Random sketching of matrices with application to learning preconditioners

O. Balabanov[2], A. Nouy[1] and A. Pasco[1,⋆]

[1] Laboratoire de Mathématiques Jean Leray, École Centrale de Nantes, Nantes Université. ⋆ EPFL - École Polytechnique Fédérale de Lausanne.
[2] International Computer Science Institute; Lawrence Berkeley National Laboratory; Department of Statistics, University of California, Berkeley.

Random sketching for matrices

Goal. Sketch high-dimensional matrices $\mathbf{X}$ accessible only via matrix-vector products.

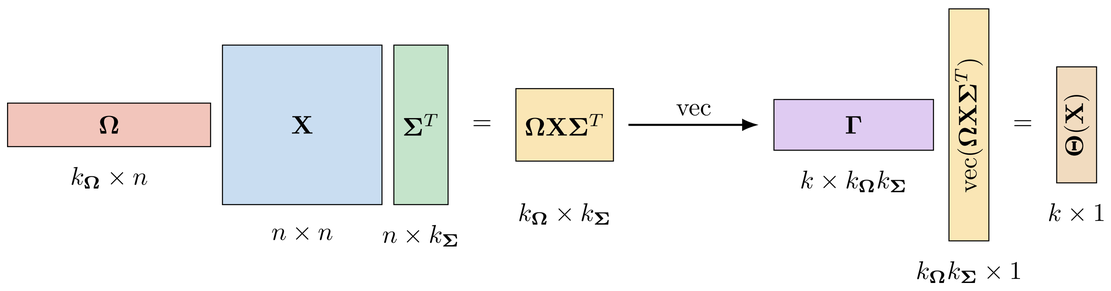
OSE in Frobenius norm. A random linear $\boldsymbol{\Theta} : \mathbb{R}^{n\times n} \to \mathbb{R}^k$ is a (ε, δ, d) oblivious subspace embedding (OSE) if for any d-dim subspace $Z \subset \mathbb{R}^{n\times n}$,

$$\mathbb{P}_{\boldsymbol{\Theta}}\left[\forall \mathbf{X} \in Z,\ (1-\varepsilon)\|\mathbf{X}\|_F^2 \le \|\boldsymbol{\Theta}(\mathbf{X})\|_2^2 \le (1+\varepsilon)\|\mathbf{X}\|_F^2\right] \ge 1-\delta.$$

Structured sketch for matrices. Consider random sketching matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Theta}(\mathbf{X}) := \boldsymbol{\Gamma}\mathrm{vec}(\boldsymbol{\Omega}\mathbf{X}\boldsymbol{\Sigma}^T) \quad \text{with} \quad \boldsymbol{\Gamma}, \boldsymbol{\Omega}, \boldsymbol{\Sigma} \text{ random matrices.}$$

Computed right-to-left using $k_{\boldsymbol{\Sigma}}$ matrix-vector products on $\mathbf{X}$. Example: structured random $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ and Gaussian $\boldsymbol{\Sigma}$.



OSE result for Θ

Let $\boldsymbol{\Sigma} \in \mathbb{R}^{k_{\boldsymbol{\Sigma}}\times n}$ with i.i.d. $\mathcal{N}(0, k_{\boldsymbol{\Sigma}}^{-1})$ entries. Let $\boldsymbol{\Omega} \in \mathbb{R}^{k_{\boldsymbol{\Omega}}\times n}$ be a $(\varepsilon/4, \delta/(3k_{\boldsymbol{\Sigma}}), d)$-OSE. Let $\boldsymbol{\Gamma} \in \mathbb{R}^{k\times k_{\boldsymbol{\Sigma}}k_{\boldsymbol{\Omega}}}$ be a $(\varepsilon/4, \delta/3, d)$-OSE. Assume that $\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are independent. If

$$k_{\boldsymbol{\Sigma}} \ge 168\varepsilon^{-2}(6.9d + \log(6/\delta)),$$

then $\boldsymbol{\Theta}$ is a (ε, δ, d)-OSE.

