A Subspace Constrained Randomized Kaczmarz Method for Structure or External Knowledge Exploitation

Jackie Lok[1] Elizaveta Rebrova[1]

[1]Princeton University

Problem setup

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, be full rank. Consider solving a consistent, overdetermined system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \text{with unique solution} \quad \mathbf{x}^* \in \mathbb{R}^n.$$

Fix a subset $I_0 \subseteq [m]$ of rows, and let $\mathbf{P} := \mathbf{I} - \mathbf{A}_{I_0}^\dagger \mathbf{A}_{I_0}$ be the orthogonal projector onto $\mathrm{Null}(\mathbf{A}_{I_0})$. Suppose that the initial iterate $\mathbf{x}^0$ satisfies $\mathbf{A}_{I_0}\mathbf{x}^0 = \mathbf{b}_{I_0}$.

$$\begin{pmatrix} \mathbf{A}_{I_0} \\ \mathbf{A}_{I_1} \end{pmatrix}$$

Subspace constrained randomized Kaczmarz (SCRK)

The SCRK method modifies the *RK method* (Strohmer, Vershynin, 2009) to confine the iterates within a *selected solution space* $\mathbf{A}_{I_0}\mathbf{x} = \mathbf{b}_{I_0}$.

Algorithm. In each iteration, a row $j \in I_1 := [m] \setminus I_0$ is sampled with probability $\|\mathbf{P}\mathbf{a}_j\|^2 / \|\mathbf{A}_{I_1}\mathbf{P}\|_F^2$, and the current iterate $\mathbf{x}^k$ is projected onto the hyperplane $\mathbf{A}_{I_0 \cup \{j\}}\mathbf{x} = \mathbf{b}_{I_0 \cup \{j\}}$ using a simple formula:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \frac{(b_j - \mathbf{a}_j^\intercal \mathbf{x}^k)}{\|\mathbf{P}\mathbf{a}_j\|} \cdot \frac{\mathbf{P}\mathbf{a}_j}{\|\mathbf{P}\mathbf{a}_j\|}. \qquad (\triangle)$$

Theorem 1. The SCRK iterates $\mathbf{x}^k$ satisfy

$$\mathbb{E}\|\mathbf{x}^k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\sigma_{\min}^+(\mathbf{A}_{I_1}\mathbf{P})^2}{\|\mathbf{A}_{I_1}\mathbf{P}\|_F^2}\right)^k \cdot \|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

- SCRK convergence rate depends on restricted singular values: e.g.
$$\sigma_{\min}^+(\mathbf{A}_{I_1}\mathbf{P}) = \min_{\mathbf{x} \in \mathrm{Null}(\mathbf{A}_{I_0}) \cap \mathbb{S}^{n-1}} \|\mathbf{A}\mathbf{x}\|.$$
- $I_0 = \emptyset \Rightarrow$ RK rate based on scaled condition number $\|\mathbf{A}\|_F / \sigma_{\min}(\mathbf{A})$.

Upshot: *If we can find a (small) subset of rows I_0 such that $\|\mathbf{A}_{I_1}\mathbf{P}\|_F \ll \|\mathbf{A}\|_F$ or $\sigma_{\min}^+(\mathbf{A}_{I_1}\mathbf{P}) \gg \sigma_{\min}(\mathbf{A})$, then SCRK $\gg$ RK.*

Exploiting (low-rank) structure

Question: Suppose that $\mathbf{A}$ is effectively low-rank in the sense that

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \sigma_i(\mathbf{A})^2 \gg \sum_{i=r+1}^n \sigma_i(\mathbf{A})^2 = \|\mathbf{A} - \mathbf{A}_{(r)}\|_F^2$$

for some $r > 0$. *How can we use this to accelerate convergence?*

Connection with approximate CX decompositions (*randomized numerical linear algebra*) $\Rightarrow$ can find a good subspace by randomly sampling rows!

