D_1 L_1^T.$$

-

$$\begin{array}{rcl} \lambda_{20}(L_1 D_1 L_1^T) & \& & \lambda_{21}(L_1 D_1 L_1^T) \\ -7.28 \times 10^{-14} & \& & -1.22 \times 10^{-15} \\ (\hat{v}_{20}, \hat{v}_{21}) & = & 1.0 \times 10^{-16} \end{array}$$

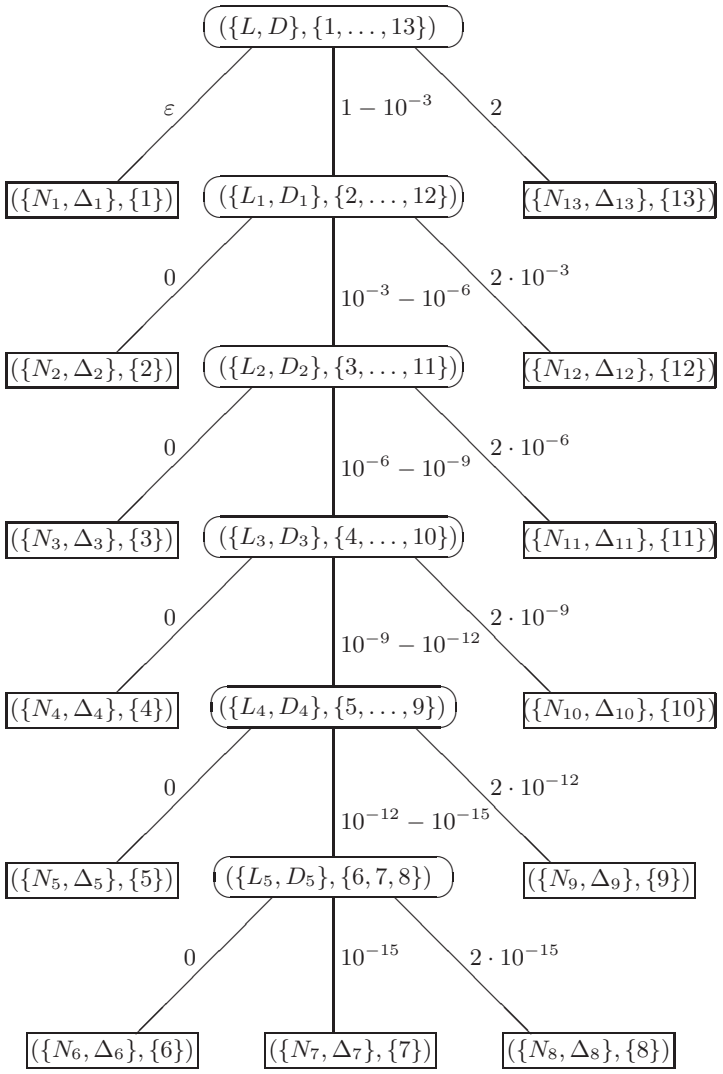
— no digits in common!

- Computed Eigenvectors $\hat{v}_{20}$ and $\hat{v}_{21}$:



Worst Case / Large Depth

13×13 matrix with eigenvalues: $0, 1, 1 \pm 10^{-15}, 1 \pm 10^{-12}, 1 \pm 10^{-9}, 1 \pm 10^{-6}, 1 \pm 10^{-3}, 2$.



Performance of MRRR on Biphenyl Matrix

For the biphenyl matrix ($n = 966$),

- the root node had 805 leaf children and 63 internal node children.
- all nodes at the next level were leaf nodes.
- 49 clusters had 2 eigenvalues,
- 13 clusters had 3-8 eigenvalues.
- one cluster had 9 eigenvalues,

Residual Norms for computed z - Part 1

- **Paper 1** guarantees small residuals at the bottom of the representation tree :
  a leaf and its parent

$$\|(\mathsf{leaf} - \delta\lambda I)z\| \;=\; |\gamma| \;=\; O(\varepsilon\delta\lambda).$$

