Note on a Randomized Krylov Subspace Method

Jonas Bresch*
Technische Universität Berlin

based on Inspiration at the ILAS 2026 at Virginia Tech in Blacksburg,
Inverse Problems and Uncertainty Quantification
through the Lens of Numerical Linear Algebra, MS . 10

The Krylov subspace for a matrix $A \in \mathbb{R}^{n\times n}$ and vector $b \in \mathbb{R}$ is given by

$$\mathcal{K}_k(A, b) \coloneqq \operatorname{span}\{b, Ab, A^2 b, \dots, A^{k-1}b\}. \tag{1}$$

The concept is named after Russian applied mathematician and naval engineer Alexei Krylov, who published a paper about the concept in 1931 [7].

Here, we will consider to solve the optimization problem [4]

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2^2 \tag{2}$$

known as the least square problem via the at most k-dimensional Krylov subspaces $\mathcal{K}_k(A, b)$. This means, we aim to solve [7]

$$x^k \in \operatorname*{arg\,min}_{x\in\mathcal{K}_k(A,b)} \|Ax - b\|_2^2. \tag{3}$$

As a first note, we observe that the solution in the kth iteration is precisely given by solving a system of linear equations for the coordinates with respect to the Krylov subspaces $\mathcal{K}_k(A, b)$, cf. (1), iteratively. Later one, we aim to solve (3) by a random optimization method similar, but different to [9]. This is inspired by [1–3] as well as [6; 8] for the eigenvalue problem and [5; 9; 10] for the randomized methods.

Therefore, lets us denote the vectors of the inducing system for the Krylov subspace (1) by

$$b^0 \coloneqq b, \qquad b^\ell \coloneqq Ab^{\ell-1}, \quad \ell \in \{1, \dots, k-1\},$$

so that

$$b^\ell = Ab^{\ell-1} = AAb^{\ell-2} = A^2 b^{\ell-2} = A^\ell b^0 = A^\ell b.$$

Then for solving (3) we rewrite the variables x in the coordinates of the inducing system of (1), i.e.

$$x = \sum_{\ell=0}^{k-1} \gamma_\ell b^\ell \in \mathcal{K}_k(A, b), \qquad \gamma_\ell \in \mathbb{R}, \quad \ell \in \{0, \dots, k-1\}. \tag{4}$$

We further define the variables

$$\beta_{i,j} \coloneqq \langle b^i, b^j\rangle, \qquad (i, j) \in \{1, \dots, k\} \times \{0, \dots, k\}.$$

*Webpage and Email

24th May 2026

Plugging in (4) in the objective from (3) yields

$$
\begin{aligned}
\|Ax-b\|_2^2 &= \|Ax\|_2^2 - 2\langle Ax, b\rangle + \text{const.} \\
&= \Big\|A\Big(\sum_{\ell=0}^{k-1}\gamma_\ell b^\ell\Big)\Big\|_2^2 - 2\Big\langle A\Big(\sum_{\ell=0}^{k-1}\gamma_\ell b^\ell\Big), b\Big\rangle + \text{const.} \\
&= \Big\|\sum_{\ell=0}^{k-1}\gamma_\ell A b^\ell\Big\|_2^2 - 2\Big\langle \sum_{\ell=0}^{k-1}\gamma_\ell A b^\ell, b\Big\rangle + \text{const.} \\
&= \Big\|\sum_{\ell=0}^{k-1}\gamma_\ell b^{\ell+1}\Big\|_2^2 - 2\Big\langle \sum_{\ell=0}^{k-1}\gamma_\ell b^{\ell+1}, b\Big\rangle + \text{const.} \\
&= \sum_{i,j=0}^{k-1}\gamma_i\gamma_j\langle b^{i+1}, b^{j+1}\rangle - 2\sum_{i=0}^{k-1}\gamma_i\langle b^{i+1}, b^0\rangle + \text{const.} \\
&= \sum_{i,j=0}^{k-1}\gamma_i\gamma_j\beta_{i+1,j+1} - 2\sum_{i=0}^{k-1}\gamma_i\beta_{i+1,0} + \text{const.} \\
&= \sum_{i=0}^{k-1}\gamma_i^2\beta_{i+1,i+1} + 2\sum_{0\le i<j\le k-1}\gamma_i\gamma_j\beta_{i+1,j+1} - 2\sum_{i=0}^{k-1}\gamma_i\beta_{i+1,0} + \text{const.}
\end{aligned}
$$

