Regularized randomized iterative algorithms for factorized linear systems

Kui Du

kuidu@xmu.edu.cn

School of Mathematical Sciences, Xiamen University

https://kuidu.github.io

Outline

The Pseudoinverse Solution of a Linear System

- 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 $\mathbf{b} \in \text{range}(\mathbf{A})$, otherwise, *inconsistent*.

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

$\mathbf{A}\mathbf{x} = \mathbf{b}$	$\text{rank}(\mathbf{A})$	$\mathbf{A}^{\dagger}\mathbf{b}$
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

Sparse (Least Squares) Solutions of a Linear System

- Sparsest solutions:

$$\text{minimize} \quad \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

- The basis pursuit problem:

$$\text{minimize} \quad \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

- The regularized basis pursuit problem

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Sparse least squares solutions: replacing $\mathbf{A}\mathbf{x} = \mathbf{b}$ with the normal equations

$$\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}.$$

Sparsity-Promoting Property of ℓ_1 Norm

- Comparison of ℓ_0, ℓ_1, and ℓ_2 norms

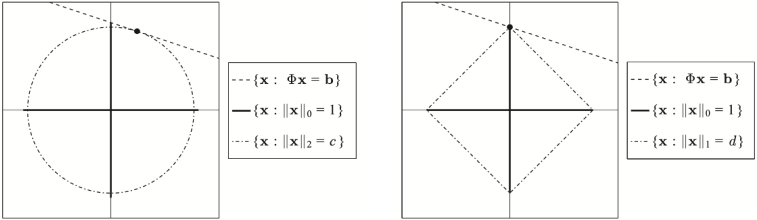


Figure 1.2 of [BL13]: Two-dimensional ℓ_0, ℓ_1, and ℓ_2 balls and the solution set $\{\mathbf{x} \mid \Phi\mathbf{x} = \mathbf{b}\}$. Here c and d are constants with c a bit less than d. Note that the set $\{\mathbf{x} \mid \|\mathbf{x}\|_0 = 1\}$ coincides with the coordinate axes.

[BL13] K. Bryan and T. Leise. *Making Do with Less: An Introduction to Compressed Sensing*. SIAM Review, 55(3):547–566, 2013

Outline

Randomized Kaczmarz (RK)

The RK algorithm for solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ [SV09]

Input: $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and maximum number of iterations `maxit`.

Output: an approximation of the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$.

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

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

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

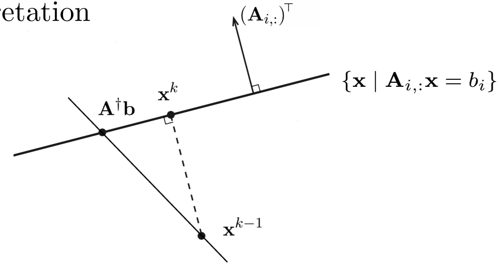
Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \dfrac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - b_i}{\|\mathbf{A}_{i,:}\|_2^2} (\mathbf{A}_{i,:})^\top$

end

[SV09] T. Strohmer and R. Vershynin. *A randomized Kaczmarz algorithm with exponential convergence*. J. Fourier Anal. Appl., 15(2):262–278, 2009.

Geometric Interpretation and Convergence of RK

- Geometric interpretation



- 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 $\rho = 1 - \dfrac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}, \quad \mathbf{x}_\star^0 = (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{x}^0 + \mathbf{A}^\dagger\mathbf{b}.$

- RK fails to find lease squares solutions for inconsistent case [Needell10].

[Needell10] D. Needell. *Randomized Kaczmarz solver for noisy linear systems*, BIT, 50(2):395–403, 2010.

Randomized Gauss–Seidel (RGS)

The RGS algorithm for solving $\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$ [LL10]

Input: $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and maximum number of iterations `maxit`.

Output: an approximation of the solution of $\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$.

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

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

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

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

end

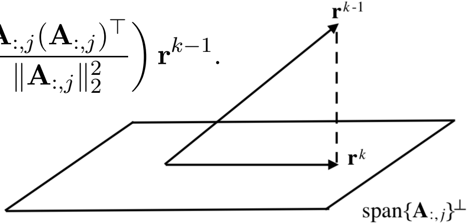
[LL10] D. J. Leventhal and A. S. Lewis. *Randomized methods for linear constraints: convergence rates and conditioning*. Math. Oper. Res., 35(3):641–654, 2010.

