Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods

Anna Ma
Institute of Mathematical Sciences
Claremont Graduate University
Claremont, CA 91711
anna.ma@cgu.edu

Deanna Needell
Claremont McKenna College
Claremont, CA 91711
dneedell@cmc.edu

Aaditya Ramdas
University of California, Berkeley
Berkeley, CA 94720
aramdas@cs.berkeley.edu

Abstract

The Kaczmarz and Gauss-Seidel methods both solve a linear system $\boldsymbol{X\beta} = \boldsymbol{y}$ by iteratively refining the solution estimate. Recent interest in these methods has been sparked by a proof of Strohmer and Vershynin which shows the *randomized* Kaczmarz method converges linearly in expectation to the solution. Lewis and Leventhal then proved a similar result for the randomized Gauss-Seidel algorithm. However, the behavior of both methods depends heavily on whether the system is under or overdetermined, and whether it is consistent or not. Here [1] we provide a unified theory of both methods, their variants for these different settings, and draw connections between both approaches. We compare convergence properties of both methods and their extended variants in all possible system settings. In doing so, we also provide an extended version of randomized Gauss-Seidel which converges linearly to the least norm solution in the underdetermined case. With this result, a complete and rigorous theory of both methods is furnished.

1 Introduction

We consider solving a linear system of equations

$$\boldsymbol{X\beta} = \boldsymbol{y}, \tag{1}$$

for a (real or complex) $m \times n$ matrix $\boldsymbol{X}$, in various problem settings. Recent interest in the topic was reignited when Strohmer and Vershynin [7] proved the linear[2] convergence rate of the Randomized Kaczmarz (RK) algorithm that works on the rows of $\boldsymbol{X}$ (data points). Following that, Leventhal and Lewis [4] proved the linear convergence of a Randomized Gauss-Seidel (RGS), i.e. Randomized Coordinate Descent, algorithm that works on the columns of $\boldsymbol{X}$ (features).

[1]Our work has been accepted for publication in SIAM Journal on Matrix Analysis and Applications [1] .
[2]Mathematicians often refer to linear convergence as exponential convergence.

1.1 Motivation and contribution

When the system of equations is inconsistent, as is typically the case when $m > n$ in real-world overconstrained systems, RK is known to not converge to the ordinary least squares solution

$$\boldsymbol{\beta_{LS}} := \arg\min_{\boldsymbol{\beta}} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \tag{2}$$

as studied by Needell [5]. Zouzias and Freris [8] extended the RK method with the modified Randomized Extended Kaczmarz (REK) algorithm, which linearly converges to $\boldsymbol{\beta_{LS}}$. Interestingly, in this setting, RGS does converge to $\boldsymbol{\beta_{LS}}$ without any special extensions.

When $m < n$, there are fewer constraints than variables, and the system has infinitely many solutions. In this case, we are often interested in finding the least Euclidean norm solution:

$$\boldsymbol{\beta_{LN}} := \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_2 \ \text{ s.t. } \ \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}. \tag{3}$$

While RGS converges to $\boldsymbol{\beta_{LS}}$ in the overdetermined setting, in the underdetermined setting it does not converge to $\boldsymbol{\beta_{LN}}$. We will also argue that RK does converge to $\boldsymbol{\beta_{LN}}$ without any extensions in this setting.

The main contribution of our paper is to provide a unified theory of these related iterative methods under different settings. We will also construct an extension to RGS that parallels REK, which unlike RGS does converge to $\boldsymbol{\beta_{LN}}$ (just as REK, unlike RK, converges to $\boldsymbol{\beta_{LS}}$). We shall see that our Randomized Extended Gauss-Seidel (REGS) method does indeed possess these desired properties. A summary of this unified theory is provided in Table 1.

Method	Overconstrained, consistent : convergence to $\boldsymbol{\beta}^\star$?	Overconstrained, inconsistent : convergence to $\boldsymbol{\beta_{LS}}$?	Underconstrained : convergence to $\boldsymbol{\beta_{LN}}$?
RK	Yes [7]	No [5, Thm. 2.1]	Yes [1]
REK	Yes [8]	Yes [8]	Yes [1]
RGS	Yes [4]	Yes [4]	No [1]
REGS	Yes (Sec. 3.1)	Yes (Sec. 3.3)	Yes (Thm. 1)

Table 1: Summary of convergence properties for the overdetermined and consistent setting, overdetermined and inconsistent setting, and underdetermined settings. We write $\boldsymbol{\beta}^\star$ to denote the solution to (1) in the overdetermined consistent setting, with $\boldsymbol{\beta_{LS}}$ and $\boldsymbol{\beta_{LN}}$ being defined in (2) and (3) for the other two settings.