Theorem 2. If I_0 contains $O(r \log r / \varepsilon^2)$ rows of $\mathbf{A}$, sampled w.p. proportional to the *leverage scores* of $\mathbf{A}$ relative to its best rank-r approximation, then with prob. ≥ 0.9 over the randomness in I_0,

$$\mathbb{E}\|\mathbf{x}^k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\sigma_{\min}(\mathbf{A})^2}{(1+\varepsilon)\sum_{i=r+1}^n \sigma_i(\mathbf{A})^2}\right)^k \cdot \|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

Corrupted linear systems

Furthermore, consider solving a corrupted linear system, where the goal is to reconstruct $\mathbf{x}^*$ given a set of *corrupted measurements*

$$\widetilde{\mathbf{b}} = \mathbf{b} + \mathbf{b}_{\mathcal{C}}, \quad \text{with } \mathbf{b}_{\mathcal{C}} \text{ a sparse vector supported on } \mathcal{C} \subseteq [m].$$

The QuantileSCRK method adapts the SCRK method to solve this problem, following the *QuantileRK method* introduced by (Haddock et al., 2022). Given a quantile parameter $q \in (0,1]$, the projection ($\triangle$) is made *only if the residual $|b_j - \mathbf{a}_j^\intercal \mathbf{x}^k|$ of the sampled row is less than the qth quantile of all the residuals.*

Exploiting external knowledge

Question: Suppose we have external knowledge about corruption-free measurements: i.e. $I_0 \subseteq [m]$ with $|I_0| = m_0$ such that $(\mathbf{b}_{\mathcal{C}})_{I_0} = \mathbf{0}$. *How can we exploit this to accelerate (or enable) convergence?*

We can show that *if the effective aspect ratio $(m - m_0)/(n - m_0)$ is tall enough, and the proportion of corruptions $\beta := |\mathcal{C}|/(m - m_0)$ is not too large, then QuantileSCRK converges robustly and quickly for generic "homogeneous" matrices $\mathbf{A}$.*

Theorem 3. Suppose that $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a *"Gaussian-like" random matrix*. There exist constants R, β_0, c such that if

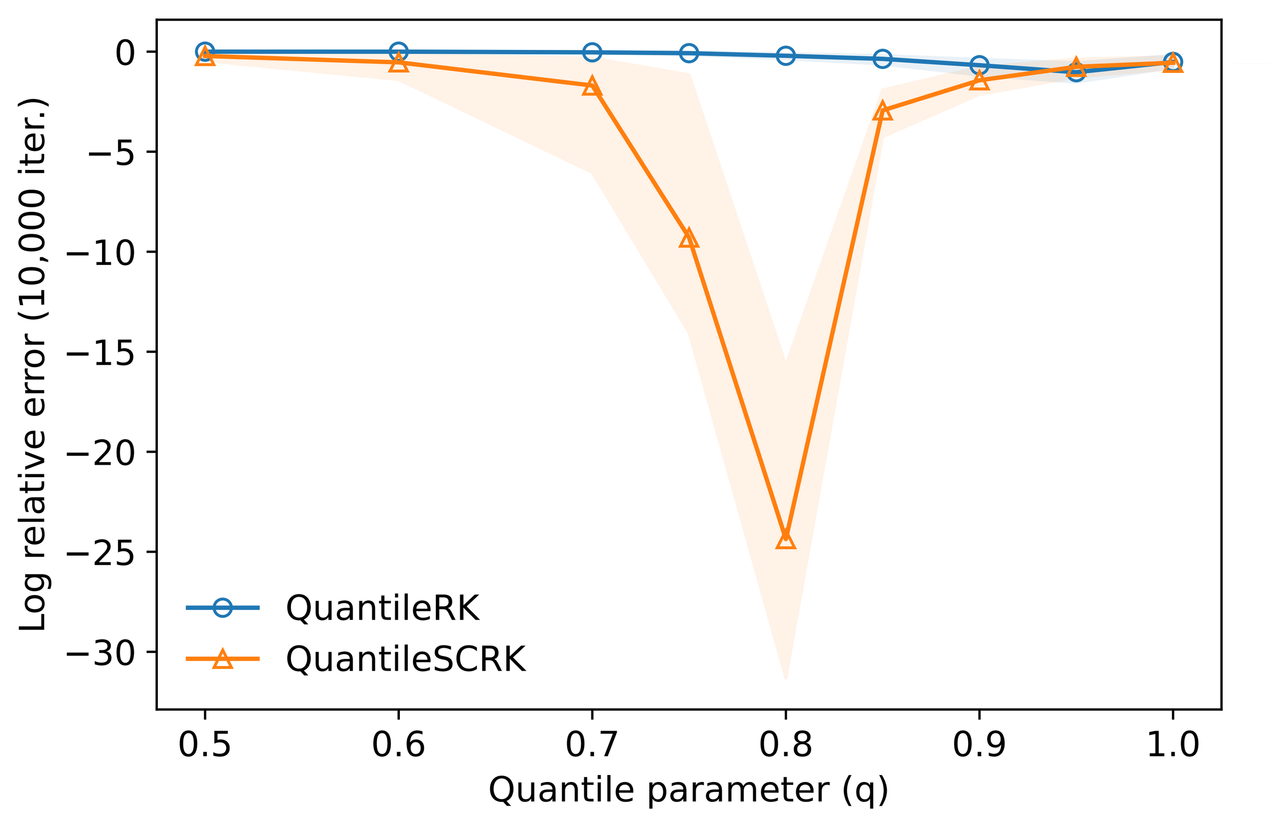
$$\frac{m - m_0}{n - m_0} \geq R \quad \text{and} \quad \beta := \frac{|\mathcal{C}|}{m - m_0} \leq \beta_0,$$

