

求解大型稀疏线性系统的贪婪双子空间随机 Kaczmarz 方法

荆燕飞¹, 李彩霞¹, 胡少亮²

(1. 电子科技大学 数学科学学院, 四川 成都 611731; 2. 中国工程物理研究院 高性能数值模拟软件中心, 北京 100088)

摘要: 基于一种有效的从系数矩阵中选取两个工作行的贪婪概率准则, 提出一类求解大型稀疏线性系统的贪婪双子空间随机 Kaczmarz 方法。理论证明该方法收敛到相容线性系统的最小范数解, 而且该方法的理论收敛因子小于原始双子空间随机 Kaczmarz 方法的收敛因子。数值实验表明, 该方法在求解性能方面较原始双子空间随机 Kaczmarz 方法更具优势。

关键词: 双子空间投影法; 随机 Kaczmarz 方法; 大型稀疏线性系统; 收敛性分析

中图分类号: O241.6

文献标志码: A

A Greedy Two-Subspace Randomized Kaczmarz Method for Solving Large Sparse Linear Systems

JING Yanfei¹, LI Caixia¹, HU Shaoliang²

(1. School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. CAEP Software Center for High Performance Numerical Simulation, Beijing 100088, China)

Abstract: Based on an effective greedy probability criterion for selecting two working rows from a coefficient matrix, a greedy two-subspace randomized Kaczmarz method for solving large sparse linear systems is proposed. The theoretical analysis shows that this method converges to the minimal-norm solution of consistent linear systems, and the convergence factor of the method is smaller than that of the original two-subspace randomized Kaczmarz method. The numerical experiments show that this method is superior to the original two-subspace randomized Kaczmarz method from

the point of view of solution performance.

Key words: two-subspace projection method; randomized Kaczmarz method; large sparse linear systems; convergence analysis

考虑求解具有如下形式的相容线性系统:

$$Ax = b \quad (1)$$

其中, 系数矩阵 $A \in \mathbb{R}^{m \times n}$ ($m > n$) 可为满秩或秩亏矩阵, 且 $b \in \mathbb{R}^m$ 。通常考虑求式(1)的最小范数解

$$x_* := \arg \min_x \|x\|_2 \quad \text{s.t. } Ax = b$$

当系数矩阵 A 为列满秩时, x_* 为式(1)的惟一解; 当线性系统(1)有无穷多组解时, x_* 为式(1)的最小范数解(这里的范数指欧式范数)。

为求解线性系统(1), 许多迭代方法^[1-5]已被开发和进一步研究。其中, Kaczmarz^[5]于1937年首次提出的 Kaczmarz 方法, 因其成本低廉、易于操作而迅速得到数值计算领域的专家学者的认可和广泛关注, 进而使其获得巨大的理论发展^[6-9]。因其是一种具有代表性的行处理迭代方法且在计算机上易于实现和并行化, 因此被广泛应用于计算机断层扫描^[10-13], 图像重建^[14-16], 分布式计算^[17-18]和信号处理^[19-21]等领域。

若以随机顺序而不是以给定顺序选用系数矩阵的行可以极大地提高 Kaczmarz 方法的收敛速度^[13, 16, 21]。尽管这些随机选行的 Kaczmarz 方法在应用中颇具吸引力, 但尚不能保证其收敛速率。Strohmer 等^[22]首次采用选取工作行的概率与其所对向量的欧式范数的平方成正比这样的选行准则, 提出能保证收敛速率的随机 Kaczmarz 方法, 并证明其

收稿日期: 2021-03-19

基金项目: 国家自然科学基金(12071062, 61772003); 科学挑战项目(TZ2016002—TZZT2019-B1.4); 电子科技大学理科实力提升计划

第一作者: 荆燕飞(1982—), 男, 研究员, 博士生导师, 理学博士, 主要研究方向为数值代数与科学计算及应用。

E-mail: yanfeijing@uestc.edu.cn; E-mail: 00jyfvictory@163.com

通信作者: 李彩霞(1995—), 女, 理学硕士, 主要研究方向为数值代数与科学计算及应用。

E-mail: licaixia5429@163.com



论文
拓展
介绍

误差的期望具有指数收敛速率。Dai等^[23]通过求解最小化收敛速度的上限这个凸优化问题来获得选择行的最佳概率分布,提出最优的随机Kaczmarz方法;Bai等^[24]结合贪婪和随机的数学思想引入一种有效的行选择准则提出贪婪随机Kaczmarz方法。此后,松弛贪婪随机Kaczmarz方法^[25]和贪婪距离随机Kaczmarz方法^[26]也被提出并深入研究。随机Kaczmarz方法的各种变种^[27-39]和收敛理论^[30-35]获得了大量研究和发展。另一方面,Leventhal等^[36]采用同样的随机思想,提出了求解最小二乘问题的随机坐标下降方法。因随机Kaczmarz方法和随机坐标下降方法存在一定的共性,后期有学者将两者同时比对研究^[37-38]。

尽管随机Kaczmarz方法适用于任何相容线性系统的求解,但当系数矩阵有许多相关行时(常出现于地球物理学中),收敛可能停滞。为克服这一缺点,Needell和Ward采用等可能地随机选择两个不同工作行的选择准则,提出了随机Kaczmarz方法的双子空间拓展-双子空间随机Kaczmarz方法^[27],并从理论和数值上说明了该方法至少与随机Kaczmarz方法有相同的指数收敛速率。此外,当系数矩阵高度相干时,它能极大地提高随机Kaczmarz方法的收敛速率。此后同时选用多个工作行的随机块方法^[39-40]也被陆续提出并加以深入研究。

Nutini等^[41]研究表明,采用非等可能概率准则选取工作行的Kaczmarz方法收敛速度至少与采用等可能概率准则选取工作行的Kaczmarz方法收敛速度一样快。在原始双子空间随机Kaczmarz方法^[27]中,等可能地随机选择两个不同的工作行,然后将当前解投影到由这两行所确定的解的超平面上获得下一个近似解。基于上一次迭代所生成的残差或基于当前解离系数矩阵各行所形成的超平面的距离来选取工作行都能较大程度地提高Kaczmarz方法的收敛速率^[41]。受该思想启发,笔者通过引入控制参数 θ 建立了一种基于残差的准则来选择工作行,导出一类贪婪双子空间随机Kaczmarz方法。理论证明新方法收敛到相容线性方程组的最小范数解,而且新方法的理论收敛因子小于原始双子空间随机Kaczmarz方法的收敛因子。数值实验验证了新方法的有效性,其在迭代步数和计算时间上均优于原始双子空间随机Kaczmarz方法。

1 贪婪双子空间随机Kaczmarz方法

在经典的随机Kaczmarz方法(RK)中,Strohmer

和Vershynin将各行所对向量的欧式范数与 $\|A\|_F$ 比值的平方作为概率,然后根据此概率分布随机选取工作行。具体地说,如果定义 $A^{(i)}$ 代表系数矩阵 A 的第 i 行, $b^{(i)}$ 代表向量 b 的第 i 个分量,初始向量为 x_0 ,则随机Kaczmarz方法的具体过程见算法1。

算法1^[22] 随机Kaczmarz方法。①置 $k:=0$ 。②根据概率 $P(r=i_k)=\frac{\|A^{(i_k)}\|_2^2}{\|A\|_F^2}$ 选取指标 $i_k \in \{1, 2, \dots, m\}$ 。③计算 $x_{k+1}=x_k + \frac{b^{(i_k)} - A^{(i_k)}x_k}{\|A^{(i_k)}\|_2} (A^{(i_k)})^T$ 。④置 $k=k+1$,转步骤②。

算法1中, $P(r=i_k)=\frac{\|A^{(i_k)}\|_2^2}{\|A\|_F^2}$ 代表选取矩阵 A 的第 i_k 行作为本次迭代工作行的概率为 $\frac{\|A^{(i_k)}\|_2^2}{\|A\|_F^2}$ 。

关于随机Kaczmarz方法,有如下收敛性定理。

定理1^[22,24] 若线性系统(1)相容,其中系数矩阵 $A \in \mathbb{R}^{m \times n}$ 且右端项 $b \in \mathbb{R}^m$ 。初始向量 $x_0 \in \text{Ran}(A^T)$,其中 $\text{Ran}(A^T)$ 表示 A^T 的列空间,令 x_k 为通过随机Kaczmarz方法生成的第 k 个迭代值,则

$$E\|x_* - x_k\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{\|A\|_F^2}\right)^k \|x_* - x_0\|_2^2$$

