Properties of extended Krylov subspaces

Carl Christian Kjelgaard Mikkelsen

Department of Mathematics and Information Technology
Umeå University

March 10th, 2010
Outline of Topics

The standard Krylov subspaces

The extended Krylov subspaces

Difficult LTI systems

Open Questions
The standard Krylov subspace $K_j(A, B)$ is given by

$$K_j(A, B) = \text{span}\{B, AB, AB, \ldots, A^{j-1}B\}, \quad j = 1, 2, \ldots.$$ 

Clearly, the sequence is monotone increasing

$$K_j(A, B) \subseteq K_{j+1}(A, B), \quad j = 1, 2, \ldots.$$
The smallest $A$ invariant subspace containing the range of $B$ is

$$K(A, B) = \bigcup_{j=1}^{\infty} K_j(A, B).$$

There is a smallest $m$ such that

$$K(A, B) = K_m(A, B),$$

called the grade of $B$ with respect to $A$. 
Elementary properties

In addition

\[ AK_j(A, B) \subseteq K_{j+1}(A, B), \quad j = 1, 2, \ldots \]

That is to say, \( A \) maps each subspace into the next.
Elementary properties

Let \( n_j \) denote the dimension of \( K_j(A, B) \),

\[
n_j = \dim_{\mathbb{R}} K_j(A, B)
\]

and let \( \{v_i\}_{i=1}^{nm} \) be an orthonormal sequence of vectors such that

\[
K_j(A, B) = \text{span}\{v_1, v_2, \ldots, v_{n_j}\}, \quad j = 1, 2, \ldots, m.
\]

Define

\[
V_j = \begin{bmatrix} v_1 & v_2 & \cdots & v_{n_j} \end{bmatrix}, \quad j = 1, 2, \ldots, m.
\]

Then

\[
AV_m = V_m H_m
\]

for some matrix \( H_m \in \mathbb{R}^{nm \times nm} \).
Elementary properties

Specifically,

\[ H_m = V_m^T AV_m. \]

In addition, \( H_m \) is upper block Hessenberg,

\[
H_m = \begin{bmatrix}
H_{11} & H_{12} & \ldots & \ldots & H_{1m} \\
H_{21} & H_{22} & \ldots & \ldots & H_{2m} \\
0 & H_{32} & \ldots & \ldots & H_{3m} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & H_{m,m-1} & H_{mm}
\end{bmatrix}
\]

simply because

\[ AK_j(A, B) \subseteq K_{j+1}(A, B), \; j = 1, 2, \ldots. \]
Relation to Lyapunov matrix equations

Let $A$ be a stable matrix and consider

$$AX + XA^T + BB^T = 0$$

This equation has a unique solution $X$ and

$$X = X^T \geq 0, \quad \text{Ran } X = K(A, B), \quad \text{Ker } X = K(A, B)^\perp.$$ 

In addition

$$X = V_m Y_m V_m^T$$

if and only if

$$H_m Y_m + Y_m H_m^T + B_m B_m^T = 0, \quad B_m = V_m^T B.$$
The basic Arnoldi method

The basic Arnoldi method [6, 2] solves the reduced order equation

\[ H_j Y_j + Y_j H_j^T + B_j B_j^T = 0, \quad B_j = V_j^T B. \]

with respect to \( Y_j \) and uses

\[ X_j = V_j Y_j V_j^T, \quad j = 1, 2, \ldots \]

as an approximation for \( X \).
The basic Arnoldi method

In general, if $A$ is negative definite,

$$A + A^T < 0$$

then

$$H = V^T AV$$

is negative definite for any matrix $V$ with orthonormal columns.

Therefore, if

$$A + A^T < 0$$

then the Arnoldi method is well defined.
In practice, the Arnoldi method converges very slowly.
In practice, the Arnoldi method converges very slowly.

