

Lecture 3: Randomized Iterative Methods for Linear Systems



School of Mathematical Sciences, Xiamen University

1. Pseudoinverse solutions of linear systems

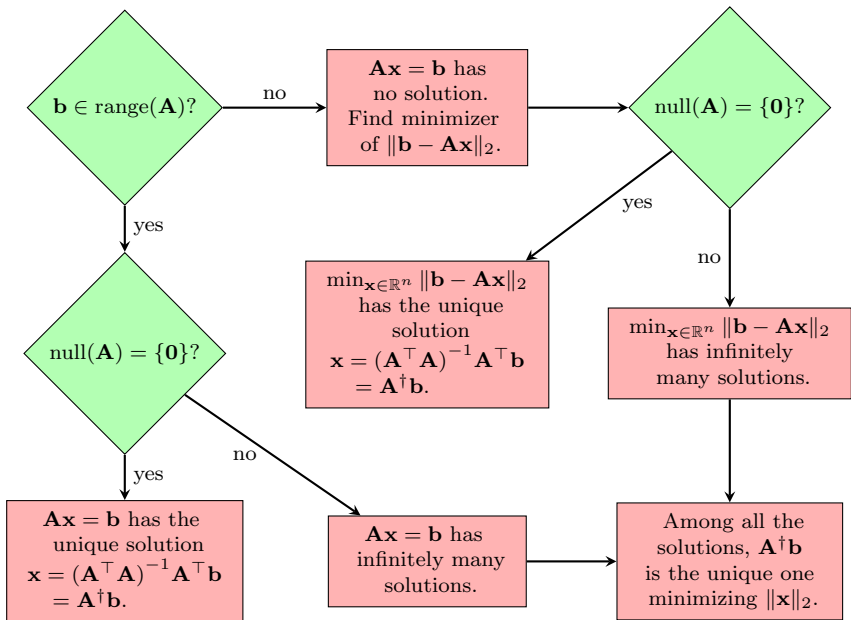
- Consider a linear system of equations

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

The system is called *consistent* if $\mathbf{b} \in \text{range}(\mathbf{A})$, otherwise, *inconsistent*.

- We are interested in the pseudoinverse solution $\mathbf{A}^\dagger \mathbf{b}$, where \mathbf{A}^\dagger denotes the Moore–Penrose pseudoinverse of \mathbf{A} .

$\mathbf{Ax} = \mathbf{b}$	$\text{rank}(\mathbf{A})$	$\mathbf{A}^\dagger \mathbf{b}$
consistent	$= n$	unique solution
consistent	$< n$	unique minimum ℓ_2 -norm solution
inconsistent	$= n$	unique least-squares (LS) solution
inconsistent	$< n$	unique minimum ℓ_2 -norm LS solution

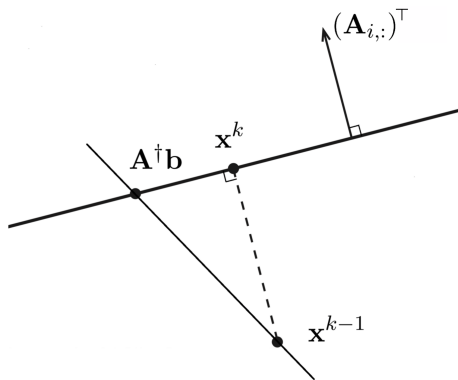


2. Randomized Kaczmarz (RK) (Strohmer & Vershynin 2009)

- Kaczmarz method projects \mathbf{x}^{k-1} onto $\{\mathbf{x} \mid \mathbf{A}_{i,:}\mathbf{x} = \mathbf{b}_i\}$,

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - \mathbf{b}_i}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^\top,$$

where $\mathbf{A}_{i,:}$ is the i th row of \mathbf{A} and \mathbf{b}_i is the i th component of \mathbf{b} .