then with prob. at least $1 - e^{-c(m - m_0)}$ over the randomness in $\mathbf{A}$, the QuantileSCRK iterates $\mathbf{x}^k$ converge to $\mathbf{x}^*$ with

$$\mathbb{E}\|\mathbf{x}^k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{c}{n - m_0}\right)^k \cdot \|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

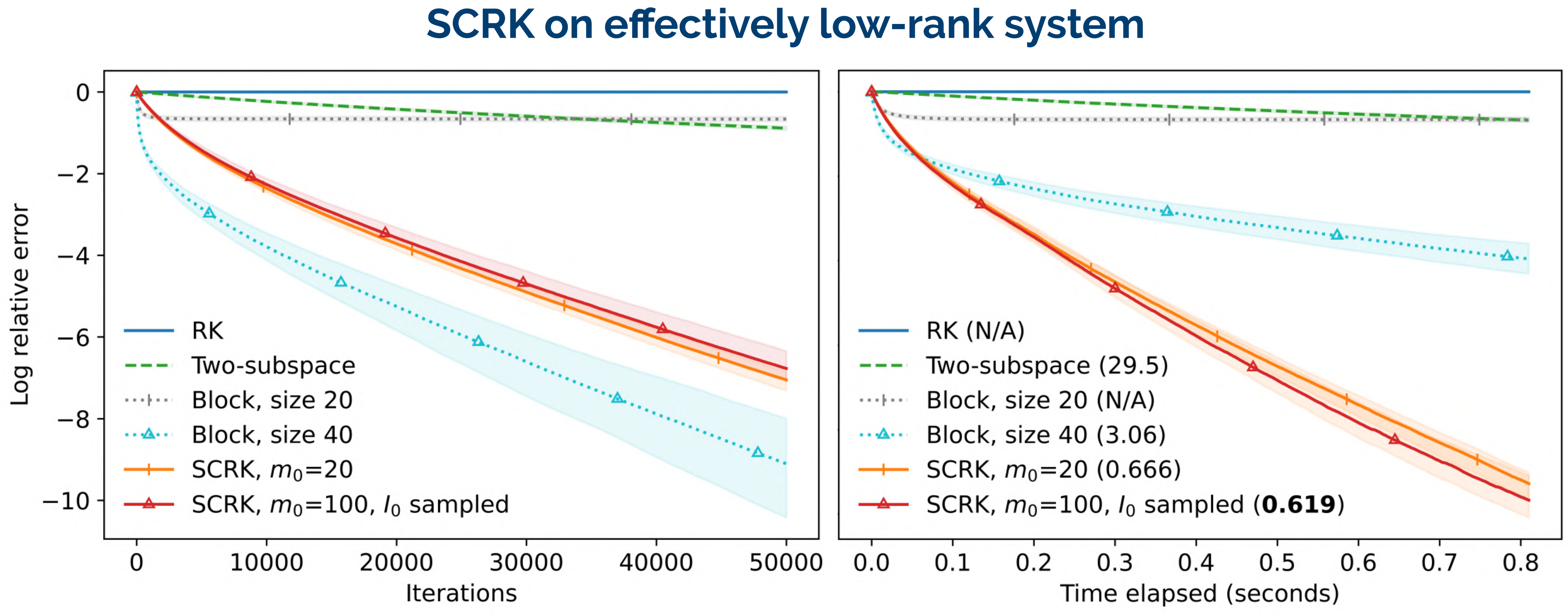
Proof ideas: combine a deterministic convergence criterion with results showing that $\sigma_{\min}^+(\mathbf{A}_{I_1}\mathbf{P}) \gtrsim (q - \beta)^{3/2}\sqrt{m - m_0}$ and $\sigma_{\max}(\mathbf{A}_{I_1}\mathbf{P}) \lesssim \sqrt{m - m_0}$ w.h.p. using tools from high-dimensional probability.