2 Existing Algorithms and Related Work

In this section, we will summarize the algorithms mentioned in the introduction, i.e. RK, RGS and REK. We will describe their iterative update rules and mention their convergence guarantees. Throughout the paper we will use the notation $\boldsymbol{X}^i$ to represent the ith row of $\boldsymbol{X}$ (or ith entry in the case of a vector) and $\boldsymbol{X}_{(j)}$ to denote the jth column of a matrix $\boldsymbol{X}$. We will write the estimation $\boldsymbol{\beta}$ as a column vector. We write vectors and matrices in boldface, and constants in standard font.

2.1 Randomized Kaczmarz (RK)

The Kaczmarz method was first introduced in the notable work of Kaczmarz [3]. Strohmer and Vershynin [7] showed that the RK method described above has an expected linear convergence rate to the solution $\boldsymbol{\beta}^\star$ of (1), and are the first to provide an explicit convergence rate in expectation which depends only on the geometric properties of the system. This work was extended by Needell [5] to the inconsistent case, analyzed almost surely by Chen and Powell [2].

We describe here the randomized variant of the Kaczmarz method put forth by Strohmer and Vershynin [7]. Taking $\boldsymbol{X}, \boldsymbol{y}$ as input and starting from an arbitrary initial estimate for $\boldsymbol{\beta}$ (for example $\boldsymbol{\beta}_0 = \boldsymbol{0}$), RK repeats the following in each iteration. First, a random row $i \in \{1, ..., m\}$ is selected with probability proportional to its Euclidean norm, i.e. $\Pr(\text{row} = i) = \frac{\|\boldsymbol{X}^i\|_2^2}{\|\boldsymbol{X}\|_F^2}$, where $\|\boldsymbol{X}\|_F$ denotes the Frobenius norm of $\boldsymbol{X}$. Then, project the current iterate onto that row, i.e.

$$\boldsymbol{\beta}_{t+1} := \boldsymbol{\beta}_t + \frac{(y^i - \boldsymbol{X}^i \boldsymbol{\beta}_t)}{\|\boldsymbol{X}^i\|_2^2} (\boldsymbol{X}^i)^*, \tag{4}$$

where here and throughout $\boldsymbol{X}^*$ denotes the (conjugate) transpose of $\boldsymbol{X}$.

Intuitively, this update can be seen as greedily satisfying the ith equation in the linear system.

2.2 Randomized Extended Kaczmarz (REK)

For inconsistent systems, the RK method does not converge to the least-squares solution as one might desire. This fact is clear since the method at each iteration projects completely onto a selected solution space, being unable to break the so-called *convergence horizon*. Recently, Zouzias and Freris [8] proposed a variant of the RK method motivated by the work of Popa [6] which instead includes a random projection to iteratively reduce the component of $\boldsymbol{y}$ which is orthogonal to the range of $\boldsymbol{X}$. This method, named Randomized Extended Kaczmarz (REK) can be described by the following iteration updates, which can be initialized with $\boldsymbol{\beta}_0 = \boldsymbol{0}$ and $\boldsymbol{z}_0 = \boldsymbol{y}$:

$$\boldsymbol{\beta}_{t+1} := \boldsymbol{\beta}_t + \frac{(y^i - z_t^i - \boldsymbol{X}^i \boldsymbol{\beta}_t)}{\|\boldsymbol{X}^i\|_2^2} (\boldsymbol{X}^i)^*, \quad \boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \frac{\langle \boldsymbol{X}_{(j)}, \boldsymbol{z}_t \rangle}{\|\boldsymbol{X}_{(j)}\|_2^2} \boldsymbol{X}_{(j)}. \tag{5}$$

Here, a column $j \in \{1, ..., n\}$ is also selected at random with probability proportional to its Euclidean norm: $\Pr(\text{column} = j) = \frac{\|\boldsymbol{X}_{(j)}\|_2^2}{\|\boldsymbol{X}\|_F^2}$, and again $\boldsymbol{X}_{(j)}$ denotes the jth column of $\boldsymbol{X}$. Here, $\boldsymbol{z}_t$ approximates the component of $\boldsymbol{y}$ which is orthogonal to the range of $\boldsymbol{X}$, allowing for the iterates $\boldsymbol{\beta}_t$ to converge to the true least-squares solution of the system. Zouzias and Freris [8] prove that REK converges linearly in expectation to this solution $\boldsymbol{\beta_{LS}}$.

