On the HZ Method for PGEP

Vjeran Hari and Josip Matejaš

Faculty of Science, Department of Mathematics, University of Zagreb
hari@math.hr

Faculty of Economics, University of Zagreb
jmatejas@efzg.hr

6'th Croatian Mathematical Congress
Zagreb, Croatia

OUTLINE

OUTLINE

- GEP and PGEP
- Derivation of the algorithm
- Convergence, global and asymptotic
- Stability and relative accuracy

This work has been fully supported by Croatian Science Foundation under the project IP-09-2014-3670.

GEP and PGEP

Let $A = A^T, \quad B = B^T.$

GEP and PGEP

Let $A = A^T, \quad B = B^T$.
We consider the Generalized Eigenvalue Problem (GEP)

$$Ax = \lambda Bx, \quad x \neq 0.$$

If $B \succ O$, GEP is usually called Positive definite GEP or shorter PGEP.

GEP and PGEP

Let $A = A^T, \quad B = B^T$.
We consider the Generalized Eigenvalue Problem (GEP)

$$Ax = \lambda Bx, \quad x \neq 0.$$

If $B \succ O$, GEP is usually called Positive definite GEP or shorter PGEP.

For such a pair there is a nonsingular matrix F such that

$$F^T AF = \Lambda_A\,, \qquad F^T BF = \Lambda_B,$$

$\Lambda_A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n), \quad \Lambda_B = \mathrm{diag}(\beta_1, \ldots, \beta_n) \succ O.$

GEP and PGEP

Let $A = A^T, \quad B = B^T$.
We consider the Generalized Eigenvalue Problem (GEP)

$$Ax = \lambda Bx, \quad x \neq 0.$$

If $B \succ O$, GEP is usually called Positive definite GEP or shorter PGEP.

For such a pair there is a nonsingular matrix F such that

$$F^T AF = \Lambda_A\,, \qquad F^T BF = \Lambda_B,$$

$\Lambda_A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n), \quad \Lambda_B = \mathrm{diag}(\beta_1, \ldots, \beta_n) \succ O.$

The eigenpairs of (A, B) are: $(\alpha_i/\beta_i, Fe_i)$, $1 \leq i \leq n$; $I_n = [e_1, \ldots, e_n]$.

How to solve PGEP?

One can try with the transformation $(A, B) \mapsto (L^{-1}AL^{-T}, I)$, $B = LL^T$ and reduce PGEP to EP for one symmetric matrix.

If L has small singular value(s), then computed $L^{-1}AL^{-T}$ will have corrupt smallest eigenvalues.

How to solve PGEP?

One can try with the transformation $(A, B) \mapsto (L^{-1}AL^{-T}, I)$, $B = LL^T$ and reduce PGEP to EP for one symmetric matrix.

If L has small singular value(s), then computed $L^{-1}AL^{-T}$ will have corrupt smallest eigenvalues. Then one can try to maximize the minimum eigenvalue of B by rotating the pair

$$(A, B) \mapsto (A_\varphi, B_\varphi) = (A\cos\varphi + B\sin\varphi, -A\sin\varphi + B\cos\varphi),$$

or derive a method which works with the initial pair (A, B).

How to solve PGEP?

One can try with the transformation $(A, B) \mapsto (L^{-1}AL^{-T}, I)$, $B = LL^T$ and reduce PGEP to EP for one symmetric matrix.

If L has small singular value(s), then computed $L^{-1}AL^{-T}$ will have corrupt smallest eigenvalues. Then one can try to maximize the minimum eigenvalue of B by rotating the pair

$$(A, B) \mapsto (A_\varphi, B_\varphi) = (A\cos\varphi + B\sin\varphi, -A\sin\varphi + B\cos\varphi),$$

or derive a method which works with the initial pair (A, B).

We follow the second path.

Jacobi methods for PGEP

We have two diagonalization methods for PGEP

- Falk-Langemeyer method (shorter: FL method)
  (Elektronische Datenverarbeitung, 1960)
- Hari-Zimmermann variant of the FL method (shorter: HZ method)
  (Hari Ph.D. 1984)

The two methods are connected: the FL method can be viewed as the HZ method with "fast scaled" transformations. So, the FL method seems to be somewhat faster and the HZ method seems to be more robust.