其中 $\sigma_{\min}(\cdot)$ 表示矩阵的最小非零奇异值。

Needell和Ward发现当线性系统是超定且高度相干时,随机Kaczmarz方法的求解效率低下甚至无效,因此提出同时启用两个工作行的双子空间随机Kaczmarz方法(2S-RK)去求解这类特殊系统。

为了简便,Needell和Ward假设系数矩阵是标准化的,这意味着其每行都具有单位欧几里得范数,在接下来的部分,仍沿用该假设。在原始双子空间随机Kaczmarz方法中,等可能地随机选取两个不同工作行,再将当前解投影到由两个工作行所确定的超平面上获得下一个估计值,具体过程见算法2。

算法2^[31] 双子空间随机Kaczmarz方法。①置 $k:=0$ 。②随机均匀地选择行指标 r_k 和 s_k 。③计算 $\mu_{r_k, s_k} = \langle A^{(r_k)}, A^{(s_k)} \rangle$ 。④计算 $y_k = x_k + (b^{(s_k)} - A^{(s_k)}x_k)(A^{(s_k)})^T$ 。⑤计算 $\nu_k = \frac{A^{(r_k)} - \mu_{r_k, s_k} A^{(s_k)}}{\sqrt{1 - |\mu_{r_k, s_k}|^2}}$ 。⑥计

算 $\beta_k = \frac{\mathbf{b}^{(r_k)} - \mu_{r_k, s_k} \mathbf{b}^{(s_k)}}{\sqrt{1 - |\mu_{r_k, s_k}|^2}}$ 。⑦ 计算 $\mathbf{x}_{k+1} = \mathbf{y}_k + (\beta_k -$

$\nu_k \mathbf{y}_k)(\nu_k)^\top$ 。置 $k = k + 1$, 转步②。

关于双子空间随机 Kaczmarz 方法, 有如下收敛性定理。

定理 2^[31] 若线性系统(1)相容, 其中系数矩阵 $A \in \mathbb{R}^{m \times n}$ 且右端项 $\mathbf{b} \in \mathbb{R}^m$ 。初始向量 $\mathbf{x}_0 \in \text{Ran}(A^\top)$, 其中 $\text{Ran}(A^\top)$ 表示 A^\top 的列空间, 令 \mathbf{x}_{k+1} 为通过双子空间随机 Kaczmarz 方法生成的第 $(k+1)$ 个迭代值, 则

$$E \|\mathbf{x}_* - \mathbf{x}_{k+1}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^2 \|\mathbf{x}_* - \mathbf{x}_k\|_2^2 - \frac{1}{m(m-1)} \sum_{r \neq s} \left| \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} \langle \mathbf{e}_k^\top, A^{(r)} \rangle - \frac{\mu_{r,s}}{\sqrt{1 - \mu_{r,s}^2}} \langle \mathbf{e}_k^\top, A^{(s)} \rangle \right|^2 \quad (2)$$

这里, $\mathbf{e}_k = \mathbf{x}_* - \mathbf{x}_k, \mu_{r,s} = \langle A^{(r)}, A^{(s)} \rangle$ 。

通过算法 2, 可知双子空间随机 Kaczmarz 方法的相应误差向量的 2-范数的平方满足

$$\begin{aligned} \|\mathbf{x}_* - \mathbf{x}_{k+1}\|_2^2 &= \|\mathbf{x}_* - \mathbf{x}_k\|_2^2 - \\ & \left| \langle \mathbf{e}_k^\top, A^{(s_k)} \rangle \right|^2 - \left| \langle (\mathbf{x}_* - \mathbf{y}_k)^\top, \nu_k \rangle \right|^2 = \\ & \|\mathbf{x}_* - \mathbf{x}_k\|_2^2 - \left| \mathbf{b}^{(s_k)} - A^{(s_k)} \mathbf{x}_k \right|^2 - \\ & \frac{\left| \mathbf{b}^{(r_k)} - A^{(r_k)} \mathbf{y}_k \right|^2}{1 - |\mu_{r_k, s_k}|^2} \end{aligned} \quad (3)$$

若式(3)的最后两项的和取到最大值, 则 \mathbf{x}_* 和 \mathbf{x}_{k+1} 之间的距离可以最小化。基于这个想法, 设计最优的贪婪距离选择准则为

$$(r_k, s_k) = \max_{1 \leq s, r \leq m} \left\{ \left| \mathbf{b}^{(s)} - A^{(s)} \mathbf{x}_k \right|^2 + \frac{\left| \mathbf{b}^{(r)} - A^{(r)} \mathbf{y}_k \right|^2}{1 - |\langle A^{(r)}, A^{(s)} \rangle|^2} \right\} \quad (4)$$

若双子空间随机 Kaczmarz 方法采用式(4)的选择准则, 则其定能以一个极快的速率收敛到线性系统(1)的解。因此, 可采用式(4)的规则来建立双子空间随机 Kaczmarz 方法的最佳版本。但在实践中计算满足式(4)行指标 r_k 和 s_k 的成本非常昂贵。为克服其耗时的缺点, 笔者将通过依次构造两个具有控制参数 θ 的行索引集来构造次优版本的贪婪双子

空间随机 Kaczmarz 方法 (2S-GRK(θ)), 具体过程见算法 3。

首先, 构造与 \mathbf{x}_k 有关的指标集 U_k

$$U_k = \left\{ s \mid \left| \mathbf{b}^{(s)} - A^{(s)} \mathbf{x}_k \right|^2 \geq (1 - \theta) \max_{1 \leq i \leq m} \left\{ \left| \mathbf{b}^{(i)} - A^{(i)} \mathbf{x}_k \right|^2 \right\} \right\}, \theta \in [0, 1] \quad (5)$$

并以概率 $P(r = s_k) = \frac{\left| \hat{\mathbf{r}}_{\mathbf{x}_k}^{(s_k)} \right|^2}{\left\| \hat{\mathbf{r}}_{\mathbf{x}_k} \right\|_2^2}$ 选取指标 $s_k \in U_k$, 其中

向量 $\hat{\mathbf{r}}_{\mathbf{x}_k}$ 的定义见算法 3 的步 3。将当前解 \mathbf{x}_k 投影到解空间 $\{\mathbf{z} \mid A^{(s_k)} \mathbf{z} = \mathbf{b}^{(s_k)}\}$ 得到中间解向量 \mathbf{y}_k

$$\mathbf{y}_k = \mathbf{x}_k + \left(\mathbf{b}^{(s_k)} - A^{(s_k)} \mathbf{x}_k \right) \left(A^{(s_k)} \right)^\top$$

由此可得, 对于 $k = 1, 2, 3 \dots$ 成立

$$\begin{aligned} \mathbf{r}_{\mathbf{y}_k}^{(s_k)} = \mathbf{b}^{(s_k)} - A^{(s_k)} \mathbf{y}_k &= \mathbf{b}^{(s_k)} - A^{(s_k)} \mathbf{x}_k - \left(\mathbf{b}^{(s_k)} - A^{(s_k)} \mathbf{x}_k \right) A^{(s_k)} \left(A^{(s_k)} \right)^\top = 0 \end{aligned}$$

$$\text{即是 } \left\| \mathbf{b} - A \mathbf{y}_k \right\|_2^2 = \sum_{j \neq s_k} \left| \mathbf{b}^{(j)} - A^{(j)} \mathbf{y}_k \right|^2$$

其次, 构造与 \mathbf{y}_k 有关的指标集 \hat{U}_k

$$\hat{U}_k = \left\{ r \mid \left| \mathbf{b}^{(r)} - A^{(r)} \mathbf{y}_k \right|^2 \geq (1 - \theta) \max_{1 \leq j \leq m} \left\{ \left| \mathbf{b}^{(j)} - A^{(j)} \mathbf{y}_k \right|^2 \right\} \right\}, \theta \in [0, 1]$$

