MatTriad'2015
Coimbra
September 7–11, 2015

Numerical Solution of Matrix Equations Arising in Control of Bilinear and Stochastic Systems

Peter Benner

Max Planck Institute for Dynamics of Complex Technical Systems
Computational Methods in Systems and Control Theory
Magdeburg, Germany

`http://www.mpi-magdeburg.mpg.de/benner`
`benner@mpi-magdeburg.mpg.de`

Overview

Overview

Introduction

Linear Matrix Equations/Men with Beards

Sylvester equation

James Joseph Sylvester
(September 3, 1814 – March 15, 1897)

$$AX + XB = C.$$

Introduction

Linear Matrix Equations/Men with Beards

Sylvester equation

James Joseph Sylvester
(September 3, 1814 – March 15, 1897)

$$AX + XB = C.$$

Lyapunov equation

Alexander Michailowitsch Ljapunow
(June 6, 1857 – November 3, 1918)

$$AX + XA^T = C, \quad C = C^T.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Generalized Sylvester equation:

$$AXD + EXB = C.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Generalized Sylvester equation:

$$AXD + EXB = C.$$

Generalized Lyapunov equation:

$$AXE^T + EXA^T = C, \quad C = C^T.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Generalized Sylvester equation:

$$AXD + EXB = C.$$

Generalized Lyapunov equation:

$$AXE^T + EXA^T = C, \quad C = C^T.$$

Stein equation:

$$X - AXB = C.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Generalized Sylvester equation:

$$AXD + EXB = C.$$

Generalized Lyapunov equation:

$$AXE^T + EXA^T = C, \quad C = C^T.$$

Stein equation:

$$X - AXB = C.$$

(Generalized) discrete Lyapunov/Stein equation:

$$EXE^T - AXA^T = C, \quad C = C^T.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Generalized Sylvester equation:

$$AXD + EXB = C.$$

Generalized Lyapunov equation:

$$AXE^T + EXA^T = C, \quad C = C^T.$$

Stein equation:

$$X - AXB = C.$$

(Generalized) discrete Lyapunov/Stein equation:

$$EXE^T - AXA^T = C, \quad C = C^T.$$

Note:

- Consider only regular cases, having a unique solution!
- Solutions of symmetric cases are symmetric, $X = X^T \in \mathbb{R}^{n\times n}$; otherwise, $X \in \mathbb{R}^{n\times \ell}$ with $n \neq \ell$ in general.

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Bilinear Lyapunov equation/Lyapunov-plus-positive equation:

$$AX + XA^T + \sum_{k=1}^{m} N_k X N_k^T = C, \quad C = C^T.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Bilinear Lyapunov equation/Lyapunov-plus-positive equation:

$$AX + XA^T + \sum_{k=1}^{m} N_k X N_k^T = C, \quad C = C^T.$$

Bilinear Sylvester equation:

$$AX + XB + \sum_{k=1}^{m} N_k X M_k = C.$$

Introduction

Generalizations of Sylvester ($AX + XB = C$) and Lyapunov ($AX + XA^T = C$) Equations

Bilinear Lyapunov equation/Lyapunov-plus-positive equation:

$$AX + XA^T + \sum_{k=1}^{m} N_k X N_k^T = C, \quad C = C^T.$$

Bilinear Sylvester equation:

$$AX + XB + \sum_{k=1}^{m} N_k X M_k = C.$$

(Generalized) discrete bilinear Lyapunov/Stein-minus-positive eq.:

$$EXE^T - AXA^T - \sum_{k=1}^{m} N_k X N_k^T = C, \quad C = C^T.$$

Note: Again consider only regular cases, symmetric equations have symmetric solutions.

Introduction

Existence of Solutions of Linear Matrix Equations I

Exemplarily, consider the generalized Sylvester equation

$$AXD + EXB = C. \tag{1}$$

Introduction

Existence of Solutions of Linear Matrix Equations I

Exemplarily, consider the generalized Sylvester equation

$$AXD + EXB = C. \tag{1}$$

Vectorization (using Kronecker product) $\rightsquigarrow$ representation as linear system:

$$\big(\underbrace{D^T \otimes A + B^T \otimes E}_{=:\mathcal{A}}\big)\underbrace{\mathrm{vec}(X)}_{=:x} = \underbrace{\mathrm{vec}(C)}_{=:c} \qquad \Longleftrightarrow \qquad \mathcal{A}x = c.$$

Introduction

Existence of Solutions of Linear Matrix Equations I

Exemplarily, consider the generalized Sylvester equation

$$AXD + EXB = C. \tag{1}$$

Vectorization (using Kronecker product) $\rightsquigarrow$ representation as linear system:

$$\big(\underbrace{D^T \otimes A + B^T \otimes E}_{=:\mathcal{A}}\big)\underbrace{\mathrm{vec}(X)}_{=:x} = \underbrace{\mathrm{vec}(C)}_{=:c} \quad \Longleftrightarrow \quad \mathcal{A}x = c.$$

$\Longrightarrow$ "(1) has a unique solution $\Longleftrightarrow$ $\mathcal{A}$ is nonsingular"

Introduction

Existence of Solutions of Linear Matrix Equations I

Exemplarily, consider the generalized Sylvester equation

$$AXD + EXB = C. \tag{1}$$

Vectorization (using Kronecker product) $\rightsquigarrow$ representation as linear system:

$$\big(\underbrace{D^T \otimes A + B^T \otimes E}_{=:\mathcal{A}}\big)\underbrace{\mathrm{vec}(X)}_{=:x} = \underbrace{\mathrm{vec}(C)}_{=:c} \qquad \Longleftrightarrow \qquad \mathcal{A}x = c.$$

$\Longrightarrow$ "(1) has a unique solution $\Longleftrightarrow$ $\mathcal{A}$ is nonsingular"

Lemma

$$\Lambda(\mathcal{A}) = \{\alpha_j + \beta_k \mid \alpha_j \in \Lambda(A, E),\ \beta_k \in \Lambda(B, D)\}.$$

Hence, (1) has unique solution $\Longleftrightarrow \Lambda(A, E) \cap -\Lambda(B, D) = \emptyset$.

Introduction

Existence of Solutions of Linear Matrix Equations I

Exemplarily, consider the generalized Sylvester equation

$$AXD + EXB = C. \tag{1}$$

Vectorization (using Kronecker product) $\rightsquigarrow$ representation as linear system:

$$(\underbrace{D^T \otimes A + B^T \otimes E}_{=:\mathcal{A}})\underbrace{\mathrm{vec}(X)}_{=:x} = \underbrace{\mathrm{vec}(C)}_{=:c} \quad \Longleftrightarrow \quad \mathcal{A}x = c.$$

$\Longrightarrow$ "(1) has a unique solution $\Longleftrightarrow$ $\mathcal{A}$ is nonsingular"

Lemma

$$\Lambda(\mathcal{A}) = \{\alpha_j + \beta_k \mid \alpha_j \in \Lambda(A, E),\ \beta_k \in \Lambda(B, D)\}.$$

Hence, (1) has unique solution $\Longleftrightarrow \Lambda(A, E) \cap -\Lambda(B, D) = \emptyset$.

Example: Lyapunov equation $AX + XA^T = C$ has unique solution $\Longleftrightarrow \nexists\, \mu \in \mathbb{C} :\ \pm\mu \in \Lambda(A)$.

Introduction

The Classical Lyapunov Theorem

Theorem (Lyapunov 1892)

Let $A \in \mathbb{R}^{n\times n}$ and consider the Lyapunov operator $\mathcal{L} : X \to AX + XA^T$. Then the following are equivalent:

(a) $\forall Y > 0\colon \exists X > 0\colon \mathcal{L}(X) = -Y$,

(b) $\exists Y > 0\colon \exists X > 0\colon \mathcal{L}(X) = -Y$,

(c) $\Lambda(A) \subset \mathbb{C}^- := \{z \in \mathbb{C} \mid \Re z < 0\}$, *i.e., A is (asymptotically) stable or Hurwitz.*

A. M. Lyapunov. *The General Problem of the Stability of Motion* (in Russian). Doctoral dissertation, Univ. Kharkov 1892. English translation: Stability of Motion, Academic Press, New-York & London, 1966.

P. Lancaster, M. Tismenetsky. *The Theory of Matrices* (2nd edition). Academic Press, Orlando, FL, 1985. [Chapter 13]

Introduction

The Classical Lyapunov Theorem

Theorem (Lyapunov 1892)

Let $A \in \mathbb{R}^{n \times n}$ and consider the Lyapunov operator $\mathcal{L} : X \to AX + XA^T$. Then the following are equivalent:

(a) $\forall Y > 0$: $\exists X > 0$: $\mathcal{L}(X) = -Y$,

(b) $\exists Y > 0$: $\exists X > 0$: $\mathcal{L}(X) = -Y$,

(c) $\Lambda(A) \subset \mathbb{C}^- := \{z \in \mathbb{C} \mid \Re z < 0\}$, *i.e., A is (asymptotically) stable or Hurwitz.*

The proof (c) $\Rightarrow$ (a) is trivial from the necessary and sufficient condition for existence and uniqueness, apart from the positive definiteness. The latter is shown by studying $z^H Yz$ for all eigenvectors z of A.

A. M. Lyapunov. *The General Problem of the Stability of Motion* (in Russian). Doctoral dissertation, Univ. Kharkov 1892. English translation: Stability of Motion, Academic Press, New-York & London, 1966.