Jacobi methods for PGEP

We have two diagonalization methods for PGEP

- Falk-Langemeyer method (shorter: FL method) (Elektronische Datenverarbeitung, 1960)
- Hari-Zimmermann variant of the FL method (shorter: HZ method) (Hari Ph.D. 1984)

The two methods are connected: the FL method can be viewed as the HZ method with "fast scaled" transformations. So, the FL method seems to be somewhat faster and the HZ method seems to be more robust.

V. Novaković, S. Singer, S. Singer (Parallel Comput., 2015):

Jacobi methods for PGEP

We have two diagonalization methods for PGEP

- Falk-Langemeyer method (shorter: FL method)
  (Elektronische Datenverarbeitung, 1960)
- Hari-Zimmermann variant of the FL method (shorter: HZ method)
  (Hari Ph.D. 1984)

The two methods are connected: the FL method can be viewed as the HZ method with “fast scaled” transformations. So, the FL method seems to be somewhat faster and the HZ method seems to be more robust.

V. Novaković, S. Singer, S. Singer (Parallel Comput., 2015):

Numerical tests on large matrices, on parallel machines, have confirmed the advantage of the HZ approach. When implemented as one-sided block algorithm for the GSVD, it is almost perfectly parallelizable, so parallel shared memory versions of the algorithm are highly scalable, and their speedup almost solely depends on the number of cores used.

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0,\ B^{(0)} = D_0 B D_0$

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0,\ B^{(0)} = D_0 B D_0$

$D_0 = [\mathrm{diag}(B)]^{-\frac{1}{2}},$ so that $b_{11}^{(0)} = b_{22}^{(0)} = \cdots = b_{nn}^{(0)} = 1.$

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0, \; B^{(0)} = D_0 B D_0$

$D_0 = [\mathrm{diag}(B)]^{-\frac{1}{2}},$ so that $b_{11}^{(0)} = b_{22}^{(0)} = \cdots = b_{nn}^{(0)} = 1.$

This property of $B^{(0)}$ is maintained during the iteration process:

$$A^{(k+1)} = Z_k^T A^{(k)} Z_k, \qquad B^{(k+1)} = Z_k^T B^{(k)} Z_k, \quad k \geq 0.$$

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0,\ B^{(0)} = D_0 B D_0$

$D_0 = [\mathrm{diag}(B)]^{-\frac{1}{2}}$, so that $b_{11}^{(0)} = b_{22}^{(0)} = \cdots = b_{nn}^{(0)} = 1.$

This property of $B^{(0)}$ is maintained during the iteration process:

$$A^{(k+1)} = Z_k^T A^{(k)} Z_k, \qquad B^{(k+1)} = Z_k^T B^{(k)} Z_k, \quad k \geq 0.$$

Each Z_k is a nonsingular elementary plane matrix

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0, \; B^{(0)} = D_0 B D_0$

$D_0 = [\mathrm{diag}(B)]^{-\frac{1}{2}}$, so that $b_{11}^{(0)} = b_{22}^{(0)} = \cdots = b_{nn}^{(0)} = 1.$

This property of $B^{(0)}$ is maintained during the iteration process:

$$A^{(k+1)} = Z_k^T A^{(k)} Z_k, \qquad B^{(k+1)} = Z_k^T B^{(k)} Z_k, \quad k \geq 0.$$

Each Z_k is a nonsingular elementary plane matrix

$$Z_k = \begin{bmatrix} I & & & & \\ & c_k & & -s_k & \\ & & I & & \\ & \tilde{s}_k & & \tilde{c}_k & \\ & & & & I \end{bmatrix} \begin{matrix} \\ i(k) \\ \\ j(k) \\ \\ \end{matrix}, \qquad i(k) < j(k) \text{ are pivot indices at step } k,$$

$$c_k^2 + s_k^2 = \tilde{c}_k^2 + \tilde{s}_k^2 = 1/\sqrt{1 - b_{i(k)j(k)}^2} \quad \text{(Gose 1979).}$$

Derivation of the HZ Method

Preliminary transformation: $A^{(0)} = D_0 A D_0,\ B^{(0)} = D_0 B D_0$

$D_0 = [\mathrm{diag}(B)]^{-\frac{1}{2}}$, so that $b_{11}^{(0)} = b_{22}^{(0)} = \cdots = b_{nn}^{(0)} = 1.$