Geometric Interpretation and Convergence of RGS

- Geometric interpretation

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



- For arbitrary $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, the convergence result:

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

- RGS finds a least squares solution, but usually not the minimum ℓ_2 norm one for rank-deficient case [MNR15].

[MNR15] A. Ma, D. Needell, and A. Ramdas. *Convergence properties of the randomized extended Gauss–Seidel and Kaczmarz methods.* SIAM J. Matrix Anal. Appl., 36(4):1590–1604, 2015.

RK and RGS

- The problem of finding the solution $\mathbf{x}_\star^0$ of the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ can be posed as the following quadratic optimization problem:
$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{x}^0\|_2^2 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$
- The corresponding dual problem is
$$\min_{\mathbf{y}\in\mathbb{R}^m} \frac{1}{2}\|\mathbf{A}^\top\mathbf{y} + \mathbf{x}^0\|_2^2 - \mathbf{y}^\top\mathbf{b},$$
where the primal variable $\mathbf{x}$ and the dual variable $\mathbf{y}$ are related via the relation
$$\mathbf{x} = \mathbf{A}^\top\mathbf{y} + \mathbf{x}^0.$$
- RK can be constructed by applying a randomized coordinate descent algorithm to the dual problem. On the other hand, the residual of RGS is just the RK iterate for $\mathbf{A}^\top\mathbf{r} = \mathbf{0}$.

Randomized Extended Kaczmarz (REK)

- 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

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

  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 [ZF13] 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 $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^k$:

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

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

[ZF13] A. Zouzias and N. M. Freris. *Randomized extended Kaczmarz for solving least squares*. SIAM J. Matrix Anal. Appl., 34(2):773–793, 2013.

Convergence of REK

- The convergence result [Du19]: $\forall\ \mathbf{z}^0 \in \mathbf{b} + \mathrm{range}(\mathbf{A})$ and $\mathbf{x}^0$

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

  where

$$\rho = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}, \quad \mathbf{x}_\star^0 = (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}^0 + \mathbf{A}^\dagger \mathbf{b}.$$

- REK works for arbitrary (consistent or inconsistent) linear systems (no assumptions about the dimensions or rank of $\mathbf{A}$).
- REK is an RK-RK approach.

[Du19] K. Du. *Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms.* Numer. Linear Algebra Appl., 26(3):e2233, 14pp, 2019.

Randomized Extended Gauss–Seidel (REGS)

- RGS with arbitrary $\mathbf{z}^0$ for $\min_{\mathbf{z}} \|\mathbf{b} - \mathbf{A}\mathbf{z}\|_2$ gives $\mathbf{z}^k$ satisfying
$$\mathbf{A}\mathbf{z}^k \to \mathbf{A}\mathbf{A}^\dagger \mathbf{b} \quad \text{as} \quad k \to \infty.$$
- REGS [MNR15] solves $\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$ via intertwining an iterate of RGS on $\min_{\mathbf{z}} \|\mathbf{b} - \mathbf{A}\mathbf{z}\|_2$ with an iterate of RK on $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z}^k$: [Du19]
$$\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top (\mathbf{A}\mathbf{z}^{k-1} - \mathbf{b})}{\|\mathbf{A}_{:,j}\|_2^2} \mathbf{I}_{:,j},$$
$$\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{A}_{i,:}(\mathbf{x}^{k-1} - \mathbf{z}^k)}{\|\mathbf{A}_{i,:}\|_2^2} (\mathbf{A}_{i,:})^\top.$$
- REGS and REK are related via $\mathbf{z}^k_{\text{REK}} = \mathbf{b} - \mathbf{A}\mathbf{z}^k_{\text{REGS}}$.

[MNR15] A. Ma, D. Needell, and A. Ramdas. *Convergence properties of the randomized extended Gauss–Seidel and Kaczmarz methods.* SIAM J. Matrix Anal. Appl., 36(4):1590–1604, 2015.

Convergence of REGS

- The convergence result [Du19]: $\forall\ \mathbf{z}^0$ and $\mathbf{x}^0$,

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

where

$$\rho = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}, \quad \mathbf{x}_\star^0 = (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}^0 + \mathbf{A}^\dagger \mathbf{b}.$$

- REGS works for arbitrary (consistent or inconsistent) linear systems (no assumptions about the dimensions or rank of $\mathbf{A}$).
- REGS is an RGS-RK approach.