QuantileSCRK on corrupted Gaussian system



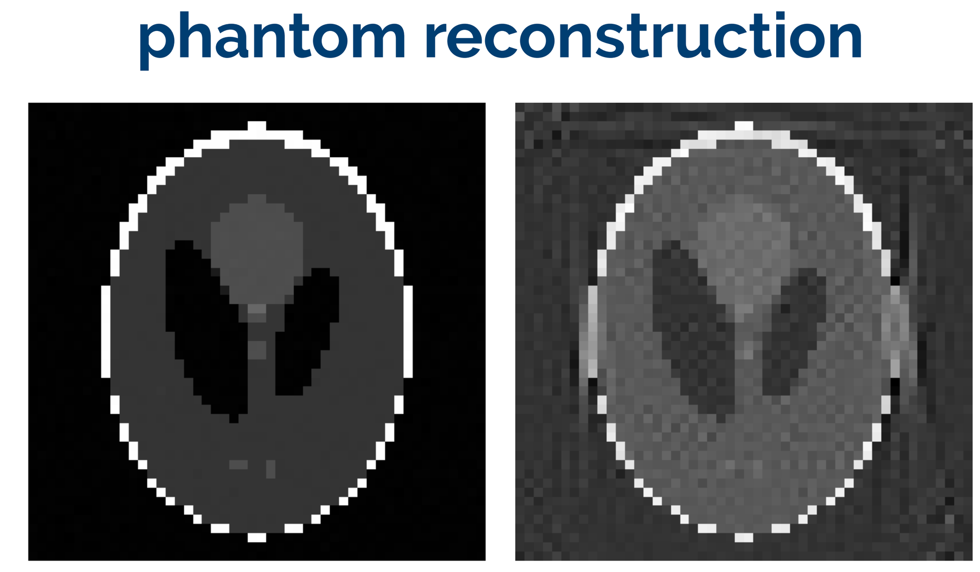
QuantileSCRK with $q \sim 0.8$ and $m_0 = 75$ converges effectively on an almost-square Gaussian system $\mathbf{A} \in \mathbb{R}^{130 \times 100}$ with $\sim 8\%$ corrupted measurements.

SCRK on effectively low-rank system



Each row of $\mathbf{A} \in \mathbb{R}^{2{,}000 \times 1{,}000}$ is given by $\mathbf{a}_j = 0.9\mathbf{a}'_j + 0.1\mathbf{c}_j$, where $\mathbf{a}'_j$ and $\mathbf{c}_j$ are unit vectors drawn from a fixed 20-dimensional subspace $\mathcal{U}$ and its orthogonal complement $\mathcal{U}^\perp$.

QuantileSCRK corrupted phantom reconstruction



QuantileSCRK (right) offers a good reconstruction of an underlying phantom (left) with sensing matrix $\mathbf{A} \in \mathbb{R}^{4{,}500 \times 2{,}500}$ and a significant number ($1{,}125$ or 25%) of corruptions using $m_0 = 500$ trusted measurements (e.g. good sensors), $q = 0.7$, and $k = 60m$ iterations.

arXiv:2309.04889; Linear Algebra Appl. 698 (2024)
https://jackielok.github.io/ jackie.lok@princeton.edu