Weighted Golub-Kahan-Lanczos Algorithms

Hongguo Xu
University of Kansas

Joint work with Hongxiu Zhong
(Jiangnan Univerisity)

NL2A, Luminy, Oct. 22-28, 2016

Outline

- Matrix bidiagonal factorizations
- Weighted Golub-Kahan-Lanczos (GKL) algorithm
- Application to eigenvalue problems
- Connection to CG

Weighted bidiagonal factorization

Given K, M, both SPD, there exist X, Y s.t.,

$$KY = XB, \qquad MX = YB^T,$$

$$X^T MX = I, \qquad Y^T KY = I,$$

and B bidiagonal.

Golub-Kahan-Lanczos (GKL) algorithm [Paige'74, Paige/Saunders'82] is based on

$$AV = UB, \qquad A^T U = VB^T$$

$$U^T U = I, \qquad V^T V = I.$$

Golub-Kahan-Lanczos (GKL) algorithm [Paige'74, Paige/Saunders'82] is based on

$$AV = UB, \qquad A^T U = VB^T$$

$$U^T U = I, \qquad V^T V = I.$$

Let $K = LL^T$, $M = RR^T$. $U = R^T X$, $V = L^T Y$.

$$\begin{array}{cccccccccc} A & V & = & U & B, & A^T & U & = & V & B^T \\ \downarrow & \downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow & \\ R^T L & L^T Y & = & R^T X & B, & L^T R & R^T X & = & L^T Y & B^T \end{array}$$

$$\Downarrow$$

$$KY = XB, \quad MX = YB^T, \quad X^T MX = I, \quad Y^T KY = I.$$

Generalized GKL algorithm [Arioli'13] is based on

$$AY = MXB, \qquad A^T X = KYB^T$$

$$X^T MX = I, \qquad Y^T KY = I.$$

Generalized GKL algorithm [Arioli'13] is based on

$$AY = MXB, \qquad A^T X = KYB^T$$

$$X^T MX = I, \qquad Y^T KY = I.$$

Let $A = MK$.

$$MKY = MXB, \qquad KMX = KYB^T$$

$$\Downarrow$$

$$KY = XB, \quad MX = YB^T, \quad X^T MX = I, \quad Y^T KY = I.$$

A weighted GKL algorithm

$$KY = XB, \quad MX = YB^T, \quad X^T MX = Y^T KY = I,$$

let

$$\begin{array}{l} X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \\ Y = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \end{array}, \quad B = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ & \alpha_2 & \ddots & \\ & & \ddots & \beta_{n-1} \\ & & & \alpha_n \end{bmatrix}.$$

$$Ky_1 = \alpha_1 x_1 \qquad Mx_1 = \alpha_1 y_1 + \beta_1 y_2$$

$$Ky_2 = \beta_1 x_1 + \alpha_2 x_2 \qquad Mx_2 = \alpha_2 y_2 + \beta_2 y_3$$

$$\vdots$$

$$Ky_j = \beta_{j-1} x_{j-1} + \alpha_j x_j \qquad Mx_j = \alpha_j y_j + \beta_j y_{j+1}$$

$$\alpha_j = \|Ky_j - \beta_{j-1}x_{j-1}\|_M, \quad x_j = (Ky_j - \beta_{j-1}x_{j-1})/\alpha_j$$
$$\beta_j = \|Mx_j - \alpha_j y_j\|_K, \qquad y_{j+1} = (Mx_j - \alpha_j y_j)/\beta_j.$$

$\|x\|_M = \sqrt{x^T M x}, \quad \|y\|_K = \sqrt{y^T K y}.$

WGKL Algorithm Choose y_1 ($\|y_1\|_K = 1$). Set $\beta_0 = 1$, $x_0 = 0$, $g_0 = Ky_1$

For $j = 1, 2 \dots$

$$s_j = g_{j-1}/\beta_{j-1} - \beta_{j-1}x_{j-1}$$
$$f_j = Ms_j$$
$$\alpha_j = \sqrt{s_j^T f_j}, \quad x_j = s_j/\alpha_j,$$
$$t_j = f_j/\alpha_j - \alpha_j y_j$$
$$g_j = Kt_j$$
$$\beta_j = \sqrt{t_j^T g_j}, \quad y_{j+1} = t_j/\beta_j$$