[Du19] K. Du. *Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms.* Numer. Linear Algebra Appl., 26(3):e2233, 14pp, 2019.

Convex Optimization Basics [Beck17]

- Subdifferential: For a function $f : \mathbb{R}^n \to \mathbb{R}$, its subdifferential at $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\partial f(\mathbf{x}) := \{\mathbf{z} \in \mathbb{R}^n \mid f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{z}, \mathbf{y} - \mathbf{x} \rangle, \quad \forall\ \mathbf{y} \in \mathbb{R}^n\}.$$

- γ-strong convexity: A function $f : \mathbb{R}^n \to \mathbb{R}$ is called γ-strongly convex for a given $\gamma > 0$ if the following inequality holds for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{z} \in \partial f(\mathbf{x})$:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{z}, \mathbf{y} - \mathbf{x} \rangle + \frac{\gamma}{2} \|\mathbf{y} - \mathbf{x}\|_2^2.$$

  As an example, the function $f(\mathbf{x}) = \dfrac{1}{2}\|\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$ with $\lambda \geq 0$ is 1-strongly convex.

[Beck17] A. Beck. *First-order methods in optimization* , volume 25 of MOS-SIAM Series on Optimization. SIAM, 2017.

Convex Optimization Basics [Beck17]

- Conjugate function: The conjugate function of $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$f^*(\mathbf{x}) := \sup_{\mathbf{y} \in \mathbb{R}^n} \{\langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{y})\}.$$

  If $f(\mathbf{x})$ is γ-strongly convex, then $f^*(\mathbf{x})$ is differentiable, and

$$\mathbf{z} \in \partial f(\mathbf{x}) \;\Leftrightarrow\; \mathbf{x} = \nabla f^*(\mathbf{z}).$$

- Bregman distance: For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the Bregman distance between $\mathbf{x}$ and $\mathbf{y}$ with respect to f and $\mathbf{z} \in \partial f(\mathbf{x})$ is defined as

$$D_{f,\mathbf{z}}(\mathbf{x}, \mathbf{y}) := f(\mathbf{y}) - f(\mathbf{x}) - \langle \mathbf{z}, \mathbf{y} - \mathbf{x} \rangle.$$

  If f is γ-strongly convex, then it holds that

$$D_{f,\mathbf{z}}(\mathbf{x}, \mathbf{y}) \geq \frac{\gamma}{2} \|\mathbf{x} - \mathbf{y}\|_2^2.$$

A Linear Equality Constrained Minimization Problem

- Consider the linear equality constrained minimization problem

$$\text{minimize } f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

where the objective function f is γ-strongly convex and the constraint $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent.

- The solution of the minimization problem is unique. The objective function f contains regularization terms for promoting certain structures of the underlying solutions.
- By combining the RK algorithm and the gradient of the conjugate function f^*, one obtains the regularized randomized Kaczmarz (RRK) algorithm [SL19][CQ21].

[SL19] F. Schöpfer and D. A. Lorenz. *Linear convergence of the randomized sparse Kaczmarz method.* Math. Program., 173(1-2,Ser.A):509–536, 2019.

[CQ21] X. Chen and J. Qin. *Regularized Kaczmarz algorithms for tensor recovery.* SIAM J. Imaging Sci., 14(4):1439–1471, 2021.

Regularized Randomized Kaczmarz (RRK)

The RRK algorithm for solving $\min_{\mathbf{x}} f(\mathbf{x})$ s.t. $\mathbf{A}\mathbf{x} = \mathbf{b}$

Input: $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and maximum number of iterations iterations `maxit`.

Output: an approximation of the solution of $\min\limits_{\mathbf{A}\mathbf{x}=\mathbf{b}} f(\mathbf{x})$.

Initialize: $\mathbf{z}^0 \in \mathrm{range}(\mathbf{A}^\top)$ and $\mathbf{x}^0 = \nabla f^*(\mathbf{z}^0)$.