Hence, the necessary optimality conditions are given by

$$
\begin{aligned}
&\partial_{\gamma_i}\|Ax-b\|_2^2 = 0 \\
&\qquad\Leftrightarrow \quad 0 = 2\gamma_i\beta_{i+1,i+1} + 2\sum_{j\neq i}\gamma_j\beta_{i+1,j+1} - 2\beta_{i+1,0} \\
&\qquad\Leftrightarrow \quad 0 = 2\Big(\sum_{j=0}^{k-1}\gamma_j\beta_{i+1,j+1} - \beta_{i+1,0}\Big), \qquad i\in\{0,\dots,k-1\}.
\end{aligned}
\tag{5}
$$

If we set

$$
B \coloneqq [\beta_{i+1,j+1}]_{i,j=0}^{k-1}\in\mathbb{R}^{k\times k}, \quad \gamma = (\gamma_0,\dots,\gamma_{k-1})^{\mathrm{T}}, \quad \beta_0 \coloneqq (\beta_{1,0},\dots,\beta_{k,0})^{\mathrm{T}}\in\mathbb{R}^k,
$$

then we obtain (5) as a system of linear k linear equation in k variables:

$$
B\gamma = \beta_0. \tag{6}
$$

Then the solution γ^* may defines the solution $x^k = \sum_{\ell=0}^{k-1}\gamma_\ell^* b^\ell$. If B is invertible, i.e. the inducing system $\{b^1,\dots,b^k\} = A\{b^0,\dots,b^{k-1}\}$ from (1) is a basis, then the solution is unique.

Now, to avoid solving (6) we propose to solve instead iteratively a fairly simple update scheme for solving the minimization problem (3) in the kth instance. Therefore, we aim to choose as random update direction w_k where the coefficients within the inducing system $\{b^0,\dots,b^{k-1}\}$ from (1) are standard normal distributed, i.e. let

$$
w_t^k = \sum_{\ell=0}^{k-1}\eta_\ell b^\ell, \qquad \eta_\ell\sim\mathcal{N}(0,1), \quad \text{iid.} \quad \ell\in\{0,\dots,k-1\}.
$$

Then for some appropriate step size $\tau \in \mathbb{R}$ we can define the next iterate

$$x_{t+1}^k = x_t^k + \tau w_t^k = \sum_{\ell=0}^{k-1} \gamma_t^k b^\ell + \tau \sum_{\ell=0}^{k-1} \eta_t^k b^\ell = \sum_{\ell=0}^{k-1} \big(\gamma_t^k + \tau\eta_t^k\big) b^\ell, \tag{7}$$

where we choose for $t = 0$ the initial coefficients for x_0^k standard normal, too, in (1). Hence, we just update the coefficients by the "optimal" step size $\tau \in \mathbb{R}$ we have to determine. Therefore, we consider (7) in the objective rom (3) which yields

$$\begin{aligned} \|Ax_{t+1}^k - b\|_2^2 &= \|A(x_t^k + \tau w_t^k) - b\|_2^2 \\ &= \|A(x_t^k + \tau w_t^k)\|_2^2 - 2\langle A(x_t^k + \tau w_t^k), b\rangle + \text{const.} \\ &= 2\tau\langle Ax_t^k, Aw_t^k\rangle + \tau^2\|Aw_t^k\|_2^2 - 2\tau\langle Aw_t^k, b\rangle + \text{const.} \end{aligned}$$

with necessary optimality criteria

$$\begin{aligned} &\partial_\tau \|Ax_{t+1}^k - b\|_2^2 = 0 \\ \Leftrightarrow \quad & 0 = \langle Ax_t^k, Aw_t^k\rangle + \tau\|Aw_t^k\|_2^2 - \langle Aw_t^k, b\rangle \\ \Leftrightarrow \quad & \tau = \frac{\langle Aw_t^k, b\rangle - \langle Ax_t^k, Aw_t^k\rangle}{\|Aw_t^k\|_2^2} \\ \Leftrightarrow \quad & \tau = \langle Aw_t^k, b - Ax_t^k\rangle \|Aw_t^k\|_2^{-2} \end{aligned}$$

and by the second derivative, which is $\partial_\tau^2 \|Ax_{t+1}^k - b\|_2^2 = 2\|Aw_t^k\|_2^2 \geq 0$, yields the minimizer. As we can express