P. Lancaster, M. Tismenetsky. *The Theory of Matrices* (2nd edition). Academic Press, Orlando, FL, 1985. [Chapter 13]

Introduction

The Classical Lyapunov Theorem

Theorem (Lyapunov 1892)

Let $A \in \mathbb{R}^{n \times n}$ and consider the Lyapunov operator $\mathcal{L} : X \to AX + XA^T$. Then the following are equivalent:

(a) $\forall Y > 0$: $\exists X > 0$: $\mathcal{L}(X) = -Y$,

(b) $\exists Y > 0$: $\exists X > 0$: $\mathcal{L}(X) = -Y$,

(c) $\Lambda(A) \subset \mathbb{C}^- := \{z \in \mathbb{C} \mid \Re z < 0\}$, *i.e., A is (asymptotically) stable or Hurwitz.*

Important in applications: the nonnegative case:

$$\mathcal{L}(X) = AX + XA^T = -WW^T, \quad \text{where} \quad W \in \mathbb{R}^{n \times n_W}, \; n_W \ll n.$$

A Hurwitz $\Rightarrow$ $\exists$ unique solution $X = ZZ^T$ for $Z \in \mathbb{R}^{n \times n_X}$ with $1 \le n_X \le n$.

A. M. Lyapunov. *The General Problem of the Stability of Motion* (in Russian). Doctoral dissertation, Univ. Kharkov 1892. English translation: Stability of Motion, Academic Press, New-York & London, 1966.

P. Lancaster, M. Tismenetsky. *The Theory of Matrices* (2nd edition). Academic Press, Orlando, FL, 1985. [Chapter 13]

Introduction

Existence of Solutions of Linear Matrix Equations II

For Lyapunov-plus-positive-type equations, the solution theory is more involved.

Introduction

Existence of Solutions of Linear Matrix Equations II

For Lyapunov-plus-positive-type equations, the solution theory is more involved. Focus on the Lyapunov-plus-positive case:

$$\underbrace{AX + XA^T}_{=:\mathcal{L}(X)} + \underbrace{\sum_{k=1}^{m} N_k X N_k^T}_{=:\mathcal{P}(X)} = C, \quad C = C^T \leq 0.$$

Note: The operator

$$\mathcal{P}(X) \mapsto \sum_{j=1}^{m} N_k X N_k^T$$

is nonnegative in the sense that $\mathcal{P}(X) \geq 0$, whenever $X \geq 0$.

Introduction

Existence of Solutions of Linear Matrix Equations II

For Lyapunov-plus-positive-type equations, the solution theory is more involved. Focus on the Lyapunov-plus-positive case:

$$\underbrace{AX + XA^T}_{=:\mathcal{L}(X)} + \underbrace{\sum_{k=1}^{m} N_k X N_k^T}_{=:\mathcal{P}(X)} = C, \quad C = C^T \leq 0.$$

Note: The operator

$$\mathcal{P}(X) \mapsto \sum_{j=1}^{m} N_k X N_k^T$$

is nonnegative in the sense that $\mathcal{P}(X) \geq 0$, whenever $X \geq 0$.

This nonnegative Lyapunov-plus-positive equation is the one occurring in applications like model order reduction.

Introduction

Existence of Solutions of Linear Matrix Equations II

For Lyapunov-plus-positive-type equations, the solution theory is more involved. Focus on the Lyapunov-plus-positive case:

$$\underbrace{AX + XA^T}_{=:\mathcal{L}(X)} + \underbrace{\sum_{k=1}^{m} N_k X N_k^T}_{=:\mathcal{P}(X)} = C, \quad C = C^T \leq 0.$$

Note: The operator

$$\mathcal{P}(X) \mapsto \sum_{j=1}^{m} N_k X N_k^T$$

is nonnegative in the sense that $\mathcal{P}(X) \geq 0$, whenever $X \geq 0$.

This nonnegative Lyapunov-plus-positive equation is the one occurring in applications like model order reduction.

If A is Hurwitz and the N_k are small enough, eigenvalue perturbation theory yields existence and uniqueness of solution.
This is related to the concept of bounded-input bounded-output (BIBO) stability of dynamical systems.

Introduction

Existence of Solutions of Linear Matrix Equations II

Theorem (Schneider 1965, Damm 2004)

Let $A \in \mathbb{R}^{n \times n}$ and consider the Lyapunov operator $\mathcal{L} : X \to AX + XA^T$ and a nonnegative operator $\mathcal{P}$ (i.e., $\mathcal{P}(X) \geq 0$ if $X \geq 0$).
The following are equivalent:

(a) $\forall Y > 0$: $\exists X > 0$: $\mathcal{L}(X) + \mathcal{P}(X) = -Y$,

(b) $\exists Y > 0$: $\exists X > 0$: $\mathcal{L}(X) + \mathcal{P}(X) = -Y$,

(c) $\exists Y \geq 0$ *with* (A, Y) *controllable:* $\exists X > 0$: $\mathcal{L}(X) + \mathcal{P}(X) = -Y$,

(d) $\Lambda(\mathcal{L} + \mathcal{P}) \subset \mathbb{C}^- := \{z \in \mathbb{C} \mid \Re z < 0\}$,

(e) $\Lambda(\mathcal{L}) \subset \mathbb{C}^-$ *and* $\rho(\mathcal{L}^{-1}\mathcal{P}) < 1$,

where $\rho(\mathcal{T}) = \max\{|\lambda| \mid \lambda \in \Lambda(\mathcal{T})\}$ = *spectral radius of* $\mathcal{T}$.

T. Damm. *Rational Matrix Equations in Stochastic Control.* Number 297 in Lecture Notes in Control and Information Sciences. Springer-Verlag, 2004.

H. Schneider. Positive operators and an inertia theorem. Numerische Mathematik, 7:11–17, 1965.

Overview

Applications

Stability Theory

From Lyapunov's theorem, immediately obtain characterization of asymptotic stability of linear dynamical systems

$$\dot{x}(t) = Ax(t). \tag{2}$$

Theorem (Lyapunov)

The following are equivalent:

- *For (2), the zero state is asymptotically stable.*
- *The Lyapunov equation $AX + XA^T = Y$ has a unique solution $X = X^T > 0$ for all $Y = Y^T < 0$.*
- *A is Hurwitz.*

A. M. Lyapunov. The General Problem of the Stability of Motion (In Russian). Doctoral dissertation, Univ. Kharkov 1892. English translation: Stability of Motion, Academic Press, New-York & London, 1966.

Classical Control Applications

Algebraic Riccati Equations (ARE)

Solving AREs by Newtons's Method

Feedback control design often involves solution of

$$A^T X + XA - XGX + H = 0, \quad G = G^T, H = H^T.$$

$\rightsquigarrow$ In each Newton step, solve Lyapunov equation

$$(A - GX_j)^T X_{j+1} + X_{j+1}(A - GX_j) = -X_j GX_j - H.$$

Classical Control Applications

Algebraic Riccati Equations (ARE)

Solving AREs by Newtons's Method

Feedback control design often involves solution of

$$A^T X + XA - XGX + H = 0, \quad G = G^T, H = H^T.$$

$\rightsquigarrow$ In each Newton step, solve Lyapunov equation

$$(A - GX_j)^T X_{j+1} + X_{j+1}(A - GX_j) = -X_j G X_j - H.$$

Decoupling of dynamical systems, e.g., in slow/fast modes, requires solution of nonsymmetric ARE

$$AX + XF - XGX + H = 0.$$

$\rightsquigarrow$ In each Newton step, solve Sylvester equation

$$(A - X_j G)X_{j+1} + X_{j+1}(F - GX_j) = -X_j G X_j - H.$$

Classical Control Applications

Model Reduction

Model Reduction via Balanced Truncation

For linear dynamical system

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx_r(t), \qquad x(t) \in \mathbb{R}^n$$

find reduced-order system

$$\dot{x}_r(t) = A_r x_r(t) + B_r u(t), \quad y_r(t) = C_r x_r(t), \qquad x(t) \in \mathbb{R}^r, \quad r \ll n$$

such that $\|y(t) - y_r(t)\| < \delta$.

The popular method balanced truncation requires the solution of the dual Lyapunov equations

$$AX + XA^T + BB^T = 0, \qquad A^T Y + YA + C^T C = 0.$$

Applications of Lyapunov-plus-Positive Equations

Bilinear control systems:

$$\Sigma : \quad \begin{cases} \dot{x}(t) = Ax(t) + \sum_{i=1}^{m} N_i x(t) u_i(t) + Bu(t), \\ y(t) = Cx(t), \quad x(0) = x_0, \end{cases}$$

where $A, N_i \in \mathbb{R}^{n\times n}, \ B \in \mathbb{R}^{n\times m}, \ C \in \mathbb{R}^{q\times n}$.

Properties:

- Approximation of (weakly) nonlinear systems by Carleman linearization yields bilinear systems.
- Appear naturally in boundary control problems, control via coefficients of PDEs, Fokker-Planck equations, ...
- Due to the close relation to linear systems, a lot of successful concepts can be extended, e.g. transfer functions, Gramians, Lyapunov equations, ...
- Linear stochastic control systems possess an equivalent structure and can be treated alike [B./Damm '11].

Applications of Lyapunov-plus-Positive Equations

The concept of balanced truncation can be generalized to the case of bilinear systems, where we need the solutions of the Lyapunov-plus-positive equations:

$$AP + PA^T + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0,$$

$$A^T Q + QA^T + \sum_{i=1}^{m} N_i^T Q A_i + C^T C = 0.$$

- Due to its approximation quality, balanced truncation is method of choice for model reduction of medium-size bilinear systems.
- For stationary iterative solvers, see [Damm 2008], extended to low-rank solutions recently by [Szyld/Shank/Simoncini 2014].

