Krylov Day 2015
Delft, Netherlands
February 2, 2015

Krylov Subspace Methods for Matrix Equations Which Include Matrix Functions

Patrick Kürschner
with
Peter Benner and Jens Saak

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

Introduction

1. Problem - Matrix Equations

Large-scale Lyapunov equations

$$AX + XA^T = -R$$

Assumptions:

- $A \in \mathbb{R}^{n \times n}$ Hurwitz ($\Lambda(A) \subset \mathbb{C}_-$), large, sparse
  - We can efficiently compute matrix-vector products,
  - solve linear systems $Au = v$ (e.g., with Krylov methods),
  - but **cannot** compute Schur-, eigen-, singular value decompositions ($\mathcal{O}(n^3)$ complexity).
- right hand side $R = R^T \in \mathbb{R}^{n \times n}$,
- $X = X^T \in \mathbb{R}^{n \times n}$ is the sought solution
  - which we **cannot** store ($\mathcal{O}(n^2)$ storage) explicitly.

Introduction

Low-rank Phenomena

Large-scale Lyapunov equations, low-rank right hand side

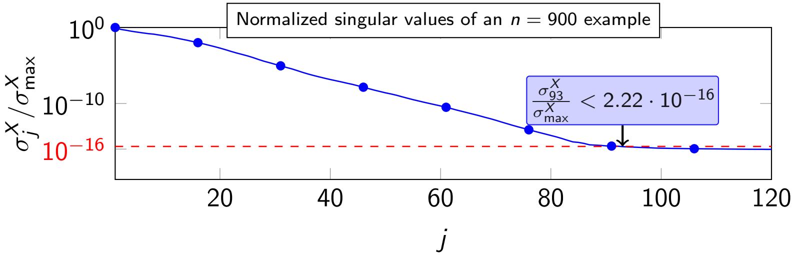
$$AX + XA^T = -BB^T, \quad B \in \mathbb{R}^{n\times r}, \quad r \ll n$$

Observation in practice, theoretical investigations

[Penzl '99, Antoulas/Sorenson/Zhou '02, Grasedyck '04]

$\rightsquigarrow$ X has small numerical rank:

$$\operatorname{rank}(X,\epsilon) := \operatorname*{argmin}_{j=1,\dots,n}\left(\frac{\sigma_j^X}{\sigma_{\max}^X} > \epsilon\right) \ll n, \quad \epsilon \ll 1$$



Introduction

Low-rank Phenomena

Large-scale Lyapunov equations, low-rank right hand side

$$AX + XA^T = -BB^T, \quad B \in \mathbb{R}^{n\times r}, \quad r \ll n$$

Observation in practice, theoretical investigations

[Penzl '99, Antoulas/Sorenson/Zhou '02, Grasedyck '04]

$\rightsquigarrow$ X has small numerical rank:

$$\operatorname{rank}(X, \epsilon) := \operatorname*{argmin}_{j=1,\ldots,n} \left(\frac{\sigma_j^X}{\sigma_{\max}^X} > \epsilon \right) \ll n, \quad \epsilon \ll 1$$

Hence, we can approximate P by a low-rank factorization:

$$X \approx X_\ell = ZDZ^T, \quad Z \in \mathbb{R}^{n\times \ell}, \ D = D^T \in \mathbb{R}^{\ell\times\ell} \quad \text{with} \quad \ell \ll n.$$

Z, D are called **low-rank solution factors**.

Krylov Subspace Methods for Lyapunov Equations

Low-rank Solvers

Algorithms for computing low-rank solutions

- Low-rank (alternating directions implicit) ADI iteration (LR-ADI)

  [Penzl '99, Li/White '02, Benner/K./Saak '14]

Krylov Subspace Methods for Lyapunov Equations

Low-rank Solvers

Algorithms for computing low-rank solutions

- Low-rank (alternating directions implicit) ADI iteration (LR-ADI)
  [Penzl '99, Li/White '02, Benner/K./Saak '14]
- Projection methods: let $\mathcal{V} = \operatorname{span}\{V\} \subset \mathbb{R}^n$, $\dim(\mathcal{V}) = h \ll n$.
  Perform Galerkin projection to $AX + XA^T = -BB^T$ and solve small Lyapunov equation

$$(V^T AV)\tilde{X} + \tilde{X}(V^T A^T V) = -(V^T B)(B^T V)$$

by standard, dense methods for $\tilde{X}$. Then $V\tilde{X}V^T := X_k \approx X$.

