Lecture 2: Randomized iterative methods for linear systems

School of Mathematical Sciences, Xiamen University

1. The pseudoinverse solution of linear systems

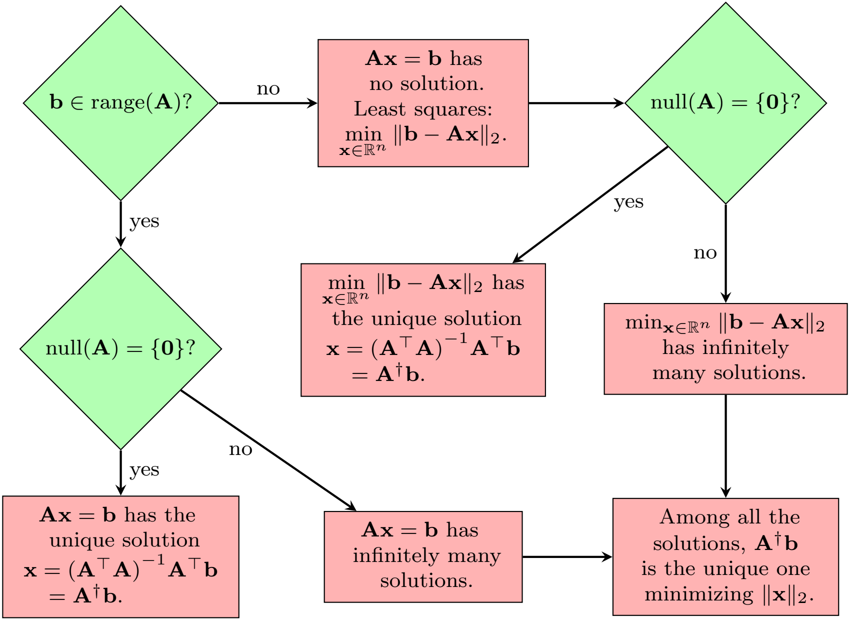
- Consider a linear system of equations

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

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

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

$\mathbf{A}\mathbf{x} = \mathbf{b}$	$\mathrm{rank}(\mathbf{A})$	$\mathbf{A}^\dagger\mathbf{b}$
consistent	$= n$	unique solution
consistent	$< n$	unique minimum 2-norm solution
inconsistent	$= n$	unique least-squares (LS) solution
inconsistent	$< n$	unique minimum 2-norm LS solution

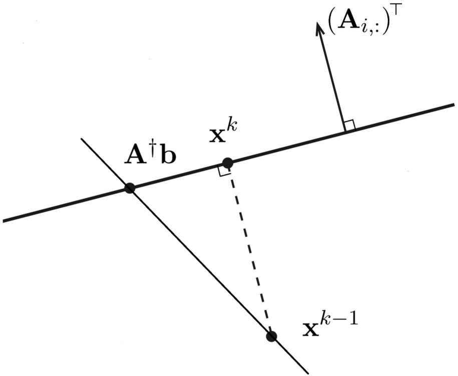

$\mathbf{b} \in \mathrm{range}(\mathbf{A})$?
no
$\mathbf{Ax} = \mathbf{b}$ has no solution. Least squares: $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|_2$.
$\mathrm{null}(\mathbf{A}) = \{\mathbf{0}\}$?
yes
$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|_2$ has the unique solution $\mathbf{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} = \mathbf{A}^\dagger\mathbf{b}$.
no
$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{Ax}\|_2$ has infinitely many solutions.
yes
$\mathrm{null}(\mathbf{A}) = \{\mathbf{0}\}$?
yes
$\mathbf{Ax} = \mathbf{b}$ has the unique solution $\mathbf{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} = \mathbf{A}^\dagger\mathbf{b}$.
no
$\mathbf{Ax} = \mathbf{b}$ has infinitely many solutions.
Among all the solutions, $\mathbf{A}^\dagger\mathbf{b}$ is the unique one minimizing $\|\mathbf{x}\|_2$.

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 ith row of $\mathbf{A}$ and $\mathbf{b}_i$ is the ith component of $\mathbf{b}$.



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

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

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

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

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 $\mathbf{b} \in \mathrm{range}(\mathbf{A})$. The convergence result:

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

where

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

and

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

2.1 A numerical example

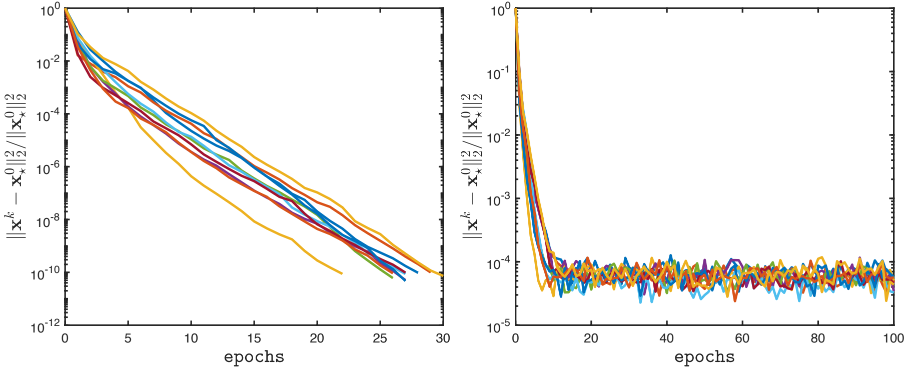


Figure: The relative error of the RK algorithm (10 independent trials) for consistent case (left) and inconsistent case (right).