Applications of Lyapunov-plus-Positive Equations

The concept of balanced truncation can be generalized to the case of bilinear systems, where we need the solutions of the Lyapunov-plus-positive equations:

$$AP + PA^T + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0,$$

$$A^T Q + QA^T + \sum_{i=1}^{m} N_i^T Q A_i + C^T C = 0.$$

Further applications:

- Analysis and model reduction for linear stochastic control systems driven by Wiener noise [B./Damm 2011], Lévy processes [B./Redmann 2011/15].

Applications of Lyapunov-plus-Positive Equations

The concept of balanced truncation can be generalized to the case of bilinear systems, where we need the solutions of the Lyapunov-plus-positive equations:

$$AP + PA^T + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0,$$

$$A^T Q + QA^T + \sum_{i=1}^{m} N_i^T Q A_i + C^T C = 0.$$

Further applications:

- Analysis and model reduction for linear stochastic control systems driven by Wiener noise [B./Damm 2011], Lévy processes [B./Redmann 2011/15].
- Model reduction of linear parameter-varying (LPV) systems using bilinearization approach [B./Breiten 2011, B./Bruns 2015].

Applications of Lyapunov-plus-Positive Equations

The concept of balanced truncation can be generalized to the case of bilinear systems, where we need the solutions of the Lyapunov-plus-positive equations:

$$AP + PA^T + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0,$$

$$A^T Q + QA^T + \sum_{i=1}^{m} N_i^T Q A_i + C^T C = 0.$$

Further applications:

- Analysis and model reduction for linear stochastic control systems driven by Wiener noise [B./Damm 2011], Lévy processes [B./Redmann 2011/15].
- Model reduction of linear parameter-varying (LPV) systems using bilinearization approach [B./Breiten 2011, B./Bruns 2015].
- Model reduction for Fokker-Planck equations [Hartmann et al. 2013].

Applications of Lyapunov-plus-Positive Equations

The concept of balanced truncation can be generalized to the case of bilinear systems, where we need the solutions of the Lyapunov-plus-positive equations:

$$AP + PA^T + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0$$

Further applications:

- Analysis and model reduction for linear stochastic control systems driven by Wiener noise [B./Damm 2011], Lévy processes [B./Redmann 2011/15].
- Model reduction of linear parameter-varying (LPV) systems using bilinearization approach [B./Breiten 2011, B./Bruns 2015].
- Model reduction for Fokker-Planck equations [Hartmann et al. 2013].
- Linear-quadratic regulators for stochastic systems require solution of AREs of the form

$$AP + PA^T - XC^TCX + \sum_{i=1}^{m} N_i P A_i^T + BB^T = 0,$$

application of Newton's method $\rightsquigarrow$ 1 L-p-P equation/iteration.

Overview

This part: joint work with Patrick Kürschner and Jens Saak (MPI Magdeburg)

Solving Large-Scale Sylvester and Lyapunov Equations

The Low-Rank Structure

Sylvester Equations

Find $X \in \mathbb{R}^{n\times m}$ solving

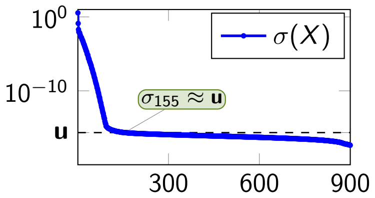
$$AX - XB = FG^T,$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$, $F \in \mathbb{R}^{n\times r}$, $G \in \mathbb{R}^{m\times r}$.

If n, m large, but $r \ll n, m$
$\rightsquigarrow$ X has a small numerical rank.
[Penzl 1999, Grasedyck 2004, Antoulas/Sorensen/Zhou 2002]

$$\operatorname{rank}(X, \tau) = f \ll \min(n, m)$$

singular values of 1600 × 900 example



$\rightsquigarrow$ Compute low-rank solution factors $Z \in \mathbb{R}^{n\times f}$, $Y \in \mathbb{R}^{m\times f}$, $D \in \mathbb{R}^{f\times f}$, such that $X \approx ZDY^T$ with $f \ll \min(n, m)$.

Solving Large-Scale Sylvester and Lyapunov Equations

The Low-Rank Structure

Lyapunov Equations

Find $X \in \mathbb{R}^{n \times n}$ solving

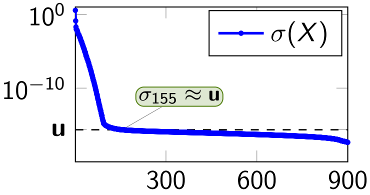
$$AX + XA^T = -FF^T,$$

where $A \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times r}$.

If n large, but $r \ll n$
$\rightsquigarrow$ X has a small numerical rank.
[Penzl 1999, Grasedyck 2004, Antoulas/Sorensen/Zhou 2002]

$$\mathrm{rank}(X, \tau) = f \ll n$$

singular values of 1600 × 900 example



$\rightsquigarrow$ Compute low-rank solution factors $Z \in \mathbb{R}^{n \times f}$, $D \in \mathbb{R}^{f \times f}$, such that $X \approx ZDZ^T$ with $f \ll n$.

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \operatorname{vec}(X) = \operatorname{vec}(FG^T).$$

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \mathrm{vec}(X) = \mathrm{vec}(FG^T).$$

This cannot be used for numerical solutions unless $nm \leq 1,000$ (or so), as

- it requires $\mathcal{O}(n^2 m^2)$ of storage;

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \operatorname{vec}(X) = \operatorname{vec}(FG^T).$$

This cannot be used for numerical solutions unless $nm \leq 1,000$ (or so), as

- it requires $\mathcal{O}(n^2 m^2)$ of storage;
- direct solver needs $\mathcal{O}(n^3 m^3)$ flops;

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \operatorname{vec}(X) = \operatorname{vec}(FG^T).$$

This cannot be used for numerical solutions unless $nm \leq 1,000$ (or so), as

- it requires $\mathcal{O}(n^2 m^2)$ of storage;
- direct solver needs $\mathcal{O}(n^3 m^3)$ flops;
- low (tensor-)rank of right-hand side is ignored;

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \mathrm{vec}(X) = \mathrm{vec}(FG^T).$$

This cannot be used for numerical solutions unless $nm \leq 1,000$ (or so), as

- it requires $\mathcal{O}(n^2 m^2)$ of storage;
- direct solver needs $\mathcal{O}(n^3 m^3)$ flops;
- low (tensor-)rank of right-hand side is ignored;
- in Lyapunov case, symmetry and possible definiteness are not respected.

Solving Large-Scale Sylvester and Lyapunov Equations

Some Basics

Sylvester equation $AX - XB = FG^T$ is equivalent to linear system of equations

$$\left(I_m \otimes A - B^T \otimes I_n\right) \operatorname{vec}(X) = \operatorname{vec}(FG^T).$$

This cannot be used for numerical solutions unless $nm \leq 1,000$ (or so), as

- it requires $\mathcal{O}(n^2m^2)$ of storage;
- direct solver needs $\mathcal{O}(n^3m^3)$ flops;
- low (tensor-)rank of right-hand side is ignored;
- in Lyapunov case, symmetry and possible definiteness are not respected.

Possible solvers:

- Standard Krylov subspace solvers in operator from [Hochbruck, Starke, Reichel, Bao, ...].
- Block-Tensor-Krylov subspace methods with truncation [Kressner/Tobler, Bollhöfer/Eppler, B./Breiten, ...].
- Galerkin-type methods based on (extended, rational) Krylov subspace methods [Jaimoukha, Kasenally, Jbilou, Simoncini, Druskin, Knizhermann,...]
- Doubling-type methods [Smith, Chu et al., B./Sadkane/El Khoury, ...].
- ADI methods [Wachspress, Reichel et al., Li, Penzl, B., Saak, Kürschner, ...].

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Sylvester and Stein equations

Let $\alpha \neq \beta$ with $\alpha \notin \Lambda(B)$, $\beta \notin \Lambda(A)$, then

$$\underbrace{AX - XB = FG^T}_{\text{Sylvester equation}} \quad \Leftrightarrow \quad \underbrace{X = \mathcal{A} X \mathcal{B} + (\beta - \alpha)\mathcal{F}\mathcal{G}^H}_{\text{Stein equation}}$$

with the Cayley like transformations

$$\begin{aligned} \mathcal{A} &:= (A - \beta I_n)^{-1}(A - \alpha I_n), & \mathcal{B} &:= (B - \alpha I_m)^{-1}(B - \beta I_m), \\ \mathcal{F} &:= (A - \beta I_n)^{-1}F, & \mathcal{G} &:= (B - \alpha I_m)^{-H}G. \end{aligned}$$

$\rightsquigarrow$ fix point iteration

$$X_k = \mathcal{A} X_{k-1} \mathcal{B} + (\beta - \alpha)\mathcal{F}\mathcal{G}^H$$

for $k \geq 1$, $X_0 \in \mathbb{R}^{n \times m}$.

