Sketched polynomial Krylov subspaces for matrix functions

Valeria Simoncini

Dipartimento di Matematica
Alma Mater Studiorum - Università di Bologna
valeria.simoncini@unibo.it

From joint works with
Davide Palitta, Marcel Schweitzer, Yihong Wang

First, an anniversary. Bari, 2005

Università di Bologna

Approssimazione dell'esponenziale di matrice: convergenza asintotica e proprietà geometriche

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

Collaborazione con Luciano Lopez, Bari

1

The approximation problem

Given $v \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$, approximate

$$x = f(A)\,v \qquad \text{e.g.} \qquad f(\lambda) = e^{\lambda}$$

- f analytic function
- Focus: A large dimension
- General approach: $x_m \in \mathbb{K}_m$ (some) Krylov subspace

Problem in context

Wide range of applications:

- Numerical solution of Time-dependent PDEs
- Complex networks
- Flows on manifolds

$$Q_t = H(Q,t)Q, \quad Q(t)|_{t=0} = Q_0 \in V_k(\mathbb{R}^n)$$

V_k Stiefel manifold (computation of a few Lyapunov exponents)

Approximation using Krylov subspace

$$\mathbb{K}_m \equiv \mathbb{K}_m(A, v) = \mathrm{span}\{v, Av, \ldots, A^{m-1}v\}$$

V_m s.t. $\mathrm{range}(V_m) = \mathbb{K}_m(A, v)$ and $V_m^* V_m = I$

Arnoldi relation

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^*$$

A common approach

$$f(A)v \approx x_m = V_m f(H_m) e_1, \qquad \|v\| = 1$$

x_m derived from interpolation problem in Hermite sense (Saad '92)

Pausing on pros and cons

Current full-orth based Krylov subspaces may be "expensive"

- "expensive" in different ways: Memory, computation, communication, etc.
- General concern :
  linear systems, eigenvalue problems, matrix equations, etc.

Imperative

Keep the Krylov recurrence short and cheap!

Several steps back. Short recurrences

Main ingredient: Krylov decomposition (Stewart, '01)

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^*$$

with

- B_k is $k \times k$, Rayleigh quotient
- $[U_k, u_{k+1}]$ are linearly independent, build a Krylov space (here, $b_{k+1} = \beta_{k+1} e_k$)

Procedures fitting this framework:

- Full orth Arnoldi (*)
- Truncated Arnoldi, restarted Arnoldi
- Chebyshev, Newton, ... iterations
- Nonsymmetric Lanczos

Except for (*), all methods suffer from lack/loss of orthogonality properties!

(Rich literature from the 1990s and early 2000s)

Several steps back. Short recurrences

Main ingredient: Krylov decomposition (Stewart, '01)

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^*$$

with

- B_k is $k \times k$, Rayleigh quotient
- $[U_k, u_{k+1}]$ are linearly independent, build a Krylov space (here, $b_{k+1} = \beta_{k+1} e_k$)

Procedures fitting this framework:

- Full orth Arnoldi (*)
- Truncated Arnoldi, restarted Arnoldi
- Chebyshev, Newton, ... iterations
- Nonsymmetric Lanczos

Except for (*), all methods suffer from lack/loss of orthogonality properties!

(Rich literature from the 1990s and early 2000s)

Towards stabilized procedures. Randomized strategies

Randomized methods have become a key tool to:

- Lower memory requirements by "oblivious" projections
- Lower computational costs
- Stabilize procedures

... with probabilistic confidence.

Random projections can be used to make bases more orthogonal

Towards stabilized procedures. Randomized strategies

Randomized methods have become a key tool to:

- Lower memory requirements by “oblivious” projections
- Lower computational costs
- Stabilize procedures

... with probabilistic confidence.