2.3 Randomized Gauss-Seidel (RGS)

The Randomized Gauss-Seidel (RGS) method (or the Randomized Coordinate Descent method) repeats the following in each iteration. First, a random column $j \in \{1, ..., n\}$ is selected with probability proportional to its Euclidean norm (as in REK). We then minimize the objective $L(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2$ with respect to this coordinate to get

$$\boldsymbol{\beta}_{t+1} := \boldsymbol{\beta}_t + \frac{\boldsymbol{X}_{(j)}^* (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_t)}{\|\boldsymbol{X}_{(j)}\|_2^2} \boldsymbol{e}_{(j)} \tag{6}$$

where $\boldsymbol{e}_{(j)}$ is the jth coordinate basis column vector (all zeros with a 1 in the jth position). It can be seen as greedily minimizing the objective with respect to the jth coordinate. Leventhal and Lewis [4] showed that this algorithm has an expected linear convergence rate.

3 REGS

We next introduce an extension of RGS, analogous to the extension REK of RK. The purpose of extending RK was to allow for convergence to the least squares solution. Now, the purpose of extending RGS is to allow for convergence to the least norm solution. We view this method as a completion to the unified analysis of these approaches, and it may also possess advantages in its own right.

3.1 The algorithm

Consider the linear system (1) with $m < n$. Let $\boldsymbol{\beta_{LN}}$ denote the least norm solution of the underdetermined system as described in (3). The REGS algorithm is described by the following pseudo-code. Analogous to the role $\boldsymbol{z}$ plays in REK, $\boldsymbol{z}$ iteratively approximates the component in $\boldsymbol{\beta}$ orthogonal to the row-span of $\boldsymbol{X}$. By iteratively removing this component, we converge to the least norm solution. Note that outputting $\boldsymbol{\beta}_t$ instead of $\boldsymbol{\beta}_t^{LN} = \boldsymbol{\beta}_t - \boldsymbol{z}_t$ in Algorithm 1 recovers the RGS algorithm. This may be preferable in the overdetermined setting.

Algorithm 1 Randomized Extended Gauss-Seidel (REGS)

procedure ($\boldsymbol{X}$, $\boldsymbol{y}$, maxIter) $\triangleright$ $m \times n$ matrix $\boldsymbol{X}$, $\boldsymbol{y} \in \mathbb{C}^m$, maximum iterations T
  Initialize $\boldsymbol{\beta}_0 = \boldsymbol{0}$, $\boldsymbol{z}_0 = \boldsymbol{0}$
  for $t = 1, 2, \ldots, T$ **do**
    Choose column $\boldsymbol{X}_{(j)}$ with probability $\frac{\|\boldsymbol{X}_{(j)}\|_2^2}{\|\boldsymbol{X}\|_F^2}$
    Choose row $\boldsymbol{X}^i$ with probability $\frac{\|\boldsymbol{X}^i\|_2^2}{\|\boldsymbol{X}\|_F^2}$
    Set $\boldsymbol{\gamma}_t = \frac{\boldsymbol{X}^*_{(j)}(\boldsymbol{X}\boldsymbol{\beta}_{t-1} - \boldsymbol{y})}{\|\boldsymbol{X}_{(j)}\|_2^2} \boldsymbol{e}_{(j)}$
    Set $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\gamma}_t$
    Set $\boldsymbol{P}_i = \boldsymbol{Id}_n - \frac{(\boldsymbol{X}^i)^* \boldsymbol{X}^i}{\|\boldsymbol{X}^i\|_2^2}$ $\triangleright$ $\boldsymbol{Id}_n$ denotes the $n \times n$ identity matrix
    Update $\boldsymbol{z}_t = \boldsymbol{P}_i(\boldsymbol{z}_{t-1} + \boldsymbol{\gamma}_t)$
    Update $\boldsymbol{\beta}_t^{LN} = \boldsymbol{\beta}_t - \boldsymbol{z}_t$
  end for
  Output $\boldsymbol{\beta}_t^{LN}$
end procedure

3.2 Main result

Our main result for the REGS method shows linear convergence to the least norm solution.

Theorem 1. *The REGS algorithm outputs an estimate* $\boldsymbol{\beta}_T^{LN}$ *such that*

$$\mathbb{E}\|\boldsymbol{\beta}_T^{LN} - \boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \leq \alpha^T \|\boldsymbol{\beta}^{LN}\|_2^2 + 2\alpha^{\lfloor T/2 \rfloor} \frac{B}{1-\alpha} \tag{7}$$

where $B = \frac{\|\boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2}{\|\boldsymbol{X}\|_F^2}$ *and* $\alpha = \left(1 - \frac{\sigma^2_{min}(\boldsymbol{X})}{\|\boldsymbol{X}\|_F^2}\right)$.

The full proof of Theorem 1 can be found in [1].