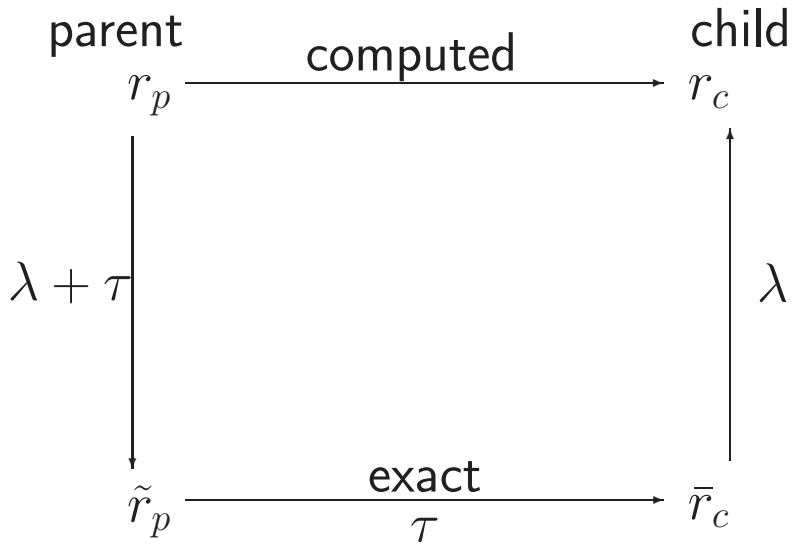
- Our problem :

$$\|(\mathsf{root} - \lambda I)z\| \;=\; |\gamma| \;=\; O(\varepsilon\,\mathsf{spdiam}(\mathsf{root})) \qquad ???$$

- In exact arithmetic, root residual is also $O(\varepsilon\delta\lambda)$.

Residual Norms for computed z - Part 2

Compare child residual with parent residual at each interval node from leaf to root.



$$r_c = (T_c - \lambda I)z$$

$$T_c = L_c D_c L_c^T$$

- By design

$$\tilde{r}_p \;=\; \bar{r}_c \qquad \text{exact}$$

- also

$$\tilde{r}_p \;=\; r_p + \delta T_p z \;\;, \qquad \tilde{T}_p \;=\; T_p + \delta T_p$$
$$\bar{r}_c \;=\; r_c + \delta T_c z \;\;, \qquad \bar{T}_c \;=\; T_c + \delta T_c$$

$$T + \delta T \;=\; (L + \delta L)(D + \delta D)(L + \delta L)^T \;\;, \qquad T \;=\; LDL^T$$

Technical Lemma

If

$$\begin{aligned} \|Dz\| &\leq c\,\mathsf{spdiam}(T_0) \\ \|\mathring{L}D\mathring{L}^T z\| &\leq c\,\mathsf{spdiam}(T_0), \qquad \mathring{L} = L - I, \end{aligned}$$