Sketch of proof: first Hanson-Wright inequality on $\|\mathbf{X}\boldsymbol{\Sigma}^T\|_F^2$, then union-bound argument on the columns of $\mathbf{X}\boldsymbol{\Sigma}^T$, finally propagate error at each sketching step.

See [2] for extension with arbitrary OSE for $\boldsymbol{\Sigma}$, control in Hilbert-Schmidt norm and matrices with complex entries.

Comparison with Khatri Rao embedding. The KR embedding,

$$\boldsymbol{\Theta}^{\mathrm{KR}}(\mathbf{X}) = \mathrm{diag}(\boldsymbol{\Omega}\mathbf{X}\boldsymbol{\Sigma}^T) = (\boldsymbol{\omega}_i^*\mathbf{X}\mathbf{s}_i)_{1\le i\le k}$$

is an OSE when $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are Gaussian with $k = \mathcal{O}(d^{3/2})$ rows [3]. KR is better suited when only vector-matrix-vector is available.

Discrepancy measures for preconditioners

Setting. Approximation of the solution $\mathbf{u} \in \mathbb{R}^n$ of a large ill-conditioned linear system,

$$\mathbf{A}\mathbf{u} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{n\times n}, \quad \mathbf{b} \in \mathbb{R}^n,$$

by a (preconditioned) Galerkin projection $\mathbf{u}_r$ in a fixed $r \ll n$ dimensional subspace $U_r \subset \mathbb{R}^n$ spanned by the orthonormal columns of $\mathbf{U}_r \in \mathbb{R}^{n\times r}$,

$$\mathbf{u}_r = \mathbf{U}_r\mathbf{a}_r \quad \text{where} \quad (\mathbf{U}_r^T\mathbf{P}\mathbf{A}\mathbf{U}_r)\mathbf{a}_r = \mathbf{U}_r^T\mathbf{P}\mathbf{b} \quad \text{and} \quad \mathbf{P} \in \mathbb{R}^{n\times n}.$$

Following [4], $\mathbf{P}$ is a linear combination of implicit $\mathbf{Y}_i \in \mathbb{R}^{n\times n}$, solution to

$$\mathbf{P} = \sum_{i=1}^p c_i^*\mathbf{Y}_i \quad \text{with} \quad \mathbf{c}^* \in \operatorname*{argmin}_{\mathbf{c}\in\mathbf{R}^p} \|\mathbf{I} - \sum_{i=1}^p c_i\mathbf{Y}_i\mathbf{A}\|^2. \tag{1}$$

Galerkin sub-optimality. The seminorm $\|\cdot\| = \|\boldsymbol{\Pi}_{U_r}\cdot\|_2$ is enough,

$$\|\mathbf{u} - \mathbf{u}_r\|_2 \le (1 - \|\mathbf{I} - \mathbf{P}\mathbf{A}\|)^{-1}\|\mathbf{u} - \boldsymbol{\Pi}_{U_r}\mathbf{u}\|_2.$$

Residual norm. For $\mathbf{v} \in U_m \supset U_r$ with $r \le m \ll n$, if $\|\mathbf{u} - \boldsymbol{\Pi}_{U_m}\mathbf{u}\|_2 \le \tau\|\mathbf{u} - \mathbf{v}\|_2$ for some $\tau \in [0, 1)$, then $\|\cdot\| = \|\boldsymbol{\Pi}_{U_m}\cdot\|_2$ is enough,

$$\sqrt{1-\tau^2} - (1+\tau)\|\mathbf{I} - \mathbf{P}\mathbf{A}\| \le \frac{\|\boldsymbol{\Pi}_{U_m}\mathbf{P}(\mathbf{A}\mathbf{v} - \mathbf{b})\|_2}{\|\mathbf{u} - \mathbf{v}\|_2} \le 1 + (1+\tau)\|\mathbf{I} - \mathbf{P}\mathbf{A}\|.$$