Random projections can be used to make bases more orthogonal

Towards stabilized procedures. Randomized strategies

$$AU_k = U_k B_k + u_{k+1} b_k^*$$

⋆ Let $\mathcal{J} : \mathbb{R}^{n\times(k+1)} \to \mathbb{R}^{s\times(k+1)}$ be a row selection operator with $s > k+1$

⋆ Let $\mathcal{J}(U) = QR$ be the reduced QR decomposition of the *subsampled* matrix (see, e.g., Woodruff, '14)

Ideal low-cost stabilization problem

Given $U_{k+1} = [U_k, u_{k+1}]$, for some s with $k+1 < s \ll n$, find $\mathcal{J}$ giving the best conditioned matrix

$$\widehat{U}_{k+1} := U_{k+1} R^{-1}$$

where $\mathcal{J}(U_{k+1}) = QR$

Unrealistic:

- Solving this problem is expensive
- Columns of U_{k+1} are assumed not to be available simultaneously!

First practical compromise

Select $\mathcal{J}$ a priori $\Rightarrow$ random subsampling, with row selection probability p_i

Towards stabilized procedures. Randomized strategies

$$AU_k = U_k B_k + u_{k+1} b_k^*$$

⋆ Let $\mathcal{J} : \mathbb{R}^{n\times(k+1)} \to \mathbb{R}^{s\times(k+1)}$ be a row selection operator with $s > k+1$

⋆ Let $\mathcal{J}(U) = QR$ be the reduced QR decomposition of the *subsampled* matrix
(see, e.g., Woodruff, '14)

Ideal low-cost stabilization problem

Given $U_{k+1} = [U_k, u_{k+1}]$, for some s with $k+1 < s \ll n$, find $\mathcal{J}$ giving the best conditioned matrix

$$\widehat{U}_{k+1} := U_{k+1} R^{-1}$$

where $\mathcal{J}(U_{k+1}) = QR$

Unrealistic:

- Solving this problem is expensive
- Columns of U_{k+1} are assumed not to be available simultaneously!

First practical compromise:

Select $\mathcal{J}$ a priori $\Rightarrow$ random subsampling, with row selection probability p_i

Towards stabilized procedures. Randomized strategies

$$AU_k = U_k B_k + u_{k+1} b_k^*$$

$\star$ Let $\mathcal{J} : \mathbb{R}^{n\times(k+1)} \to \mathbb{R}^{s\times(k+1)}$ be a row selection operator with $s > k+1$

$\star$ Let $\mathcal{J}(U) = QR$ be the reduced QR decomposition of the *subsampled* matrix (see, e.g., Woodruff, '14)

Ideal low-cost stabilization problem

Given $U_{k+1} = [U_k, u_{k+1}]$, for some s with $k+1 < s \ll n$, find $\mathcal{J}$ giving the best conditioned matrix

$$\widehat{U}_{k+1} := U_{k+1} R^{-1}$$

where $\mathcal{J}(U_{k+1}) = QR$

Unrealistic:

- Solving this problem is expensive
- Columns of U_{k+1} are assumed not to be available simultaneously!

First practical compromise:

Select $\mathcal{J}$ a priori $\Rightarrow$ random subsampling, with row selection probability p_i

Towards stabilized procedures. Randomized strategies

$$AU_k = U_k B_k + u_{k+1} b_k^*$$

⋆ Let $\mathcal{J} : \mathbb{R}^{n\times(k+1)} \to \mathbb{R}^{s\times(k+1)}$ be a row selection operator with $s > k+1$

⋆ Let $\mathcal{J}(U) = QR$ be the reduced QR decomposition of the *subsampled* matrix
(see, e.g., Woodruff, '14)

Ideal low-cost stabilization problem

Given $U_{k+1} = [U_k, u_{k+1}]$, for some s with $k+1 < s \ll n$, find $\mathcal{J}$ giving the best conditioned matrix

$$\widehat{U}_{k+1} := U_{k+1} R^{-1}$$

where $\mathcal{J}(U_{k+1}) = QR$

Unrealistic:

- Solving this problem is expensive
- Columns of U_{k+1} are assumed not to be available simultaneously!