Algorithm: RK for $\mathbf{Ax} = \mathbf{b}$

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$

for $k = 1, 2, \dots$, **do**

Pick $i \in [m]$ with probability $\frac{\|\mathbf{A}_{i,:}\|_2^2}{\|\mathbf{A}\|_F^2}$

Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - \mathbf{b}_i}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^\top$

end

- Suppose that $\mathbf{b} \in \text{range}(\mathbf{A})$. The convergence result:

$$\mathbb{E} [\|\mathbf{x}^k - \mathbf{x}_\star^0\|_2^2] \leq \rho^k \|\mathbf{x}^0 - \mathbf{x}_\star^0\|_2^2,$$

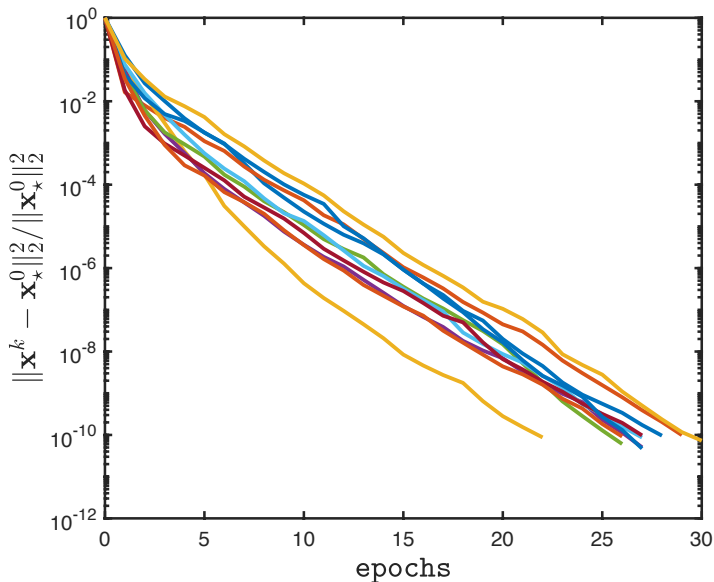
where

$$\mathbf{x}_\star^0 = (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}^0 + \mathbf{A}^\dagger \mathbf{b}$$

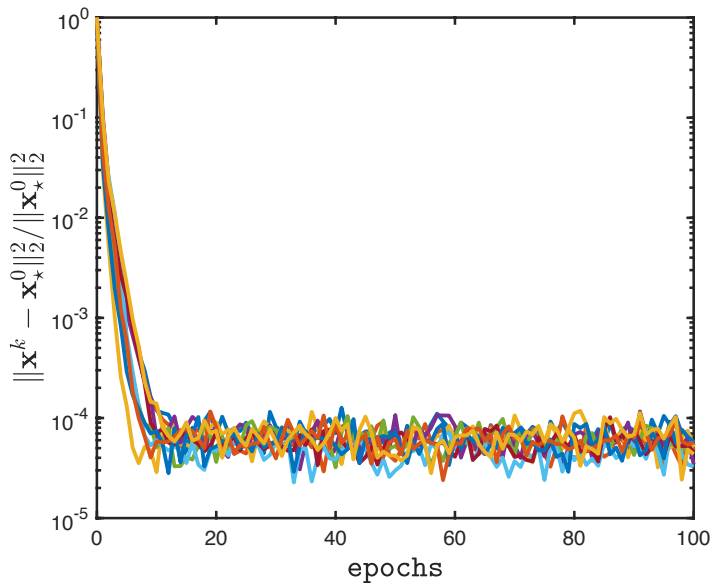
and

$$\rho = 1 - \frac{\sigma_r^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}.$$