Sketched Frobenius norms. Solving (1) is challenging when $\|\cdot\|$ is an operator norm. Replacing with Frobenius norms yields a l.s. problem of size $\mathbb{R}^{nm\times p}$ and is equivalent up to a factor m. Replacing again with sketched Frobenius norms is equivalent up to an additional factor $\mathcal{O}(1+\varepsilon)$ with probability at least $1-\delta$, and reduces (1) to

$$\min_{\mathbf{c}\in\mathbb{R}^p} \|\mathbf{h} - \mathbf{W}\mathbf{c}\|_2^2, \quad \mathbf{W} := (\boldsymbol{\Theta}(\mathbf{Y}_1\mathbf{A}), \cdots, \boldsymbol{\Theta}(\mathbf{Y}_p\mathbf{A})) \in \mathbb{R}^{k\times p}, \quad \mathbf{h} := \boldsymbol{\Theta}(\mathbf{I}).$$

Computing $\boldsymbol{\Theta}(\mathbf{Y}_i\mathbf{A})$ costs $\mathcal{O}(k) = \mathcal{O}(\varepsilon^{-2}(p + \log(1/\delta)))$ matrix-vector with $\mathbf{Y}_i\mathbf{A}$.

Parameter dependent problem

For parameter separable problems we take parameter separable $\mathbf{P}$,

$$\mathbf{A}(\xi) = \sum_{j=1}^{m_{\mathbf{A}}} \theta_j^{\mathbf{A}}(\xi)\mathbf{A}_j, \quad \mathbf{b}(\xi) = \sum_{j=1}^{m_{\mathbf{b}}} \theta_j^{\mathbf{b}}(\xi)\mathbf{b}_j, \quad \mathbf{P}(\xi) = \sum_{j=1}^{p} c_i(\xi)\mathbf{Y}_i.$$

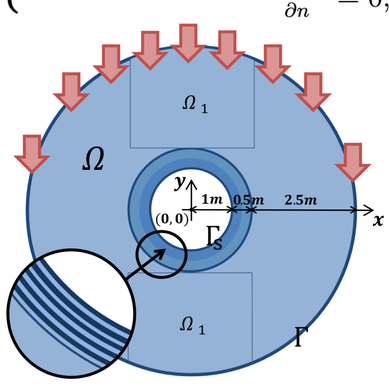
Then the columns of $\mathbf{W}(\xi)$ are also parameter separable with $m_{\mathbf{A}}$ terms. Offline cost: $\min(k, m)$ left matrix-vector products for each $\mathbf{Y}_1, \cdots, \mathbf{Y}_p$. Online cost: assemble and solve the sketched least-squares, $\mathcal{O}(kp(p + m_{\mathbf{A}}))$.

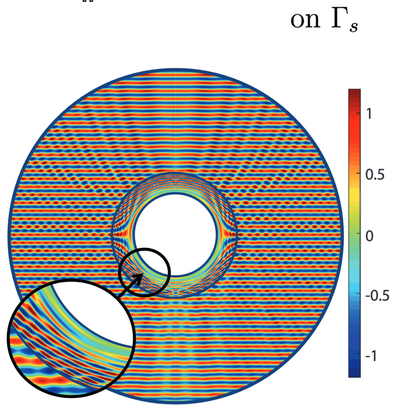
Numerics: acoustic invisibility cloak

Problem setting

An acoustic wave scattering in 2D with a perfect scatterer covered in an invisibility cloak composed of layers of homogeneous isotropic materials [1] is used as illustration. The parameters are the frequency, the density ρ_i and the bulk modulus b_i of 10 liquid metal layers, gathered in a parameter $\xi \in \mathbb{R}^{21}$.

$$\begin{cases} \nabla\cdot(\rho^{-1}\nabla u) + \rho^{-1}\kappa^2 u = 0, & \text{in } \Omega \\ (j\kappa - \frac{1}{2R_\Omega})u + \frac{\partial u}{\partial n} = (j\kappa - \frac{1}{2R_\Omega})u^{in} + \frac{\partial u^{in}}{\partial n}, & \text{on } \Gamma \\ \frac{\partial u}{\partial n} = 0, & \text{on } \Gamma_s \end{cases}$$





Discretization with piecewise quadratic FEM, resulting in around $4 \cdot 10^5$ complex degrees of freedom.