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Sylvester and Stein equations

Let $\alpha_\mathbf{k} \neq \beta_\mathbf{k}$ with $\alpha_\mathbf{k} \notin \Lambda(B)$, $\beta_\mathbf{k} \notin \Lambda(A)$, then

$$\underbrace{AX - XB = FG^T}_{\text{Sylvester equation}} \quad \Leftrightarrow \quad \underbrace{X = \mathcal{A}_\mathbf{k} X \mathcal{B}_\mathbf{k} + (\beta_\mathbf{k} - \alpha_\mathbf{k})\mathcal{F}_\mathbf{k} {\mathcal{G}_\mathbf{k}}^H}_{\text{Stein equation}}$$

with the Cayley like transformations

$$\begin{aligned} \mathcal{A}_\mathbf{k} &:= (A - \beta_\mathbf{k} I_n)^{-1}(A - \alpha_\mathbf{k} I_n), & \mathcal{B}_\mathbf{k} &:= (B - \alpha_\mathbf{k} I_m)^{-1}(B - \beta_\mathbf{k} I_m), \\ \mathcal{F}_\mathbf{k} &:= (A - \beta_\mathbf{k} I_n)^{-1} F, & \mathcal{G}_\mathbf{k} &:= (B - \alpha_\mathbf{k} I_m)^{-H} G. \end{aligned}$$

$\rightsquigarrow$ **alternating directions implicit (ADI)** iteration

$$X_k = \mathcal{A}_\mathbf{k} X_{k-1} \mathcal{B}_\mathbf{k} + (\beta_\mathbf{k} - \alpha_\mathbf{k})\mathcal{F}_\mathbf{k} {\mathcal{G}^H}_\mathbf{k}$$

for $k \geq 1$, $X_0 \in \mathbb{R}^{n \times m}$. [Wachspress 1988]

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Sylvester ADI iteration [Wachspress 1988]

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H,$$
$$\mathcal{A}_k := (A - \beta_k I_n)^{-1}(A - \alpha_k I_n), \quad \mathcal{B}_k := (B - \alpha_k I_m)^{-1}(B - \beta_k I_m),$$
$$\mathcal{F}_k := (A - \beta_k I_n)^{-1} F \in \mathbb{R}^{n \times r}, \quad \mathcal{G}_k := (B - \alpha_k I_m)^{-H} G \in \mathbb{C}^{m \times r}.$$

Now set $X_0 = 0$ and find factorization $X_k = Z_k D_k Y_k^H$

$$X_1 = \mathcal{A}_1 X_0 \mathcal{B}_1 + (\beta_1 - \alpha_1) \mathcal{F}_1 \mathcal{G}_1^H$$

,

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Sylvester ADI iteration [Wachspress 1988]

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H,$$
$$\mathcal{A}_k := (A - \beta_k I_n)^{-1}(A - \alpha_k I_n), \quad \mathcal{B}_k := (B - \alpha_k I_m)^{-1}(B - \beta_k I_m),$$
$$\mathcal{F}_k := (A - \beta_k I_n)^{-1} F \in \mathbb{R}^{n \times r}, \quad \mathcal{G}_k := (B - \alpha_k I_m)^{-H} G \in \mathbb{C}^{m \times r}.$$

Now set $X_0 = 0$ and find factorization $X_k = Z_k D_k Y_k^H$

$$X_1 = (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1}$$
$$\Rightarrow V_1 := Z_1 = (A - \beta_1 I_n)^{-1} F \in \mathbb{R}^{n \times r},$$
$$D_1 = (\beta_1 - \alpha_1) I_r \in \mathbb{R}^{r \times r},$$
$$W_1 := Y_1 = (B - \alpha_1 I_m)^{-H} G \in \mathbb{C}^{m \times r}.$$

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Sylvester ADI iteration [Wachspress 1988]

$$\begin{aligned} X_k &= \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H, \\ \mathcal{A}_k &:= (A - \beta_k I_n)^{-1}(A - \alpha_k I_n), \quad \mathcal{B}_k := (B - \alpha_k I_m)^{-1}(B - \beta_k I_m), \\ \mathcal{F}_k &:= (A - \beta_k I_n)^{-1} F \in \mathbb{R}^{n \times r}, \quad \mathcal{G}_k := (B - \alpha_k I_m)^{-H} G \in \mathbb{C}^{m \times r}. \end{aligned}$$

Now set $X_0 = 0$ and find factorization $X_k = Z_k D_k Y_k^H$

$$\begin{aligned} X_2 &= \mathcal{A}_2 X_1 \mathcal{B}_2 + (\beta_2 - \alpha_2) \mathcal{F}_2 \mathcal{G}_2^H = \ldots = \\ V_2 &= V_1 + (\beta_2 - \alpha_1)(A + \beta_2 I)^{-1} V_1 \in \mathbb{R}^{n \times r}, \\ W_2 &= W_1 + \overline{(\alpha_2 - \beta_1)}(B + \alpha_2 I)^{-H} W_1 \in \mathbb{R}^{m \times r}, \\ Z_2 &= [Z_1, \ V_2], \\ D_2 &= \operatorname{diag}(D_1, (\beta_2 - \alpha_2) I_r), \\ Y_2 &= [Y_1, \ W_2]. \end{aligned}$$

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Algorithm

[B. 2005, Li/Truhar 2008, B./Li/Truhar 2009]

Algorithm 1: Low-rank Sylvester ADI / factored ADI (fADI)

Input : Matrices defining $AX - XB = FG^T$ and shift parameters $\{\alpha_1, \dots, \alpha_{k_{\max}}\}$, $\{\beta_1, \dots, \beta_{k_{\max}}\}$.

Output: Z, D, Y such that $ZDY^H \approx X$.

$Z_1 = V_1 = (A - \beta_1 I_n)^{-1} F$,

$Y_1 = W_1 = (B - \alpha_1 I_m)^{-H} G$.

$D_1 = (\beta_1 - \alpha_1) I_r$

for $k = 2, \dots, k_{\max}$ **do**

$V_k = V_{k-1} + (\beta_k - \alpha_{k-1})(A - \beta_k I_n)^{-1} V_{k-1}$.

$W_k = W_{k-1} + \overline{(\alpha_k - \beta_{k-1})}(B - \alpha_k I_n)^{-H} W_{k-1}$.

Update solution factors

$$Z_k = [Z_{k-1}, V_k], \quad Y_k = [Y_{k-1}, W_k], \quad D_k = \operatorname{diag}(D_{k-1}, (\beta_k - \alpha_k) I_r).$$

Solving Large-Scale Sylvester and Lyapunov Equations

ADI Shifts

Optimal Shifts

Solution of rational optimization problem

$$\min_{\substack{\alpha_j \in \mathbb{C} \\ \beta_j \in \mathbb{C}}} \max_{\substack{\lambda \in \Lambda(A) \\ \mu \in \Lambda(B)}} \prod_{j=1}^{k} \left| \frac{(\lambda - \alpha_j)(\mu - \beta_j)}{(\lambda - \beta_j)(\mu - \alpha_j)} \right|,$$

for which no analytic solution is known in general.

Some shift generation approaches:

- generalized Bagby points, [Levenberg/Reichel 1993]
- adaption of Penzl's cheap heuristic approach available [Penzl 1999, Li/Truhar 2008]
  $\rightsquigarrow$ approximate $\Lambda(A)$, $\Lambda(B)$ by small number of Ritz values w.r.t. A, A^{-1}, B, B^{-1} via Arnoldi,
- just taking these Ritz values alone also works well quite often.

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Disadvantages of Low-Rank ADI as of 2012:

1. No efficient stopping criteria:
   - Difference in iterates $\rightsquigarrow$ norm of added columns/step: not reliable, stops often too late.
   - Residual is a full dense matrix, can not be calculated as such.
2. Requires complex arithmetic for real coefficients when complex shifts are used.
3. Expensive (only semi-automatic) set-up phase to precompute ADI shifts.

Solving Large-Scale Sylvester and Lyapunov Equations

LR-ADI Derivation

Disadvantages of Low-Rank ADI as of 2012:

1. No efficient stopping criteria:
   - Difference in iterates $\rightsquigarrow$ norm of added columns/step: not reliable, stops often too late.
   - Residual is a full dense matrix, can not be calculated as such.
2. Requires complex arithmetic for real coefficients when complex shifts are used.
3. Expensive (only semi-automatic) set-up phase to precompute ADI shifts.

None of these disadvantages exists as of today
$\Longrightarrow$ speed-ups old vs. new LR-ADI can be up to 20!

Projection-Based Lyapunov Solvers. . .

. . . for Lyapunov equation $0 = AX + XA^T + BB^T$

Projection-based methods for Lyapunov equations with $A + A^T < 0$:

1. Compute orthonormal basis $\operatorname{range}(Z)$, $Z \in \mathbb{R}^{n \times r}$, for subspace $\mathcal{Z} \subset \mathbb{R}^n$, $\dim \mathcal{Z} = r$.
2. Set $\hat{A} := Z^T A Z$, $\hat{B} := Z^T B$.
3. Solve small-size Lyapunov equation $\hat{A}\hat{X} + \hat{X}\hat{A}^T + \hat{B}\hat{B}^T = 0$.
4. Use $X \approx Z\hat{X}Z^T$.

Projection-Based Lyapunov Solvers. . .

. . . for Lyapunov equation $0 = AX + XA^T + BB^T$

Projection-based methods for Lyapunov equations with $A + A^T < 0$:

1. Compute orthonormal basis $\mathrm{range}\,(Z)$, $Z \in \mathbb{R}^{n \times r}$, for subspace $\mathcal{Z} \subset \mathbb{R}^n$, $\dim \mathcal{Z} = r$.
2. Set $\hat{A} := Z^T AZ$, $\hat{B} := Z^T B$.
3. Solve small-size Lyapunov equation $\hat{A}\hat{X} + \hat{X}\hat{A}^T + \hat{B}\hat{B}^T = 0$.
4. Use $X \approx Z\hat{X}Z^T$.