3.3 Comparison

Theorem 1 shows that, like the RK and REK methods, REGS converges linearly to the least-norm solution in the underdetermined case. We believe it serves to complement existing analysis and completes the theory of these iterative methods in all three cases of interest. For that reason, we compare the three approaches for the underdetermined setting here. For ease of comparison, set α as in Theorem 1, and write $\kappa = \sigma_{\max}(\boldsymbol{X})/\sigma_{\min}(\boldsymbol{X})$ for the condition number of $\boldsymbol{X}$. From the convergence rate bounds for RK [7] and REK [8], and after applying elementary bounds to (7) of Theorem 1, we have:

$$\text{(RK)} \quad \mathbb{E}\|\boldsymbol{\beta}_t - \boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \leq \alpha^t \|\boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \tag{8}$$

$$\text{(REK)} \quad \mathbb{E}\|\boldsymbol{\beta}_{2t} - \boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \leq \alpha^t (1 + 2\kappa^2) \|\boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \tag{9}$$

$$\text{(REGS)} \quad \mathbb{E}\|\boldsymbol{\beta}_{2t} - \boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2 \leq \alpha^t (1 + 2\kappa^2) \|\boldsymbol{\beta}_{\boldsymbol{LN}}\|_2^2. \tag{10}$$

We find similar results in the overdetermined, inconsistent setting. Using the convergence rate bounds for RGS [4], REK [8], and REGS (Theorem 1), we have:

$$\text{(RGS)} \quad \mathbb{E}\|\boldsymbol{\beta}_t - \boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2 \leq \alpha^t \|\boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2 \tag{11}$$

$$\text{(REK)} \quad \mathbb{E}\|\boldsymbol{\beta}_{2t} - \boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2 \leq \alpha^t (1 + 2\kappa^2) \|\boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2 \tag{12}$$

$$\text{(REGS)} \quad \mathbb{E}\|\boldsymbol{\beta}_{2t} - \boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2 \leq \alpha^t (1 + 2\kappa^2) \|\boldsymbol{\beta}_{\boldsymbol{LS}}\|_2^2. \tag{13}$$

Thus, up to constant terms (which are likely artifacts of the proofs), the bounds provide the same convergence rate α, which is not surprising in light of the connections between the methods. Detailed experiments can be found in the full version of the paper [1].

4 Conclusion

The Kaczmarz and Gauss-Seidel methods operate in two different spaces (i.e. row versus column space), but share many parallels. In this paper we drew connections between these two methods, highlighting the similarities and differences in convergence analysis. The approaches possess conflicting convergence properties; RK converges to the desired solution in the underdetermined case but not the inconsistent overdetermined setting, while RGS does the exact opposite. The extended method REK in the Kaczmarz framework fixes this issue, converging to the solution in both scenarios. Here, we present the REGS method, a natural extension of RGS, which completes the overall picture. We hope that our unified analysis of all four methods will assist researchers working with these approaches.

References

[1] D. Needell A. Ma and A. Ramdas. Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods. *SIAM Journal on Matrix Analysis and Applications*, 2015. To appear.

[2] X. Chen and A. Powell. Almost sure convergence of the Kaczmarz algorithm with random measurements. *J. Fourier Anal. Appl.*, pages 1–20, 2012. ISSN 1069-5869. URL `http://dx.doi.org/10.1007/s00041-012-9237-2`. 10.1007/s00041-012-9237-2.

[3] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull. Int. Acad. Polon. Sci. Lett. Ser. A*, pages 335–357, 1937.

[4] D. Leventhal and A. S. Lewis. Randomized methods for linear constraints: convergence rates and conditioning. *Math. Oper. Res.*, 35(3):641–654, 2010. ISSN 0364-765X. doi: 10.1287/moor.1100.0456. URL `http://dx.doi.org/10.1287/moor.1100.0456`.

[5] D. Needell. Randomized Kaczmarz solver for noisy linear systems. *BIT*, 50(2):395–403, 2010. ISSN 0006-3835. doi: 10.1007/s10543-010-0265-5. URL `http://dx.doi.org/10.1007/s10543-010-0265-5`.

[6] C. Popa. Extensions of block-projections methods with relaxation parameters to inconsistent and rank-deficient least-squares problems. *BIT*, 38(1):151–176, 1998. ISSN 0006-3835. doi: 10.1007/BF02510922. URL `http://dx.doi.org/10.1007/BF02510922`.

[7] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009. ISSN 1069-5869. doi: 10.1007/s00041-008-9030-4. URL `http://dx.doi.org/10.1007/s00041-008-9030-4`.

[8] A. Zouzias and N. M. Freris. Randomized extended Kaczmarz for solving least squares. *SIAM Journal on Matrix Analysis and Applications*, 34(2):773–793, 2013.