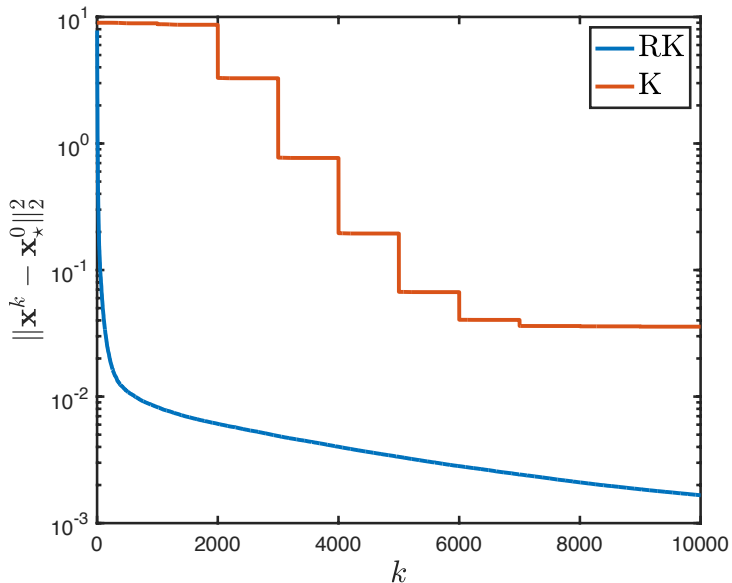
- Consistent case: relative error. $\mathbf{x}_\star^0 := \mathbf{A}^\dagger \mathbf{b} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}^0$.



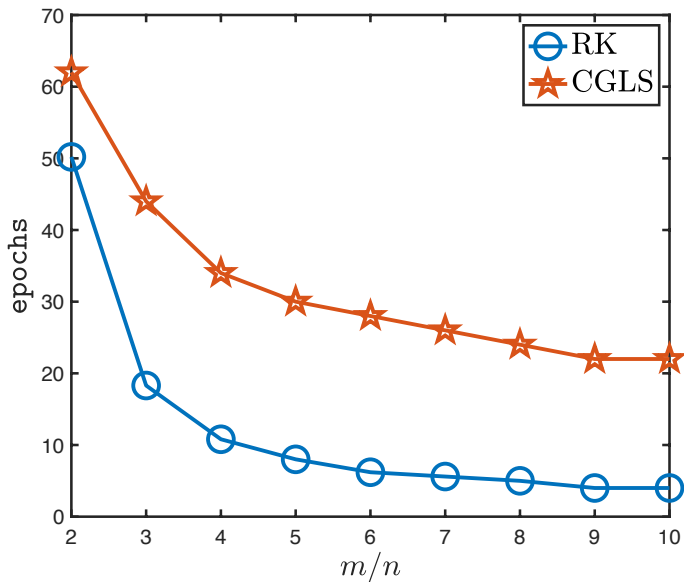
- Inconsistent case: relative error



- RK vs Kaczmarz (phillips)



• RK vs CGLS



3. Randomized coordinate descent (RCD)

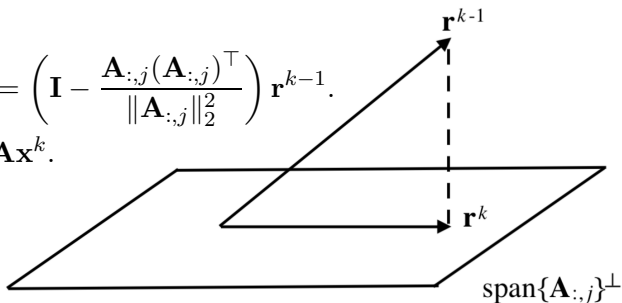
- RCD or RGS: (Leventhal & Lewis 2010)

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top (\mathbf{A}\mathbf{x}^{k-1} - \mathbf{b})}{\|\mathbf{A}_{:,j}\|_2^2} \mathbf{I}_{:,j},$$

where $\mathbf{A}_{:,j}$ is the j th column of \mathbf{A} and $\mathbf{I}_{:,j}$ is the j th column of the $n \times n$ identity matrix \mathbf{I} .