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

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

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

Set $\mathbf{x}^k = \nabla f^*(\mathbf{z}^k)$

end

Convergence of RRK [CQ21]

- Assume that the objective function f is γ-strongly convex. Let $\mathbf{x}_\star$ be the unique solution. If $\gamma > 1/2$, and for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \partial f(\mathbf{x}) \cap \mathrm{range}(\mathbf{A}^\top)$,

$$D_{f,\mathbf{z}}(\mathbf{x}, \mathbf{x}_\star) \le \frac{1}{\nu_0} \|\mathbf{A}(\mathbf{x} - \mathbf{x}_\star)\|_2^2,$$

then for all $\mathbf{z}^0 \in \mathrm{range}(\mathbf{A}^\top)$, the sequences $\{\mathbf{x}^k\}$ and $\{\mathbf{z}^k\}$ in the RRK algorithm satisfy

$$\mathbb{E}\left[D_{f,\mathbf{z}^k}(\mathbf{x}^k, \mathbf{x}_\star)\right] \le \beta_0^k D_{f,\mathbf{z}^0}(\mathbf{x}^0, \mathbf{x}_\star)$$

with

$$\beta_0 = 1 - \frac{(2\gamma - 1)\nu_0}{2\gamma \|\mathbf{B}\|_\mathrm{F}^2}.$$

It follows that

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{x}_\star\|_2^2\right] \le \beta_0^k \frac{2}{\gamma} D_{f,\mathbf{z}^0}(\mathbf{x}^0, \mathbf{x}_\star).$$

Special Cases of RRK: RK and RSK

- Case 1: $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$. We have

$$\nabla f^*(\mathbf{x}) = \mathbf{x}.$$

  The RRK algorithm becomes the RK algorithm.

- Case 2: $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$ with $\lambda > 0$. We have

$$\nabla f^*(\mathbf{x}) = S_\lambda(\mathbf{x}),$$

  where $S_\lambda(\mathbf{x})$ is the soft shrinkage function defined component-wise as

$$(S_\lambda(\mathbf{x}))_i = \max\{|x_i| - \lambda, 0\}\mathrm{sign}(x_i).$$

  The RRK algorithm becomes the randomized sparse Kaczmarz (RSK) algorithm [SL19].

[SL19] F. Schöpfer and D. A. Lorenz. *Linear convergence of the randomized sparse Kaczmarz method*. Math. Program., 173(1-2,Ser.A):509–536, 2019.

RRK: RCD for Dual Problem [Petra15]

- The dual problem of $\min_{\mathbf{Ax}=\mathbf{b}} f(\mathbf{x})$ with $\mathbf{b} \in \mathrm{range}(\mathbf{A})$ is the unconstrained problem

$$\min_{\mathbf{y}\in\mathbb{R}^m} g(\mathbf{y}) := f^*(\mathbf{A}^\top \mathbf{y}) - \langle \mathbf{y}, \mathbf{b} \rangle.$$

  The gradient of $g(\mathbf{y})$ is $\nabla g(\mathbf{y}) = \mathbf{A}\nabla f^*(\mathbf{A}^\top \mathbf{y}) - \mathbf{b}$.
  The strong duality holds. The primal variable $\mathbf{x}$ and the dual variable $\mathbf{y}$ are related through the relation $\mathbf{x} = \nabla f^*(\mathbf{A}^\top \mathbf{y})$.
- Randomized coordinate descent (RCD) algorithm:

$$\mathbf{y}^k = \mathbf{y}^{k-1} - \frac{\mathbf{A}_{i,:}\nabla f^*(\mathbf{A}^\top \mathbf{y}^{k-1}) - b_i}{\|\mathbf{A}_{i,:}\|_2^2} \mathbf{I}_{:,i}$$

  Introducing $\mathbf{z}^k = \mathbf{A}^\top \mathbf{y}^k$ and $\mathbf{x}^k = \nabla f^*(\mathbf{z}^k)$ yields RRK.

[Petra15] S. Petra. *Randomized sparse block Kaczmarz as randomized dual block-coordinate descent.* An. Ştiinţ. Univ. "Ovidius" Constanţa Ser. Mat., 23(3):129–149, 2015.