This property of $B^{(0)}$ is maintained during the iteration process:

$$A^{(k+1)} = Z_k^T A^{(k)} Z_k, \qquad B^{(k+1)} = Z_k^T B^{(k)} Z_k, \quad k \geq 0.$$

Each Z_k is a nonsingular elementary plane matrix

$$Z_k = \begin{bmatrix} I & & & & \\ & c_k & & -s_k & \\ & & I & & \\ & \tilde{s}_k & & \tilde{c}_k & \\ & & & & I \end{bmatrix} \begin{matrix} \\ i(k) \\ \\ j(k) \\ \\ \end{matrix}, \qquad i(k) < j(k) \text{ are pivot indices at step } k,$$

$$c_k^2 + s_k^2 = \tilde{c}_k^2 + \tilde{s}_k^2 = 1/\sqrt{1 - b_{i(k)j(k)}^2} \quad \text{(Gose 1979).}$$

The selection of pivot pairs $(i(k), j(k))$ defines pivot strategy.

Derivation of the HZ Method

To describe step k, we assume: $A = A^{(k)}$, $A' = A^{(k+1)}$, $Z_k = Z$,

$$\hat{Z} = \begin{bmatrix} c_k & -s_k \\ \tilde{s}_k & \tilde{c}_k \end{bmatrix}$$ the pivot submatrix of Z.

Derivation of the HZ Method

To describe step k, we assume: $A = A^{(k)}$, $A' = A^{(k+1)}$, $Z_k = Z$,

$$\hat{Z} = \begin{bmatrix} c_k & -s_k \\ \tilde{s}_k & \tilde{c}_k \end{bmatrix} \qquad \text{the pivot submatrix of } Z.$$

We have

$$A' = Z^T A Z, \quad B' = Z^T B Z \qquad \left(\hat{A}' = \hat{Z}^T \hat{A} \hat{Z}, \quad \hat{B}' = \hat{Z}^T \hat{B} \hat{Z} \right).$$

Derivation of the HZ Method

To describe step k, we assume: $A = A^{(k)}$, $A' = A^{(k+1)}$, $Z_k = Z$,

$$\hat{Z} = \begin{bmatrix} c_k & -s_k \\ \tilde{s}_k & \tilde{c}_k \end{bmatrix} \qquad \text{the pivot submatrix of } Z.$$

We have

$$A' = Z^T A Z, \quad B' = Z^T B Z \qquad \left(\hat{A}' = \hat{Z}^T \hat{A} \hat{Z}, \quad \hat{B}' = \hat{Z}^T \hat{B} \hat{Z} \right).$$

Z is chosen to annihilate the pivot elements $a_{ij}^{(k)}$ and $b_{ij}^{(k)}$.

Derivation of the HZ Method

To describe step k, we assume: $A = A^{(k)}$, $A' = A^{(k+1)}$, $Z_k = Z$,

$$\hat{Z} = \begin{bmatrix} c_k & -s_k \\ \tilde{s}_k & \tilde{c}_k \end{bmatrix} \qquad \text{the pivot submatrix of } Z.$$

We have

$$A' = Z^T A Z, \quad B' = Z^T B Z \qquad \left(\hat{A}' = \hat{Z}^T \hat{A} \hat{Z}, \quad \hat{B}' = \hat{Z}^T \hat{B} \hat{Z} \right).$$

Z is chosen to annihilate the pivot elements $a_{ij}^{(k)}$ and $b_{ij}^{(k)}$.

$\hat{Z}$ is sought in the form of a product of two Jacobi rotations and one diagonal matrix. We have two possibilities:

$\hat{Z}$ is sought in the form:

$$\text{(a)}\quad \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1+b_{ij}}} & 0 \\ 0 & \frac{1}{\sqrt{1-b_{ij}}} \end{bmatrix} \begin{bmatrix} \cos(\theta-\frac{\pi}{4}) & -\sin(\theta-\frac{\pi}{4}) \\ \sin(\theta-\frac{\pi}{4}) & \cos(\theta-\frac{\pi}{4}) \end{bmatrix}$$

$$\text{(b)}\quad \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1-b_{ij}}} & 0 \\ 0 & \frac{1}{\sqrt{1+b_{ij}}} \end{bmatrix} \begin{bmatrix} \cos(\theta+\frac{\pi}{4}) & -\sin(\theta+\frac{\pi}{4}) \\ \sin(\theta+\frac{\pi}{4}) & \cos(\theta+\frac{\pi}{4}) \end{bmatrix}$$