- The residual $\mathbf{r}^k = \left(\mathbf{I} - \frac{\mathbf{A}_{:,j}(\mathbf{A}_{:,j})^\top}{\|\mathbf{A}_{:,j}\|_2^2} \right) \mathbf{r}^{k-1}$.

Here $\mathbf{r}^k := \mathbf{b} - \mathbf{A}\mathbf{x}^k$.



We have $\mathbf{A}(\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b}) \rightarrow \mathbf{0}$ and $\mathbf{r}^k \rightarrow (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}$.

Algorithm: RCD for $\mathbf{Ax} = \mathbf{b}$

Initialize $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{r}^0 = \mathbf{b}$

for $k = 1, 2, \dots$ **do**

Select $j \in [n]$ randomly with probability $\frac{\|\mathbf{A}_{:,j}\|^2}{\|\mathbf{A}\|_F^2}$

Compute $w_k = \frac{(\mathbf{A}_{:,j})^\top \mathbf{r}^{k-1}}{\|\mathbf{A}_{:,j}\|^2}$

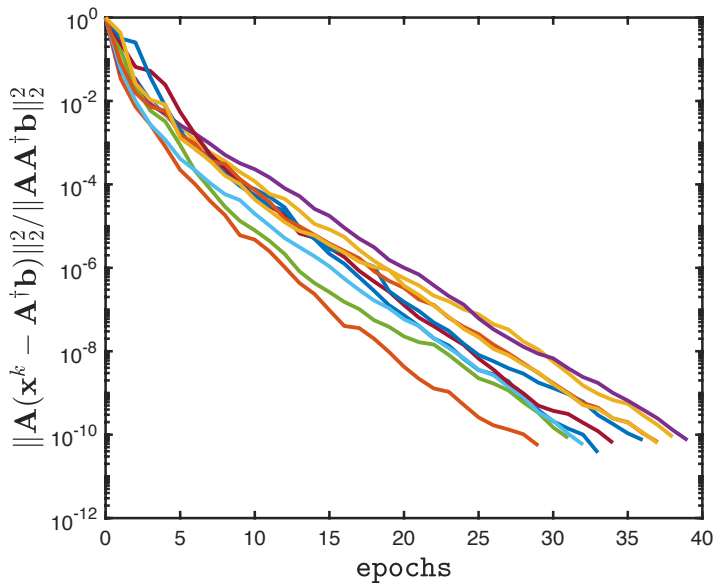
Update $\mathbf{x}_j^k = \mathbf{x}_j^{k-1} + w_k$ and $\mathbf{r}^k = \mathbf{r}^{k-1} - w_k \mathbf{A}_{:,j}$

end for

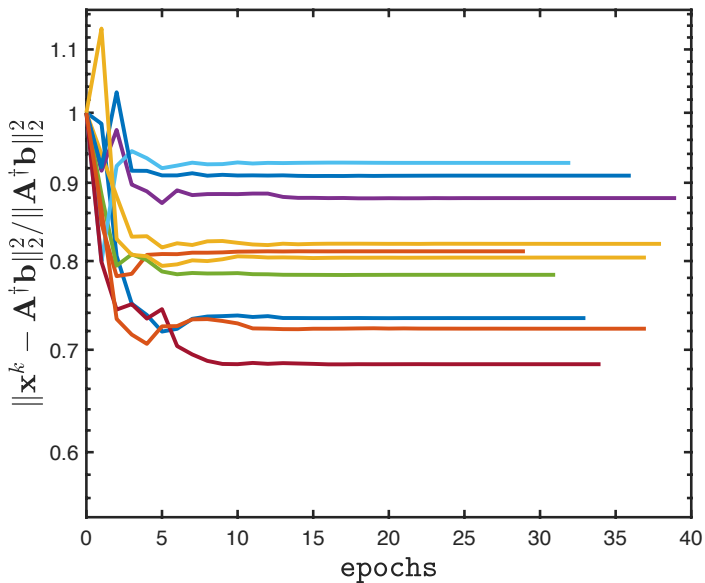
- The convergence result:

$$\mathbb{E} [\|\mathbf{A}(\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b})\|^2] \leq \left(1 - \frac{\sigma_r^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\right)^k \|\mathbf{A}(\mathbf{x}^0 - \mathbf{A}^\dagger \mathbf{b})\|^2.$$

