Fast and Accurate Randomized Algorithms for Linear Systems and Eigenvalue Problems

Yuji Nakatsukasa Joel A. Tropp

SIAM J. Matrix Analysis & Applications
Vol. 45, Iss. 2 (2024)
DOI: 10.1137/23M1565413

Motivation and Background

- In scientific computing and machine learning, solving large-scale linear systems $\mathbf{Ax} = \mathbf{f}$ and eigenvalue problems $\mathbf{Ax} = \lambda\mathbf{x}$ is fundamental.
- Classical algorithms such as GMRES (for linear systems) and Rayleigh–Ritz / eigs (for eigenvalue problems) are accurate but costly for large n.
- **Randomized algorithms** (e.g., sketching) enable fast dimension reduction, making traditional solvers scalable.

Goal: Develop projection-based solvers accelerated by random sketching, retaining accuracy while reducing cost.

Sketching: A Primer

- Sketching: project a high-dimensional problem onto a lower-dimensional subspace using a random matrix.
- Let $\mathbf{S} \in \mathbb{C}^{s \times n}$ be a random sketching matrix with $s \ll n$, such that:

$$\mathbb{E}_{\mathbf{S}} \left[\|\mathbf{S}\mathbf{x}\|_2^2 \right] = \|\mathbf{x}\|_2^2 \quad \text{for all } \mathbf{x} \in \mathbb{C}^n,$$

  where $\mathbb{E}_{\mathbf{S}}$ denotes expectation over the randomness of $\mathbf{S}$, i.e., the average taken over multiple independent sketching matrices sampled from a certain distribution (e.g., Gaussian).
- Leads to small least-squares or eigenvalue subproblems with much lower cost.

Sketching + GMRES (sGMRES)

- Classic GMRES solves $\mathbf{A}\mathbf{x} = \mathbf{f}$ ($\mathbf{A} \in \mathbb{C}^{n\times n}, \quad \mathbf{f} \in \mathbb{C}^{n}$) by finding an approximate solution $\mathbf{x}_B = \mathbf{B}\mathbf{y}$ in the Krylov subspace $\mathcal{K}_d(\mathbf{A}, \mathbf{f})$, where $\mathbf{B} \in \mathbb{C}^{n\times d}, \quad \mathbf{y} \in \mathbb{C}^{d}$.
- It minimizes the residual $\|\mathbf{A}\mathbf{x}_B - \mathbf{f}\|_2$ by solving:

$$\min_{\mathbf{y}\in\mathbb{C}^d} \|\mathbf{A}\mathbf{B}\mathbf{y} - \mathbf{f}\|_2$$

- sGMRES replaces this with a sketched version using a random matrix $\mathbf{S} \in \mathbb{C}^{s\times n}$ $(s \ll n)$:

$$\min_{\mathbf{y}\in\mathbb{C}^d} \|\mathbf{S}(\mathbf{A}\mathbf{B}\mathbf{y} - \mathbf{f})\|_2 \,,$$

- where
$\mathbf{A} \in \mathbb{C}^{n\times n}, \quad \mathbf{B} \in \mathbb{C}^{n\times d}, \quad \mathbf{S} \in \mathbb{C}^{s\times n}, \quad \mathbf{y} \in \mathbb{C}^{d}, \quad \mathbf{f} \in \mathbb{C}^{n}.$

GMRES: Standard Arnoldi Process

- Build a full orthonormal basis for the Krylov subspace $\mathcal{K}_d(\mathbf{A}, \mathbf{f})$

Standard Arnoldi Process

Input: Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, vector $\mathbf{f} \in \mathbb{C}^n$, target dim. d
Output: Basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_d]$