$\downarrow$ $\hat{B} \longrightarrow \text{diag}$ | $\downarrow$ $\hat{B} \longrightarrow I_2$ | $\downarrow$ $\hat{A} \longrightarrow \text{diag}$

Essential Part of the Algorithm

$$\xi = \frac{b_{ij}}{\sqrt{1+b_{ij}} + \sqrt{1-b_{ij}}}, \quad \rho = \xi + \sqrt{1-b_{ij}}, \quad \xi^2 + \rho^2 = 1,$$

Essential Part of the Algorithm

$$\xi = \frac{b_{ij}}{\sqrt{1+b_{ij}}+\sqrt{1-b_{ij}}}, \quad \rho = \xi + \sqrt{1-b_{ij}}, \quad \xi^2+\rho^2 = 1,$$

$$\tan(2\theta) = \frac{2a_{ij} - (a_{ii}+a_{jj})\, b_{ij}}{\sqrt{1-(b_{ij})^2}\,(a_{ii}-a_{jj})}, \qquad -\frac{\pi}{4} \le \theta \le \frac{\pi}{4},$$

Essential Part of the Algorithm

$$\xi = \frac{b_{ij}}{\sqrt{1+b_{ij}}+\sqrt{1-b_{ij}}}, \quad \rho = \xi + \sqrt{1-b_{ij}}, \quad \xi^2+\rho^2 = 1,$$

$$\tan(2\theta) = \frac{2a_{ij} - (a_{ii}+a_{jj})\, b_{ij}}{\sqrt{1-(b_{ij})^2}\,(a_{ii}-a_{jj})}, \qquad -\frac{\pi}{4} \le \theta \le \frac{\pi}{4},$$

$$\begin{aligned}
\cos\phi &= \rho\cos\theta - \xi\sin\theta,\\
\sin\phi &= \rho\sin\theta + \xi\cos\theta,\\
\cos\psi &= \rho\cos\theta + \xi\sin\theta,\\
\sin\psi &= \rho\sin\theta - \xi\cos\theta.
\end{aligned}$$

Essential Part of the Algorithm

$$\xi = \frac{b_{ij}}{\sqrt{1+b_{ij}} + \sqrt{1-b_{ij}}}, \quad \rho = \xi + \sqrt{1-b_{ij}}, \quad \xi^2 + \rho^2 = 1,$$

$$\tan(2\theta) = \frac{2a_{ij} - (a_{ii} + a_{jj})\, b_{ij}}{\sqrt{1-(b_{ij})^2}\,(a_{ii} - a_{jj})}, \qquad -\frac{\pi}{4} \le \theta \le \frac{\pi}{4},$$

$$\begin{aligned}
\cos\phi &= \rho\cos\theta - \xi\sin\theta, \\
\sin\phi &= \rho\sin\theta + \xi\cos\theta, \\
\cos\psi &= \rho\cos\theta + \xi\sin\theta, \\
\sin\psi &= \rho\sin\theta - \xi\cos\theta.
\end{aligned}$$

Then

$$\hat{Z} = \frac{1}{\sqrt{1-b_{ij}^2}} \begin{bmatrix} \cos\phi & \sin\phi \\ \cos\psi & \sin\psi \end{bmatrix}.$$

Global Convergence

We have used the following measure in the convergence analysis:

$$S(A,B) = \left[S^2(A) + S^2(B)\right]^{\frac{1}{2}}, \quad S^2(A) = \|A - \mathrm{diag}(A)\|_F^2.$$

The HZ method converges globally if

$$A^{(k)} \to \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad B^{(k)} \to I_n \qquad \text{as} \quad k \to \infty,$$

holds for any initial pair of symmetric matrices (A, B) with $B \succ O$.

Global Convergence

We have used the following measure in the convergence analysis:

$$S(A,B) = \left[S^2(A) + S^2(B)\right]^{\frac{1}{2}}, \quad S^2(A) = \|A - \mathrm{diag}(A)\|_F^2.$$

The HZ method converges globally if

$$A^{(k)} \to \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad B^{(k)} \to I_n \qquad \text{as} \quad k \to \infty,$$

holds for any initial pair of symmetric matrices (A, B) with $B \succ O$.