A Combined Optimization Problem

- Consider the combined optimization problem:

$$\text{minimize } f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{x} \in \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{z}\|_2.$$

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

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

- An RK-RRK approach:

$$\begin{aligned}
\mathbf{y}^k &= \mathbf{y}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top \mathbf{y}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2}\mathbf{A}_{:,j}, \\
\mathbf{z}^k &= \mathbf{z}^{k-1} - \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - b_i + y_i^k}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^\top, \\
\mathbf{x}^k &= \nabla f^*(\mathbf{z}^k),
\end{aligned}$$

  with initial iterates

$$\mathbf{y}^0 \in \mathbf{b} + \operatorname{range}(\mathbf{A}), \quad \mathbf{z}^0 \in \operatorname{range}(\mathbf{A}^\top), \quad \mathbf{x}^0 = \nabla f^*(\mathbf{z}^0).$$

Special Cases: REK and ExSRK

- Case 1: For $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, by $\nabla f^*(\mathbf{x}) = \mathbf{x}$, we obtain the REK algorithm.
- Case 2: For $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$ with $\lambda > 0$, by

$$\nabla f^*(\mathbf{x}) = S_\lambda(\mathbf{x}),$$

  we obtain the extended sparse randomized Kaczmarz (ExSRK) algorithm [SLTW22]:

$$\begin{aligned}
\mathbf{y}^k &= \mathbf{y}^{k-1} - \frac{(\mathbf{A}_{:,j})^\top \mathbf{y}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2}\mathbf{A}_{:,j}, \\
\mathbf{z}^k &= \mathbf{z}^{k-1} - \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - b_i + y_i^k}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^\top, \\
\mathbf{x}^k &= S_\lambda(\mathbf{z}^k)
\end{aligned}$$

[SLTW22] F. Schöpfer, D. A. Lorenz, L. Tondji, and M. Winkler. *Extended randomized Kaczmarz method for sparse least squares and impulsive noise problems.* arXiv:2201.08620, 2022.

Randomized Sparse Extended Gauss–Seidel

- RGS with arbitrary $\mathbf{y}^0$ for $\min_{\mathbf{y}} \|\mathbf{b} - \mathbf{A}\mathbf{y}\|_2$ gives $\mathbf{y}^k$ satisfying

$$\mathbf{A}\mathbf{y}^k \to \mathbf{A}\mathbf{A}^{\dagger}\mathbf{b} \quad \text{as} \quad k \to \infty.$$

- A randomized sparse extended Gauss–Seidel (RSEGS) algorithm:

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

with initial iterates

$$\mathbf{y}^0 \in \mathbb{R}^n, \quad \mathbf{z}^0 \in \operatorname{range}(\mathbf{A}^\top), \quad \mathbf{x}^0 = \nabla f^*(\mathbf{z}^0).$$

Outline

A Factorized Linear System

- Consider the following factorized linear system

$$\mathbf{ABx} = \mathbf{b},$$

where

$$\mathbf{A} \in \mathbb{R}^{m\times\ell}, \quad \mathbf{B} \in \mathbb{R}^{\ell\times n}, \quad \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}) = \ell, \quad \mathbf{b} \in \mathbb{R}^{m}.$$

- The factorized linear system can be written as two individual subsystems

$$\mathbf{Ay} = \mathbf{b} \quad \text{(possibly inconsistent)}$$

and

$$\mathbf{Bx} = \mathbf{y}. \quad \text{(always consistent)}$$

- Is it feasible to solve each subsystem separately?

RIAs for Factorized Linear Systems

- Motivated by REK and REGS, interlaced randomized algorithms are proposed for solving factorized linear systems.
- The approach: ALG1-ALG2, where ALG1 is the algorithm used to solve subsystem $\mathbf{Ay} = \mathbf{b}$ and ALG2 is the algorithm used to solve subsystem $\mathbf{Bx} = \mathbf{y}$. For example,

  (1) The RK-RK algorithm [MNR18]

  (2) The REK-RK algorithm [MNR18]

  (3) The RGS-RK algorithm [ZWZ22]

  All find the minimum ℓ_2 norm (least squares) solution.
- How to find sparse solutions?

[MNR18] A. Ma, D. Needell, and A. Ramdas. *Iterative methods for solving factorized linear systems*. SIAM J. Matrix Anal. Appl., 39(1):104–122, 2018.

[ZWZ22] J. Zhao, X. Wang, and J. Zhang. *A randomised iterative method for solving factorised linear systems*. Linear Multilinear Algebra, to appear, 2022.