Examples:

- Krylov subspace methods, i.e., for $m = 1$:

$$\mathcal{Z} = \mathcal{K}(A, B, r) = \mathrm{span}\{B, AB, A^2B, \ldots, A^{r-1}B\}$$

  [Saad 1990, Jaimoukha/Kasenally 1994, Jbilou 2002–2008].

- Extended Krylov subspace method (EKSM) [Simoncini 2007],

$$\mathcal{Z} = \mathcal{K}(A, B, r) \cup \mathcal{K}(A^{-1}, B, r).$$

- Rational Krylov subspace methods (RKSM) [Druskin/Simoncini 2011].

The New LR-ADI Applied to Lyapunov Equations

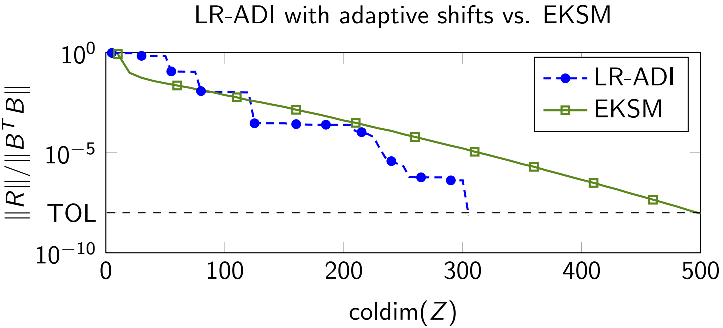
Example: an ocean circulation problem [Van Gijzen et al. 1998]

- FEM discretization of a simple 3D ocean circulation model (barotropic, constant depth) $\rightsquigarrow$ stiffness matrix $-A$ with $n = 42,249$, choose artificial constant term $B =$ `rand(n,5)`.

The New LR-ADI Applied to Lyapunov Equations

Example: an ocean circulation problem [Van Gijzen et al. 1998]

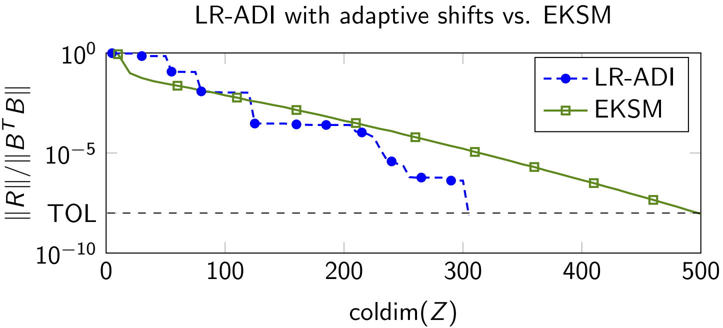
- FEM discretization of a simple 3D ocean circulation model (barotropic, constant depth) $\rightsquigarrow$ stiffness matrix $-A$ with $n = 42,249$, choose artificial constant term $B = \texttt{rand(n,5)}$.
- **Convergence history:**



The New LR-ADI Applied to Lyapunov Equations

Example: an ocean circulation problem [Van Gijzen et al. 1998]

- FEM discretization of a simple 3D ocean circulation model (barotropic, constant depth) $\rightsquigarrow$ stiffness matrix $-A$ with $n = 42,249$, choose artificial constant term $B = \texttt{rand(n,5)}$.
- **Convergence history:**



- CPU times: LR-ADI $\approx$ 110 sec, EKSM $\approx$ 135 sec.

Solving Large-Scale Sylvester and Lyapunov Equations

Summary & Outlook

- Numerical enhancements of low-rank ADI for large Sylvester/Lyapunov equations:
  1. low-rank residuals, reformulated implementation,
  2. compute real low-rank factors in the presence of complex shifts,
  3. self-generating shift strategies (quantification in progress).

For diffusion-convection-reaction example:
332.02 sec. down to **17.24 sec.** $\rightsquigarrow$ acceleration by factor almost **20**.

- Generalized version enables derivation of low-rank solvers for various generalized Sylvester equations.
- Ongoing work:
  - Apply LR-ADI in Newton methods for algebraic Riccati equations

$$\mathcal{R}(X) = AX + XA^T + GG^T - XSS^TX = 0,$$
$$\mathcal{D}(X) = AXA^T - EXE^T + GG^T + A^TXF(I_r + F^TXF)^{-1}F^TXA = 0.$$

For nonlinear AREs see

P. Benner, P. Kürschner, J. Saak. *Low-rank Newton-ADI methods for large nonsymmetric algebraic Riccati equations.* J. Franklin Inst., 2015.

Overview

This part: joint work with Tobias Breiten (KFU Graz, Austria)

Solving Large-Scale Lyapunov-plus-Positive Equations

Some basic facts and assumptions

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0. \qquad (3)$$

- Need a positive semi-definite symmetric solution X.

Solving Large-Scale Lyapunov-plus-Positive Equations

Some basic facts and assumptions

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0. \tag{3}$$

- Need a positive semi-definite symmetric solution X.
- As discussed before, solution theory for Lyapuonv-plus-positive equation is more involved then in standard Lyapuonv case.
  Here, existence and uniqueness of positive semi-definite solution $X = X^T$ is assumed.

Solving Large-Scale Lyapunov-plus-Positive Equations

Some basic facts and assumptions

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0. \tag{3}$$

- Need a positive semi-definite symmetric solution X.
- As discussed before, solution theory for Lyapuonv-plus-positive equation is more involved then in standard Lyapuonv case. Here, existence and uniqueness of positive semi-definite solution $X = X^T$ is assumed.
- Want: solution methods for large scale problems, i.e., only matrix-matrix multiplication with A, N_j, solves with (shifted) A allowed!

Solving Large-Scale Lyapunov-plus-Positive Equations

Some basic facts and assumptions

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0. \tag{3}$$

- Need a positive semi-definite symmetric solution X.
- As discussed before, solution theory for Lyapuonv-plus-positive equation is more involved then in standard Lyapuonv case. Here, existence and uniqueness of positive semi-definite solution $X = X^T$ is assumed.
- Want: solution methods for large scale problems, i.e., only matrix-matrix multiplication with A, N_j, solves with (shifted) A allowed!
- Requires to compute data-sparse approximation to generally dense X; here: $X \approx ZZ^T$ with $Z \in \mathbb{R}^{n \times n_Z}$, $n_Z \ll n$!

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \ ?$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \text{ ?}$$

Standard Lyapunov case: [Grasedyck '04]

$$AX + XA^T + BB^T = 0 \iff \underbrace{(I_n \otimes A + A \otimes I_n)}_{=:\mathcal{A}} \operatorname{vec}(X) = -\operatorname{vec}(BB^T).$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Standard Lyapunov case: [Grasedyck '04]

$$AX + XA^T + BB^T = 0 \iff \underbrace{(I_n \otimes A + A \otimes I_n)}_{=:\mathcal{A}} \operatorname{vec}(X) = -\operatorname{vec}(BB^T).$$

Apply

$$M^{-1} = -\int_0^\infty \exp(tM)\mathrm{d}t$$

to $\mathcal{A}$ and approximate the integral via (sinc) quadrature $\Rightarrow$

$$\mathcal{A}^{-1} \approx -\sum_{i=-k}^{k} \omega_i \exp(t_k \mathcal{A}),$$

with error $\sim \exp(-\sqrt{k})$ ($\exp(-k)$ if $A = A^T$), then an approximate Lyapunov solution is given by

$$\operatorname{vec}(X) \approx \operatorname{vec}(X_k) = \sum_{i=-k}^{k} \omega_i \exp(t_i \mathcal{A}) \operatorname{vec}(BB^T).$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Standard Lyapunov case: [Grasedyck '04]

$$AX + XA^T + BB^T = 0 \iff \underbrace{(I_n \otimes A + A \otimes I_n)}_{=:\mathcal{A}} \operatorname{vec}(X) = -\operatorname{vec}(BB^T).$$

$$\operatorname{vec}(X) \approx \operatorname{vec}(X_k) = \sum_{i=-k}^{k} \omega_i \exp(t_i \mathcal{A}) \operatorname{vec}(BB^T).$$

Now observe that

$$\exp(t_i \mathcal{A}) = \exp\left(t_i (I_n \otimes A + A \otimes I_n)\right) \equiv \exp(t_i A) \otimes \exp(t_i A).$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Standard Lyapunov case: [Grasedyck '04]

$$AX + XA^T + BB^T = 0 \iff \underbrace{(I_n \otimes A + A \otimes I_n)}_{=:\mathcal{A}} \operatorname{vec}(X) = -\operatorname{vec}(BB^T).$$

$$\operatorname{vec}(X) \approx \operatorname{vec}(X_k) = \sum_{i=-k}^{k} \omega_i \exp(t_i \mathcal{A}) \operatorname{vec}(BB^T).$$

Now observe that

$$\exp(t_i \mathcal{A}) = \exp\left(t_i (I_n \otimes A + A \otimes I_n)\right) \equiv \exp(t_i A) \otimes \exp(t_i A).$$