Simoncini and Druskin [8] have bounded the convergence rate in terms of the numerical range of $A$,

$$nr(A) = \{ x^*Ax : x \in \mathbb{C}^n, \|x\|_2 \leq 1 \}$$
Convergence of the Arnoldi method

- In practice, the Arnoldi method converges very slowly.
- Simoncini and Druskin [8] have bounded the convergence rate in terms of the numerical range of $A$,
  \[ \text{nr}(A) = \{ x^* A x : x \in \mathbb{C}^n, \|x\|_2 \leq 1 \} \]
- Mikkelsen [4] has shown that any positive residual history is possible, even for symmetric negative definite systems.
The solution of

$$AX + XA^T + BB^T = 0 \quad (1)$$

is given by

$$X = \int_0^\infty e^{tA}BB^T e^{tA^T} dt$$

Therefore, solving (1) is really a question of approximating

$$t \rightarrow e^{tA}B, \quad t > 0.$$
The Arnoldi algorithm gives us

\[ AV_j = V_j H_j + H_{j+1,j} Q_{j+1} E_j^T \]

and we approximate

\[ e^{tA} B \approx V_j e^{tH_j} B_j, \quad B_j = V_j^T B \]

simply because

\[ e^{tA} B = V_m e^{tH_m} B_m, \quad B_m = V_m^T B. \]
The extended Krylov subspace $\mathbf{EK}_j(A, B)$ is given by

$$\mathbf{EK}_j(A, B) = \text{span}\{A^{-j}B, A^{-j+1}B, \ldots, B, AB, \ldots, A^{i-1}B\}$$

Druskin and Knizhnerman [1] found that $\mathbf{EK}_j(A, B)$ is superior to $\mathbf{K}_{2j}(A, B)$ when it comes to approximating certain matrix functions.
Clearly,
\[ \mathbf{EK}_j(A, B) \subseteq \mathbf{EK}_{j+1}(A, B), \quad j = 1, 2, \ldots \]
and
\[ K(A, B) = \bigcup_{j=1}^{\infty} \mathbf{EK}_j(A, B) \]

In addition
\[ A \mathbf{EK}_j(A, B) \subseteq \mathbf{EK}_{j+1}(A, B), \quad j = 1, 2, \ldots \]
and
\[ A^{-1} \mathbf{EK}_j(A, B) \subseteq \mathbf{EK}_{j+1}(A, B), \quad j = 1, 2, \ldots \]
Let $n_j$ denote the dimension of $\text{EK}_j(A, B)$,

$$n_j = \dim \mathbb{R} K_j(A, B)$$

and let $\{v_i\}_{i=1}^{nm}$ be any orthonormal sequence of vectors such that

$$\text{EK}_j(A, B) = \text{span}\{v_1, v_2, \ldots, v_{n_j}\}, \quad j = 1, 2, \ldots, m.$$

Define

$$V_j = \begin{bmatrix} v_1 & v_2 & \ldots & v_{n_j} \end{bmatrix}, \quad j = 1, 2, \ldots, m.$$

Then

$$AV_m = V_m H_m$$

for some matrix $H_m \in \mathbb{R}^{nm \times nm}$. 
Elementary properties

Specifically,

\[ H_m = V_m^T A V_m. \]

In addition, \( H_m \) is upper block Hessenberg,

\[
H_m = \begin{bmatrix}
H_{11} & H_{12} & \cdots & \cdots & H_{1m} \\
H_{21} & H_{22} & \cdots & \cdots & H_{2m} \\
0 & H_{32} & \cdots & \cdots & H_{3m} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & H_{m,m-1} & H_{mm}
\end{bmatrix}
\]

simply because

\[ A \text{EK}_j(A, B) \subseteq \text{EK}_{j+1}(A, B), \quad j = 1, 2, \ldots. \]
Elementary properties

In addition

\[ A^{-1}V_m = V_mK_m, \quad K_m = V_m^T A^{-1}V_m = H_m^{-1} \]

and \( K_m \) is upper block Hessenberg,

\[
K_m = \begin{bmatrix}
K_{11} & K_{12} & \ldots & \ldots & K_{1m} \\
K_{21} & K_{22} & \ldots & \ldots & K_{2m} \\
0 & K_{32} & \ldots & \ldots & K_{3m} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & K_{m,m-1} & K_{mm}
\end{bmatrix}
\]

simply because

\[ A^{-1} EK_j(A, B) \subseteq EK_{j+1}(A, B), \quad j = 1, 2, \ldots \]
A fairly general special case