- Rank-deficient case: relative residual



- Rank-deficient case: relative error



4. Randomized extended Kaczmarz (REK) (Zouzias & Freris 2013)

- The normal equations $\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{A}^\top \mathbf{b}$ can be written as

$$\mathbf{A}^\top \mathbf{z} = \mathbf{0}, \quad \mathbf{A} \mathbf{x} = \mathbf{b} - \mathbf{z}.$$

- RK for $\mathbf{A}^\top \mathbf{z} = \mathbf{0}$ with $\mathbf{z}^0 \in \mathbf{b} + \text{range}(\mathbf{A})$ yields $\{\mathbf{z}^k\}_0^\infty$ satisfying

$$\mathbf{z}^k \rightarrow (\mathbf{I} - \mathbf{A} \mathbf{A}^\dagger) \mathbf{b} \quad \text{as } k \rightarrow \infty.$$

Then $\mathbf{A} \mathbf{x} = \mathbf{b} - \mathbf{z}^k \rightarrow \mathbf{A} \mathbf{x} = \mathbf{A} \mathbf{A}^\dagger \mathbf{b}$, which is consistent.

- REK solves $\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{A}^\top \mathbf{b}$ via intertwining an iterate of RK on $\mathbf{A}^\top \mathbf{z} = \mathbf{0}$ with an iterate of RK on $\mathbf{A} \mathbf{x} = \mathbf{b} - \mathbf{z}$.

$$\begin{aligned} \mathbf{z}^k &= \mathbf{z}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top \mathbf{z}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2} \mathbf{A}_{:,j}, \\ \mathbf{x}^k &= \mathbf{x}^{k-1} - \frac{\mathbf{A}_{i,:} \mathbf{x}^{k-1} - \mathbf{b}_i + \mathbf{z}_i^k}{\|\mathbf{A}_{i,:}\|_2^2} (\mathbf{A}_{i,:})^\top. \end{aligned}$$

Algorithm: REK for $\mathbf{A}^\top \mathbf{A} \mathbf{x} = \mathbf{A}^\top \mathbf{b}$

Initialize $\mathbf{z}^0 \in \mathbf{b} + \text{range}(\mathbf{A})$ and $\mathbf{x}^0 \in \mathbb{R}^n$

for $k = 1, 2, \dots$, **do**

Pick $j \in [n]$ with probability $\frac{\|\mathbf{A}_{:,j}\|_2^2}{\|\mathbf{A}\|_F^2}$

Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top \mathbf{z}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2} \mathbf{A}_{:,j}$

Pick $i \in [m]$ with probability $\frac{\|\mathbf{A}_{i,:}\|_2^2}{\|\mathbf{A}\|_F^2}$

Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{A}_{i,:} \mathbf{x}^{k-1} - \mathbf{b}_i + \mathbf{z}_i^k}{\|\mathbf{A}_{i,:}\|_2^2} (\mathbf{A}_{i,:})^\top$

end

- The convergence result:

$$\mathbb{E} [\|\mathbf{x}^k - \mathbf{x}_\star^0\|^2] \leq \rho^k \|\mathbf{x}^0 - \mathbf{x}_\star^0\|^2 + \frac{k\rho^k}{\|\mathbf{A}\|_F^2} \|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\|^2.$$

5. Summary of randomized iterative algorithms

- Randomized iterative algorithms are preferable if the coefficient matrix \mathbf{A} is too large to fit in memory, or the matrix-vector product $\mathbf{A}\mathbf{v}$ is considerably expensive.
- Consistent: RK and its variants
- Inconsistent, full-column rank: RCD and its variants
- Inconsistent, rank-deficient: REK and its variants