On low-rank updates of matrix functions

Ana Šušnjara, Daniel Kressner

EPF Lausanne, MATHICSE, ANCHP

Delft, February 2, 2015

Outline

Outline

Motivation

Given $A \in \mathbb{C}^{n \times n}$ and a function f, compute low-rank update

$$f(A + BC^*) - f(A),$$

where $B, C \in \mathbb{C}^{n \times r}$, $r \ll n$.

Special cases

- Sherman-Morrison-Woodbury formula

$$(A + \text{rank-}k)^{-1} - A^{-1} = \text{rank-}k.$$

- Rank-1 update of a rational matrix function [Bernstein/van Loan]

$$r(A + \text{rank-}1) - r(A) = \text{rank-}k,$$

where $r(z) = p(z)/q(z)$, $k = \max\{\deg p, \deg q\}$.

Krylov subspaces

Given a matrix $A \in \mathbb{C}^{n\times n}$, and a vector $b \in \mathbb{C}^n$

- the **polynomial Krylov subspace** of order k is defined as

$$\mathcal{K}_k(A, b) = \{P(A)b : P \in \mathcal{P}_{k-1}\},$$

  where $\mathcal{P}_{k-1}$ denotes the space of polynomials of deg $\leq k-1$;

- the **rational Krylov subpace** of order k is defined as

$$\mathcal{K}_k^{(d)}(A, b) = \{R(A)b : R \in \mathcal{P}_{k-1}/Q\},$$

  where Q is a fixed polynomial with d poles, $d \leq k-1$, disjoint from the spectrum of A.

The **tensor Krylov subspace** for (A, b) and (A^*, c) is defined as

$$\mathcal{K}_{\mathcal{I}}^{\otimes} = \operatorname{span}\{\mathcal{S}_{k_1}(A, b) \otimes \mathcal{S}_{k_2}(A^*, c)\},$$

where $\mathcal{I} = (k_1, k_2) \in \mathbb{N} \times \mathbb{N}$, $\mathcal{S}_{k_1}(A, b)$ and $\mathcal{S}_{k_2}(A^*, c)$ are either polynomial or rational Krylov subspaces for (A, b) and (A^*, c).

Polynomial Krylov method for rank-1 matrix update

[Beckermann/Kressner]

Given $A \in \mathbb{C}^{n\times n}$ and a function f, compute low-rank update

$$f(A + bc^*) - f(A), \quad b, c \in \mathbb{C}^n.$$

I. Use the Arnoldi algorithm to compute
   - orthonormal basis U_k of $\mathcal{K}_k(A, b)$.
   - orthonormal basis V_k of $\mathcal{K}_k(A^*, c)$.

II. Compress $H_k = U_k^* A V_k$, $\tilde{b} = U_k^* b$ and $\tilde{c} = V_k^* c$.

III. Compute the compressed problem $\tilde{D}_k = f(H_k + \tilde{b}\tilde{c}^*) - f(H_k)$.

IV. Return the approximate update $D_k = U_k^* \tilde{D}_k V_k$.

Outline

Corrected approximation

Motivation: The Arnoldi algorithm computes $k+1$ vectors, but only the first k are used. How to exploit extra information to obtain

- a better approximation,
- an error estimate for

$$\exp(A + bb^*) - \exp(A)?$$

Basic idea [Saad]: replace the standard Arnoldi approximation of $\exp(A)b$ with the corrected approximation using φ-functions.

Definition

The φ-functions are defined by the integral representation

$$\varphi_0(z) = e^z, \quad \varphi_l(z) = \frac{1}{l!} \int_0^1 e^{(1-t)z} t^{l-1} \mathrm{d}t, \quad l \in \mathbb{N}, z \in \mathbb{C}.$$

Corrected approximation

- Corrected approximation

$$\exp(A + bb^*) - \exp(A) \approx D_k + E_k,$$

  E_k involves computation of φ_1– function.
- E_k is an estimate for the approximation error

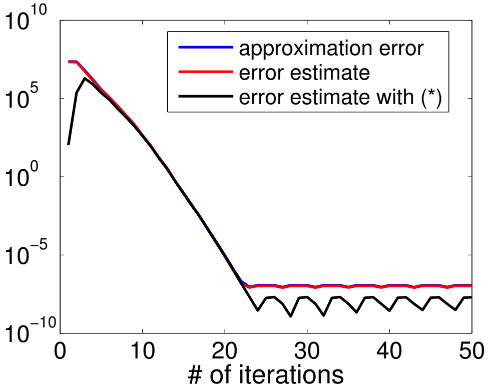
$$\exp(A + bb^*) - (\exp(A) + D_k).$$

- **Problem**: in the computation of E_k, a function evaluation of a large matrix is needed.

  $\Rightarrow$ use φ_2– function instead $(\star)$

Numerical experiment

- Adjacency matrix $A \in \mathbb{R}^{492\times 492}$ in Erdös collaboration graph
- $A = \tilde{A} + bb^*$, where bb^* is a rank-1 perturbation resulting from removal of an edge



Results for $\exp(A)$ using the polynomial Krylov method.