First practical compromise:

Select $\mathcal{J}$ a priori $\Rightarrow$ random subsampling, with row selection probability p_i

A randomly subsampled Krylov decomposition

Let $\widehat{U}_k = U_k R_k^{-1}$

Randomized Krylov decomposition v.1 (Palitta, Schweitzer, Simoncini, '25)

It holds that

$$\mathcal{J}(A\widehat{U}_k) = \mathcal{J}(\widehat{U}_k)(\widehat{B}_k + d_k e_k^*) + q_{k+1}\chi_k e_k^*, \quad q_{k+1} \perp \mathcal{J}(\widehat{U}_k)$$

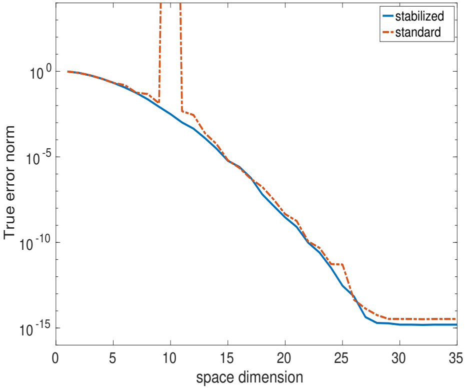
1. $\widehat{B}_k + d_k e_k^*$ rank-one modification (last column) of $\widehat{B}_k = R_k^{-1} B_k R_k$
2. Randomized Krylov decomposition corresponds to $\mathcal{J}^*\mathcal{J}$-orthogonalization of original Krylov decomposition

A simple example. Matrix function evaluation.

$$\exp(A)v \approx \widehat{U}_k \exp(\widehat{B}_k + d_k e_k^*) e_1 \chi_0$$

$n = 4900$, $s = 200$ – similar results for $s = 80$ ($k_{\rm max}$=40)

Nonsymmetric Lanczos iteration:



Carrying on

Why non-symmetric Lanczos iterations?

- Pros: Inherently short-term recurrence (no truncation parameter!)
- Pros: Builds same Krylov subspace as all Arnoldi-type methods
- Cons: Requires A^T
- Cons: Breakdown possible

Is row subsampling enough?

- Row sampling cheap and easy fix
- Row sampling is often not enough as stabilizer
- Conditioning not necessarily low

$$\Downarrow$$

Sketching strategies: Subspace embedding

Carrying on

Why non-symmetric Lanczos iterations?

- Pros: Inherently short-term recurrence (no truncation parameter!)
- Pros: Builds same Krylov subspace as all Arnoldi-type methods
- Cons: Requires A^T
- Cons: Breakdown possible

Is row subsampling enough?

- Row sampling cheap and easy fix
- Row sampling is often not enough as stabilizer
- Conditioning not necessarily low

$$\Downarrow$$

Sketching strategies: Subspace embedding

Carrying on

Why non-symmetric Lanczos iterations?

- Pros: Inherently short-term recurrence (no truncation parameter!)
- Pros: Builds same Krylov subspace as all Arnoldi-type methods
- Cons: Requires A^T
- Cons: Breakdown possible

Is row subsampling enough?

- Row sampling cheap and easy fix
- Row sampling is often not enough as stabilizer
- Conditioning not necessarily low

$$\Downarrow$$

Sketching strategies: Subspace embedding

Sketching strategies. Subspace embedding.

A $(1 \pm \varepsilon)$ ℓ_2-subspace embedding for $V \in \mathbb{R}^{n \times k}$ is an operator $\mathcal{S}$ such that

$$(1-\varepsilon)\|Vx\|_2^2 \leq \|\mathcal{S}(Vx)\|_2^2 \leq (1+\varepsilon)\|Vx\|_2^2, \qquad \forall x \in \mathbb{R}^k$$

with high probability.

The Subsampled Randomized Hadamard Transform

A convenient such choice (Rademacher operator)

$$\mathcal{S}(v) := \frac{1}{\sqrt{sn}} PCDv, \qquad \mathcal{S}(\cdot) \text{ is an } s \times n \text{ matrix}$$

with

D "rotation" (diagonal matrix from random distr. in $\{-1, 1\}$)

C fast cosine transform

P coordinate sampling

See, e.g., David Woodruff (2014), Martinsson and Tropp, Acta Num. (2020)

Sketching strategies. Subspace embedding.