End

$$X_k = \begin{bmatrix} x_1 & x_2 & \dots & x_k \end{bmatrix}$$

$$Y_k = \begin{bmatrix} y_1 & y_2 & \dots & y_k \end{bmatrix}, \quad B_k = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ & \alpha_2 & \ddots & \\ & & \ddots & \beta_{k-1} \\ & & & \alpha_k \end{bmatrix},$$

$$\hat{B}_k = \begin{bmatrix} B_k & \beta_k e_k \end{bmatrix}$$

Then

$$\begin{aligned} KY_k &= X_k B_k, \\ MX_k &= Y_k B_k^T + \beta_k y_{k+1} e_k^T = Y_{k+1} \hat{B}_k^T \end{aligned}$$

Approximations of eigenpairs of MK (KM)

$$MKY_k = Y_k B_k^T B_k + \alpha_k \beta_k y_{k+1} e_k^T,$$
$$KMX_k = X_k \hat{B}_k \hat{B}_k^T + \alpha_{k+1} \beta_k x_{k+1} e_k^T$$

SVD: $B_k \begin{bmatrix} \nu_1 & \dots & \nu_k \end{bmatrix} = \begin{bmatrix} \mu_1 & \dots & \mu_k \end{bmatrix} \mathrm{diag}(\sigma_1, \dots, \sigma_k)$

$$B_k^T B_k \begin{bmatrix} \nu_1 & \dots & \nu_k \end{bmatrix} = \begin{bmatrix} \nu_1 & \dots & \nu_k \end{bmatrix} \mathrm{diag}(\sigma_1^2, \dots, \sigma_k^2)$$

Matrix $B_k^T B_k$: Ritz-value: $\sigma_1^2 \geq \dots \geq \sigma_k^2$

R-Ritz vector: $Y_k \nu_1, \dots, Y_k \nu_k$ (K-orth.)

L-Ritz vector: $X_k \mu_1, \dots, X_k \mu_k$ (M-orth.)

$$\|(MK - \sigma_j^2 I) Y_k \nu_j\|_K = |e_k^T \nu_j| \alpha_k \beta_k,$$
$$\|(KM - \sigma_j^2 I) X_k \mu_j\|_M = |e_k^T \mu_j| \beta_k \sqrt{\beta_k^2 + \alpha_{k+1}^2}.$$

MK: E-value: $\lambda_1^2 \geq \ldots \geq \lambda_n^2, \quad \lambda_j > 0$
R-E-vector: $\xi_1, \ldots, \xi_n$ (K-orth.)
L-E-vector: $\eta_1, \ldots, \eta_n$ ($\eta_j = \frac{1}{\lambda_j} K\xi_j$, M-orth.)

Theorem 1[Convergence theory]

$$0 \leq \lambda_j^2 - \sigma_j^2 \leq (\lambda_1^2 - \lambda_n^2) \left(\frac{\pi_{j,k} \tan \theta_K(y_1, \xi_j)}{C_{k-j}(1 + 2\gamma_j)} \right)^2 \qquad (1)$$

with $\pi_{j,k} = \prod_{i=1}^{j-1} \frac{\sigma_i^2 - \lambda_n^2}{\sigma_i^2 - \lambda_j^2}$, $\theta_K(y_1, \xi_j) = \arccos |y_1^T K \xi_j|$. *and*

$$\gamma_j = \frac{\lambda_j^2 - \lambda_{j+1}^2}{\lambda_{j+1}^2 - \lambda_n^2}$$

$$
\begin{aligned}
&\sin\theta_K(Y_k\nu_j,\xi_j)\\
&=\sqrt{\left(\frac{\sigma_j}{\lambda_j}\right)^2\sin^2\theta_M(X_k\mu_j,\eta_j)+1-\left(\frac{\sigma_j}{\lambda_j}\right)^2}\\
&\le\frac{\pi_j\sqrt{1+(\alpha_k\beta_k)^2/\delta_j^2}}{C_{k-j}(1+2\gamma_j)}\sin\theta_K(y_1,\xi_j) \qquad (2)
\end{aligned}
$$