Consider the special case of

\[ B \in \mathbb{R}^n, \quad n = 2m, \quad K(A, B) = \mathbb{R}^n. \]

Let \( W_m \) be the matrix given by

\[ W_m(A, B) = \begin{bmatrix} B & A^{-1}B & AB & A^{-2}B & \ldots & A^{m-1}B & A^{-m}B \end{bmatrix} \]

and let \( \{v_i\}_{i=1}^{2m} \) be any sequence of orthonormal vectors such that

\[ \text{span}\{w_1, w_2, \ldots, w_i\} = \text{span}\{v_1, v_2, \ldots, v_i\}, \quad i = 1, 2, \ldots, n. \]

Then

\[ H_m = V_m^T A V_m, \quad \text{and} \quad K_m = H_m^{-1} \]

are both block Hessenberg with block size 2.
In addition, the subdiagonal blocks have a particular nonzero pattern

$$
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  * & x & x & x & x \\
  x & x & x & x & x \\
  * & x & x & x & x
\end{bmatrix}
$$

$$
\begin{bmatrix}
  x & x & x & x & x \\
  x & x & x & x & x \\
  * & x & x & x & x \\
  x & x & x & x & x \\
  * & x & x & x & x
\end{bmatrix}
$$

where $x$ indicates a possible nonzero and $*$ is a nonzero.
KPIK/EKSM algorithm for Lyapunov matrix equations

Simoncini [7] applied the extended Krylov subspaces to the Lyapunov matrix equation

$$AX + XA^T + BB^T = 0.$$  

The reduced order equations

$$H_jY_j + Y_jH_j^T + B_j^TB_j^T = 0, \quad B_j = V_j^TB, \quad j = 1, 2, \ldots, m$$  

are solved and

$$X_j = V_jY_jV_j^T, \quad j = 1, 2, \ldots, m$$  

is used as an approximation for $X$. 

Kjelgaard Mikkelsen  
Properties of extended Krylov subspaces
Knizhnerman and Simoncini [3] have bounded the convergence rate in terms of the numerical range of $A$. 
Convergence of KPIK/EKSM

- Knizhnerman and Simoncini [3] have bounded the convergence rate in terms of the numerical range of $A$.
- In practice, the new algorithm is vastly superior to the old.
Knizhnerman and Simoncini [3] have bounded the convergence rate in terms of the numerical range of $A$.

In practice, the new algorithm is vastly superior to the old.

However, Mikkelsen [5] has shown that any positive residual history is still possible, but highly unlikely.
It is now possible to create a linear time invariant system
\[ \dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \]
such that the Gramians \( P \) and \( Q \), i.e. the solutions of
\[ AP + PA^T + BB^T = 0, \quad A^T Q + QA + C^T C = 0 \]
are nonsingular, but if
\[ P_i, Q_j \]
are the EKSM approximations, then
\[ P_i Q_j = 0, \quad i + j \leq m \]
Implications

- If the algorithm converges rapidly, then this property is passed to the exact Gramians, $P$ and $Q$

and nothing can be said about the product $PQ$ and $QP$!
Implications

- If the algorithm converges rapidly, then this property is passed to the exact Gramians, $P$ and $Q$,

and nothing can be said about the product $PQ$ and $QP$!

- It is impossible to prove convergence for a specific class of model reduction algorithms!
Open questions

- What are the systems for which the action of $PQ$ and $QP$ can be approximated?
Open questions

- What are the systems for which the action of $PQ$ and $QP$ can be approximated?
- Can we find well-conditioned Lyapunov equations for which EKSM converges slowly?
Open questions

- What are the systems for which the action of $PQ$ and $QP$ can be approximated?
- Can we find well-conditioned Lyapunov equations for which EKSM converges slowly?
- Is there an iterative method which converges rapidly if the exact solution admits a good low rank approximation?
Questions?
Thank you for your attention!


Bibliography II