并以概率 $P(r = r_k) = \frac{\left| \hat{\mathbf{r}}_{\mathbf{y}_k}^{(r_k)} \right|^2}{\left\| \hat{\mathbf{r}}_{\mathbf{y}_k} \right\|_2^2}$ 选取指标 $r_k \in \hat{U}_k$, 其中

向量 $\hat{\mathbf{r}}_{\mathbf{y}_k}$ 的定义见算法 3 的步 8。将当前解 \mathbf{y}_k 投影到解空间 $\{\mathbf{z} \mid A^{(r_k)} \mathbf{z} = \mathbf{b}^{(r_k)}\}$ 得到满足 $A^{(s_k)} \mathbf{x}_{k+1} = \mathbf{b}^{(s_k)}$ 的解向量 \mathbf{x}_{k+1}

$$\begin{aligned} \mathbf{x}_{k+1} = \mathbf{y}_k + & \left[\frac{\mathbf{b}^{(r_k)} - \langle A^{(r_k)}, A^{(s_k)} \rangle \mathbf{b}^{(s_k)}}{\sqrt{1 - |\langle A^{(r_k)}, A^{(s_k)} \rangle|^2}} - \frac{A^{(r_k)} - \langle A^{(r_k)}, A^{(s_k)} \rangle A^{(s_k)}}{\sqrt{1 - |\langle A^{(r_k)}, A^{(s_k)} \rangle|^2}} \mathbf{y}_k \right] \times \\ & \left[\frac{A^{(r_k)} - \langle A^{(r_k)}, A^{(s_k)} \rangle A^{(s_k)}}{\sqrt{1 - |\langle A^{(r_k)}, A^{(s_k)} \rangle|^2}} \right]^\top \end{aligned}$$

由指标集 \hat{U}_k 的定义, 对于 $\forall r \in \hat{U}_k$, 可得

$$\begin{aligned} \left| \mathbf{r}_{\mathbf{y}_k}^{(r)} \right|^2 = \left| \mathbf{b}^{(r)} - A^{(r)} \mathbf{y}_k \right|^2 &\geq \\ (1 - \theta) \max_{1 \leq j \leq m} \left\{ \left| \mathbf{b}^{(j)} - A^{(j)} \mathbf{y}_k \right|^2 \right\} &= \end{aligned}$$

$$\begin{aligned}
 (1-\theta) \|b - Ay_k\|_2^2 &= \frac{\max_{1 \leq j \leq m} \{ |b^{(j)} - A^{(j)} y_k|^2 \}}{\|b - Ay_k\|_2^2} \\
 (1-\theta) \|b - Ay_k\|_2^2 &\geq \frac{\max_{1 \leq j \leq m} \{ |b^{(j)} - A^{(j)} y_k|^2 \}}{\sum_{j \neq s_k} |b^{(j)} - A^{(j)} y_k|^2} \\
 (1-\theta) \|b - Ay_k\|_2^2 &= \frac{\max_{1 \leq j \leq m} \{ |b^{(j)} - A^{(j)} y_k|^2 \}}{(m-1) \max_{1 \leq j \leq m} \{ |b^{(j)} - A^{(j)} y_k|^2 \}} = \\
 \frac{1-\theta}{m-1} \|b - Ay_k\|_2^2 & \quad (6)
 \end{aligned}$$

对于 $k=1, 2, 3, \dots$, 同样可得 $r_{x_{k+1}}^{(s_k)} = b^{(s_k)} - A^{(s_k)} x_{k+1} = 0$ 和 $r_{x_{k+1}}^{(r_k)} = b^{(r_k)} - A^{(r_k)} x_{k+1} = 0$, 即是 $\|b - Ax_{k+1}\|_2^2 = \sum_{i \neq s_k, r_k} |b^{(i)} - A^{(i)} x_{k+1}|^2$ 。同理, 由指标集 U_k 的定义, 对于 $\forall s \in U_k$, 可得

$$\begin{aligned}
 |r_{x_k}^{(s)}|^2 &= |b^{(s)} - A^{(s)} x_k|^2 \geq \\
 (1-\theta) \max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \} &= \\
 (1-\theta) \|b - Ax_k\|_2^2 &= \frac{\max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \}}{\|b - Ax_k\|_2^2} = \\
 (1-\theta) \|b - Ax_k\|_2^2 &\geq \frac{\max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \}}{\sum_{i \neq s_{k-1}, r_{k-1}} |b^{(i)} - A^{(i)} x_k|^2} \\
 (1-\theta) \|b - Ax_k\|_2^2 &= \frac{\max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \}}{(m-2) \max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \}} = \\
 \frac{1-\theta}{m-2} \|b - Ax_k\|_2^2 & \quad (7)
 \end{aligned}$$

算法3 贪婪双子空间随机Kaczmarz方法。

步1 置 $k:=0$ 。计算 $\epsilon_k = (1-\theta) \max_{1 \leq i \leq m} \{ |b^{(i)} - A^{(i)} x_k|^2 \}$ 。定义正整数指标集 $U_k = \{s \mid |b^{(s)} - A^{(s)} x_k|^2 \geq \epsilon_k\}$ 。计算 \hat{r}_{x_k} 的第 s 个分量 $\hat{r}_{x_k}^{(s)} = \begin{cases} b^{(s)} - A^{(s)} x_k, & s \in U_k \\ 0, & s \notin U_k \end{cases}$

步2 根据概率 $P(r=s_k) = \frac{|\hat{r}_{x_k}^{(s_k)}|^2}{\|\hat{r}_{x_k}\|_2^2}$ 选取指标

$s_k \in U_k$ 。

步3 计算 $y_k = x_k + (b^{(s_k)} - A^{(s_k)} x_k)(A^{(s_k)})^T$, $\hat{\epsilon}_k = (1-\theta) \max_{1 \leq j \leq m} \{ |b^{(j)} - A^{(j)} y_k|^2 \}$ 。

步4 定义正整数指标集 $\hat{U}_k = \{r \mid |b^{(r)} - A^{(r)} y_k|^2 \geq \hat{\epsilon}_k\}$ 。计算 \hat{r}_{y_k} 的第 r 个分量 $\hat{r}_{y_k}^{(r)} = \begin{cases} b^{(r)} - A^{(r)} y_k, & r \in \hat{U}_k \\ 0, & r \notin \hat{U}_k \end{cases}$

步5 根据概率 $P(r=r_k) = \frac{|\hat{r}_{y_k}^{(r_k)}|^2}{\|\hat{r}_{y_k}\|_2^2}$ 选取指标

$r_k \in \hat{U}_k$ 。

步6 计算 $\mu_{r_k, s_k} = \langle A^{(r_k)}, A^{(s_k)} \rangle$, $\nu_k = \frac{A^{(r_k)} - \mu_{r_k, s_k} A^{(s_k)}}{\sqrt{1 - |\mu_{r_k, s_k}|^2}}$;

$\beta_k = \frac{b^{(r_k)} - \mu_{r_k, s_k} b^{(s_k)}}{\sqrt{1 - |\mu_{r_k, s_k}|^2}}$ 。

步7 计算 $x_{k+1} = y_k + (\beta_k - \nu_k y_k)(\nu_k)^T$ 。

步8 置 $k=k+1$, 转步1。

此外, 由 y_k 和 x_{k+1} 的递推式可以推得相对应的残量的递推公式如下:

$$\begin{aligned}
 r_{y_k} &= b - Ay_k = \\
 b - A(x_k + (b^{(s_k)} - A^{(s_k)} x_k)(A^{(s_k)})^T) &= \\
 b - Ax_k - (b^{(s_k)} - A^{(s_k)} x_k)A(A^{(s_k)})^T &= \\
 r_{x_k} - r_{x_k}^{(s_k)} A(A^{(s_k)})^T &= \\
 r_k - r_{x_k}^{(s_k)} B_{(s_k)} & \\
 r_{x_{k+1}} &= b - Ax_{k+1} =
 \end{aligned}$$

$$b - A(y_k + (\beta_k - \nu_k y_k)(\nu_k)^T) =$$

$$b - Ay_k - (\beta_k - \nu_k y_k)A(\nu_k)^T =$$

$$r_{y_k} - \frac{r_{y_k}^{(r_k)} - \mu_k r_{y_k}^{(s_k)}}{\sqrt{1 - |\mu_k|^2}} \frac{AA^T_{(r_k)} - \mu_k AA^T_{(s_k)}}{\sqrt{1 - |\mu_k|^2}} =$$

$$r_{y_k} - \frac{r_{y_k}^{(r_k)} - \mu_k r_{y_k}^{(s_k)}}{1 - |\mu_k|^2} (B_{(r_k)} - \mu_k B_{(s_k)})$$