with

$$
\delta_j=\min_{i\neq j}|\lambda_j^2-\sigma_i^2|,\quad \pi_j=\prod_{i=1}^{j-1}\frac{\lambda_i^2-\lambda_n^2}{\lambda_i^2-\lambda_j^2}.
$$

$$0 \leq \sigma_j^2 - \lambda_{n-k+j}^2 \leq (\lambda_1^2 - \lambda_n^2)\left(\frac{\tilde{\pi}_{j,k} \tan \theta_K(y_1, \xi_{n-k+j})}{C_{j-1}(1+2\tilde{\gamma}_j)}\right)^2$$

with

$$\tilde{\pi}_{j,k} = \prod_{i=n-k+j+1}^{n} \frac{\sigma_i^2 - \lambda_1^2}{\sigma_i^2 - \lambda_{n-k+j}^2};$$

and

$$\tilde{\gamma}_j = \frac{\lambda_{n-k+j-1}^2 - \lambda_{n-k+j}^2}{\lambda_1^2 - \lambda_{n-k+j-1}^2};$$

$$sin\theta_K(Y_k\nu_j, \xi_{n-k+j})$$
$$= \sqrt{\left(\frac{\sigma_j}{\lambda_{n-k+j}}\right)^2 \sin^2\theta_M(X_k\mu_j, \eta_{n-k+j}) + 1 - \left(\frac{\sigma_j}{\lambda_{n-k+j}}\right)^2}$$
$$\leq \frac{\tilde{\pi}_j\sqrt{1+(\alpha_k\beta_k)^2/\tilde{\delta}_j^2}}{C_{j-1}(1+2\tilde{\gamma}_j)} \sin\theta_K(y_1, \xi_{n-k+j})$$

with

$$\tilde{\delta}_j = \min_{i\neq j} |\lambda_{n-k+j}^2 - \sigma_i^2|, \quad \tilde{\pi}_j = \prod_{i=n-k+j+1}^{n} \frac{\lambda_i^2 - \lambda_1^2}{\lambda_i^2 - \lambda_{n-k+j}^2}.$$

Linear response eigenvalue problem

[Bai&Li/12,13] Compute extreme positive eigenvalues of

$$\mathbf{H} = \begin{bmatrix} 0 & M \\ K & 0 \end{bmatrix}, \quad M, K \text{ positive definite.}$$

$$KY = XB, \quad MX = YB^T \Rightarrow$$

$$\mathbf{HX} = \mathbf{XB}, \quad \mathbf{X}^T\mathbf{KX} = \mathbf{I},$$

$$\mathbf{X} = \begin{bmatrix} Y & 0 \\ 0 & X \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & B^T \\ B & 0 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix}.$$

Let $B = \Phi\Lambda\Psi^T$, $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, $\mathbf{P} = \frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}$.

$$\mathbf{H}\tilde{\mathbf{X}} = \tilde{\mathbf{X}}\begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{bmatrix},$$

$$\begin{aligned}\tilde{\mathbf{X}} &= \mathbf{X}\begin{bmatrix} \Psi & 0 \\ 0 & \Phi \end{bmatrix}\mathbf{P} = \frac{1}{\sqrt{2}}\begin{bmatrix} \xi_1 & \dots & \xi_n & \xi_1 & \dots & \xi_n \\ \eta_1 & \dots & \eta_n & -\eta_1 & \dots & -\eta_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_1^+ & \dots & \mathbf{x}_n^+ & \mathbf{x}_1^- & \dots & \mathbf{x}_n^- \end{bmatrix}\end{aligned}$$

M-orth. L-E-vectors: $\mathbf{y}_j^\pm = \frac{1}{\sqrt{2}}\begin{bmatrix} \pm\eta_j \\ \xi_j \end{bmatrix}$, $\mathbf{M} = \begin{bmatrix} M & 0 \\ 0 & K \end{bmatrix}$.