Reduced spaces

U_r and U_m are the first $r = 50$ and $m = 250$ randomized POD modes from a sample of 2000 random snapshots.

Preconditioner space

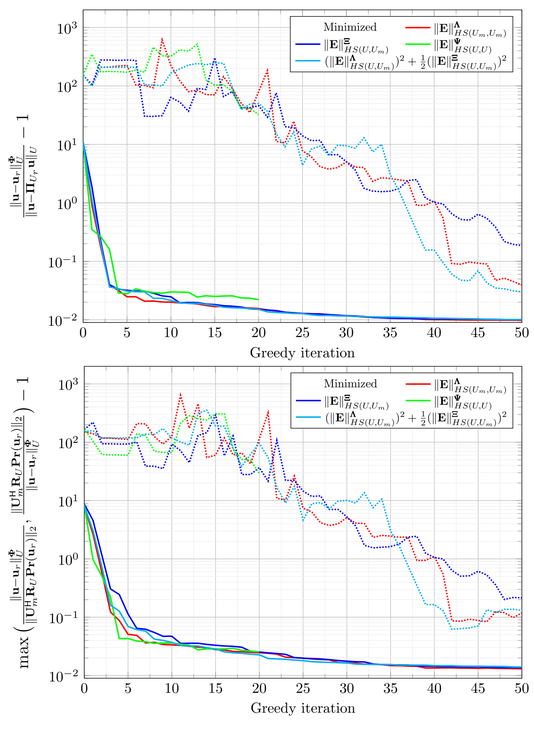
The space of preconditioners is constructed as the span of $\mathbf{Y}_i = \mathbf{A}(\xi_i)^{-1}$ using a greedy algorithm. At iteration i we select

$$\xi_i \in \operatorname*{argmax}_{\xi\in\mathcal{P}_{train}} \min_{\mathbf{P}\in\mathrm{span}\{\mathbf{Y}_0,\cdots,\mathbf{Y}_{i-1}\}} \|\mathbf{I} - \mathbf{P}\mathbf{A}(\xi)\|^{\Theta},$$

with $\#\mathcal{P}_{train} = 99000$. We compare different choices of sketched norm. Sketches sizes $k = 1024$.

Results

Three quantities are monitored along the greedy iterations on an independent random test set of 2000 parameter samples: the error to the solution, the suboptimality of the preconditioned Galerkin projection and the accuracy of the preconditioned residual-based error estimator.



References

[1] O. Balabanov and A. Nouy. Randomized linear algebra for model reduction—part II: Minimal residual methods and dictionary-based approximation. *Advances in Computational Mathematics*, 47(2):26, Apr. 2021.

[2] O. Balabanov, A. Nouy, and A. Pasco. Random sketching of operators with application to learning preconditioners, 2026.

[3] Z. Bujanović, L. Grubišić, D. Kressner, and H. Yin Lam. Subspace embedding with random Khatri–Rao products and its application to eigensolvers. *IMA Journal of Numerical Analysis*, page draf043, June 2025.

[4] O. Zahm and A. Nouy. Interpolation of inverse operators for preconditioning parameter-dependent equations. *SIAM Journal on Scientific Computing*, 38(2):A1044–A1074, 2016.

Conclusion

Take-away:

- Sketched-and-solve in subspace of $\mathbb{R}^{n\times n}$.
- Adapt the norm to purpose of $\mathbf{P}$.
- Implementation available at `https://github.com/alexandre-pasco/rla4mor`

Perspectives:

- Runtime comparisons.
- Sketch for low-rank tensor formats.

LaTeX TikZposter