其中 $B=AA^T$ 且 $B_{(s_k)}, B_{(r_k)}$ 表示矩阵 B 的第 s_k, r_k 列。若迭代前已知 AA^T , 则可进一步提高贪婪双子空间随机Kaczmarz方法的求解性能, 参见文献[24]。关于采用随机策略近似计算矩阵 AA^T , 参见文献[42]。

若每次迭代仅选择一个工作行, 则算法3成为

一类广义贪婪随机 Kaczmarz (GRK(θ)) 方法。通过改变参数 θ , 可得到一系列常用算法作为特殊情况。

$$\text{取 } \theta=0, \theta=\frac{1}{2}, \theta=\frac{1}{2}\left(1-\frac{\|b-Ax_k\|}{m \max_{1 \leq i \leq m}\{|b^{(i)}-A^{(i)}x_k|^2\}}\right),$$

$$\theta=(1-\vartheta)\left(1-\frac{\|b-Ax_k\|}{m \max_{1 \leq i \leq m}\{|b^{(i)}-A^{(i)}x_k|^2\}}\right) (\vartheta \in [0, 1])$$

时, GRK(θ) 方法分别简化为贪婪 Kaczmarz 方法^[41], 贪婪距离随机 Kaczmarz 方法^[25], 贪婪随机 Kaczmarz 方法^[24]和松弛贪婪随机 Kaczmarz 方法^[33]。

2 贪婪双子空间随机 Kaczmarz 方法的收敛性分析

相容线性系统(1)的最小范数解 x_* 形如

$$x_* := \arg \min_{Ax=b} \|x\|_2 \triangleq A^\dagger b \in \text{Ran}(A^T)$$

其中 $\text{Ran}(A^T)$ 表示 A^T 的列空间。若初始向量 $x_0 \in \text{Ran}(A^T)$, 由算法 3 的步 5 和步 13 可知, 算法 3 生成的迭代序列 $\{x_k\}_{k=0}^\infty$ 一定在 $\text{Ran}(A^T)$ 中, 故若算法 3 收敛, 则一定收敛到相容线性系统(1)的最小范数解。

关于算法 3 特殊情况(每次迭代仅选一个工作行)——广义贪婪随机 Kaczmarz 方法, 有如下收敛性定理。

定理 3 若线性系统(1)相容, 其中系数矩阵 $A \in \mathbb{R}^{m \times n}$ 且右端项 $b \in \mathbb{R}^m$ 。初始向量 $x_0 \in \text{Ran}(A^T)$, 令 x_{k+1} 为通过广义贪婪随机 Kaczmarz 方法生成的第 $(k+1)$ 个迭代值, 则有

$$E\|x_* - x_{k+1}\|_2^2 \leq \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) \|x_* - x_k\|_2^2 \quad (8)$$

和

$$E\|x_* - x_{k+1}\|_2^2 \leq \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right)^{k+1} \|x_* - x_0\|_2^2, \quad k=0, 1, 2, \dots$$

证明: 由广义贪婪随机 Kaczmarz 方法(算法 3 的前 5 步)可得

$$x_{k+1} = x_k + (b^{(s_k)} - A^{(s_k)}x_k)(A^{(s_k)})^T$$

根据概率 $P(r=s_k) = \frac{|\hat{r}_{x_k}^{(s_k)}|^2}{\|\hat{r}_{x_k}\|_2^2}$ 选指标 $s_k \in U_k$ 。由

$A^{(s_k)}(x_* - x_{k+1}) = 0$, 即 $x_* - x_k$ 与 $x_{k+1} - x_k$ 正交, 可得

$$\|x_* - x_{k+1}\|_2^2 = \|x_* - x_k\|_2^2 - \|x_k - x_{k+1}\|_2^2 \quad (9)$$

对等式(9)两侧取期望, 有

$$E\|x_* - x_{k+1}\|_2^2 =$$

$$\|x_* - x_k\|_2^2 - E\|x_k - x_{k+1}\|_2^2 =$$

$$\|x_* - x_k\|_2^2 - \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}\|_2^2} \|(b^{(s)} - A^{(s)}x_k)(A^{(s)})^T\|_2^2 =$$

$$\|x_* - x_k\|_2^2 - \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}\|_2^2} |b^{(s)} - A^{(s)}x_k|^2 \quad (10)$$

由 $r_{x_{k+1}}^{(s_k)} = b^{(s_k)} - A^{(s_k)}x_{k+1} = 0, k=0, 1, 2, \dots$, 则

$$\|b - Ax_{k+1}\|_2^2 = \sum_{i \neq s_k} |b^{(i)} - A^{(i)}x_{k+1}|^2$$

(6), 对于 $\forall r \in \hat{U}_k$, 可以推得

$$|r_{x_k}^{(s)}|^2 = |b^{(s)} - A^{(s)}x_k|^2 \geq \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \|x_* - x_k\|_2^2 \quad (11)$$

其中式(11)最后一个不等式需要用到不等式(12), 即对于任意 $u \in \text{Ran}(A^T)$, 成立

$$\|Au\|_2^2 \geq \sigma_{\min}^2(A) \|u\|_2^2 \quad (12)$$

联立等式(11)和不等式(12)可得

$$E\|x_* - x_{k+1}\|_2^2 = \|x_* - x_k\|_2^2 -$$

$$\sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}\|_2^2} |b^{(s)} - A^{(s)}x_k|^2 \leq$$

$$\|x_* - x_k\|_2^2 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \|x_* - x_k\|_2^2 =$$

$$\|x_* - x_k\|_2^2 \sum_{s \in U_k} \frac{|b^{(s)} - A^{(s)}x_k|^2}{\|\hat{r}_{x_k}\|_2^2} =$$

$$\|x_* - x_k\|_2^2 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \|x_* - x_k\|_2^2 =$$

$$\left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) \|x_* - x_k\|_2^2$$

通过归纳法, 可得

$$E\|x_* - x_{k+1}\|_2^2 \leq \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right)^{k+1} \|x_* - x_0\|_2^2$$

$$K=0, 1, 2, \dots$$

关于贪婪双子空间随机Kaczmarz方法,有如下的收敛性定理。

定理4 若线性系统(1)相容,其中系数矩阵 $A \in \mathbf{R}^{m \times n}$ 且右端项 $b \in \mathbf{R}^m$ 。初始向量 $x_0 \in \text{Ran}(A^T)$, 令 x_{k+1} 为通过贪婪双子空间随机Kaczmarz方法生成的第 $(k+1)$ 个迭代值,则

$$E \|x_* - x_{k+1}\|_2^2 \leq M(\theta) \|x_* - x_k\|_2^2 - \sum_{r \in U_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}^{(s)}\|_2^2} |\mu_{r,s}|^2 |\langle e_k^T, \nu_{r,s} \rangle|^2 \quad (13)$$

这里, $M(\theta) = \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) \left(1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A)\right)$, $e_k = x_* - x_k$, $\mu_{r,s} = \langle A^{(r)}, A^{(s)} \rangle$, $\nu_{r,s} = \frac{A^{(r)} - \mu_k A^{(s)}}{\sqrt{1 - |\mu_{r,s}|^2}}$ 。进一步放缩可得一个松弛但更简单的收敛界

$$E \|x_* - x_{k+1}\|_2^2 \leq M_k(\theta) \|x_* - x_k\|_2^2$$

其中 $M_k(\theta) = \left(1 - \frac{(1+D_k)(1-\theta)}{m-1} \sigma_{\min}^2(A)\right) \left(1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A)\right)$, $D_k = \min_{r \in U_k, s \in U_k} \left\{ \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} \right\}$ 。