Outline

1. Solutions of Linear Systems
2. Randomized Iterative Algorithms
3. Factorized Linear Systems
4. The Proposed Algorithms
5. Computed Examples
6. Summary

A Combined Optimization Problem

- Consider the combined optimization problem:

$$\text{minimize } f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{x} \in \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{B}\mathbf{z}\|_2.$$

- We consider the ALG1-ALG2 approach. Specifically, we interlace the RK algorithm or the RGS algorithm for the subsystem

$$\mathbf{A}\mathbf{y} = \mathbf{b}$$

  with the RRK algorithm for the linear equality constrained minimization problem

$$\text{minimize } f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{B}\mathbf{x} = \mathbf{y}.$$

- The proposed algorithms become the RK-RK algorithm and the RGS-RK algorithm if $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$.

The RK-RRK algorithm: $\mathbf{b} \in \mathrm{range}(\mathbf{AB})$

The RK-RRK algorithm for solving $\min_{\mathbf{ABx}=\mathbf{b}} f(\mathbf{x})$

Input: $\mathbf{A} \in \mathbb{R}^{m\times\ell}$, $\mathbf{B} \in \mathbb{R}^{\ell\times n}$, $\mathbf{b} \in \mathbb{R}^{m}$, and maximum number of iterations `maxit`.

Output: an approximation of the solution of $\min\limits_{\mathbf{ABx}=\mathbf{b}} f(\mathbf{x})$.

Initialize: $\mathbf{y}^0 = \mathbf{0}$, $\mathbf{z}^0 \in \mathrm{range}(\mathbf{B}^\top)$, and $\mathbf{x}^0 = \nabla f^*(\mathbf{z}^0)$.

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

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

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

Pick $i \in [\ell]$ with probability $\|\mathbf{B}_{i,:}\|_2^2 / \|\mathbf{B}\|_{\mathrm{F}}^2$

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

Set $\mathbf{x}^k = \nabla f^*(\mathbf{z}^k)$

end

RK-RSK and ExSRK

- The RK-RRK algorithm becomes the RK-RSK algorithm if
$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$
- The iterates of the ExSRK algorithm [SLTW22] for $\mathbf{C}\mathbf{x} = \mathbf{b}$ are
$$\begin{aligned}
\mathbf{y}^k &= \mathbf{y}^{k-1} - \frac{(\mathbf{C}_{:,j})^\top \mathbf{y}^{k-1}}{\|\mathbf{C}_{:,j}\|_2^2}\mathbf{C}_{:,j},\\
\mathbf{z}^k &= \mathbf{z}^{k-1} - \frac{\mathbf{C}_{i,:}\mathbf{x}^{k-1} - b_i + y_i^k}{\|\mathbf{C}_{i,:}\|_2^2}(\mathbf{C}_{i,:})^\top,\\
\mathbf{x}^k &= S_\lambda(\mathbf{z}^k),
\end{aligned}$$
with initial iterates $\mathbf{y}^0 = \mathbf{b}$, $\mathbf{z}^0 \in \mathrm{range}(\mathbf{C}^\top)$, and $\mathbf{x}^0 = S_\lambda(\mathbf{z}^0)$.

[SLTW22] F. Schöpfer, D. A. Lorenz, L. Tondji, and M. Winkler. *Extended randomized Kaczmarz method for sparse least squares and impulsive noise problems.* arXiv:2201.08620, 2022.

RK-RSK and ExSRK

- Note that the normal equations

$$\mathbf{C}^\top \mathbf{C}\mathbf{x} = \mathbf{C}^\top \mathbf{b}$$

can be viewed as the factorized linear system

$$\widehat{\mathbf{A}}\widehat{\mathbf{B}}\mathbf{x} = \widehat{\mathbf{b}}$$

with

$$\widehat{\mathbf{A}} = \mathbf{C}^\top, \quad \widehat{\mathbf{B}} = \mathbf{C}, \quad \widehat{\mathbf{b}} = \mathbf{C}^\top \mathbf{b}.$$