$$KY_k = X_k B_k, \quad MX_k = Y_k B_k^T + \beta_k y_{k+1} e_k^T.$$

$$\mathbf{H} \begin{bmatrix} Y_k & 0 \\ 0 & X_k \end{bmatrix} = \begin{bmatrix} Y_k & 0 \\ 0 & X_k \end{bmatrix} \begin{bmatrix} 0 & B_k^T \\ B_k & 0 \end{bmatrix} + \beta_k \begin{bmatrix} y_{k+1} \\ 0 \end{bmatrix} e_{2k}^T.$$

Let $B_k \begin{bmatrix} \nu_1 & \dots & \nu_k \end{bmatrix} = \begin{bmatrix} \mu_1 & \dots & \mu_k \end{bmatrix} \mathrm{diag}(\sigma_1, \dots, \sigma_k).$

Ritz-values: $\pm\sigma_j$

R-Ritz-vectors: $\mathbf{v}_j^{\pm} = \frac{1}{\sqrt{2}} \begin{bmatrix} Y_k \nu_j \\ \pm X_k \mu_j \end{bmatrix}$ (**K**-orthonormal)

L-Ritz-vectors: $\mathbf{u_j}^{\pm} = \frac{1}{\sqrt{2}} \begin{bmatrix} \pm X_k \mu_j \\ Y_k \nu_j \end{bmatrix}$ (**M**-orthonormal)

$$\|\mathbf{H}\mathbf{v}_j^+ - \sigma_j \mathbf{v}_j^+\|_{\mathbf{K}} = \|\mathbf{H}^T \mathbf{u}_j^+ - \sigma_j \mathbf{u}_j^+\|_{\mathbf{M}} = \frac{\beta_j |e_k^T \mu_j|}{\sqrt{2}}.$$

Theorem 2 [Convergence theory for **H**]

$$0 \le \lambda_j - \sigma_j = (-\sigma_j) - (-\lambda_j) \le \frac{\lambda_1^2 - \lambda_n^2}{\lambda_j + \sigma_j} \left(\frac{\pi_{j,k} \tan \theta_K(y_1, \xi_j)}{C_{k-j}(1+2\gamma_j)} \right)^2$$

$$\sin \theta_{\mathbf{K}} \left(\mathbf{v}_j^{\pm}, , \mathbf{x}_j^{\pm} \right) = \sin \theta_{\mathbf{M}} \left(\mathbf{u}_j^{\pm}, \mathbf{y}_j^{\pm} \right)$$
$$\le \frac{1}{\cos \varrho_j} \sqrt{\frac{\pi_j^2 (1 + (\alpha_k \beta_k)^2 / \delta_j^2)}{C_{k-j}^2 (1+2\gamma_j)} \sin^2 \theta_K(y_1, \xi_j) - \sin^2 \varrho_j}$$

where $\cos \varrho_j = \frac{2\sigma_j}{\lambda_j + \sigma_j}$, $\sin \varrho_j = \sqrt{(\lambda_j - \sigma_j) \frac{\lambda_j + 3\sigma_j}{\lambda_j + \sigma_j}}$;

$$
\begin{aligned}
0 &\leq \sigma_j - \lambda_{n-k+j} = (-\lambda_{n-k+j}) - (-\sigma_j) \\
&\leq \frac{\lambda_1^2 - \lambda_n^2}{\lambda_{n-k+j} + \sigma_j} \left(\frac{\tilde{\pi}_{j,k} \tan \theta_K(y_1, \xi_{n-k+j})}{C_{j-1}(1 + 2\tilde{\gamma}_j)} \right)^2;
\end{aligned}
$$

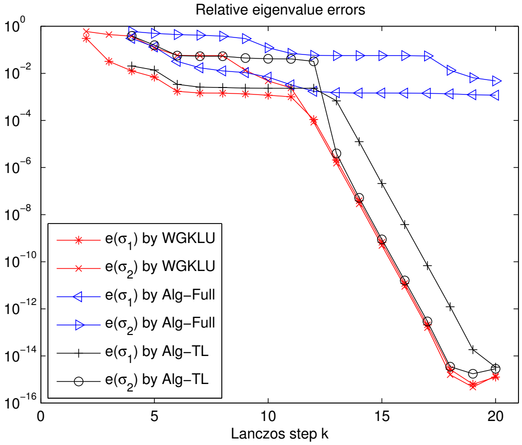
$$
\sin \theta_{\mathbf{K}} \left(\mathbf{v}_j^{\pm}, \mathbf{x}_{n-k+j}^{\pm} \right) = \sin \theta_{\mathbf{M}} \left(\mathbf{u}_j^{\pm}, \mathbf{y}_{n-k+j}^{\pm} \right)
$$