证明: 由算法3的步5和步13可得

$$\begin{aligned} y_k &= x_k + (b^{(s_k)} - A^{(s_k)} x_k) (A^{(s_k)})^T = \\ & x_k + \langle e_k^T, A^{(s_k)} \rangle (A^{(s_k)})^T \\ x_{k+1} &= y_k + (\beta_k - \nu_k y_k) (\nu_k)^T = \\ & x_k + \langle e_k^T, A^{(s_k)} \rangle (A^{(s_k)})^T + \langle e_k^T, \nu_k \rangle (\nu_k)^T \end{aligned}$$

记 $\xi_k = \langle A^{(r_k)}, \nu_k \rangle$ 。为了便利,在后续的证明中, $A^{(r_k)}$ 、 $A^{(s_k)}$ 、 μ_k 、 ν_k 、 ξ_k 分别简记为 $A^{(r)}$ 、 $A^{(s)}$ 、 μ 、 ν 、 ξ 。由于 $\langle A^{(s)}, \nu \rangle = 0$, 即向量 ν 正交于向量 $A^{(s)}$, 展开 $\|x_* - x_{k+1}\|_2^2$ 可得

$$\|x_* - x_{k+1}\|_2^2 = \|e_k\|_2^2 - |\langle e_k^T, A^{(s)} \rangle|^2 - |\langle e_k^T, \nu \rangle|^2 \quad (14)$$

在算法3中,一次迭代需执行两次投影。因此,将式(14)中的误差与广义贪婪随机Kaczmarz方法两次迭代后获得的误差进行比较。设 z 是广义贪婪随机Kaczmarz方法基于 y_k 的下一代迭代值,则有

$$\begin{aligned} z &= y_k + (b^{(r)} - A^{(r)} y_k) (A^{(r)})^T = \\ & x_k + \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T + \\ & \langle x_*^T - y_k^T, A^{(r)} \rangle (A^{(r)})^T \end{aligned} \quad (15)$$

由 y_k 的递推式可得

$$\begin{aligned} \langle x_*^T - y_k^T, A^{(r)} \rangle &= \\ \langle e_k^T, A^{(r)} \rangle - \langle e_k^T, A^{(s)} \rangle \langle A^{(s)}, A^{(r)} \rangle &= \\ \langle e_k^T, A^{(r)} \rangle - \mu \langle e_k^T, A^{(s)} \rangle &= \\ \langle e_k^T, A^{(r)} - \mu A^{(s)} \rangle \end{aligned} \quad (16)$$

由 μ 、 ν 、 ξ 的定义可得

$$A^{(r)} = \mu A^{(s)} + \nu, \xi^2 + \mu^2 = 1 \quad (17)$$

联立式(15)~(17)可得

$$\begin{aligned} z &= x_k + \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T + \\ & \langle e_k^T, A^{(r)} - \mu A^{(s)} \rangle (A^{(r)})^T = \\ & x_k + \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T + \\ & \langle e_k^T, \xi \nu \rangle (\mu A^{(s)} + \xi \nu)^T = \\ & x_k + \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T + \\ & \xi \mu \langle e_k^T, \nu \rangle (A^{(s)})^T + \\ & \xi^2 \langle e_k^T, \nu \rangle (\nu)^T = \\ & x_k + \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T + \\ & \xi \mu \langle e_k^T, \nu \rangle (A^{(s)})^T + \langle e_k^T, \nu \rangle (\nu)^T - \\ & \mu^2 \langle e_k^T, \nu \rangle (\nu)^T \end{aligned}$$

利用向量 ν 与向量 $A^{(s)}$ 的正交性,展开 $\|x_* - z\|_2^2$ 可得

$$\begin{aligned} \|x_* - z\|_2^2 &= \|e_k - \langle e_k^T, A^{(s)} \rangle (A^{(s)})^T - \\ & \xi \mu \langle e_k^T, \nu \rangle (A^{(s)})^T - \\ & |\langle e_k^T, \nu \rangle|^2 \nu^T + \mu^2 \langle e_k^T, \nu \rangle \nu^T\|_2^2 = \\ & \|e_k\|_2^2 - |\langle e_k^T, A^{(s)} \rangle|^2 - \\ & |\langle e_k^T, \nu \rangle|^2 + \mu^2 |\langle e_k^T, \nu \rangle|^2 \end{aligned} \quad (18)$$

联立式(14)和式(18)可得

$$\|x_* - x_{k+1}\|_2^2 = \|x_* - z\|_2^2 - |\mu \langle e_k^T, \nu \rangle|^2 \quad (19)$$

对式(19)两端取期望可得

$$E \|x_* - x_{k+1}\|_2^2 = E \|x_* - z\|_2^2 - E |\mu \langle e_k^T, \nu \rangle|^2 \quad (20)$$

由定理3可得

$$E \|x_* - z\|_2^2 \leq \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) E \|x_* - y_k\|_2^2 \quad (21)$$

再由

$$E\|x_* - y_k\|_2^2 = \|x_* - x_k\|_2^2 - E|\mathbf{b}^{(s_k)} - A^{(s_k)}x_k|^2 = \|x_* - x_k\|_2^2 - \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}^{(s)}\|_2^2} |\mathbf{b}^{(s)} - A^{(s)}x_k|^2 \leq \|x_* - x_k\|_2^2 - \frac{1-\theta}{m-2} \|\mathbf{b} - Ax_k\|_2^2 \text{ (由等式 (7))} \leq \left(1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A)\right) \|x_* - x_k\|_2^2$$

因此, 可得

$$E\|x_* - z\|_2^2 \leq \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) \left(1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A)\right) \|x_* - x_k\|_2^2 \quad (22)$$

再由期望的定义可得

$$E|\mu \langle e_k^T, \nu \rangle|^2 = \sum_{r \in \hat{U}_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}^{(s)}\|_2^2} |\mu_{r,s}|^2 |\langle e_k^T, \nu_{r,s} \rangle|^2 \quad (23)$$

把式(22)和式(23)带入式(20)可得式(13)。

把式(16)和式(17)带入式(23), 同时再结合式(6)可得

$$E|\mu \langle e_k^T, \nu \rangle|^2 = \sum_{r \in \hat{U}_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}^{(s)}\|_2^2} |\mu_{r,s}|^2 |\langle e_k^T, \nu_{r,s} \rangle|^2 = \sum_{r \in \hat{U}_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} |\langle x_*^T - y_k^T, A^{(r)} \rangle|^2 \geq D_k \sum_{r \in \hat{U}_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} |\langle x_*^T - y_k^T, A^{(r)} \rangle|^2 \geq D_k \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \|x_* - y_k\|_2^2 \quad (24)$$

其中 $D_k = \min_{r \in \hat{U}_k, s \in U_k} \left\{ \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} \right\}$ 。把式(21)和式(24)

带入式(20)可得

$$E\|x_* - x_{k+1}\|_2^2 = \left(1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A)\right) \|x_* - y_k\|_2^2 - D_k \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \|x_* - y_k\|_2^2 =$$

$$\left(1 - \frac{(1+D_k)(1-\theta)}{m-1} \sigma_{\min}^2(A)\right) \|x_* - y_k\|_2^2 = \left(1 - \frac{(1+D_k)(1-\theta)}{m-1} \sigma_{\min}^2(A)\right) E\|x_* - y_k\|_2^2 \leq \left(1 - \frac{(1+D_k)(1-\theta)}{m-1} \sigma_{\min}^2(A)\right) \left(1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A)\right) \|x_* - x_k\|_2^2$$

收敛因子越小, 方法收敛得越快。通过定理 4 收敛界可知, 参数 θ 越小, 收敛因子 $M_k(\theta)$ 越小, 即贪婪双子空间随机 Kaczmarz 方法收敛速度越快。为便于比较, 用 θ 取值特殊的贪婪双子空间随机 Kaczmarz 方法与原双子空间随机 Kaczmarz 方法作比较, 但对实际问题, 因贪婪双子空间随机 Kaczmarz 方法选进行概率更具优势性, 对于 $\forall \theta \in [0, 1]$, 贪婪双子空间随机 Kaczmarz 方法都比原双子空间随机 Kaczmarz 方法收敛更快。

