Lanczos tridiagonalization, Golub-Kahan bidiagonalization and the core problem

I. Hnětynková[*,], Z. Strakoš[*]**

[*] Institute of Computer Science, Academy of Sciences, Prague, Czech republic

[**] Charles University, Faculty of Mathematics and Physics, Department of Numerical Mathematics, Prague

SNA'06 - Seminar on numerical analysis
January 16 - 20, 2006, Sedlec-Prčice

Overview:

- Motivation
- Golub-Kahan bidiagonalization
- Connection with Lanczos tridiagonalization
- The sketch of the proof

1. Motivation

Consider an orthogonally invariant approximation problem

$$(1)\qquad \boldsymbol{Ax} \approx \boldsymbol{b}, \quad \boldsymbol{A} \in \mathbb{R}^{n\times m}, \boldsymbol{b} \in \mathbb{R}^{n},$$

where $\approx$ means, e.g., using data corrections of the prescribed type in order to get the nearest compatible system (Least Squares, (Scaled) Total/Data Least squares).

For simplicity let $\boldsymbol{A}^T\boldsymbol{b} \neq \mathbf{0}$.

C.C. Paige, Z. Strakoš (2006) proposed transformation

$$P^T [\, b \,\|\, AQ \,] = \left[\begin{array}{c||c|c} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right], \quad P^{-1} = P^T,\ Q^{-1} = Q^T,$$

where $b_1 = \beta_1 e_1$ and A_{11} is a lower bidiagonal with nonzero bidiagonal elements. The original problem is thus decomposed into two subproblems

$$A_{11} x_1 \approx b_1, \qquad A_{22} x_2 \approx 0$$

and the original problem solution is taken $x \equiv Q \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$.

Properties:

a) $\boldsymbol{A}_{11}$ has no zero or multiple singular values. Any zero singular values or repeats that $\boldsymbol{A}$ has must appear in $\boldsymbol{A}_{22}$.
b) $\boldsymbol{A}_{11}$ has minimal dimensions and $\boldsymbol{A}_{22}$ has maximal dimensions, over all orthogonal transformations giving the presented block structure.
c) All components of $\boldsymbol{b}_1 = \beta_1 \boldsymbol{e}_1$ in the left singular vector subspaces of $\boldsymbol{A}_{11}$ (i.e. the first elements of all left singular vectors of $\boldsymbol{A}_{11}$) are nonzero.

Consequence: The TLS solution of $A_{11}x_1 \approx b_1$ *always exists and is unique.* Thus the system is called the core problem in (1).

Computation of the core problem:

– direct partial upper bidiagonalization of $[b, A]$ by using the Householder reflections (small A)

– partial Golub-Kahan bidiagonalization (large A)

We will use the Golub-Kahan bidiagonalization, in relation with the Lanczos tridiagonalization and the properties of Jacobi matrices, to prove the core problem properties a) – c).

2. Golub-Kahan bidiagonalization

Let $v_0 \equiv 0, u_1 \equiv b/\beta_1$, where $\beta_1 = \|b\|$. The algorithm computes for $i = 1, 2, \ldots$

$$\alpha_i v_i = A^T u_i - \beta_i v_{i-1}\,, \qquad \|v_i\| = 1$$

$$\beta_{i+1} u_{i+1} = A v_i - \alpha_i u_i\,, \qquad \|u_i\| = 1$$

until $\alpha_i = 0$ or $\beta_{i+1} = 0$ or $i = \min\{n, m\}$.

Let $\alpha_i \beta_i \neq 0$ for $1 \leq i \leq k+1$ and denote by

$U_k = (u_1, \ldots, u_k), \; V_k = (v_1, \ldots, v_k),$

$$L_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \cdots & & \\ & \beta_k & \alpha_k & \end{pmatrix}, \quad L_{k+} = \begin{pmatrix} L_k \\ \beta_{k+1} e_k^T \end{pmatrix}.$$

The matrices U_k, V_k have orthonormal columns and

(2) $$A^T U_k = V_k L_k^T ,$$

(3) $$A V_k = [U_k, u_{k+1}] L_{k+} .$$

The bidiagonalization of $\boldsymbol{A}$ with $\boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$ stops with:

- $\boldsymbol{\alpha_i \beta_i \neq 0}$ for $\boldsymbol{i = 1, \ldots, p}$; $\boldsymbol{\beta_{p+1} = 0}$ or $\boldsymbol{p = n}$. Then (2) gives

$$\boldsymbol{U_p^T A V_p = L_p}\,,$$

$$\boldsymbol{U_p^T [b, AV_p]} = \left[\begin{array}{c|cccc} \boldsymbol{\beta_1} & \boldsymbol{\alpha_1} & & & \\ & \boldsymbol{\beta_2} & \boldsymbol{\alpha_2} & & \\ & & \cdots & & \\ & & \boldsymbol{\beta_p} & \boldsymbol{\alpha_p} & \end{array}\right],$$

and $\boldsymbol{L_p x_1 \approx \beta_1 e_1}$ is the *compatible* core problem. $\boldsymbol{U_p, V_p}$ represent the first $\boldsymbol{p}$ columns of the matrices $\boldsymbol{P, Q}$ respectively.

- $\alpha_i \beta_i \neq 0$ for $i = 1, \ldots, p$, and $\beta_{p+1} \neq 0$; $\alpha_{p+1} = 0$ or $p = m$. Then (3) gives

$$[U_p, u_{p+1}]^T A V_p = L_{p+},$$

$$[U_p, u_{p+1}]^T [b, AV_p] = \left[\begin{array}{c|cccc} \beta_1 & \alpha_1 & & & \\ & \beta_2 & \alpha_2 & & \\ & & \ldots & & \\ & & \beta_p & \alpha_p & \\ & & & \beta_{p+1} & \end{array}\right]$$

and $L_{p+} x_1 \approx \beta_1 e_1$ is the *incompatible* core problem. U_{p+1}, V_p represent the first $(p+1)$ and p columns of the matrices P, Q respectively.

3. Connection with the Lanczos tridiagonalization

Let $B \in \mathbb{R}^{\nu\times\nu}$, $w_1 \in R^{\nu}$ $(\|w_1\| = 1)$. The algorithm computes in i steps

$$BW_i = W_i Z_i + \gamma_i w_{i+1} e_i^T, \qquad W_i^T w_{i+1} = 0,$$

where $W_i^T W_i = I$, $Z_i \in \mathbb{R}^{i\times i}$ is symmetric tridiagonal with positive subdiagonal elements.

We considere a possibly incomplete tridiagonalization stoped whenever $\gamma_i = 0$.

Properties of the Jacobi matrix $\boldsymbol{Z_i}$:

- All its eigenvalues are simple.

- If $\boldsymbol{B}$ is symmetric positive semidefinite and $\boldsymbol{w_1} \perp \mathbf{ker}\,(\boldsymbol{B})$, then all its eigenvalues are positive.

- The first components of all its eigenvectors are nonzero.

Bidiagonalization $\times$ tridiagonalization:

- augmented matrix formulation

- normal matrix formulation

Augmented matrix formulation:

Tridiagonalization of

$$\boldsymbol{B} = \begin{pmatrix} \mathbf{0} & \boldsymbol{A} \\ \boldsymbol{A}^T & \mathbf{0} \end{pmatrix}$$

with $\boldsymbol{w}_1 = (\boldsymbol{u}_1, 0)^T$ yields in $\boldsymbol{2k}$ steps the matrix

$$\boldsymbol{W}_{2k} = \begin{pmatrix} \boldsymbol{u}_1 & \mathbf{0} & \ldots & \boldsymbol{u}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{v}_1 & \ldots & \mathbf{0} & \boldsymbol{v}_k \end{pmatrix}$$

and $\boldsymbol{Z}_{2k}$ with zeros on the diagonal and the subdiagonals $(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\alpha}_k)$ see [Paige - 1974].

Normal matrix formulation:

Assume that the Lanczos algorithm does not stop for $i = 1, \ldots, k+1$.

Using
$$A V_k = U_{k+1} L_{k+},$$
$$A^T U_{k+1} = V_k L_{k+}^T + v_{k+1} \alpha_{k+1} e_{k+1}^T$$

we obtain

$$A^T A V_k = A^T U_{k+1} L_{k+} = V_k L_{k+}^T L_{k+} + \alpha_{k+1} \beta_{k+1} v_{k+1} e_k^T,$$

where $V_k^T V_k = I$, $V_k^T v_{k+1} = 0$ and

$$L_{k+}^T L_{k+} = \begin{pmatrix} \alpha_1^2 + \beta_2^2 & \alpha_2\beta_2 & & & \\ \alpha_2\beta_2 & \alpha_2^2 + \beta_3^2 & \ddots & & \\ & \ddots & \ddots & \alpha_k\beta_k \\ & & \alpha_k\beta_k & \alpha_k^2 + \beta_{k+1}^2 \end{pmatrix}.$$

Summarizing, k steps of the Lanczos tridiagonalization of $B \equiv A^T A$ with the starting vector $v_1 = A^T b / \|A^T b\|$ produces $Z_k = L_{k+}^T L_{k+}$.