A $(1 \pm \varepsilon)$ ℓ_2-subspace embedding for $V \in \mathbb{R}^{n\times k}$ is an operator $\mathcal{S}$ such that

$$(1-\varepsilon)\|Vx\|_2^2 \le \|\mathcal{S}(Vx)\|_2^2 \le (1+\varepsilon)\|Vx\|_2^2, \qquad \forall x \in \mathbb{R}^k$$

with high probability.

The Subsampled Randomized Hadamard Transform

A convenient such choice (Rademacher operator)

$$\mathcal{S}(v) := \frac{1}{\sqrt{sn}} P C D v, \qquad \mathcal{S}(\cdot) \text{ is an } s \times n \text{ matrix}$$

with

D "rotation" (diagonal matrix from random distr. in $\{-1, 1\}$)

C fast cosine transform

P coordinate sampling

See, e.g., David Woodruff (2014), Martinsson and Tropp, Acta Num. (2020)

Subspace embedding in Krylov decomposition

Sketched Krylov decomposition:

$$\mathcal{S}(A\widehat{U}_k) = \mathcal{S}(\widehat{U}_k)(\widehat{B}_k + d_k e_k^*) + q_{k+1}\chi_k e_k^*, \quad q_{k+1} \perp \mathcal{S}(\widehat{U}_k)$$

with

$$\kappa_2(\widehat{U}_k) \leq \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$$

Contributions within the "Krylov world", Balabanov, Cortinovis, Grigori, Guettel, Kressner, Nakatsukasa, Nouy, Palitta, Schweitzer, Timsit, Tropp, etc.

This leads in fact to an "equivalent" Krylov decomposition (in Stewart's sense):

(Simoncini, Wang, 2025)

Assume U_{k+1} full rank, $\mathcal{S}(U_{k+1}) = QR_{k+1}$, and let

$$\widehat{U}_{k+1} = [\widehat{U}_k, \widehat{u}_{k+1}] = U_{k+1}R_{k+1}^{-1} \quad R_{k+1} = \begin{bmatrix} R_k & r_{k+1} \\ 0 & \rho_{k+1} \end{bmatrix}$$

Any Krylov decomp. is transformed by sketching into the equivalent Krylov decomp.:

$$A\widehat{U}_k = \widehat{U}_k\widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1}e_k^T, \qquad \widehat{B}_k = R_k B_k R_k^{-1} + r_{k+1} b_{k+1,k} e_k^T R_k^{-1}, \quad \widehat{\beta}_{k+1} = \rho_{k+1} b_{k+1,k} r_{k,k}^{-1}.$$

Subspace embedding in Krylov decomposition

Sketched Krylov decomposition:

$$\mathcal{S}(A\widehat{U}_k) = \mathcal{S}(\widehat{U}_k)(\widehat{B}_k + d_k e_k^*) + q_{k+1}\chi_k e_k^*, \quad q_{k+1} \perp \mathcal{S}(\widehat{U}_k)$$

with

$$\kappa_2(\widehat{U}_k) \le \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$$

Contributions within the "Krylov world", Balabanov, Cortinovis, Grigori, Guettel, Kressner, Nakatsukasa, Nouy, Palitta, Schweitzer, Timsit, Tropp, etc.