事实 ① 令 t_1, t_2, \dots, t_l 是在概率分别为 p_1, p_2, \dots, p_l 的某个概率空间上定义的一组随机变量。若越大的 $|t_i|$ 所对应的概率 p_i 越大, 则值 $\sum_{i=1}^l |t_i| p_i$ 越大; ② 对于任意数组 $\Gamma = \{t_1, t_2, \dots, t_l\}$, $t_i \in \mathbf{R}$, $\text{ave}(\Gamma)$ 表示数组 Γ 的平均值。令 $\hat{\Gamma} = \{t_i \geq \text{ave}(\Gamma), t_i \in \Gamma\}$, 有 $\text{ave}(\hat{\Gamma}) \geq \text{ave}(\Gamma)$ 。

当 $\theta \leq \frac{1}{m}$ 时, 显然成立

$$1 - \frac{1-\theta}{m-2} \sigma_{\min}^2(A) \leq 1 - \frac{1-\theta}{m-1} \sigma_{\min}^2(A) \leq 1 - \frac{\sigma_{\min}^2(A)}{m}$$

$$\text{即 } M(\theta) \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^2。$$

同时, 当 $\theta \leq \frac{1}{m}$ 时, 对于任意 $r \in \hat{U}_k$, 显然成立

$$|\mathbf{r}_{y_k}^{(r)}|^2 \geq (1-\theta) \max_{1 \leq j \leq m} \{|\mathbf{b}^{(j)} - A^{(j)}y_k|^2\} \geq \frac{m-1}{m} \max_{1 \leq j \leq m} \{|\mathbf{b}^{(j)} - A^{(j)}y_k|^2\} \geq \frac{\|\mathbf{b} - Ay_k\|_2^2}{m}。$$

基于事实①和②可得

$$E|\mu \langle e_k^T, \nu \rangle|^2 = \sum_{r \in \hat{U}_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}^{(r)}\|_2^2} \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}^{(s)}\|_2^2} |\mu_{r,s}|^2 |\langle e_k^T, \nu_{r,s} \rangle|^2 =$$

$$\begin{aligned} & \sum_{r \in U_k} \frac{|\hat{r}_{y_k}^{(r)}|^2}{\|\hat{r}_{y_k}\|_2^2} \sum_{s \in U_k} \frac{|\hat{r}_{x_k}^{(s)}|^2}{\|\hat{r}_{x_k}\|_2^2} \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} |\langle x_*^T - y_k^T, A^{(r)} \rangle|^2 \geq \\ & \frac{1}{|\hat{U}_k|} \frac{1}{|U_k|} \sum_{r \in U_k} \sum_{s \in U_k} \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} |\langle x_*^T - y_k^T, A^{(r)} \rangle|^2 \geq \\ & \frac{1}{m-1} \frac{1}{m} \sum_{r \neq s} \sum_{s \in m} \frac{|\mu_{r,s}|^2}{1 - |\mu_{r,s}|^2} |\langle x_*^T - y_k^T, A^{(r)} \rangle|^2 = \\ & \frac{1}{m(m-1)} \sum_{r \neq s} \left| \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} \langle e_k^T, A^{(r)} \rangle - \right. \\ & \left. \frac{\mu_{r,s}}{\sqrt{1 - \mu_{r,s}^2}} \langle e_k^T, A^{(s)} \rangle \right|^2 \end{aligned}$$

因此,贪婪双子空间随机 Kaczmarz 方法收敛因子小于原双子空间随机 Kaczmarz 方法的收敛因子。

浮点运算量方面,原双子空间随机 Kaczmarz 方法每次迭代需要 $12n + 4$ 个浮点运算量。若残差由递推公式计算,则贪婪双子空间随机 Kaczmarz 方法每次迭代需要 $16m + 6n + 7$ 个浮点运算量。因此,当 $m < \frac{6n-3}{16}$ 时,贪婪双子空间随机 Kaczmarz 方法计算量小于原双子空间随机 Kaczmarz 方法。

3 数值实验

在本节中,分别通过数值实验比较随机 Kaczmarz 方法(RK)、双子空间随机 Kaczmarz 方法(2S-RK)和贪婪双子空间随机 Kaczmarz 方法(2S-GRK(θ)),并显示后者无论在迭代步数(IT)还是计算时间(CPU)上均优于前两者。这里,迭代步数(IT)和计算时间(CPU)取 30 次计算结果的平均值。

在实验中,通过 Matlab 函数 randn 随机生成解 $x_* \in \mathbb{R}^n$,右端项 $b \in \mathbb{R}^m$ 由 Ax_* 给出。此外,需要指出 2S-GRK(θ)方法按照算法 3 中定义的过程精确执行,并没有显式地计算矩阵 $B = AA^T$ 来计算残量。本节的所有数值实验,均取初始向量为 $x_0 = 0$,停机准则为近似解的相对误差 R_{SE} 满足

$$R_{SE} = \frac{\|x_* - x_k\|_2^2}{\|x_*\|_2^2} < 10^{-6}$$

或者迭代步数超过 30 万。在实验表格中,若在 30 万步内未达到指定精度,即用“—”表示。

为了测试这些方法,选取了两类矩阵进行实验。一类是通过 Matlab 函数 unifrnd 生成的 500×100 的相干矩阵。矩阵元素是在区间 $[d, 1]$ 上独立且均匀分布的随机变量。通过改变 d 的值,可以得到具有不同相干系数对的矩阵,相干系数对 (σ, Δ) 定义如下 $\sigma = \sigma(A) = \min_{i \neq j} |\langle A^{(i)}, A^{(j)} \rangle|, \Delta = \Delta(A) =$

$$\max_{i \neq j} |\langle A^{(i)}, A^{(j)} \rangle|$$

表 1 列出了测试的相干矩阵的相关信息。包含不同相干程度的 500×100 矩阵。

表 1 相干矩阵信息

Tab.1 Information of coherent matrices

矩阵名称	阶数	取值区间	相干系数对
A1	500×100	$[-0.4, 1]$	(0.008 4, 1)
A2	500×100	$[-0.1, 1]$	(0.477 9, 1)
A3	500×100	$[0.2, 1]$	(0.799 3, 1)
A4	500×100	$[0.5, 1]$	(0.944 8, 1)
A5	500×100	$[0.8, 1]$	(0.993 5, 1)

对于测试的矩阵,表 2 给出了 RK, 2S-RK, 2S-GRK(θ) 3 种方法的迭代步数(IT)与计算时间(CPU),以及 2S-GRK(θ)方法对于 2S-RK 方法的加速比。

表 2 相干矩阵数值结果

Tab.2 Numerical results of coherent matrices

算法	A1		A2		A3		A4		A5	
	IT	CPU	IT	CPU	IT	CPU	IT	CPU	IT	CPU
RK	2 530.0	0.133 8	5 020.0	0.255 2	12 120.0	0.618 7	43 870.0	2.236 4	—	—
2S-RK	1 130.6	0.034 5	1 474.7	0.036 0	1 644.8	0.039 1	1 752.6	0.041 5	1 745.8	0.041 1
2S-GRK(θ)	124.0	0.013 7	136.0	0.011 0	142.0	0.011 3	150	0.011 7	141.0	0.011 2
加速比	2.48		3.27		3.46		3.55		3.64	

由表 2 可知,2S-RK 与 2S-GRK(θ)方法总能计算出符合精度的解,而 RK 方法有时在迭代步数达到 30 万步后仍不能计算出符合精度的解。而且即便 RK 方法收敛,2S-RK 与 2S-GRK(θ)方法在迭代步数与计算时间上也均少于 RK 方法。此外,2S-

GRK(θ)方法进一步优化了 2S-RK 方法,其关于 2S-RK 方法迭代时间的加速比最大可以达到 3.64,最小可以达到 2.48。值得注意的是,无论矩阵的相干程度如何,2S-GRK(θ)方法均收敛,且其迭代步数与计算时间均优于原始的 2S-RK 方法。

