Università di Bologna

Sketching strategies as NLA/Krylov-space companion

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

Some large-scale NLA problems

Typical problems encountered in NLA

- Linear systems, (non-)linear matrix equations
- (Non-)linear eigenvalue problems
- Matrix function evaluations

Common strategy

- Determine a rich "dictionary"
- Compute an approximation by imposing some condition

Our dictionary: Krylov subspaces

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

Some large-scale NLA problems

Typical problems encountered in NLA

- Linear systems, (non-)linear matrix equations
- (Non-)linear eigenvalue problems
- Matrix function evaluations

Common strategy

- Determine a rich "dictionary"
- Compute an approximation by imposing some condition

Our dictionary: Krylov subspaces

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

Some large-scale NLA problems

Typical problems encountered in NLA

- Linear systems, (non-)linear matrix equations
- (Non-)linear eigenvalue problems
- Matrix function evaluations

Common strategy

- Determine a rich "dictionary"
- Compute an approximation by imposing some condition

Our dictionary: Krylov subspaces

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

Well-known bottlenecks

Full-orth based Krylov subspaces may be “expensive”

- “expensive” in different ways: Memory, computation, communication, etc.
- General concern :
  linear systems, eigenvalue problems, matrix function evaluations, etc.

Imperative

Keep the Krylov recurrence short and cheap!

A general framework. 1

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 (oblique projection of A)
- $[U_k, u_{k+1}]$ 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

All methods suffer from lack/loss of orthogonality properties!
(in exact or finite precision arithmetic)

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

A general framework. 1

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 (oblique projection of A)
- $[U_k, u_{k+1}]$ 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

All methods suffer from lack/loss of orthogonality properties!
(in exact or finite precision arithmetic)

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

A general framework. 2

Main ingredient: Krylov decomposition (Stewart, '01)

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^* \qquad (*)$$

Krylov decompositions are very flexible ("invariant"):

- Closed wrto translations: Set $\eta_k \widehat{u}_{k+1} := u_{k+1} - U_k g_k, \eta_k \neq 0$. Substituting into (*)

$$AU_k = U_k(B_k + g_k b_{k+1}^*) + \widehat{u}_{k+1} \eta_k b_{k+1}^*$$

(rank-one modification of Rayleigh quotient matrix)

- Closed wrto similarity transformations: Given $R \in \mathbb{R}^{k \times k}$ nonsingular, (*) becomes

$$AU_k R^{-1} = U_k R^{-1}(RB_k R^{-1}) + u_{k+1}(b_{k+1}^* R^{-1})$$

A general framework. 2

Main ingredient: Krylov decomposition (Stewart, '01)

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^* \qquad (*)$$

Krylov decompositions are very flexible ("invariant"):

- Closed wrto translations: Set $\eta_k \widehat{u}_{k+1} := u_{k+1} - U_k g_k, \eta_k \neq 0$. Substituting into (*)

$$AU_k = U_k(B_k + g_k b_{k+1}^*) + \widehat{u}_{k+1} \eta_k b_{k+1}^*$$

(rank-one modification of Rayleigh quotient matrix)

- Closed wrto similarity transformations: Given $R \in \mathbb{R}^{k \times k}$ nonsingular, (*) becomes

$$AU_k R^{-1} = U_k R^{-1}(R B_k R^{-1}) + u_{k+1}(b_{k+1}^* R^{-1})$$

A general framework. 2

Main ingredient: Krylov decomposition (Stewart, '01)

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^* \qquad (*)$$

Krylov decompositions are very flexible ("invariant"):

- Closed wrto translations: Set $\eta_k \widehat{u}_{k+1} := u_{k+1} - U_k g_k, \eta_k \neq 0$. Substituting into (*)

  $$AU_k = U_k(B_k + g_k b_{k+1}^*) + \widehat{u}_{k+1} \eta_k b_{k+1}^*$$

  (rank-one modification of Rayleigh quotient matrix)

- Closed wrto similarity transformations: Given $R \in \mathbb{R}^{k \times k}$ nonsingular, (*) becomes

  $$AU_k R^{-1} = U_k R^{-1}(RB_k R^{-1}) + u_{k+1}(b_{k+1}^* R^{-1})$$

Krylov decompositions

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

Question:

How can we exploit this invariance to make Krylov-based methods more effective?