We observe that the iterates $\mathbf{x}^k$, $\mathbf{y}^k$, and $\mathbf{z}^k$ of the ExSRK algorithm for

$$\mathbf{C}\mathbf{x} = \mathbf{b}$$

are equal to $\widehat{\mathbf{x}}^k$, $\mathbf{b} - \widehat{\mathbf{y}}^k$, and $\widehat{\mathbf{z}}^k$, respectively, where $\widehat{\mathbf{x}}^k$, $\widehat{\mathbf{y}}^k$, and $\widehat{\mathbf{z}}^k$ are the iterates of the RK-RSK algorithm for

$$\widehat{\mathbf{A}}\widehat{\mathbf{B}}\mathbf{x} = \widehat{\mathbf{b}}.$$

The RGS-RRK algorithm

The RGS-RRK algorithm for solving $\min_{\mathbf{x} \in \operatorname{argmin}_{\mathbf{z}} \|\mathbf{b}-\mathbf{ABz}\|_2} f(\mathbf{x})$

Input: $\mathbf{A} \in \mathbb{R}^{m \times \ell}$, $\mathbf{B} \in \mathbb{R}^{\ell \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and maximum number of iterations `maxit`.

Output: an approximation of the solution of $\min\limits_{\mathbf{x} \in \operatorname{argmin}_{\mathbf{z}} \|\mathbf{b}-\mathbf{ABz}\|_2} f(\mathbf{x})$.

Initialize: $\mathbf{y}^0 = \mathbf{0}$, $\mathbf{r}^0 = \mathbf{b}$, $\mathbf{z}^0 \in \operatorname{range}(\mathbf{B}^\top)$, and $\mathbf{x}^0 = \nabla f^*(\mathbf{z}^0)$.

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

- Pick $j \in [\ell]$ with probability $\|\mathbf{A}_{:,j}\|_2^2 / \|\mathbf{A}\|_\mathrm{F}^2$
- Compute $d_k = (\mathbf{A}_{:,j})^\top \mathbf{r}^{k-1} / \|\mathbf{A}_{:,j}\|_2^2$
- Set $y_j^k = y_j^{k-1} + d_k$, $y_l^k = y_l^{k-1}$ for $l \neq j$
- Set $\mathbf{r}^k = \mathbf{r}^{k-1} - d_k \mathbf{A}_{:,j}$
- Pick $i \in [\ell]$ with probability $\|\mathbf{B}_{i,:}\|_2^2 / \|\mathbf{B}\|_\mathrm{F}^2$
- Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \dfrac{\mathbf{B}_{i,:}\mathbf{x}^{k-1} - y_i^k}{\|\mathbf{B}_{i,:}\|_2^2} (\mathbf{B}_{i,:})^\top$
- Set $\mathbf{x}^k = \nabla f^*(\mathbf{z}^k)$

end

Outline

Example 1

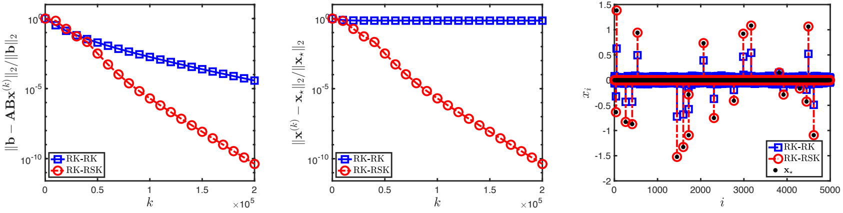
- `A=randn(m,l)`, `B=randn(l,n)`.
- $\mathbf{x}_\star$ is an s sparse vector with normally distributed non-zero entries, whose support is randomly generated.
- $\mathbf{b} = \mathbf{A}\mathbf{B}\mathbf{x}_\star$ for $\mathbf{b} \in \mathrm{range}(\mathbf{A}\mathbf{B})$.
- $\mathbf{b} = \widehat{\mathbf{b}} + \widehat{\mathbf{b}}_\perp$ for $\mathbf{b} \notin \mathrm{range}(\mathbf{A}\mathbf{B})$ with $\widehat{\mathbf{b}} = \mathbf{A}\mathbf{B}\mathbf{x}_\star$ and

$$\widehat{\mathbf{b}}_\perp = \mathbf{N}\mathbf{v}\|\widehat{\mathbf{b}}\|_2/\|\mathbf{N}\mathbf{v}\|_2 \in \mathrm{null}(\mathbf{B}^\top\mathbf{A}^\top) = \mathrm{null}(\mathbf{A}^\top),$$