Actually, it is sufficient to show that $S(A, B) \to 0$ as $k \to \infty$.

Global Convergence

We have used the following measure in the convergence analysis:

$$S(A,B) = \left[S^2(A) + S^2(B)\right]^{\frac{1}{2}}, \quad S^2(A) = \|A - \mathrm{diag}(A)\|_F^2.$$

The HZ method converges globally if

$$A^{(k)} \to \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad B^{(k)} \to I_n \qquad \text{as} \quad k \to \infty,$$

holds for any initial pair of symmetric matrices (A, B) with $B \succ O$.

Actually, it is sufficient to show that $S(A,B) \to 0$ as $k \to \infty$.

We have proved the global convergence for the serial pivot strategies.

Global Convergence

We have used the following measure in the convergence analysis:

$$S(A,B) = \left[S^2(A) + S^2(B)\right]^{\frac{1}{2}}, \quad S^2(A) = \|A - \mathrm{diag}(A)\|_F^2.$$

The HZ method converges globally if

$$A^{(k)} \to \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \quad B^{(k)} \to I_n \qquad \text{as} \quad k \to \infty,$$

holds for any initial pair of symmetric matrices (A, B) with $B \succ O$.

Actually, it is sufficient to show that $S(A, B) \to 0$ as $k \to \infty$.

We have proved the global convergence for the serial pivot strategies.

We are adapting the proof to hold for a new larger class of generalized serial strategies which includes the known weak wavefront strategies.

Asymptotic Convergence

Let (A, B) have simple eigenvalues:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n, \qquad \mu = \max\{|\lambda_1|, |\lambda_n|\},$$

$$3\delta_i = \min_{\substack{1\le i\le n \\ j\ne i}} |\lambda_i - \lambda_j|, \quad 1 \le i \le n; \qquad \delta = \min_{1\le i\le n} \delta_i .$$

Asymptotic Convergence

Let (A, B) have simple eigenvalues:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n, \qquad \mu = \max\{|\lambda_1|\,,\,|\lambda_n|\},$$

$$3\delta_i = \min_{\substack{1\le i\le n\\ j\ne i}} |\,\lambda_i - \lambda_j\,|, \quad 1 \le i \le n; \qquad \delta = \min_{1\le i\le n} \delta_i\,.$$

Theorem

If $S(B^{(0)}) < \dfrac{1}{n(n-1)}$ *and* $S(A^{(0)}, B^{(0)}) < \dfrac{\delta}{2\sqrt{1+\mu^2}}$,
then for the general cyclic and for the serial strategies it holds, respectively:

$$S(A^{(N)}, B^{(N)}) \le \sqrt{N(1+\mu^2)}\frac{S^2(A^{(0)}, B^{(0)})}{\delta}, \quad N = n(n-1)/2$$

$$S(A^{(N)}, B^{(N)}) \le \sqrt{1+\mu^2}\frac{S^2(A^{(0)}, B^{(0)})}{\delta}.$$

Asymptotic Convergence

Let (A, B) have simple eigenvalues:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n, \qquad \mu = \max\{|\lambda_1|, |\lambda_n|\},$$

$$3\delta_i = \min_{\substack{1\le i\le n \\ j\ne i}} |\lambda_i - \lambda_j|, \quad 1 \le i \le n; \qquad \delta = \min_{1\le i\le n} \delta_i .$$

Theorem

If $S(B^{(0)}) < \dfrac{1}{n(n-1)}$ *and* $S(A^{(0)}, B^{(0)}) < \dfrac{\delta}{2\sqrt{1+\mu^2}}$,
then for the general cyclic and for the serial strategies it holds, respectively:

$$S(A^{(N)}, B^{(N)}) \le \sqrt{N(1+\mu^2)}\frac{S^2(A^{(0)}, B^{(0)})}{\delta}, \quad N = n(n-1)/2$$

$$S(A^{(N)}, B^{(N)}) \le \sqrt{1+\mu^2}\frac{S^2(A^{(0)}, B^{(0)})}{\delta}.$$

In the case of multiple eigenvalues, the method is not quadratically convergent, but can be modified to be such.