Popular choices for $\mathcal{V}$:

- ordinary (block) Krylov subspaces $\mathcal{K}_k(A, B)$
  [Jaimoukha/Kasenally '94, ...]
- rational Krylov subspaces [Druskin/Knizherman/Simoncini '07/'11]

$$\mathcal{RK}_k(A, B, \xi) = d_{k-1}(A)^{-1}\mathcal{K}_k(A, B), \quad d_{k-1}(z) = \prod_{j=1}^{k-1}(1 - \tfrac{z}{\xi_j})$$

Krylov Methods for Matrix Functions

2. Problem - Matrix Function Times Vectors

Problem Statement

Given: $B \in \mathbb{R}^{n \times r}$, (smooth), nonlinear function f
Sought: Approximation to $B_f = f(A)B$.

Direct computation of $f(A)$ uses Schur-, eigen decomposition of A

$\rightsquigarrow$ not possible!

Again, a projection approach saves the day!
Let $B \in \mathcal{V} = \operatorname{span}\{V\} \subset \mathbb{C}^n$, $\dim(\mathcal{V}) = h \ll n$, $V^H V = I$.
Approximation: $B_f \approx VV^H B_f = V\hat{B}_f \in \mathcal{V}$, $\hat{B}_f := V^H B_f$.
Impose Ritz-Galerkin condition:

$$V\hat{B}_f - f(A)VV^H B \quad \perp \quad \mathcal{V} \quad \Longrightarrow \quad B_f \approx Vf(V^H AV)V^H B$$

$V^H AV \in \mathbb{C}^{h \times h} \Rightarrow f(V^H AV)$ can be computed.

Krylov Methods for Matrix Functions

2. Problem - Matrix Function Times Vectors

Popular choices for $\mathcal{V}$:

- Block Krylov subspace $\mathcal{V} = \mathcal{K}_k(A,\ B) \quad \Rightarrow f \approx p_{k-1} \in \mathcal{P}_{r(k-1)}$.
  Slow convergence for general functions f!
- Rational Krylov subspace $\mathcal{RK}_k(A, B, \xi) = d_{k-1}(A)^{-1}\mathcal{K}_k(A, B)$
  [Druskin/Knizherman '98, Beckermann/Reichel '09, Güttel '10]
  $\Rightarrow\ f \approx r_k = p_{k-1}/d_{k-1} \rightsquigarrow$ much better convergence for most f!

Krylov Methods for Matrix Functions

2. Problem - Matrix Function Times Vectors

Popular choices for $\mathcal{V}$:

- Block Krylov subspace $\mathcal{V} = \mathcal{K}_k(A,\ B) \quad \Rightarrow f \approx p_{k-1} \in \mathcal{P}_{r(k-1)}$.
  Slow convergence for general functions f!
- Rational Krylov subspace $\mathcal{RK}_k(A, B, \xi) = d_{k-1}(A)^{-1}\mathcal{K}_k(A, B)$
  [Druskin/Knizherman '98, Beckermann/Reichel '09, Güttel '10]
  $\Rightarrow\ f \approx r_k = p_{k-1}/d_{k-1} \rightsquigarrow$ much better convergence for most f!
- The shifts ξ_i (roots of d_{k-1}, poles of r_k) heavily influence the convergence speed.
- E.g., adaptive computation via [Druskin/Simoncini '11]

  $$\xi_{k+1} = \underset{s \in \mathcal{D}}{\operatorname{argmax}} |r_k(s)|^{-1},$$

  where $\mathcal{D}$ contains discrete points from boundary of the convex hull of $\Lambda(-V^H AV)$. $\rightsquigarrow$ RKSM$_{\mathcal{D}}$
- Choosing $\xi_{2k} = \infty$ and $\xi_{2k-1} = 0$ yields extended Krylov subspace
  $\mathcal{EK}_k(A, B) := \mathcal{K}_k(A,\ B) \cup \mathcal{K}_k(A^{-1},\ A^{-1}B)$ $\rightsquigarrow$ EKSM

Krylov Methods for Matrix Functions

Consider the Lyapunov equation

$$AX_f + X_f A^T = -f(A)BB^T - BB^T f(A).$$

Application: [Gawronski/Juang '90, Petterson '13]

Frequency-limited balanced truncation model order reduction of

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t).$$

Krylov Methods for Matrix Functions

Consider the Lyapunov equation

$$AX_f + X_f A^T = -f(A)BB^T - BB^T f(A).$$

Application: [GAWRONSKI/JUANG '90, PETTERSON '13]

Frequency-limited balanced truncation model order reduction of

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t).$$

There, $$f(A) = \frac{1}{2\pi}\int_{\Omega}(\mathrm{i}\omega I - A)^{-1}\mathrm{d}\omega = \frac{1}{\pi}\mathrm{Re}\left(\int_{\omega_1}^{\omega_2}(\mathrm{i}\omega I - A)^{-1}\mathrm{d}\omega\right)$$
$$= \mathrm{Re}\left(\frac{\mathrm{i}}{\pi}\log\left((A+\mathrm{i}\omega_1 I)^{-1}(A+\mathrm{i}\omega_2 I)\right)\right)$$

with frequency region $\Omega = -[\omega_1,\ \omega_2] \cup [\omega_1,\ \omega_2]$, $0 \leq \omega_1 < \omega_2 < \infty$.