图 1 描绘了相干矩阵 A2(图 1a)和 A4(图 1b)的近似解的相对误差 R_{SE} 以 10 为底的对数随着迭代步

数变化的曲线, 进一步验证了 2S-GRK(θ)方法比经典的 2S-RK 方法收敛更快。

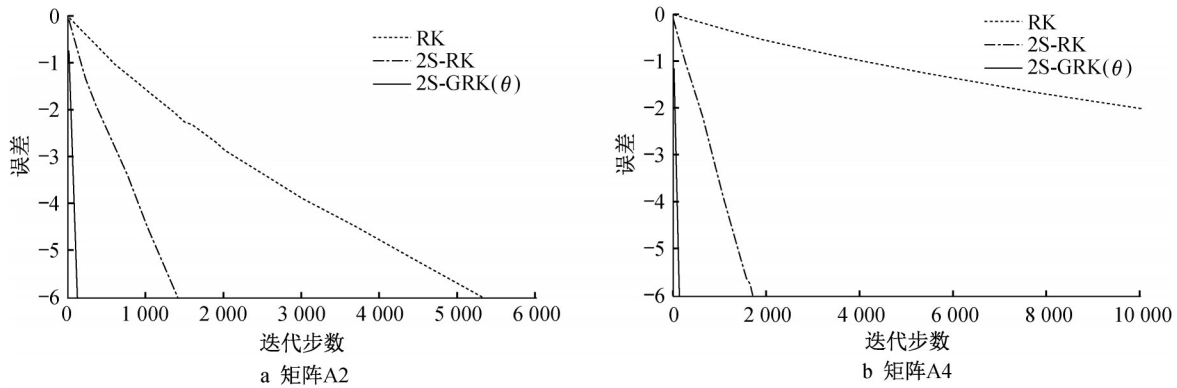


图 1 以 A2 和 A4 为系数矩阵的线性系统的近似解的相对误差随着迭代步数变化的曲线

Fig.1 $\lg(R_{SE})$ versus IT for linear systems with coefficient matrices: A2 and A4

表 3 列出了测试的稀疏矩阵的相关信息。

表 3 稀疏矩阵信息

Tab.3 Information of sparse matrices

矩阵名称	阶数	非零元个数	相干系数对
Abtaha1	$14\ 596 \times 209$	51 307	(0.816 5, 1)
Ash219	219×85	438	(0, 1)
Ash331	331×104	662	(0, 1)
Ash608	608×188	1 216	(0, 1)
Ash958	958×292	1 916	(0, 1)

对于测试的稀疏矩阵, 表 4 给出了 RK、2S-RK、2S-GRK(θ) 3 种方法的迭代步数 (IT) 与计算时间

(CPU), 以及 2S-GRK(θ) 方法对于 2S-RK 方法的加速比。

由表 4 可得与表 2 类似结论, 即 2S-GRK(θ) 与 2S-RK 方法在迭代步数与计算时间上也均少于 RK 方法。此外, 2S-GRK(θ) 方法进一步优化了 2S-RK 方法, 其关于 2S-RK 方法迭代时间的加速比最大可以达到 6.17, 最小可以达到 1.75。因此, 无论对于相干矩阵还是稀疏 (不相干) 矩阵, 2S-GRK(θ) 方法均收敛, 且其迭代步数与计算时间均优于传统的 2S-RK 方法。

表 4 稀疏矩阵数值结果

Tab.4 Numerical results of sparse matrices

算法	Abtaha1		Ash219		Ash331		Ash608		Ash958	
	IT	CPU	IT	CPU	IT	CPU	IT	CPU	IT	CPU
RK	81 883	26. 572 9	1 896	0. 123 1	2 183	0. 145 1	4 119	0. 302 7	5. 885	0. 512 8
2S-RK	24 931	12. 346 3	901	0. 065 0	1 050	0. 078 4	2 056	0. 184 3	3. 005	0. 316 3
2S-GRK(θ)	254	7. 059 6	127	0. 010 6	129	0. 012 7	237	0. 030 2	353	0. 055 7
加速比	1. 75		6. 13		6. 17		6. 10		5. 68	

图 2 描绘了矩阵 Ash219(图 2a)和 Ash958(图 2b)的近似解的相对误差以 10 为底的对数随着迭代步数变化的曲线, 进一步验证了 2S-GRK(θ)方法比经典的 2S-RK 方法收敛更快。

对于 4 个测试矩阵 A2、A4、Ash219 和 Ash958, 图 3 描绘了 2S-GRK(θ) 方法的计算时间(图 3a)和迭代步数(图 3b)随控制参数 θ 变化的曲线。由图 3 可知, 只要在计算时采用合适的控制参数 θ , 2S-GRK(θ) 方法的性能就会得到很大提升。结合表 2 和表 4 可知, 无论 $\theta \in [0, 1]$ 取何值, 2S-GRK(θ) 方法总比经典的 2S-RK 方法收敛更快。

4 结论

提出一类贪婪双子空间随机 Kaczmarz 方法, 理论分析证明新方法的收敛性, 还表明收敛因子小于传统的双子空间随机 Kaczmarz 方法的收敛因子。数值实验结果也表明所提出的新方法在迭代步数和计算时间上均优于传统的双子空间随机 Kaczmarz 方法。贪婪双子空间随机 Kaczmarz 方法能够快速求解大规模稀疏相容线性方程组, 主要原因在于新方法定义的随机选取工作行的概率更具优势。如何随机选取工作行以加速 Kaczmarz 方法及其应用^[43] 仍然是一个值得研究的问题。

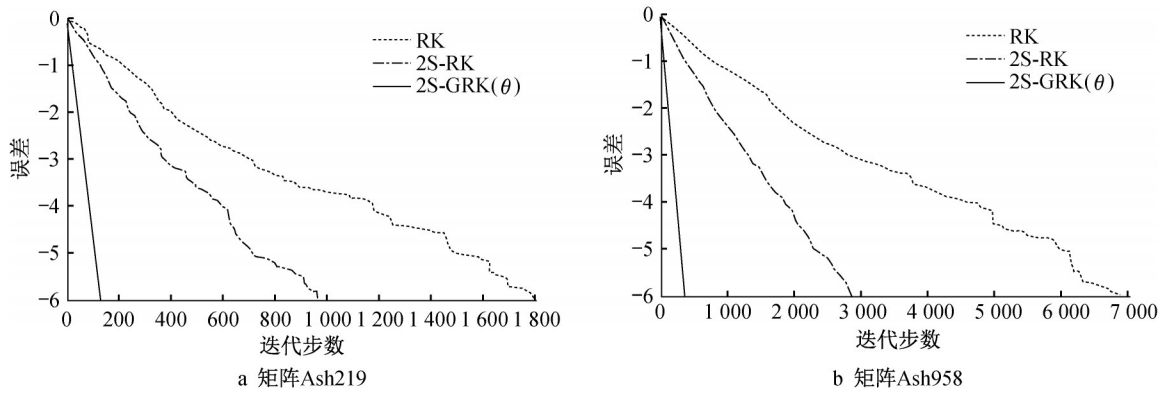


图2 以 Ash219 和 Ash958 为系数矩阵的线性系统的近似解的相对误差随着迭代步数变化的曲线
 Fig.2 $\lg(R_{SE})$ versus IT for linear systems with coefficient matrices: Ash219 and Ash958

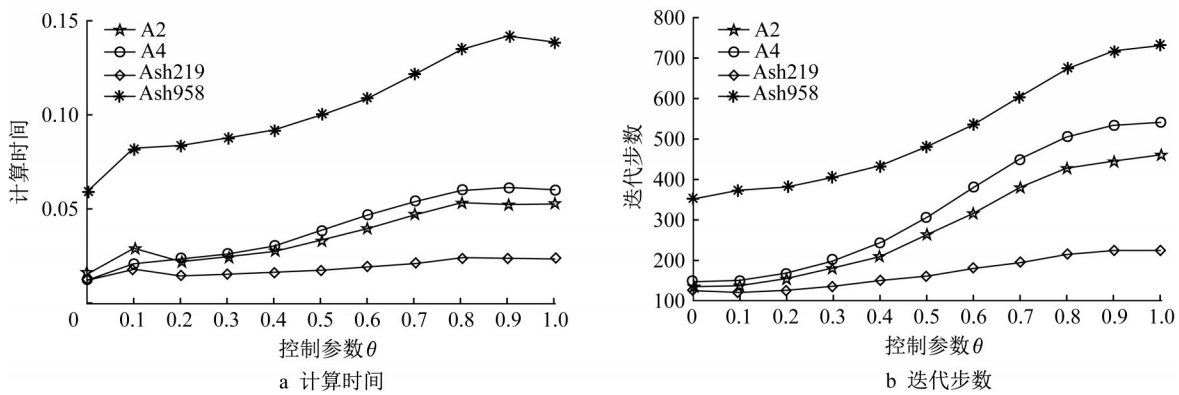


图3 2S-GRK(theta)方法的计算时间和迭代步数随控制参数theta变化的曲线
 Fig.3 CPU and IT versus theta of 2S-GRK(theta)