Initialize: $\mathbf{b}_1 \leftarrow \mathbf{f}/\|\mathbf{f}\|_2$
for $j = 2, 3, \ldots, d$ **do**
  $\mathbf{v} \leftarrow \mathbf{A}\mathbf{b}_{j-1}$
  for $i = 1, \ldots, j-1$ **do**
    $\mathbf{v} \leftarrow \mathbf{v} - \langle \mathbf{v}, \mathbf{b}_i \rangle \mathbf{b}_i$
  end for
  $\mathbf{b}_j \leftarrow \mathbf{v}/\|\mathbf{v}\|_2$
end for

Note: Full orthogonalization ensures numerical stability but is expensive.

sGMRES: Truncated Arnoldi Process

- **Key idea:** Only orthogonalize against the most recent k vectors to reduce computational cost.

Truncated Arnoldi Process

Input: Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, vector $\mathbf{f} \in \mathbb{C}^{n}$, target dim. d, truncation k
Output: Basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_d]$

Initialize: $\mathbf{b}_1 \leftarrow \mathbf{f}/\|\mathbf{f}\|_2$
for $j = 2, 3, \ldots, d$ **do**
  $\mathbf{v} \leftarrow \mathbf{A}\mathbf{b}_{j-1}$
  for $i = \max(1, j-k), \ldots, j-1$ **do**
    $\mathbf{v} \leftarrow \mathbf{v} - \langle \mathbf{v}, \mathbf{b}_i \rangle \mathbf{b}_i$
  end for
  $\mathbf{b}_j \leftarrow \mathbf{v}/\|\mathbf{v}\|_2$
end for

Theoretical Guarantees: sGMRES

- With high probability, the sketching matrix $\mathbf{S} \in \mathbb{C}^{s \times n}$ (e.g., Gaussian, SRHT) satisfies:

$$(1-\varepsilon)\|\mathbf{r}\|_2 \leq \|\mathbf{Sr}\|_2 \leq (1+\varepsilon)\|\mathbf{r}\|_2 \quad \forall \mathbf{r} \in \text{range}(\mathbf{AB})$$

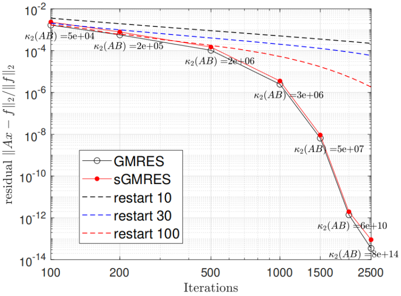
- Implies approximate solution $\mathbf{x}_B = \mathbf{B}\mathbf{y}^\star$ from sGMRES satisfies:

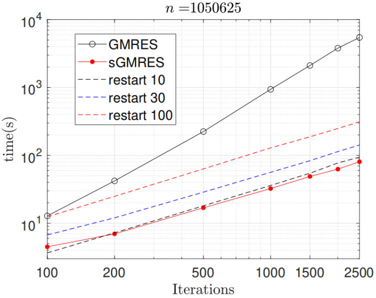
$$\|\mathbf{A}\mathbf{x}_B - \mathbf{f}\|_2 \leq (1+\varepsilon) \min_{\mathbf{y}} \|\mathbf{ABy} - \mathbf{f}\|_2$$

- Sketch size $s = \mathcal{O}(d \log d)$ suffices for $(1 \pm \varepsilon)$-accuracy with high probability.

Numerical Comparison: GMRES vs sGMRES

- **GMRES:** $\mathcal{O}(nd^2)$ operations
- **sGMRES:** $\mathcal{O}(d^3 + nd \log d)$ operations
- Up to $70\times$ faster for PDE discretizations (e.g., convection-diffusion)





Rayleigh–Ritz

- Given: matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, and subspace basis $\mathbf{B} \in \mathbb{C}^{n \times d}$ with orthonormal columns.