Similarly,
$$\boldsymbol{A^T U_k = V_k L_k^T,}$$
$$\boldsymbol{A V_k = U_k L_k + u_{k+1} \beta_{k+1} e_k^T}$$

gives

$$\boldsymbol{A A^T U_k = U_k L_k L_k^T + \alpha_k \beta_{k+1} u_{k+1} e_k^T,} \tag{4}$$

where $\boldsymbol{U_k^T U_k = I}$, $\boldsymbol{U_k^T u_{k+1} = 0}$ and

$$\boldsymbol{L_k L_k^T = \begin{pmatrix} \alpha_1^2 & \alpha_1 \beta_2 & & \\ \alpha_1 \beta_2 & \alpha_2^2 + \beta_2^2 & \ddots & \\ & \ddots & \ddots & \alpha_{k-1} \beta_k \\ & & \alpha_{k-1} \beta_k & \alpha_k^2 + \beta_k^2 \end{pmatrix}.}$$

Summarizing, $\boldsymbol{k}$ steps of the Lanczos tridiagonalization of $\boldsymbol{B} \equiv \boldsymbol{A}\boldsymbol{A}^T$ with the starting vector $\boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$ produces $\boldsymbol{Z}_k = \boldsymbol{L}_k\boldsymbol{L}_k^T$.

4. Sketch of the proof

I. Case $\beta_{p+1} = 0$ or $p = n$, i.e. $A_{11} = L_p$:

L_p is the Cholesky factor of $Z_p = L_p L_p^T$ that results from p steps of the tridiagonalization of AA^T with $u_1 = b/\|b\|$.

Let $L_p = R\Sigma S^T$ be the SVD ($\Sigma =$diag$(\sigma_1, \ldots, \sigma_p)$), then

$$Z_p \equiv L_p L_p^T = R\Sigma^2 R^T.$$

Consequently, (σ_i^2, Re_i) are the eigenpairs of Z_p.

The properties of Z_p yield:

- $\sigma_1^2, \ldots, \sigma_p^2$ are nonzero and distinct $\Rightarrow$ the singular values of L_p are simple and nonzero
- $e_1^T R e_i \neq 0$ for $i = 1, 2, \ldots, p \; \Rightarrow$

$$b_1^T R e_i = \beta_1 e_1^T R e_i \neq 0 \quad \text{for} \quad i = 1, 2, \ldots, p$$

and thus b_1 has no zero components in the direction of the left singular vectors of L_p

Consequence: The system $L_p x_1 = \beta_1 e_1$ is compatible and satisfies the core problem properties a) and c).

II. Case $\alpha_{p+1} = 0$ or $p = m$, i.e. $A_{11} = L_{p+}$ **:**
Then $Z_p = L_{p+}^T L_{p+}$ results from p steps of the tridiagonalization of $A^T A$ with the starting vector $v_1 = A^T b / \|A^T b\|$.

Let $L_{p+} = R\Sigma S^T$ be the SVD, then

$$Z_p \equiv L_{p+}^T L_{p+} = S\Sigma^2 S^T.$$

Consequently, (σ_i^2, Se_i) are the eigenpairs of Z_p.

Similarly to the previous, the singular values of L_{p+} are distinct and nonzero and $e_1^T S e_i \neq 0$.

Since $\boldsymbol{L_{p+}S} = \boldsymbol{R\Sigma}$ and $\boldsymbol{L_{p+}}$ is lower bidiagonal,
$\boldsymbol{e_1^T R e_i \neq 0} \;\Rightarrow$

$$\boldsymbol{b_1^T R e_i \neq 0} \quad \text{for} \quad \boldsymbol{i = 1, \ldots, p.}$$

Consequence: The system $\boldsymbol{L_{p+}x_1 \approx \beta_1 e_1}$ is incompatible and satisfies the properties a) and c).

Condition b) (the minimal dimensions of $\boldsymbol{A_{11}}$):
Follows by contradictions similarly to [Paige, Strakoš - 2006].

Thank you for your attention!