Hence,

$$\operatorname{vec}(X_k) = \sum_{i=-k}^{k} \omega_i \left(\exp(t_i A) \otimes \exp(t_i A)\right) \operatorname{vec}(BB^T)$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Standard Lyapunov case: [Grasedyck '04]

$$AX + XA^T + BB^T = 0 \iff \underbrace{(I_n \otimes A + A \otimes I_n)}_{=:\mathcal{A}} \operatorname{vec}(X) = -\operatorname{vec}(BB^T).$$

Hence,

$$\begin{aligned} \operatorname{vec}(X_k) &= \sum_{i=-k}^{k} \omega_i \left(\exp(t_i A) \otimes \exp(t_i A)\right) \operatorname{vec}(BB^T) \\ \implies X_k &= \sum_{i=-k}^{k} \omega_i \exp(t_i A) B B^T \exp(t_i A^T) \equiv \sum_{i=-k}^{k} \omega_i B_i B_i^T, \end{aligned}$$

so that $\operatorname{rank}(X_k) \leq (2k+1)m$ with

$$\|X - X_k\|_2 \lesssim \exp(-\sqrt{k}) \qquad (\ \exp(-k) \text{ for } A = A^T\ )!$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \text{ ?}$$

Problem: in general,

$$\exp\left(t_i(I \otimes A + A \otimes I + \sum_{j=1}^{m} N_j \otimes N_j)\right) \neq (\exp(t_i A) \otimes \exp(t_i A)) \exp\left(t_i(\sum_{j=1}^{m} N_j \otimes N_j)\right).$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \ ?$$

Assume that $m = 1$ and $N_1 = UV^T$ with $U, V \in \mathbb{R}^{n \times r}$ and consider

$$(\underbrace{I_n \otimes A + A \otimes I_n}_{=\mathcal{A}} + N_1 \otimes N_1)\operatorname{vec}(X) = \underbrace{-\operatorname{vec}(BB^T)}_{=:y}.$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \text{ ?}$$

Assume that $m = 1$ and $N_1 = UV^T$ with $U, V \in \mathbb{R}^{n\times r}$ and consider

$$(\underbrace{I_n \otimes A + A \otimes I_n}_{=\mathcal{A}} + N_1 \otimes N_1)\,\mathrm{vec}(X) = \underbrace{-\,\mathrm{vec}(BB^T)}_{=:y}.$$

Sherman-Morrison-Woodbury $\Longrightarrow$

$$\begin{aligned} \left(I_r \otimes I_r + (V^T \otimes V^T)\mathcal{A}^{-1}(U \otimes U)\right) w &= (V^T \otimes V^T)\mathcal{A}^{-1} y, \\ \mathcal{A}\,\mathrm{vec}(X) &= y - (U \otimes U) w. \end{aligned}$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Question

Can we expect low-rank approximations $ZZ^T \approx X$ to the solution of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0 \ ?$$

Assume that $m = 1$ and $N_1 = UV^T$ with $U, V \in \mathbb{R}^{n \times r}$ and consider

$$(\underbrace{I_n \otimes A + A \otimes I_n}_{=\mathcal{A}} + N_1 \otimes N_1)\,\mathrm{vec}(X) = \underbrace{-\,\mathrm{vec}(BB^T)}_{=:y}.$$

Sherman-Morrison-Woodbury $\Longrightarrow$

$$\begin{aligned}\left(I_r \otimes I_r + (V^T \otimes V^T)\mathcal{A}^{-1}(U \otimes U)\right) w &= (V^T \otimes V^T)\mathcal{A}^{-1} y,\\ \mathcal{A}\,\mathrm{vec}(X) &= y - (U \otimes U)w.\end{aligned}$$

Matrix rank of RHS $-BB^T - U\,\mathrm{vec}^{-1}(w)\,U^T$ is $\leq\ r + 1$!

$\rightsquigarrow$ Apply results for linear Lyapunov equations with r.h.s of rank $r + 1$.

Solving Large-Scale Lyapunov-plus-Positive Equations

Existence of Low-Rank Approximations

Theorem [B./Breiten 2012]

Assume existence and uniqueness with stable A and $N_j = U_j V_j^T$, with $U_j, V_j \in \mathbb{R}^{n \times r_j}$. Set $r = \sum_{j=1}^m r_j$.
Then the solution X of

$$AX + XA^T + \sum_{j=1}^{m} N_j X N_j^T + BB^T = 0$$

can be approximated by X_k of rank $(2k+1)(m+r)$, with an error satisfying

$$\|X - X_k\|_2 \lesssim \exp(-\sqrt{k}).$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Let us again consider the Lyapunov-plus-positive equation

$$AP + PA^T + NPN^T + BB^T = 0.$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Let us again consider the Lyapunov-plus-positive equation

$$AP + PA^T + NPN^T + BB^T = 0.$$

For a fixed parameter p, we can rewrite the linear Lyapunov operator as

$$AP + PA^T = \frac{1}{2p}\left((A + pI)P(A + pI)^T - (A - pI)P(A - pI)^T\right)$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Let us again consider the Lyapunov-plus-positive equation

$$AP + PA^T + NPN^T + BB^T = 0.$$

For a fixed parameter p, we can rewrite the linear Lyapunov operator as

$$AP + PA^T = \frac{1}{2p}\left((A + pI)P(A + pI)^T - (A - pI)P(A - pI)^T\right)$$

leading to the fix point iteration [Damm 2008]

$$\begin{aligned} P_j = &(A - pI)^{-1}(A + pI)P_{j-1}(A + pI)^T(A - pI)^{-T} \\ &+ 2p(A - pI)^{-1}(NP_{j-1}N^T + BB^T)(A - pI)^{-T}. \end{aligned}$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Let us again consider the Lyapunov-plus-positive equation

$$AP + PA^T + NPN^T + BB^T = 0.$$

For a fixed parameter p, we can rewrite the linear Lyapunov operator as

$$AP + PA^T = \frac{1}{2p}\left((A+pI)P(A+pI)^T - (A-pI)P(A-pI)^T\right)$$

leading to the fix point iteration [Damm 2008]

$$\begin{aligned} P_j = {} & (A-pI)^{-1}(A+pI)P_{j-1}(A+pI)^T(A-pI)^{-T} \\ & + 2p(A-pI)^{-1}(NP_{j-1}N^T + BB^T)(A-pI)^{-T}. \end{aligned}$$

$P_j \approx Z_j Z_j^T$ $(\mathrm{rank}\,(Z_j) \ll n)$ $\rightsquigarrow$ factored iteration

$$\begin{aligned} Z_j Z_j^T = {} & (A-pI)^{-1}(A+pI)Z_{j-1}Z_{j-1}^T(A+pI)^T(A-pI)^{-T} \\ & + 2p(A-pI)^{-1}(NZ_{j-1}Z_{j-1}^T N^T + BB^T)(A-pI)^{-T}. \end{aligned}$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Hence, for a given sequence of shift parameters $\{p_1, \ldots, p_q\}$, we can extend the linear ADI iteration as follows:

$$
\begin{aligned}
Z_1 &= \sqrt{2p_1}\,(A - p_1 I)^{-1}\, B,\\
Z_j &= (A - p_j I)^{-1} \left[(A + p_j I)\, Z_{j-1} \quad \sqrt{2p_j} B \quad \sqrt{2p_j} N Z_{j-1}\right], \quad j \leq q.
\end{aligned}
$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalized ADI Iteration

Hence, for a given sequence of shift parameters $\{p_1, \dots, p_q\}$, we can extend the linear ADI iteration as follows:

$$
\begin{aligned}
Z_1 &= \sqrt{2p_1}\,(A - p_1 I)^{-1} B, \\
Z_j &= (A - p_j I)^{-1} \left[(A + p_j I)\, Z_{j-1} \quad \sqrt{2p_j} B \quad \sqrt{2p_j} N Z_{j-1} \right], \quad j \le q.
\end{aligned}
$$

Problems:

- A and N in general do not commute $\rightsquigarrow$ we have to operate on full preceding subspace Z_{j-1} in each step.
- Rapid increase of $\operatorname{rank}(Z_j)$ $\rightsquigarrow$ perform some kind of column compression.
- Choice of shift parameters? $\rightsquigarrow$ No obvious generalization of minimax problem.
  Here, we will use shifts minimizing a certain $\mathcal{H}_2$-optimization problem, see [B./Breiten 2011/14].