$$Aw_t^k = A\Big(\sum_{\ell=0}^{k-1} \eta_\ell b^\ell\Big) = \sum_{\ell=0}^{k-1} \eta_\ell A b^\ell = \sum_{\ell=0}^{k-1} \eta_\ell b^{\ell+1}$$

as well as

$$Ax_t^k = A\Big(\sum_{\ell=0}^{k-1} \gamma_\ell b^\ell\Big) = \sum_{\ell=0}^{k-1} \gamma_\ell A b^\ell = \sum_{\ell=0}^{k-1} \gamma_\ell b^{\ell+1}$$

it holds

$$\begin{aligned} \langle Aw_t^k, b - Ax_t^k\rangle &= \Big\langle \sum_{\ell=0}^{k-1} \eta_\ell b^{\ell+1}, b - \sum_{\ell=0}^{k-1} \gamma_\ell b^{\ell+1}\Big\rangle \\ &= \sum_{i=0}^{k-1} \eta_i \langle b^{i+1}, b^0\rangle - \sum_{i,j=0}^{k-1} \eta_i \gamma_i \langle b^{i+1}, b^{j+1}\rangle \\ &= \sum_{i=0}^{k-1} \eta_i \beta_{i+1,0} - \sum_{i,j=0}^{k-1} \eta_i \gamma_j \beta_{i+1,j+1}. \end{aligned}$$

and furthermore

$$\begin{aligned} \|Aw_t^k\|_2^2 &= \Big\|\sum_{\ell=0}^{k-1} \eta_\ell b^{\ell+1}\Big\|_2^2 = \sum_{i,j=0}^{k-1} \eta_i \eta_j \langle b^{i+1}, b^{j+1}\rangle \\ &= \sum_{i,j=0}^{k-1} \eta_i \eta_j \beta_{i+1,j+1} = \sum_{i=0}^{k-1} \eta_i \beta_{i+1,i+1} + 2 \sum_{0 \leq i < j \leq k-1} \eta_i \eta_j \beta_{i+1,j+1} \end{aligned}$$

for an efficient computation.

References

[1] Bresch, J., *Generalization of zeroth-order method for quotients of quadratic functions*, arXiv:2604.26913, submitted to J. Optim. Theory Appl. (April, 2026), 2026. doi: 10.48550/arXiv.2604.26913.

[2] Bresch, J., Lorenz, D. A., Schneppe, F., and Winkler, M., *Matrix-free stochastic calculation of operator norms without using adjoints*, arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (March, 2026), 2024. doi: 10.48550/arXiv.2410.08297.

[3] Bresch, J., Melnyk, O., Schoen, M., and Steidl, G., *Stochastic zeroth-order method for computing the generalized Rayleigh quotient*, arXiv:2512.05520, submitted to Linear Algebra Appl., November, 2025. doi: 10.48550/arXiv.2512.05520.

[4] Golub, G. H., and Van Loan, C. F., *Matrix Computations* (Johns Hopkins Studies in the Mathematical Sciences; 4th, Baltimore, MD, USA: JHU Press, 2013), xiv+756. doi: 10.56021/9781421407944.

[5] Halko, N., Martinsson, P. G., and Tropp, J. A., "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions", *SIAM Review*, 53/2 (2011), 217–88. doi: 10.1137/090771806.

[6] Knyazev, A. V., "Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method", *SIAM J. Sci. Comput.* 23/2 (2001), 517–41. doi: 10.1137/S1064827500366124.

[7] Krylov, A. N., "О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем", *Izv. Akad. Nauk. SSSR Otd Mat. Estest.* 4 (1931), 491–539, http://mi.mathnet.ru//eng/im5215.

[8] Lanczos, C., "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators", *J. Res. Natl. Bur. Stand. B* 45 (1950), 255–82. doi: 10.6028/jres.045.026.

[9] Lorenz, D. A., Schneppe, F., and Tondji, L., "Linearly convergent adjoint free solution of least squares problems by random descent", *Inverse Probl.* 39/12 (2023), No. 125019. doi: 10.1088/1361-6420/ad08ed.

[10] Saibaba, A. K., Lee, J., and Kitanidis, P. K., "Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen–Loève expansion", *Numer. Linear Algebra Appl.* 23/2 (2016), 314–39. doi: 10.1002/nla.2026.