作者贡献声明:

荆燕飞:核心思想提炼、论文修改。
 李彩霞:论文撰写。
 胡少亮:论文修改。

参考文献:

[1] BREZINSKI C. Projection methods for systems of equations [M]. Amsterdam: North-Holland Publishing Co, 1997.
 [2] SAAD Y. Iterative methods for sparse linear systems [M]. Philadelphia: SIAM Publisher, 2003.
 [3] GALÁNTAI A. Projectors and projection methods [M]. Boston: Kluwer Academic Publishers, 2003.
 [4] GOLUB G H. Matrix computations [M]. 3rd ed. Baltimore: Johns Hopkins UP, 2008.
 [5] KACZMARZ S. Angenäherte auflösung von systemen linearer gleichungen [J]. Bulletin International de l'Academie Polonaise des Sciences et des Lettres, 1937, 35: 355.
 [6] ANSORGE R. Connections between the Cimmino-method and the Kaczmarz-method for the solution of singular and regular systems of equations [J]. Computing, 1984, 33: 367.

[7] CENSOR Y. Row-action methods for huge and sparse systems and their applications [J]. SIAM Review, 1981, 23: 444.
 [8] KNIGHT P A. Error analysis of stationary iteration and associated problems [D]. Manchester: Manchester University, 1993.
 [9] BAI Z Z, LIU X G. On the meany inequality with applications to convergence analysis of several row-action iteration methods [J]. Numerische Mathematik, 2013, 124: 215.
 [10] CENSOR Y. Parallel application of block-iterative methods in medical imaging and radiation therapy [J]. Mathematical Programming, 1988, 42: 307.
 [11] HERMAN G T. Fundamentals of computerized tomography: image reconstruction from projection [M]. 2nd ed. London: Springer, 2009.
 [12] KAK A C, SLANEY M. Principles of computerized tomographic imaging [M]. Philadelphia: SIAM Publisher, 2003.
 [13] NATTERER F. The mathematics of computerized tomography [M]. Philadelphia: SIAM Publisher, 2001.
 [14] GORDON R, BENDER R, HERMAN G T. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography [J]. Journal of Theoretical Biology, 1970, 29(3): 471.

- [15] HERMAN G T, DAVIDI R. Image reconstruction from a small number of projections [J]. *Inverse Problems*, 2008, 24(4): 45011.
- [16] HERMEN G T, MEYER L B. Algebraic reconstruction techniques can be made computationally efficient [J]. *IEEE Trans Medical Imaging*, 1993, 12(3): 600.
- [17] ELBLE J M, SAHINIDIS N V, VOUZIS P. GPU computing with Kaczmarz's and other iterative algorithms for linear systems [J]. *Parallel Computing*, 2010, 36(5): 215.
- [18] PASQUALETTI F, CARLI R, BULLO F. Distributed estimation via iterative projections with application to power network monitoring [J]. *Automatica*, 2012, 48(5): 747.
- [19] BYRNE C. A unified treatment of some iterative algorithms in signal processing and image reconstruction [J]. *Inverse Problems*, 2004, 20(1): 103.
- [20] LORENZ D, MAGNOR M, WENGER S, *et al.* A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing [C]// *IEEE International Conference on Image Processing (ICIP)*. [S.l.]:ICIP, 2014: 1347-1351.
- [21] CENKER C, FEICHTINGER H G, MAYER M, *et al.* New variants of the POCS method using affine subspaces of finite codimension, with applications to irregular sampling [J]. *Proceedings of SPIE - The International Society for Optical Engineering*, 1992, 1818: 299.
- [22] STROHMER T, VERSHYNIN R. A randomized Kaczmarz algorithm with exponential convergence [J]. *Journal of Fourier Analysis and Applications*, 2009, 15(2): 262.
- [23] DAI L, SOLTANALIAN M, PELCKMANS K. On the randomized Kaczmarz algorithm [J]. *IEEE Signal Processing Letters*, 2014, 21(3): 330.
- [24] BAI Z Z, WU W T. On greedy randomized Kaczmarz method for solving large sparse linear systems [J]. *SIAM Journal on Scientific Computing*, 2018, 40(1): 592.
- [25] BAI Z Z, WU W T. On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems [J]. *Applied Mathematics Letters*, 2018, 83: 21.
- [26] 杜亦疏, 殷俊锋, 张科. 求解大型稀疏线性方程组的贪婪距离随机Kaczmarz方法[J]. *同济大学学报(自然科学版)*, 2020, 48(8): 1224.
DU Yishu, YIN Junfeng, ZHANG Ke. Greedy randomized-distance Kaczmarz method for solving large sparse linear systems [J]. *Journal of Tongji University (Natural Science)*, 2020, 48(8): 1224.
- [27] NEEDELL D, WARD R. Two-subspace projection method for coherent over-determined systems [J]. *Journal of Fourier Analysis and Applications*, 2013, 19(2): 256.
- [28] ZOUZIAS A, FRERIS N. M. Randomized extended Kaczmarz for solving least squares [J]. *SIAM Journal on Matrix Analysis and Applications*, 2013, 34(2): 773.
- [29] BAI Z Z, WU W T. On partially randomized extended Kaczmarz method for solving large sparse overdetermined inconsistent linear systems [J]. *Linear Algebra and Its Applications*, 2019, 578: 225.
- [30] DAI L, SCHON T. On the exponential convergence of the kaczmarz algorithm [J]. *IEEE Signal Processing Letters*, 2015, 22(10): 1571.
- [31] OSWALD P, ZHOU W Q. Convergence analysis for Kaczmarz-type methods in a Hilbert space framework [J]. *Linear Algebra and Its Applications*, 2015, 478: 131.
- [32] BAI Z Z, WU W T. On convergence rate of the randomized Kaczmarz method [J]. *Linear Algebra and Its Applications*, 2018, 553: 252.
- [33] POPA C. Convergence rates for Kaczmarz-type algorithms [J]. *Numerical Algorithms*, 2018, 79(1): 1.
- [34] ELDAR Y C, NEEDELL D. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma [J]. *Numerical Algorithms*, 2011, 58(2): 163.
- [35] XIANG X, CHENG L Z. An accelerated randomized Kaczmarz method via low-rank approximation [J]. *International Journal of Computer Mathematics*, 2015, 92(7): 1413.
- [36] LEVENTHAL D, LEWIS A S. Randomized methods for linear constraints: convergence rates and conditioning [J]. *Mathematics of Operations Research*, 2010, 35(3): 641.
- [37] MA A, NEEDELL D, RAMDAS A. Convergence properties of the randomized extended Gauss - Seidel and Kaczmarz methods [J]. *SIAM Journal on Matrix Analysis and Applications*, 2015, 36(4): 1590.
- [38] DU K. Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss-Seidel algorithms [J]. *Numerical Linear Algebra with Applications*, 2019, 26(3): e2233.
- [39] NEEDELL D, TROPP J A. Paved with good intentions: analysis of a randomized block Kaczmarz method [J]. *Linear Algebra and Its Applications*, 2014, 441: 199.
- [40] NEEDELL D, ZHAO R, ZOUZIAS A. Randomized block Kaczmarz method with projection for solving least squares [J]. *Linear Algebra and Its Applications*, 2015, 484: 322.
- [41] NUTINI J, SEPEHRY B, LARADJI I, *et al.* Convergence rates for greedy Kaczmarz algorithms, and faster randomized Kaczmarz rules using the orthogonality graph [J]. *arXiv*, 2016, 1612.07838.
- [42] HOLODNAK J T, IPSEN I C F. Randomized approximation of the Gram matrix: Exact computation and probabilistic bounds [J]. *SIAM Journal on Matrix Analysis and Applications*, 2015, 36(1): 110.
- [43] DU Y S, HAYAMI K, ZHENG N, *et al.* Kaczmarz-type inner-iteration preconditioned flexible GMRES methods for consistent linear systems [J]. *arXiv*, 2020, 2006.10818.