Here, $\log M$ is the principal branch of the complex, matrix-valued, natural logarithm for $M \in \mathbb{C}^{n\times n}$ with $\Lambda(M) \in \mathbb{C}\backslash\mathbb{R}_-$.

Matrix Equations with Matrix Functions

Consider the Lyapunov equation

$$AX_f + X_f A^T = -f(A)BB^T - BB^T f(A).$$

Before we can compute a low-rank solution $Z_f D_f Z_f^T \approx X_f$ we have to approximate $f(A)B =: B_f$!

Matrix Equations with Matrix Functions

Consider the Lyapunov equation

$$AX_f + X_f A^T = -f(A)BB^T - BB^T f(A).$$

Before we can compute a low-rank solution $Z_f D_f Z_f^T \approx X_f$ we have to approximate $f(A)B =: B_f$!

Procedure part 1. [Benner/K./Saak '14/'15]

Employ rational (block) Krylov subspace method (RKSM) to get

- $\tilde{B}_f \approx B_f$ and orthonormal basis matrix V $(\mathrm{span}\,\{V\} = \mathcal{RK}_k(A, B, \xi))$.
- Recall $f(A) = \frac{1}{\pi}\,\mathrm{Re}\left(\int_{\omega_1}^{\omega_2} (\mathrm{i}\omega I - A)^{-1}\mathrm{d}\omega\right) = \mathrm{Re}\,\left(\frac{\mathrm{i}}{\pi}\log\ldots\right)$
- $\rightsquigarrow$ better adaptive shifts

$$\xi_{k+1} = \underset{s\in\mathrm{i}\hat{\Omega}}{\mathrm{argmax}}\,|r_k(s)|^{-1},$$

where $\hat{\Omega}$ contains discrete points from $[\omega_1,\ \omega_2]$ $\rightsquigarrow$ RKSM$_\Omega$

Matrix Equations with Matrix Functions

Consider the Lyapunov equation

$$AX_f + X_f A^T = -f(A)BB^T - BB^T f(A).$$

Before we can compute a low-rank solution $Z_f D_f Z_f^T \approx X_f$ we have to approximate $f(A)B =: B_f$!

Procedure part 1. [Benner/K./Saak '14/'15]

Employ rational (block) Krylov subspace method (RKSM) to get

- $\tilde{B}_f \approx B_f$ and orthonormal basis matrix V $(\mathrm{span}\{V\} = \mathcal{RK}_k(A, B, \xi))$.

Procedure part 2.

- Reuse V to solve Lyap.eqn. via Galerkin, i.e., solve

$$(V^T AV)\tilde{X}_f + \tilde{X}_f(V^T A^T V) = -(V^T \tilde{B}_f)(B^T V) - (V^T B)(\tilde{B}_f^T V)$$

  for $\tilde{X}_f$ and obtain $X_f \approx X_{f,k} := V\tilde{X}_f V^T$.
- If required (often not), continue RKSM until $\|AX_{f,k} + X_{f,k}A^T + \tilde{B}_f B^T + B\tilde{B}_f^T\| < \mathrm{tol}$

Numerical Examples

$AX_f \quad + \quad X_f A^T = -B_f B^T - BB_f^T, \quad f(z) = \mathrm{Re}\left(\frac{\mathrm{i}}{\pi}\log\frac{z+\mathrm{i}\omega_2}{z+\mathrm{i}\omega_1}\right)$

Example `fdm_2d`: $A \hat{=}$ FDM Discretization of

$$L(x) := \Delta x - 100\xi_1 \frac{\partial x}{\partial \xi_1} - 1000\xi_2 \frac{\partial x}{\partial \xi_2} \quad \text{on} \quad (0,1)^2$$

for $x = x(\xi_1,\ \xi_2)$, homogeneous Dirichlet BC.
350 grid points $\Rightarrow n = 122\ 500$, $B =$ `rand(n,5)`,
frequency interval limits $\omega_1 = 10$, $\omega_2 = 10^3$.