$\Rightarrow$ Use randomized methods (sketching):

i) Determine $S \in \mathbb{R}^{s \times n}$, $s \ll n$ but $s > k$
ii) Reduce space as SU_k

$\Rightarrow$ Provide theoretical ground for their use

Krylov decompositions

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

Question:

How can we exploit this invariance to make Krylov-based methods more effective?

$\Rightarrow$ Use randomized methods (sketching):

i) Determine $S \in \mathbb{R}^{s \times n}$, $s \ll n$ but $s > k$
ii) Reduce space as SU_k

$\Rightarrow$ Provide theoretical ground for their use

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

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

Oblivious subspace samplings (not associated to a specific subspace)

A typical choice of randomization operator (Rademacher)

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

with

D "rotation" (diag. matrix of random distr. ± 1 with prob. $1/2$)

C fast cosine transform

P coordinate sampling

$\star$ For notational simplicity, $\mathcal{S}(v) = Sv$ (S never constructed explicitly)

See, e.g., 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

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

Oblivious subspace samplings (not associated to a specific subspace)

A typical choice of randomization operator (Rademacher)

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

with

D "rotation" (diag. matrix of random distr. ± 1 with prob. $1/2$)

C fast cosine transform

P coordinate sampling

$\star$ For notational simplicity, $\mathcal{S}(v) = Sv$ (S never constructed explicitly)

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

Subspace embedding in Krylov decomposition

Let $SU_k = Q_k R_k$ be the reduced QR decomp.

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

- Reduced Krylov relation (Palitta, Schweitzer, Simoncini, 2025)

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

- Conditioning properties

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

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

Paradigm: Stabilize while constructing

At each iteration k

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

Subspace embedding in Krylov decomposition

Let $SU_k = Q_k R_k$ be the reduced QR decomp.

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

- Reduced Krylov relation (Palitta, Schweitzer, Simoncini, 2025)

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

- Conditioning properties

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

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

Paradigm: Stabilize while constructing

At each iteration k

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

Sketched basis as a Krylov decomposition

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

With the QR decomposition $SU_{k+1} = Q_{k+1} R_{k+1}$ and

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

(whitening)

Proposition (Simoncini, Wang 2025)

Assume that $U_{k+1} = [U_k, u_{k+1}]$ is full rank.
Let $R_{k+1} = [R_k, r_{k+1}; 0, \rho_{k+1}]$, and $\widehat{U}_{k+1} = [\widehat{U}_k, \widehat{u}_{k+1}] = U_{k+1} R_{k+1}^{-1}$.
Then any Krylov decomposition can be transformed by sketching and whitening in the following equivalent Krylov decomposition

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

Sketched vs ideal quantities (Simoncini, Wang 2025)

Standard full Krylov (ideal):

- $[U_k, u_{k+1}]$ orthonormal columns
- B_k such that $\mathcal{W}(B_k) \subseteq \mathcal{W}(A)$ (fov)

Krylov decomposition via sketching:

$$A\widehat{U}_k = \widehat{U}_k \widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1} e_k^T$$

- Let $\Theta_k(\widehat{U}_k, \widehat{u}_{k+1}) = \min_{v \in \widehat{\mathcal{U}}_k, \|v\|=1} \angle(v, \widehat{u}_{k+1})$. Then

$$\cos(\Theta_k) \leq \varepsilon.$$

- FoV property:

Let λ be an eigenvalue of $\widehat{B}_k$. Then with high probability there exists a unit norm vector $y \in \mathbb{C}^n$ such that

$$|\lambda - y^* Ay| \leq \frac{\sqrt{1+\varepsilon}}{\sqrt{1-\varepsilon}} \varepsilon \, \|A\|.$$

Sketched vs ideal quantities (Simoncini, Wang 2025)

Standard full Krylov (ideal):

- $[U_k, u_{k+1}]$ orthonormal columns
- B_k such that $\mathcal{W}(B_k) \subseteq \mathcal{W}(A)$ (fov)

Krylov decomposition via sketching:

$$A\widehat{U}_k = \widehat{U}_k \widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1} e_k^T$$

- Let $\Theta_k(\widehat{U}_k, \widehat{u}_{k+1}) = \min_{v \in \widehat{\mathcal{U}}_k, \|v\|=1} \angle(v, \widehat{u}_{k+1})$. Then

$$\cos(\Theta_k) \leq \varepsilon.$$

- FoV property:

Let λ be an eigenvalue of $\widehat{B}_k$. Then with high probability there exists a unit norm vector $y \in \mathbb{C}^n$ such that

$$|\lambda - y^* A y| \leq \frac{\sqrt{1+\varepsilon}}{\sqrt{1-\varepsilon}} \varepsilon \|A\|.$$

Sketched vs ideal quantities (Simoncini, Wang 2025)

Standard full Krylov (ideal):

- $[U_k, u_{k+1}]$ orthonormal columns
- B_k such that $\mathcal{W}(B_k) \subseteq \mathcal{W}(A)$ (fov)

Krylov decomposition via sketching:

$$A\widehat{U}_k = \widehat{U}_k \widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1}e_k^T$$

- Let $\Theta_k(\widehat{U}_k, \widehat{u}_{k+1}) = \min_{v\in\widehat{\mathcal{U}}_k, \|v\|=1} \angle(v, \widehat{u}_{k+1})$. Then

$$\cos(\Theta_k) \leq \varepsilon.$$

- FoV property:

Let λ be an eigenvalue of $\widehat{B}_k$. Then with high probability there exists a unit norm vector $y \in \mathbb{C}^n$ such that

$$|\lambda - y^*Ay| \leq \frac{\sqrt{1+\varepsilon}}{\sqrt{1-\varepsilon}}\varepsilon\,\|A\|.$$

Sketched vs ideal quantities (Simoncini, Wang 2025)

Standard full Krylov (ideal):

- $[U_k, u_{k+1}]$ orthonormal columns
- B_k such that $\mathcal{W}(B_k) \subseteq \mathcal{W}(A)$ (fov)

Krylov decomposition via sketching:

$$A\widehat{U}_k = \widehat{U}_k \widehat{B}_k + \widehat{u}_{k+1}\widehat{\beta}_{k+1} e_k^T$$

- Let $\Theta_k(\widehat{U}_k, \widehat{u}_{k+1}) = \min_{v \in \widehat{\mathcal{U}}_k, \|v\|=1} \angle(v, \widehat{u}_{k+1})$. Then

$$\cos(\Theta_k) \leq \varepsilon.$$

- FoV property:

Let λ be an eigenvalue of $\widehat{B}_k$. Then with high probability there exists a unit norm vector $y \in \mathbb{C}^n$ such that

$$|\lambda - y^* A y| \leq \frac{\sqrt{1+\varepsilon}}{\sqrt{1-\varepsilon}} \varepsilon \|A\|.$$

Warning on finite precision arithmetic failures

⋆ In exact arithmetic: U_k, $\widehat{U}_k$ full rank and

$$\mathrm{span}(U_k) = \mathrm{span}(\widehat{U}_k) = \mathcal{K}_k(A, c)$$

⋆ In finite precision arithmetic spurious sketched vectors may arise:

Original basis

Assume that U_k is *not* numerically full rank. Then $\mathrm{Range}(U_k) \subset \mathcal{K}_k(A, c)$

Sketched basis

$\widehat{U}_k$ is better conditioned, with high probability, but it will partially build a different subspace than a Krylov subspace

Indeed, let $U_k = \widetilde{U}_k M_k$ be the reduced QR, with M_k numerically singular
For the sketched basis

$$\widehat{U}_k = U_k R_k^{-1} = \widetilde{U}_k (M_k R_k^{-1}),$$

so that $\mathrm{Range}(\widehat{U}_k) = \mathrm{Range}(\widetilde{U}_k)$, where $\mathrm{Range}(\widetilde{U}_k)$ may contain spurious vectors, that is vectors that do not belong to $\mathcal{K}_k(A, c)$

Warning on finite precision arithmetic failures

⋆ In exact arithmetic: U_k, $\widehat{U}_k$ full rank and

$$\mathrm{span}(U_k) = \mathrm{span}(\widehat{U}_k) = \mathcal{K}_k(A, c)$$

⋆ In finite precision arithmetic spurious sketched vectors may arise:

Original basis

Assume that U_k is *not* numerically full rank. Then $\mathrm{Range}(U_k) \subset \mathcal{K}_k(A, c)$

Sketched basis

$\widehat{U}_k$ is better conditioned, with high probability, but it will partially build a different subspace than a Krylov subspace

Indeed, let $U_k = \widetilde{U}_k M_k$ be the reduced QR, with M_k numerically singular
For the sketched basis

$$\widehat{U}_k = U_k R_k^{-1} = \widetilde{U}_k (M_k R_k^{-1}),$$

so that $\mathrm{Range}(\widehat{U}_k) = \mathrm{Range}(\widetilde{U}_k)$, where $\mathrm{Range}(\widetilde{U}_k)$ may contain spurious vectors, that is vectors that do not belong to $\mathcal{K}_k(A, c)$