3. Randomized coordinate descent (RCD)

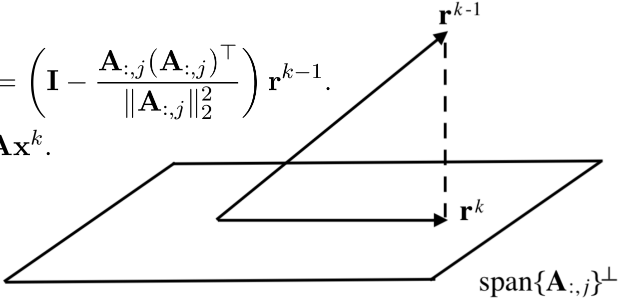
- RCD or RGS: (Leventhal & Lewis 2010)

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

where $\mathbf{A}_{:,j}$ is the jth column of $\mathbf{A}$ and $\mathbf{I}_{:,j}$ is the jth 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 $\mathbf{A}(\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}) \to \mathbf{0}$ and $\mathbf{r}^k \to (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}$.

Algorithm: RCD for $\mathbf{A}\mathbf{x} = \mathbf{b}$

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$ and $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0$

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

Select $j \in [n]$ randomly with probability $\dfrac{\|\mathbf{A}_{:,j}\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}$

Compute $w_k = \dfrac{(\mathbf{A}_{:,j})^\top \mathbf{r}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2}$

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:

$$\mathbb{E}\left[\|\mathbf{A}(\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b})\|_2^2\right] \leq \left(1 - \frac{\sigma_r^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k \|\mathbf{A}(\mathbf{x}^0 - \mathbf{A}^\dagger \mathbf{b})\|_2^2.$$

3.1 A numerical example

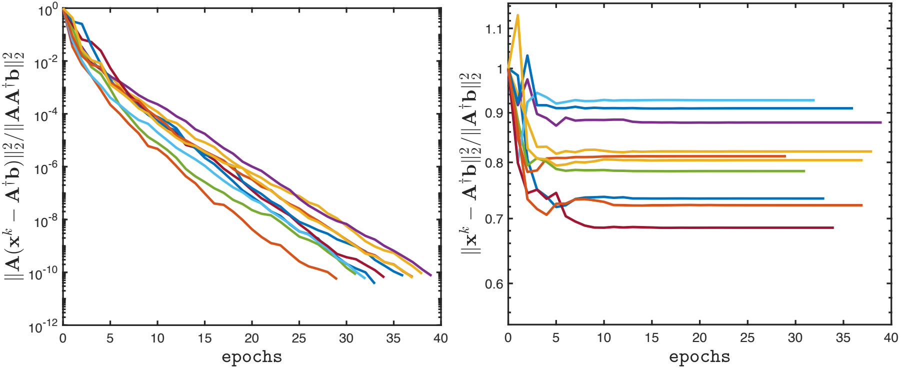


Figure: Convergence history of the RCD algorithm (10 independent trials) for rank-deficient case. Left: the relative residual. Right: the relative error.

4. Randomized extended Kaczmarz (REK) (Zouzias & Freris 2013, Du 2019)

- 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 $\mathbf{z}^0 \in \mathbf{b} + \mathrm{range}(\mathbf{A})$ yields $\{\mathbf{z}^k\}_0^\infty$ satisfying

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

  Then $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^k \to \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^\dagger\mathbf{b}$, 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 $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}$.

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

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

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

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

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

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

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

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

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

end

- The convergence result:

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{x}_\star^0\|_2^2\right] \le \rho^k \|\mathbf{x}^0 - \mathbf{x}_\star^0\|_2^2 + \frac{k\rho^k}{\|\mathbf{A}\|_\mathrm{F}^2}\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\|_2^2.$$

5. Summary of randomized iterative methods

- Randomized iterative methods are preferable if the coefficient matrix $\mathbf{A}$ is too large to fit in memory, or the matrix-vector product $\mathbf{Av}$ is considerably expensive.
- Consistent: RK and its variants

  Inconsistent, full-column rank: RCD and its variants

  Inconsistent, rank-deficient: REK and its variants

6. References

- T. Strohmer and R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, JFAA, 2009
- D. Leventhal and A.S. Lewis, Randomized methods for linear constraints: convergence rates and conditioning, MOR, 2010
- A. Zouzias and N.M. Freris, Randomized extended Kaczmarz for solving least squares, SIMAX, 2013
- K. Du, Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms, NLAA, 2019