This leads in fact to an "equivalent" Krylov decomposition (in Stewart's sense):

(Simoncini, Wang, 2025)

Assume U_{k+1} full rank, $\mathcal{S}(U_{k+1}) = QR_{k+1}$, and let

$$\widehat{U}_{k+1} = [\widehat{U}_k, \widehat{u}_{k+1}] = U_{k+1}R_{k+1}^{-1} \quad R_{k+1} = \begin{bmatrix} R_k & r_{k+1} \\ 0 & \rho_{k+1} \end{bmatrix}$$

Any Krylov decomp. is transformed by sketching into the equivalent Krylov decomp.:

$$A\widehat{U}_k = \widehat{U}_k\widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1}e_k^T, \qquad \widehat{B}_k = R_k B_k R_k^{-1} + r_{k+1}b_{k+1,k}e_k^T R_k^{-1},$$

$$\widehat{\beta}_{k+1} = \rho_{k+1}b_{k+1,k}r_{k,k}^{-1}.$$

Another example with nonsymmetric Lanczos

Paradigm: Stabilize while constructing

At each iteration k

- Compute next Lanczos vectors u_k, w_k
- Compute embedded vector $\mathcal{S}(u_k)$
- Update QR of embedded basis (i.e. stabilization matrix R_k)
- Update and use $\widehat{B}_k + d_k e_k^*$

Enhanced stabilization within non-sym Lanczos:

♣ Weak biorthogonality in U_m, V_m (no parameters)

♣ Strong subsampled orthogonality in U_m

Shared step: Two-pass strategy to recover problem solution (quick basis recostruction)

Another example with nonsymmetric Lanczos

Paradigm: Stabilize while constructing

At each iteration k

- Compute next Lanczos vectors u_k, w_k
- Compute embedded vector $\mathcal{S}(u_k)$
- Update QR of embedded basis (i.e. stabilization matrix R_k)
- Update and use $\widehat{B}_k + d_k e_k^*$

Enhanced stabilization within non-sym Lanczos:

♣ Weak biorthogonality in U_m, V_m (no parameters)

♣ Strong subsampled orthogonality in U_m

Shared step: Two-pass strategy to recover problem solution (quick basis recostruction)

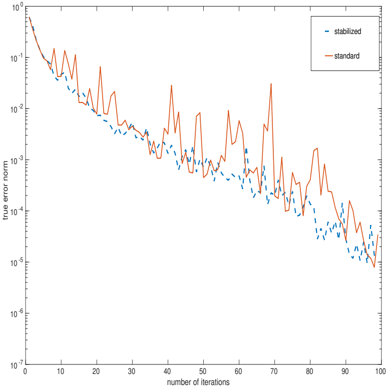
Another example with nonsymmetric Lanczos

FD discretization of the operator

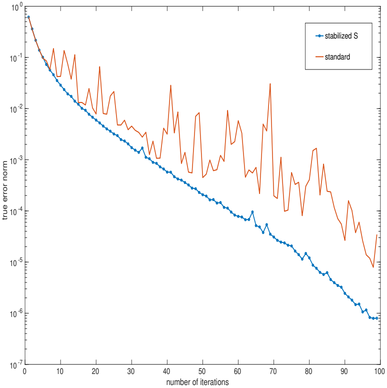
$$\mathcal{L}(u) = -(\exp(-xy)u_x)_x - (\exp(xy)u_y)_y - 100(x+y)u_y + 500u,$$

such that $n = 4900$, $v = \texttt{randn}$ (norm'd), $s = 200$

$$\exp(A^{-\frac{1}{2}})v \approx \widehat{U}_k \exp((\widehat{B}_k + d_k e_k^*)^{-\frac{1}{2}})e_1\chi_0$$



$\mathcal{J}(U_k)$



$\mathcal{S}(U_k)$

Conclusions

- Randomized subsampling is a good compromise
- Sketching as a practical tool for core NLA solvers

REFERENCES

- Davide Palitta , Marcel Schweitzer , and V. Simoncini *Sketched and truncated polynomial Krylov methods: matrix Sylvester equations*, Math. Comp, 2025.
- Davide Palitta , Marcel Schweitzer , and V. Simoncini *Sketched and truncated polynomial Krylov methods: Evaluation of Matrix functions*, Num. Linear Algebra and Appl, 2025.
- V. Simoncini and Yihong Wang *Stabilized Krylov subspace recurrences via randomized sketching*, Num. Linear Algebra and Appl, 2025.

Final note

Buon Compleanno, Luciano