Warning on finite precision arithmetic failures

⋆ In exact arithmetic: U_k, $\widehat{U}_k$ full rank and

$$\mathrm{span}(U_k) = \mathrm{span}(\widehat{U}_k) = \mathcal{K}_k(A, c)$$

⋆ In finite precision arithmetic spurious sketched vectors may arise:

Original basis

Assume that U_k is *not* numerically full rank. Then $\mathrm{Range}(U_k) \subset \mathcal{K}_k(A, c)$

Sketched basis

$\widehat{U}_k$ is better conditioned, with high probability, but it will partially build a different subspace than a Krylov subspace

Indeed, let $U_k = \widetilde{U}_k M_k$ be the reduced QR, with M_k numerically singular
For the sketched basis

$$\widehat{U}_k = U_k R_k^{-1} = \widetilde{U}_k (M_k R_k^{-1}),$$

so that $\mathrm{Range}(\widehat{U}_k) = \mathrm{Range}(\widetilde{U}_k)$, where $\mathrm{Range}(\widetilde{U}_k)$ may contain spurious vectors, that is vectors that do not belong to $\mathcal{K}_k(A, c)$

Warning on finite precision arithmetic failures

⋆ In exact arithmetic: U_k, $\widehat{U}_k$ full rank and

$$\mathrm{span}(U_k) = \mathrm{span}(\widehat{U}_k) = \mathcal{K}_k(A, c)$$

⋆ In finite precision arithmetic spurious sketched vectors may arise:

Original basis

Assume that U_k is *not* numerically full rank. Then $\mathrm{Range}(U_k) \subset \mathcal{K}_k(A, c)$

Sketched basis

$\widehat{U}_k$ is better conditioned, with high probability, but it will partially build a different subspace than a Krylov subspace

Indeed, let $U_k = \widetilde{U}_k M_k$ be the reduced QR, with M_k numerically singular
For the sketched basis

$$\widehat{U}_k = U_k R_k^{-1} = \widetilde{U}_k (M_k R_k^{-1}),$$

so that $\mathrm{Range}(\widehat{U}_k) = \mathrm{Range}(\widetilde{U}_k)$, where $\mathrm{Range}(\widetilde{U}_k)$ may contain spurious vectors, that is vectors that do not belong to $\mathcal{K}_k(A, c)$

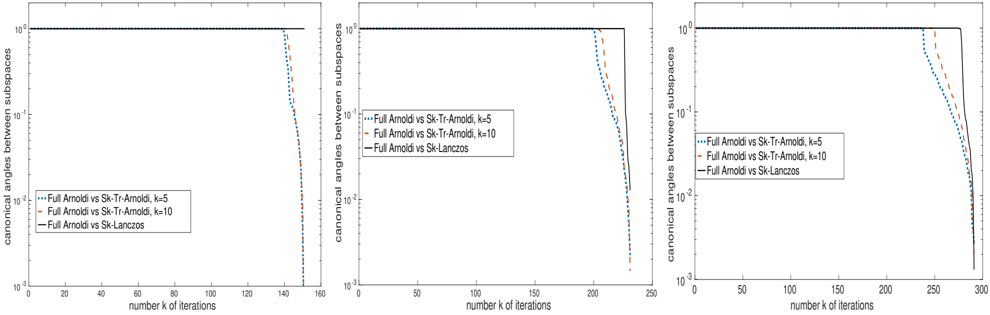
Spurious space. A computational example.

Data: $A \in \mathbb{R}^{n\times n}$ with $n = 16\,641$ stemming from FE discretization of

$$\mathcal{L}(u) = -\epsilon\Delta u + 2y(1-x^2)u_x + 2x(1-y^2)u_y$$

convection diffusion with recirculating wind with $\epsilon = 0.1$, on $[0,1]^2$ and homogeneous bc, IFISS

Methods: Full Krylov vs Sketched k-truncated Arnoldi and Sketched Lanczos



Cosine of all canonical angles after m iterations.
Left: $m = 150$; Middle: $m = 230$; Right: $m = 290$.

Conclusions

- Randomized sketching is a good companion to classical cost-reducing strategies
- 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.

Visit: `www.dm.unibo.it/~simoncin`
Email address: `valeria.simoncini@unibo.it`