Generalized ADI Iteration

Numerical Example: A Heat Transfer Model with Uncertainty

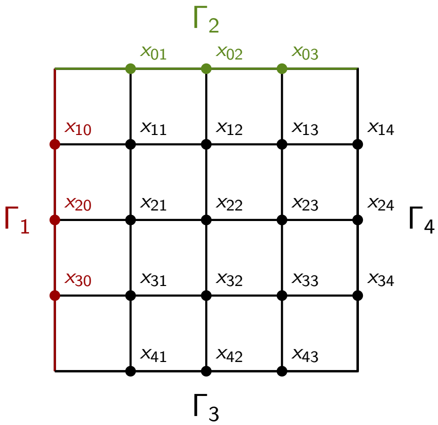
- 2-dimensional heat distribution motivated by [Benner/Saak '05]
- boundary control by a cooling fluid with an uncertain spraying intensity

$$
\begin{aligned}
\Omega &= (0,1) \times (0,1) & \\
x_t &= \Delta x & \text{in } \Omega \\
n \cdot \nabla x &= (0.5 + d\omega_1)x & \text{on } \Gamma_1 \\
x &= u & \text{on } \Gamma_2 \\
x &= 0 & \text{on } \Gamma_3, \Gamma_4
\end{aligned}
$$

- spatial discretization $k \times k$-grid

$$\Rightarrow dx \approx Axdt + Nxd\omega_i + Budt$$

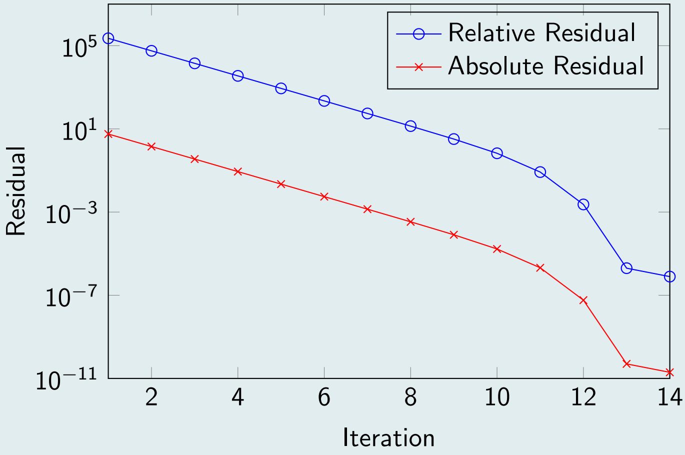
- output: $C = \frac{1}{k^2}\begin{bmatrix}1 & \dots & 1\end{bmatrix}$



Generalized ADI Iteration

Numerical Example: A Heat Transfer Model with Uncertainty

Conv. history for bilinear low-rank ADI method ($n = 40,000$)



Solving Large-Scale Lyapunov-plus-Positive Equations

Generalizing the Extended Krylov Subspace Method (EKSM) [Simoncini '07]

Low-rank solutions of the Lyapunov-plus-positive equation may be obtained by projecting the original equation onto a suitable smaller subspace $\mathcal{V} = \operatorname{span}(V)$, $V \in \mathbb{R}^{n\times k}$, with $V^T V = I$.

In more detail, solve

$$\left(V^T A V\right)\hat{X} + \hat{X}\left(V^T A^T V\right) + \left(V^T N V\right)\hat{X}\left(V^T N^T V\right) + \left(V^T B\right)\left(V^T B\right)^T = 0$$

and prolongate $X \approx V\hat{X}V^T$.

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalizing the Extended Krylov Subspace Method (EKSM) [Simoncini '07]

Low-rank solutions of the Lyapunov-plus-positive equation may be obtained by projecting the original equation onto a suitable smaller subspace $\mathcal{V} = \mathrm{span}(V)$, $V \in \mathbb{R}^{n \times k}$, with $V^T V = I$.

In more detail, solve

$$\left(V^T A V\right) \hat{X} + \hat{X} \left(V^T A^T V\right) + \left(V^T N V\right) \hat{X} \left(V^T N^T V\right) + \left(V^T B\right) \left(V^T B\right)^T = 0$$

and prolongate $X \approx V \hat{X} V^T$.

For this, one might use the extended Krylov subspace method (EKSM) algorithm in the following way:

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalizing the Extended Krylov Subspace Method (EKSM) [SIMONCINI '07]

Low-rank solutions of the Lyapunov-plus-positive equation may be obtained by projecting the original equation onto a suitable smaller subspace $\mathcal{V} = \operatorname{span}(V)$, $V \in \mathbb{R}^{n\times k}$, with $V^TV = I$.

In more detail, solve

$$\left(V^TAV\right)\hat{X} + \hat{X}\left(V^TA^TV\right) + \left(V^TNV\right)\hat{X}\left(V^TN^TV\right) + \left(V^TB\right)\left(V^TB\right)^T = 0$$

and prolongate $X \approx V\hat{X}V^T$.

For this, one might use the extended Krylov subspace method (EKSM) algorithm in the following way:

$$\begin{aligned} V_1 &= \begin{bmatrix} B & A^{-1}B \end{bmatrix}, \\ V_r &= \begin{bmatrix} AV_{r-1} & A^{-1}V_{r-1} & NV_{r-1} \end{bmatrix}, \quad r = 2,3,\ldots \end{aligned}$$

Solving Large-Scale Lyapunov-plus-Positive Equations

Generalizing the Extended Krylov Subspace Method (EKSM) [Simoncini '07]

Low-rank solutions of the Lyapunov-plus-positive equation may be obtained by projecting the original equation onto a suitable smaller subspace $\mathcal{V} = \operatorname{span}(V)$, $V \in \mathbb{R}^{n \times k}$, with $V^T V = I$.

In more detail, solve

$$\left(V^T A V\right) \hat{X} + \hat{X} \left(V^T A^T V\right) + \left(V^T N V\right) \hat{X} \left(V^T N^T V\right) + \left(V^T B\right) \left(V^T B\right)^T = 0$$

and prolongate $X \approx V \hat{X} V^T$.

For this, one might use the extended Krylov subspace method (EKSM) algorithm in the following way:

$$\begin{aligned} V_1 &= \begin{bmatrix} B & A^{-1}B \end{bmatrix}, \\ V_r &= \begin{bmatrix} AV_{r-1} & A^{-1}V_{r-1} & NV_{r-1} \end{bmatrix}, \quad r = 2, 3, \ldots \end{aligned}$$

However, criteria like dissipativity of A for the linear case which ensure solvability of the projected equation have to be further investigated.

Bilinear EKSM

Residual Computation in $\mathcal{O}(k^3)$

Theorem (B./Breiten 2012)

Let $V_i \in \mathbb{R}^{n \times k_i}$ be the extend Krylov matrix after i generalized EKSM steps. Denote the residual associated with the approximate solution $X_i = V_i \hat{X}_i V_i^T$ by

$$R_i := AX_i + X_i A^T + NX_i N^T + BB^T,$$

where $\hat{X}_i$ is the solution of the reduced Lyapunov-plus-positive equation

$$V_i^T A V_i \hat{X}_i + \hat{X}_i V_i^T A^T V_i + V_i^T N V_i \hat{X}_i V_i^T N^T V_i + V_i^T BB^T V_i = 0.$$

Then:

- $\mathrm{range}\,(R_i) \subset \mathrm{range}\,(V_{i+1})$,
- $\|R_i\| = \|V_{i+1}^T R_i V_{i+1}\|$ *for the Frobenius and spectral norms.*

Bilinear EKSM

Residual Computation in $\mathcal{O}(k^3)$

Theorem (B./Breiten 2012)

Let $V_i \in \mathbb{R}^{n \times k_i}$ be the extend Krylov matrix after i generalized EKSM steps. Denote the residual associated with the approximate solution $X_i = V_i \hat{X}_i V_i^T$ by

$$R_i := AX_i + X_i A^T + NX_i N^T + BB^T,$$

where $\hat{X}_i$ is the solution of the reduced Lyapunov-plus-positive equation

$$V_i^T A V_i \hat{X}_i + \hat{X}_i V_i^T A^T V_i + V_i^T N V_i \hat{X}_i V_i^T N^T V_i + V_i^T BB^T V_i = 0.$$

Then:

- $\operatorname{range}(R_i) \subset \operatorname{range}(V_{i+1})$,
- $\|R_i\| = \|V_{i+1}^T R_i V_{i+1}\|$ *for the Frobenius and spectral norms.*

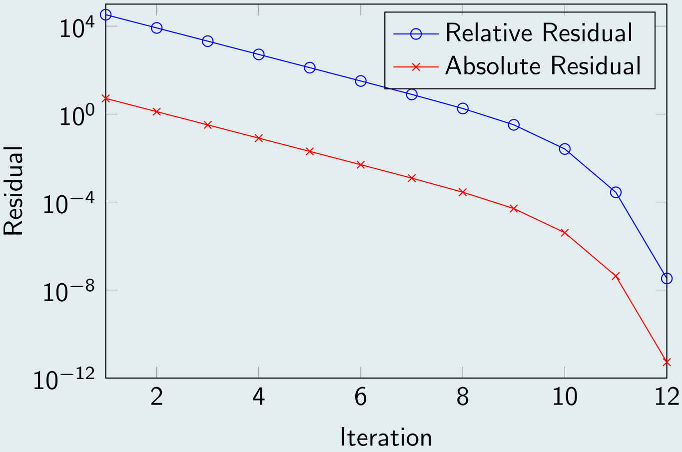
Remarks:

- Residual evaluation only requires quantities needed in $i+1$st projection step plus $\mathcal{O}(k_{i+1}^3)$ operations.
- No Hessenberg structure of reduced system matrix that allows to simplify residual expression as in standard Lyapunov case!

Bilinear EKSM

Numerical Example: A Heat Transfer Model with Uncertainty

Convergence history for bilinear EKSM variant ($n = 6,400$)



Solving Large-Scale Lyapunov-plus-Positive Equations

Tensorized Krylov Subspace Methods

Another possibility is to iteratively solve the linear system

$$(I_n \otimes A + A \otimes I_n + N \otimes N)\,\mathrm{vec}(X) = -\,\mathrm{vec}(BB^T),$$

with a fixed number of ADI iteration steps used as a preconditioner $\mathcal{M}$

$$\mathcal{M}^{-1}(I_n \otimes A + A \otimes I_n + N \otimes N)\,\mathrm{vec}(X) = -\mathcal{M}^{-1}\,\mathrm{vec}(BB^T).$$

We implemented this approach for PCG and BiCGstab.

Updates like $X_{k+1} \leftarrow X_k + \omega_k P_k$ require truncation operator to preserve low-order structure.

Note, that the low-rank factorization $X \approx ZZ^T$ has to be replaced by $X \approx ZDZ^T$, D possibly indefinite.