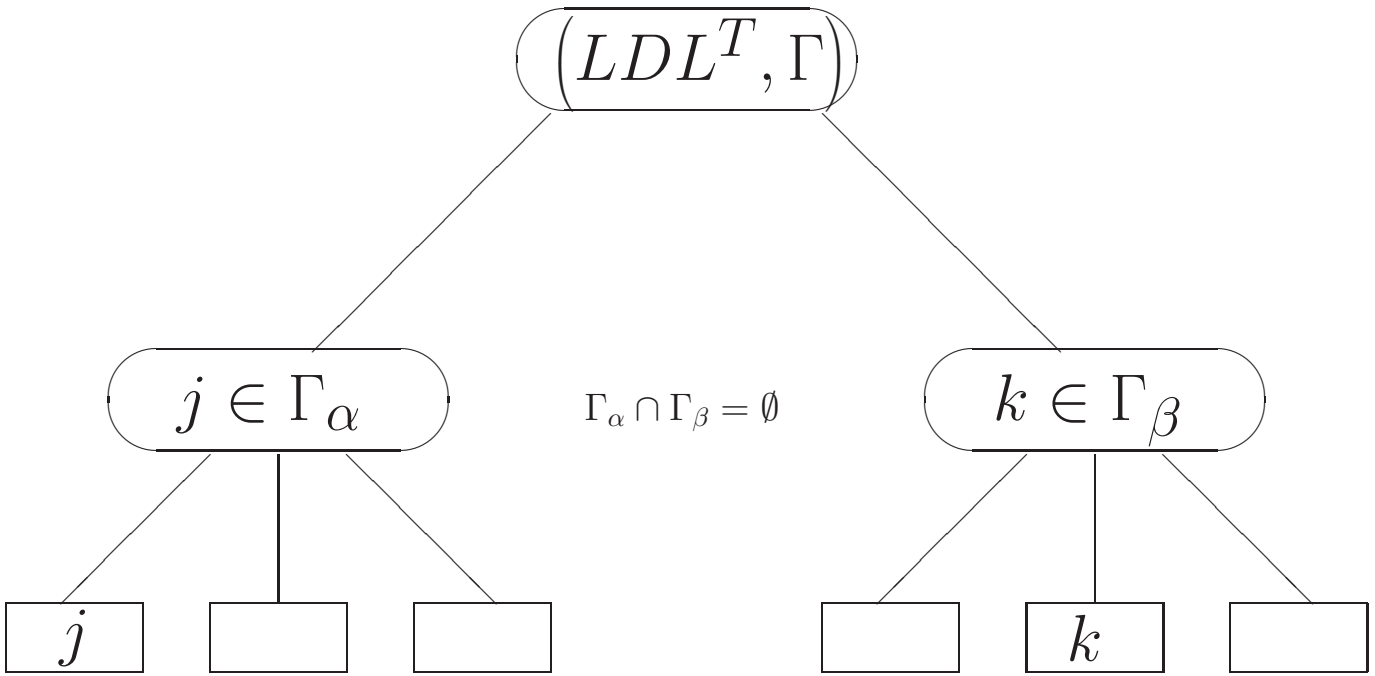
then

$$(*) \quad \|\delta T z\| \leq (2c + \frac{1}{2})(9\varepsilon)\,\mathsf{spdiam}(T_0) + O(n\varepsilon^2)$$

NOTE: large values in D and LDL^T can be neutralized by small entries in z.

(*) gives a bound on increase in residual norm at each internal node on path from **leaf** to **root**.

Orthogonality



- By **Paper 1**, eigenvectors with same parent are orthogonal to working accuracy.
- Our problem : $|z_j^T z_k| = O(n\varepsilon)$????
- $\mathcal{S}_\Gamma = \mathcal{S}_\Gamma^{LDL^T}$ = subspace invariant under LDL^T for eigenvalues in Γ.

Two Angles

$$\Psi_{k,\Gamma} := \angle(z_k, \mathcal{S}_\Gamma)$$
$$\Phi_{\Gamma_\alpha} := \angle(\mathcal{S}_{\Gamma_\alpha}^{\text{parent}}, \mathcal{S}_{\Gamma_\alpha}^{\text{child}}), \qquad \mathcal{S}_{\Gamma_\alpha}^{\text{parent}} \subset \mathcal{S}_\Gamma^{\text{parent}}$$

Lemma 1. $\sin \Psi_{j,\Gamma} \leq \sin \Psi_{j,\Gamma_\alpha} + \sin \Phi_{\Gamma_\alpha}$

Lemma 2. $\sin \Phi_{\Gamma\alpha} \leq Rn\varepsilon,$

R depends on tolerance for relgap.

Theorem

Let (LDL^T, Γ) be the least common ancestor of z_j and $z_k, j \neq k$. If all internal nodes on the paths from leaves $< j >$ and $< k >$ to Γ, (in the representation tree) are RRRs, then

$$\cos \angle(z_j, z_k) \leq 2 \text{ leafbound} + \{\text{depth}(\Gamma, j) + \text{depth}(\Gamma, k) - 2\} Rn\varepsilon.$$