Outline

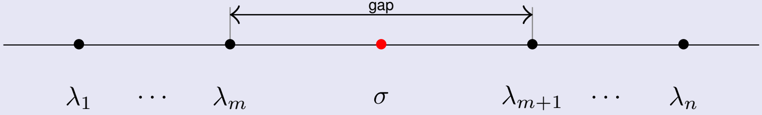
Spectral projector

Definition

If $A = VDV^*$ *is the spectral decomposition arranged so that* $D = \mathrm{diag}(D_1, D_2)$, *with* $D_1 \in \mathbb{C}^{p\times p}$, $\mathrm{diag}(D_1) \in \mathbb{C}^-$ *and* $D_2 \in \mathbb{C}^{q\times q}$, $\mathrm{diag}(D_2) \in \mathbb{C}^+$, *then* $\mathrm{sign}(A) = V \begin{bmatrix} -I_p & 0 \\ 0 & I_q \end{bmatrix} V^*$.

Definition

Let the eigenvalues $\lambda_1, \ldots, \lambda_n$ *of* A *be ordered as*



The ***spectral projector*** *onto the invariant subspace associated with eigenvalues* $\lambda_1, \ldots, \lambda_m$ *is defined as*

$$\Pi_{<\sigma}(A) = \frac{1}{2}(I - \mathrm{sign}(A - \sigma I)).$$

Divide-and-conquer method

Consider a sparse matrix $A \in \mathbb{C}^{n\times n}$ with a decomposition

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} + BB^*, \quad B \in \mathbb{C}^{n\times r}, \quad r \ll n.$$

$\Rightarrow$ divide-and-conquer method to approximate $\Pi_{<\sigma}(A)$

$$\Pi_{<\sigma}(A) \approx \begin{pmatrix} \Pi_{<\sigma}(A_{11}) & 0 \\ 0 & \Pi_{<\sigma}(A_{22}) \end{pmatrix} + D_k$$

Data-sparse representation

Idea: recursively apply one step of divide-and-conquer.

- By definition, $\operatorname{rank}(D_k) \leq kr \Longrightarrow D_k$ can be stored in a low-rank format.
- Data structure similar to $\mathcal{H}$–matrices.

Advantages: computational and storage demands significantly reduced.

A and A^2 based methods

Lets assume that A and $A + bb^*$ are

- invertible, and
- Hermitian.

Two approaches:

I. Krylov subspaces associated with A, as before;

II. Krylov subspaces associated with A^2 and its rank-1 update.

Motivation: to avoid singularity we consider different representation of sign function. [Frommer et al.]

$$
\begin{aligned}
&\mathrm{sign}(A + bb^*) - \mathrm{sign}(A) \\
&= A\left[f\left((A + bb^*)^2\right) - f\left(A^2\right)\right] + bb^* f\left((A + bb^*)^2\right),
\end{aligned}
$$

where $f(x) = x^{-1/2}$.

A^2 based approach

- Two-step approach

$$f\left((A+bb^*)^2\right) - f\left(A^2\right) = \underbrace{f\left(A^2+X\right) - f\left(A^2+aa^*\right)}_{\mathrm{I}} + \underbrace{f\left(A^2+aa^*\right) - f\left(A^2\right)}_{\mathrm{II}}, \quad X = aa^* - dd^*.$$

To obtain the approximation, compute:

- V_k^{I}, H_k^{I} corresponding to the Krylov subspace with respect to $A^2 + aa^*$ and d,
- V_k^{II}, H_k^{II} corresponding to the Krylov subspace with respect to A^2 and a.

Convergence result for A^2 approach

Theorem

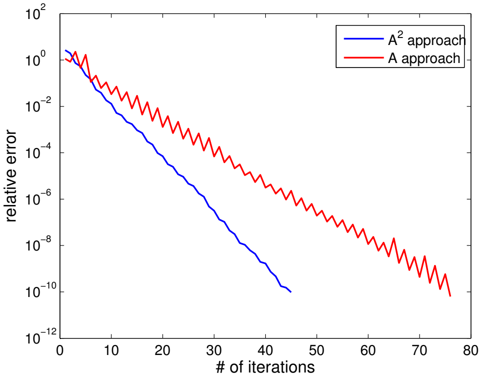
If $A = UDU^*$ and $H_k = \tilde{U}\tilde{D}\tilde{U}^*$, then

$$\left\| A \left[f\left(A^2 + aa^*\right) - f\left(A^2\right) - V_k^{\mathrm{II}} \left(f\left(H_k^{\mathrm{II}} + \tilde{a}\tilde{a}^*\right) - f\left(H_k^{\mathrm{II}}\right)\right) (V_k^{\mathrm{II}})^* \right] \right\|_2 \leq C \cdot \|r_k\|_2,$$

where r_k is residual in the $k-$th step of the CG method applied to the linear system $A^2 x = a$ with initial iterate zero and $C = C(a, U, \tilde{U})$.

Numerical experiment

- $A \in \mathbb{R}^{500 \times 500}$ tridiagonal
- $\sigma(A) \subset [-10, -0.25] \cup [0.25, 10]$



Comparison of A and A^2 based approach

Summary

- Krylov method for low–rank updates
- Corrected approximation for exponential function
- Computation of spectral projector via DC method

Thanks! :)