$$
\leq \sqrt{\sin^2 \tilde{\varrho}_j + \cos^2 \tilde{\varrho}_j \frac{\tilde{\pi}_j^2 (1 + (\alpha_k \beta_k)^2 / \tilde{\delta}_j^2)}{C_{j-1}^2 (1 + 2\tilde{\gamma}_j)} \sin^2 \theta_K(y_1, \xi_{n-k+j})},
$$

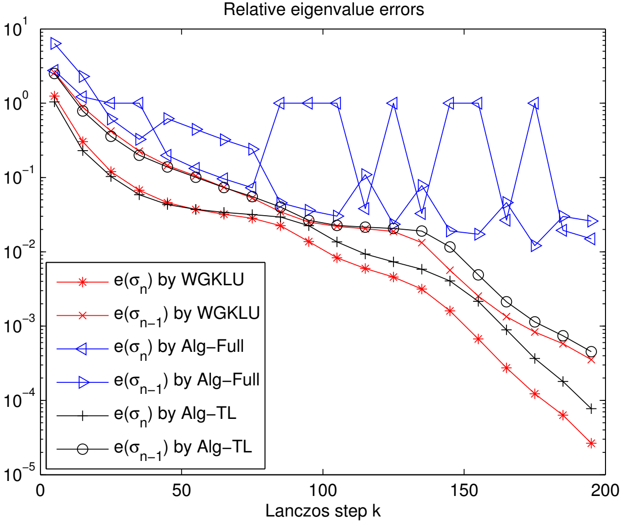
where $\tilde{\varrho}_j = \arccos \frac{\sigma_j + \lambda_{n-k+j}}{2\sigma_j}$, $\sin \tilde{\varrho}_j = \frac{\sqrt{(\sigma_j - \lambda_{n-k+j})(3\sigma_j + \lambda_{n-k+j})}}{2\sigma_j}$.

Numerical Example

Example 1 [Davis/Hu'11, Teng/Li'13] K, M are 9604×9604.
Largest E-values: 9.8, 9.75 ($k = 20$)



Smallest E-values: 1.15, 1.17 ($k = 200$)



Connection with CG

$$Mz = b, \quad M \text{ positive definite.}$$

Let z_0 be an initial guess, $r_0 = b - Mz_0$.
Take $y_1 = r_0/\|r_0\|_K$ in WGKL:

$$KY_k = X_k B_k, \quad MX_k = Y_k B_k^T + \beta_k y_{k+1} e_k^T = Y_{k+1} \hat{B}_k^T.$$

Let $z_k = z_0 + X_k w_k$ be the minimizer of

$$\min_{z=z_0+X_k w} (z_* - z)^T M(z_* - z), \qquad z_* = M^{-1} b.$$

$$z_k = z_{k-1} + \varphi_k x_k, \qquad r_k = -\beta_k \varphi_k y_{k+1}.$$

$$\varphi_k = -\frac{\beta_{k-1}}{\alpha_k}\varphi_{k-1}; \quad \beta_0 = 1,\ \varphi_0 = -\|r_0\|_K.$$

Define

$$p_{k-1} = \alpha_k^2 \varphi_k x_k, \quad p_0 = \alpha_1^2 \varphi_1 x_1 = K r_0.$$

We have the weighted CG (standard CG if $K = I$):

$$\begin{aligned}
z_k &= z_{k-1} + \gamma_{k-1} p_{k-1} \\
r_k &= r_{k-1} - \gamma_{k-1} M p_{k-1}, \quad \gamma_{k-1} = \frac{r_{k-1}^T K r_{k-1}}{p_{k-1}^T M p_{k-1}} \\
p_k &= K r_k + \vartheta_{k-1} p_{k-1}, \quad \vartheta_{k-1} = \frac{r_k^T K r_k}{r_{k-1}^T K r_{k-1}}
\end{aligned}$$

Thank you