- **Rayleigh–Ritz as projection:** Find the best approximation to $\mathbf{AB}$ in $\mathrm{range}(\mathbf{B})$ by minimizing:

$$\min_{\mathbf{M} \in \mathbb{C}^{d \times d}} \quad \|\mathbf{AB} - \mathbf{BM}\|_F^2$$

- **Closed-form solution:**

$$\mathbf{M} = \mathbf{B}^H \mathbf{AB}$$

- Solve eigenproblem: $\quad \mathbf{Mu}_i = \theta_i \mathbf{u}_i$
- Back-project eigenvectors: $\quad \mathbf{x}_i = \mathbf{Bu}_i, \quad$ with $\lambda_i \approx \theta_i$

Sketching + Rayleigh–Ritz (sRR)

- Let $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times d}$, $\mathbf{S} \in \mathbb{C}^{s \times n}$ with $s \ll n$.
- **sRR as sketched projection:** Approximate $\mathbf{AB}$ within $\mathrm{range}(\mathbf{B})$ by solving:

$$\min_{\mathbf{M} \in \mathbb{C}^{d \times d}} \|\mathbf{S}(\mathbf{AB} - \mathbf{BM})\|_F$$

- **Closed-form solution:**

$$\widehat{\mathbf{M}} = (\mathbf{SB})^{\dagger}(\mathbf{SAB})$$

- Solve eigenproblem: $\widehat{\mathbf{M}}\mathbf{u}_i = \theta_i \mathbf{u}_i$; back-project eigenvectors:

$$\mathbf{x}_i \approx \mathbf{B}\mathbf{u}_i$$

- Accurate even if $\mathbf{B}$ is poorly conditioned; sketch size $s = \mathcal{O}(d \log d)$ suffices with high probability.

Theoretical Guarantees

- Suppose $\mathbf{B} \in \mathbb{C}^{n \times d}$ spans a good approximate invariant subspace of $\mathbf{A} \in \mathbb{C}^{n \times n}$.
- Let $\mathbf{S} \in \mathbb{C}^{s \times n}$ be a random matrix with i.i.d. sub-Gaussian entries (or SRHT), and define:

$$\widehat{\mathbf{M}} = (\mathbf{SB})^{\dagger}(\mathbf{SAB})$$

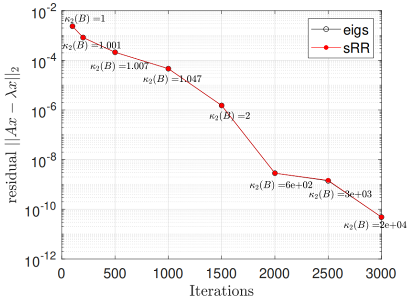
- Then for any $0 < \varepsilon < 1$, if $s \geq C \cdot d \log(d/\delta)/\varepsilon^2$, we have with probability at least $1 - \delta$:

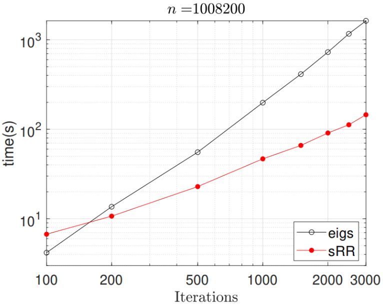
$$\left\| \widehat{\mathbf{M}} - \mathbf{B}^{*}\mathbf{A}\mathbf{B} \right\| \leq \varepsilon \|\mathbf{A}\|_2$$

- Consequently, the eigenvalues of $\widehat{\mathbf{M}}$ approximate those of $\mathbf{A}$ in range$(\mathbf{B})$ up to $\mathcal{O}(\varepsilon)$ error.

Comparison: eigs vs sRR

- **eigs:** classic Arnoldi + RR, expensive orthogonalization
- **sRR:** fast basis + sketching = 10× speedup
- Accuracy preserved, suitable for optimization subproblems





Conclusion

- Presented **sGMRES** and **sRR** for efficient solution of linear systems and eigenproblems.
- Sketching reduces dimension with little accuracy loss.
- Scalable tools for modern large-scale computations.