Numerical Examples

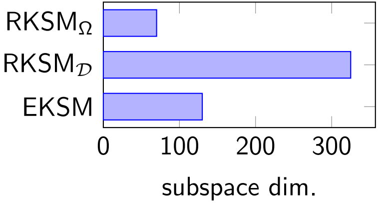
$AX_f \quad + \quad X_f A^T = -B_f B^T - BB_f^T, \quad f(z) = \mathrm{Re}\left(\frac{\mathrm{i}}{\pi} \log \frac{z+\mathrm{i}\omega_2}{z+\mathrm{i}\omega_1}\right)$

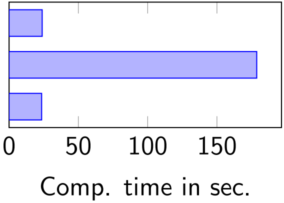
Example `fdm_2d`: $A \hat{=}$ FDM Discretization of

$$L(x) := \Delta x - 100\xi_1 \frac{\partial x}{\partial \xi_1} - 1000\xi_2 \frac{\partial x}{\partial \xi_2} \quad \text{on} \quad (0,1)^2$$

for $x = x(\xi_1,\ \xi_2)$, homogeneous Dirichlet BC.
350 grid points $\Rightarrow n = 122\,500$, $B =$ `rand(n,5)`,
frequency interval limits $\omega_1 = 10$, $\omega_2 = 10^3$.

We apply EKSM, RKSM$_\mathcal{D}$, RKSM$_\Omega$ to compute low-rank solutions of X_f





Ranks of low-rank solutions: $\mathrm{rank}(Z_f D_f Z_f^T) = 38$.

Numerical Examples

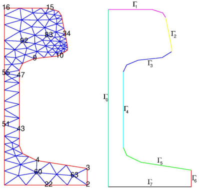
$AX_fE + EX_fA^T = -B_fB^T - BB_f^T, \quad f(z) = \mathrm{Re}\left(\frac{\mathrm{i}}{\pi}\log\frac{z+\mathrm{i}\omega_2}{z+\mathrm{i}\omega_1}\right)$

Example `rail`: Cooling of steel profiles,

A, E from FEM discretization of heat equation

[Benner/Saak '04]

$\Rightarrow n = 79\,841$, $B \in \mathbb{R}^{n\times 7}$,
frequency interval limits $\omega_1 = 10^{-2}$, $\omega_2 = 10$.



Numerical Examples

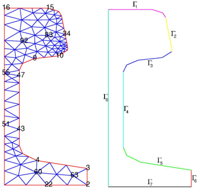
$AX_fE + EX_fA^T = -B_fB^T - BB_f^T, \quad f(z) = \mathrm{Re}\left(\frac{\mathrm{i}}{\pi}\log\frac{z+\mathrm{i}\omega_2}{z+\mathrm{i}\omega_1}\right)$

Example `rail`: Cooling of steel profiles,

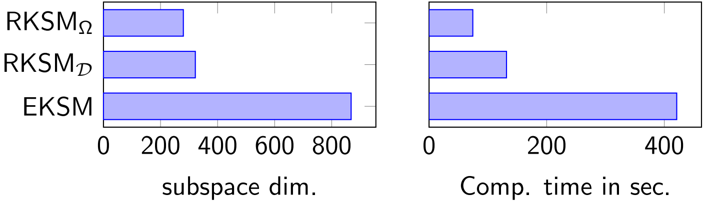
A, E from FEM discretization of heat equation

[Benner/Saak '04]

$\Rightarrow n = 79\,841$, $B \in \mathbb{R}^{n\times 7}$,
frequency interval limits $\omega_1 = 10^{-2}$, $\omega_2 = 10$.



We apply EKSM, RKSM$_{\mathcal{D}}$, RKSM$_{\Omega}$ to compute low-rank solutions of X_f



Ranks of low-rank solutions: $\mathrm{rank}(Z_fD_fZ_f^T) = 170, \ldots, 190$.

Summary / Outlook

Conclusion:

- Large-scale Lyapunov equations, where $f(A)$ appears in the right hand side, can be solved efficiently for low-rank solutions.
- Approximation of $f(A)B$ by, e.g., rational Krylov subspace methods $\rightsquigarrow$ recycling data to approximate X_f leads to very efficient *all-at-once* algorithms.
- Intelligent choice of (adaptive) shifts mandatory for fast convergence.

Further topics:

- Theoretical results w.r.t. influence of $f(A)$ to $\sigma_j(X_f)$. ✓
- Other similar applications lead to
  - $f(A) = \exp(A)$,
  - $f(A) = A^m$, $m \in \mathbb{N}$, (✓)
  - $f(A) = \frac{1}{2\pi}\,\mathrm{Re}\,(\omega_1 I - 2\mathrm{i}\log(I - A\exp(-\mathrm{i}\omega_1)))$

  or products of thereof.
- Todo: Usage of tangential rational Krylov subspaces.