Similar to more general tensorized Krylov solvers, see [Kressner/Tobler 2010/12].

Tensorized Krylov Subspace Methods

Vanilla Implementation of Tensor-PCG for Matrix Equations

Algorithm 2: Preconditioned CG method for $\mathcal{A}(X) = \mathcal{B}$

Input : Matrix functions $\mathcal{A}, \mathcal{M} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, low rank factor B of right-hand side $\mathcal{B} = -BB^T$. Truncation operator $\mathcal{T}$ w.r.t. relative accuracy ϵ_{rel}.

Output: Low rank approximation $X = LDL^T$ with $||\mathcal{A}(X) - \mathcal{B}||_F \leq$ tol.

$X_0 = 0,\ R_0 = \mathcal{B},\ Z_0 = \mathcal{M}^{-1}(R_0),\ P_0 = Z_0,\ Q_0 = \mathcal{A}(P_0),\ \xi_0 = \langle P_0, Q_0\rangle, k = 0$

while $||R_k||_F > tol$ **do**

- $\omega_k = \frac{\langle R_k, P_k\rangle}{\xi_k}$
- $X_{k+1} = X_k + \omega_k P_k, \quad X_{k+1} \leftarrow \mathcal{T}(X_{k+1})$
- $R_{k+1} = \mathcal{B} - \mathcal{A}(X_{k+1}), \quad$ *Optionally:* $R_{k+1} \leftarrow \mathcal{T}(R_{k+1})$
- $Z_{k+1} = \mathcal{M}^{-1}(R_{k+1})$
- $\beta_k = -\frac{\langle Z_{k+1}, Q_k\rangle}{\xi_k}$
- $P_{k+1} = Z_{k+1} + \beta_k P_k, \quad P_{k+1} \leftarrow \mathcal{T}(P_{k+1})$
- $Q_{k+1} = \mathcal{A}(P_{k+1}), \quad$ *Optionally:* $Q_{k+1} \leftarrow \mathcal{T}(Q_{k+1})$
- $\xi_{k+1} = \langle P_{k+1}, Q_{k+1}\rangle$
- $k = k + 1$

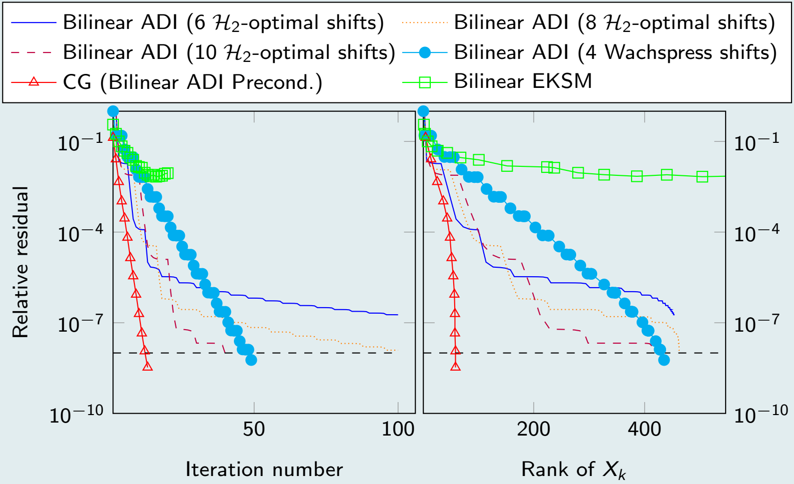
$X = X_k$

Here, $\mathcal{A} : X \to AX + XA^T + NXN^T$, $\mathcal{M}$: ℓ steps of (bilinear) ADI, both in low-rank ("ZDZ^T" format).

Comparison of Methods

Heat Equation with Boundary Control

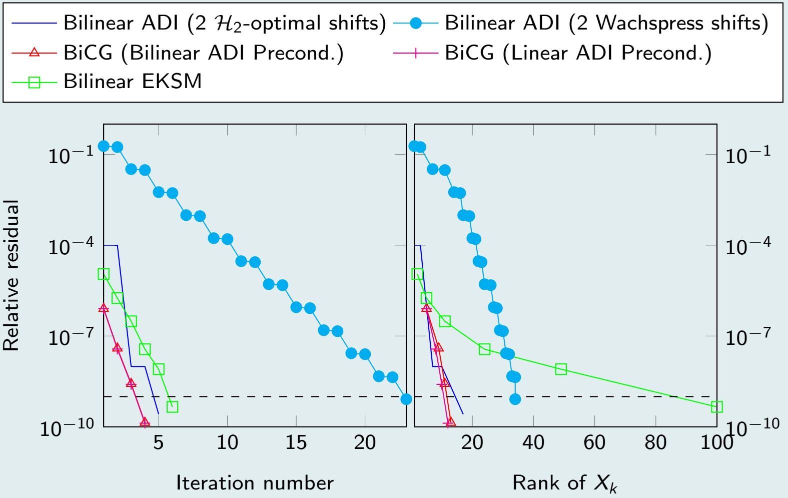
Comparison of low rank solution methods for $n = 562,500$.



Comparison of Methods

Fokker-Planck Equation

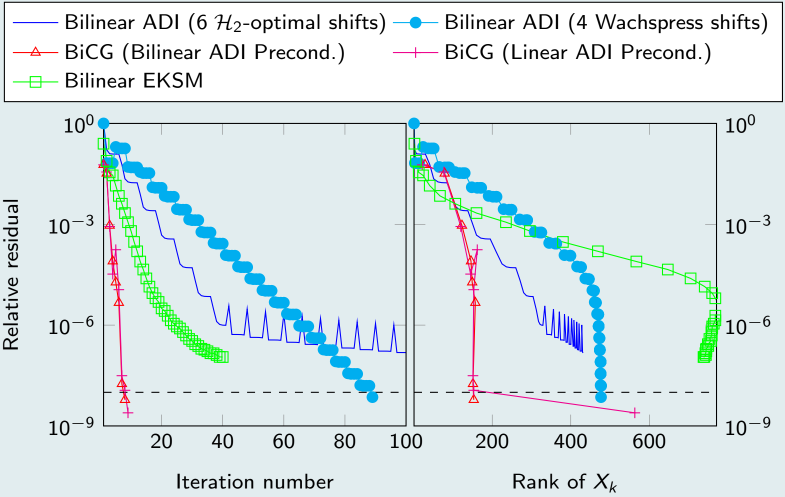
Comparison of low rank solution methods for $n = 10,000$.



Comparison of Methods

RC Circuit Simulation

Comparison of low rank solution methods for $n = 250,000$.



Comparison of Methods

Comparison of CPU times

	Heat equation	RC circuit	Fokker-Planck
Bilin. ADI 2 $\mathcal{H}_2$ shifts	-	-	1.733 (1.578)
Bilin. ADI 6 $\mathcal{H}_2$ shifts	144,065 (2,274)	20,900 (3091)	-
Bilin. ADI 8 $\mathcal{H}_2$ shifts	135,711 (3,177)	-	-
Bilin. ADI 10 $\mathcal{H}_2$ shifts	33,051 (4,652)	-	-
Bilin. ADI 2 Wachspress shifts	-	-	6.617 (4.562)
Bilin. ADI 4 Wachspress shifts	41,883 (2,500)	18,046 (308)	-
CG (Bilin. ADI precond.)	15,640	-	-
BiCG (Bilin. ADI precond.)	-	16,131	11.581
BiCG (Linear ADI precond.)	-	12,652	9.680
EKSM	7,093	19,778	8.555

Numbers in brackets: computation of shift parameters.

Solving Large-Scale Lyapunov-plus-Positive Equations

Summary & Outlook

- Under certain assumptions, we can expect the existence of low-rank approximations to the solution of Lyapunov-plus-positive equations.
- Solutions strategies via extending the ADI iteration to bilinear systems and EKSM as well as using preconditioned iterative solvers like CG or BiCGstab up to dimensions $n \sim 500,000$ in MATLAB®.
- Optimal choice of shift parameters for ADI is a nontrivial task.
- Other "tricks" (realification, low-rank residuals) not adapted from standard case so far.
- What about the singular value decomposition decay in case of N being full rank?
- Need efficient implementation!

Further Reading

P. Benner and T. Breiten.
On $\mathcal{H}_2$ model reduction of linear parameter-varying systems.
Proceedings in Applied Mathematics and Mechanics 11:805-806, 2011.

P. Benner and T. Breiten.
Low rank methods for a class of generalized Lyapunov equations and related issues.
Numerische Mathematik 124(3):441–470, 2013.

P. Benner and T. Damm.
Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems.
SIAM Journal on Control and Optimization 49(2):686–711, 2011.

P. Benner and P. Kürschner.
Computing real low-rank solutions of Sylvester equations by the factored ADI method.
Computers and Mathematics with Applications 67:1656–1672, 2014.

P. Benner, P. Kürschner, and J. Saak.
Efficient handling of complex shift parameters in the low-rank Cholesky factor ADI method.
Numerical Algorithms 62(2):225–251, 2013.

P. Benner, P. Kürschner, and J. Saak.
Self-generating and efficient shift parameters in ADI methods for large Lyapunov and Sylvester equations.
Electronic Transactions on Numerical Analysis, 43:142–162, 2014.

P. Benner and J. Saak.
Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey.
GAMM Mitteilungen 36(1):32–52, 2013.

(Upcoming) preprints available at

http://www.mpi-magdeburg.mpg.de/preprints/index.php