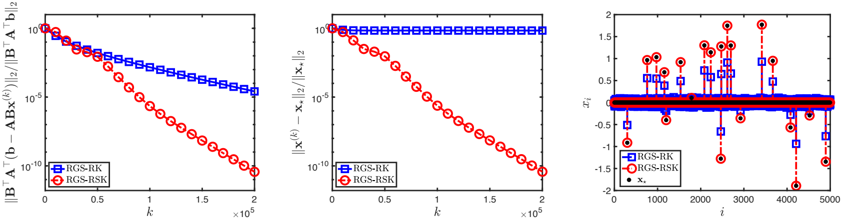
  where the columns of $\mathbf{N}$ form an orthonormal basis of $\mathrm{null}(\mathbf{A}^\top)$ and $\mathbf{v}$ is a Gaussian vector generated by `v=randn(m-l,1)`.
- For the proposed algorithms, we use $\lambda = 1$, $\mathbf{y}^0 = \mathbf{0}$, $\mathbf{z}^0 = \mathbf{0}$, and the maximum number of iterations `maxit=20`m.
- `m=10000, l=2500, n=5000, s=20`.

Example 1: Results

- Comparison of RK-RK and RK-RSK



- Comparison of RGS-RK and RGS-RSK



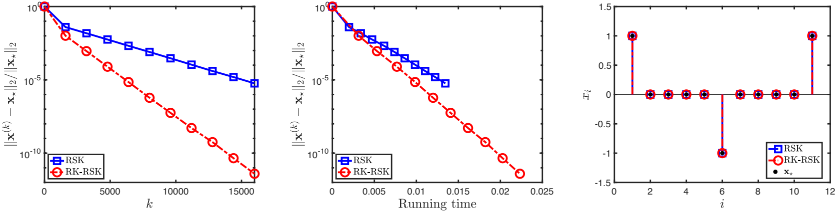
Example 2

- Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ denote the wine quality data matrix (a sample of $m = 1599$ red wines with $n = 11$ physio-chemical properties of each wine) obtained from the UCI Machine Learning Repository [uci].
- The matrices $\mathbf{A}$ and $\mathbf{B}$ are obtained as follows: `[A,B]=nnmf(X,5)`. We compute $\mathbf{C} = \mathbf{AB}$ in MATLAB.
- The condition number of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are 23.5, 4.4, and 45.4, respectively.
- Let $\mathbf{x}_\star \in \mathbb{R}^{11}$ be a 3-sparse vector with support $\{1, 6, 11\}$. The three nonzero entries of $\mathbf{x}_\star$ are set to be 1.
- For the proposed algorithms, we use $\lambda = 1$, $\mathbf{y}^0 = \mathbf{0}$, $\mathbf{z}^0 = \mathbf{0}$, and the maximum number of iterations `maxit=10`m.

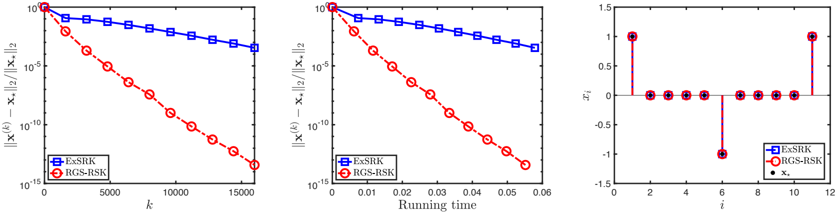
[uci] D. Dua and C. Graff. *UCI machine learning repository*, 2017. `http://archive.ics.uci.edu/ml`.

Example 2: Results

- Comparison of RSK and RK-RSK



- Comparison of ExSRK and RGS-RSK



Outline

Summary

- Two new regularized randomized iterative algorithms using the ALG1-ALG2 approach are proposed to find (least squares) solutions with certain structures of factorized linear systems.
- Computed examples are given to illustrate that the new algorithms can find sparse (least squares) solutions of $\mathbf{ABx} = \mathbf{b}$ and can be better than the existing randomized iterative algorithms for the corresponding full linear system $\mathbf{Cx} = \mathbf{b}$ with $\mathbf{C} = \mathbf{AB}$.
- Existing acceleration strategies for RK and RGS can be integrated into our algorithms easily and the corresponding convergence analysis is straightforward.
- The extension to a factorized linear system with rank-deficient $\mathbf{A}$ and $\mathbf{B}$ will be the future work.

Thanks!