

求解线性方程组稀疏解的稀疏贪婪随机 Kaczmarz 算法

王 泽, 殷俊锋

(同济大学 数学科学学院, 上海 200092)

摘要: 求解线性方程组的稀疏解在图像重构、信号处理和机器学习等领域中具有广泛的应用, 通过引入 l_1 -范数正则化, 可以转化为求解一个约束优化问题。基于一种选择系数矩阵工作行的概率准则, 提出了稀疏贪婪随机 Kaczmarz 算法, 并给出了有噪声干扰和无噪声干扰情况下该算法的收敛性分析。理论表明本文算法的收缩因子小于随机稀疏 Kaczmarz 算法的收缩因子。数值实验验证了本文算法的有效性。

关键词: Kaczmarz 算法; 稀疏解; 收敛性; 随机迭代

中图分类号: O241.6

文献标志码: A

Sparse Greedy Randomized Kaczmarz Method for Sparse Solutions to Linear Equations

WANG Ze, YIN Junfeng

(School of Mathematical Sciences, Tongji University, Shanghai 200092, China)

Abstract: The sparse solution to linear equations has been widely used in image reconstruction, signal processing, and machine learning. By introducing l_1 norm regularization, it can be transformed into solving a constrained optimization problem. Based on a novel probability criterion for selecting the working rows from the coefficient matrix, a sparse greedy randomized Kaczmarz method was proposed, and the convergence analysis of the novel method with and without noise interference were conducted, which showed that the convergence factor of the novel method was smaller than that of the randomized sparse Kaczmarz method. The numerical experiments verified the effectiveness of the proposed method.

Key words: Kaczmarz method; sparse solution; convergence; randomized iteration

求解线性方程组 $Ax = b$ 的稀疏解在图像重构、信号处理和机器学习中具有广泛应用^[1-4]。通过引入 l_1 -范数, 求解线性方程组的稀疏解可以转化为求解式(1)正则化最小二乘问题:

$$\min_{x \in \mathbb{R}^n} f(x) = \lambda \|x\|_1 + \frac{1}{2} \|x\|_2^2 \quad \text{s.t. } Ax = b \quad (1)$$

求解该问题常见的算法包括基追踪方法、Bregman 方法、共轭梯度方法等^[5-7]。

自 Strohmer 和 Vershynin^[8] 证明了随机 Kaczmarz 算法具有线性收敛率以来, 许多专家学者对随机 Kaczmarz 算法进行了深入的研究^[9-11]。Kaczmarz 算法求解的是最小二乘解, 通常是稠密的, 为了求解线性方程组的稀疏解, 文献[12]提出稀疏 Kaczmarz 算法, 该算法在原有的 Kaczmarz 算法基础之上引入了软阈值函数, 解决了最小二乘解稠密性的问题。

在线性方程组相容的情况下, 文献[13]证明了稀疏 Kaczmarz 算法迭代收敛到问题(1)的唯一解。令右端噪声项 b^δ 满足 $\|b - b^\delta\|_2 \leq \delta$ 时, 线性方程组可能不相容, 文献[14-15]证明了随机 Kaczmarz 算法的迭代值依期望收敛, 并且与无噪声情况下有相同的收缩因子。文献[16]在稀疏 Kaczmarz 算法基础之上, 提出了随机稀疏 Kaczmarz 算法, 并在有噪声干扰和无噪声干扰的情况下给出了随机稀疏 Kaczmarz 算法的收敛性分析。当系数矩阵的行范数相差较小时, 随机 Kaczmarz 算法将等概率选取 A 的行, 此时随机 Kaczmarz 算法的收敛速率较慢, 因此

收稿日期: 2021-05-12

基金项目: 国家自然科学基金(11971354)

第一作者: 王 泽(1998—), 男, 博士生, 主要研究方向为数值代数与科学计算。

E-mail: math_wangze@tongji.edu.cn

通信作者: 殷俊锋(1979—), 男, 教授, 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。

E-mail: yinjf@tongji.edu.cn



论文
拓展
介绍

Bai和Wu^[17]结合贪婪和随机思想,提出一种新的概率准则来选取工作行,形成贪婪随机Kaczmarz算法。更多贪婪随机Kaczmarz算法的研究参见文献[18-21]。

在稀疏随机Kaczmarz算法的基础之上^[21],受贪婪算法的启发,本文提出稀疏贪婪随机Kaczmarz算法,给出稀疏贪婪随机Kaczmarz算法在有噪声干扰和无噪声干扰情况下的收敛性分析,并以大量的数值实验验证本文算法的计算效率。

符号标记如下,定义 \hat{x} 为线性方程组 $Ax=b$ 的解,其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, a_i 表示系数矩阵 A 的第 i 行。支持集 $S = \text{supp}(\hat{x}) = \{j \in \{1, \dots, n\} | \hat{x}_j \neq 0\}$ 是列向量 \hat{x} 中非零元的下标所构成的集合, A_S 是 A 的子矩阵,由支持集 S 中的指标选取系数矩阵 A 的列构成, a_{i_S} 表示的是 A_S 的第 i 行, x_S 是在支持集 S 限制下的列向量;定义 $\hat{\sigma}_{\min}(A) = \min\{\sigma_{\min}(A_J) | J \subset \{1, \dots, n\}, A_J \neq 0\}$ 为子矩阵 A_J 的最小奇异值,其中 A_J 为指标集 J 中 A 的列构成的子矩阵,令 $\bar{\kappa} = \frac{\|A\|_F}{\hat{\sigma}_{\min}(A)}$; \odot 为哈达玛积(Hadamard product); \hat{k} 为对支持集大小的估计值; w_j 为稀疏随机Kaczmarz算法第 j 次迭代的权重值, $w_j \in \mathbb{R}^{m \times 1}$ 。

1 稀疏贪婪随机Kaczmarz算法

当 $Ax=b$ 的解 x 稀疏时,文献[21]提出了稀疏随机Kaczmarz算法,加快了算法的收敛速率(算法1)。

算法1 稀疏随机Kaczmarz算法^[21]。①输入 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$,最大迭代数 M 和估计的支持集的大小 \hat{k} 。②输出 x_j 。③初始化 $S = \{1, \dots, n\}$, $x_0 = 0$, $j = 0$ 。④当 $j \leq M$ 时,置 $j = j + 1$ 。⑤选择行向量 $a_i, i \in \{1, \dots, n\}$,每一行对应的概率为 $\frac{\|a_i\|_2^2}{\|A\|_F^2}$ 。⑥确定估计的支持集 $S, S = \text{supp}(x_{j-1}|_{\max\{\hat{k}, n-j+1\}})$ 。

⑦生成权重值 $w_j, w_j(l) = \begin{cases} 1 & l \in S \\ \frac{1}{\sqrt{j}} & l \in S^c \end{cases}$ ⑧ $x_j = x_{j-1} + \frac{b_i - \langle w_j \odot a_i, x_{j-1} \rangle}{(w_j \odot a_i)_i} (w_j \odot a_i)^T$ 。⑨转步骤④。

稀疏随机Kaczmarz算法可以用比随机Kaczmarz算法更少的迭代步数找到最小二乘解。由于支持集和稀疏度都是未知的,因此稀疏随机

Kaczmarz算法从支持集中所有元素的稀疏度的初始估计开始,然后在每一次迭代中,稀疏随机Kaczmarz算法通过去掉向量 x 中数量级最小的元素下标来更新估计的支持集。该算法第 j 次迭代的加权准则为

$$w_j(l) = \begin{cases} 1 & l \in S \\ \frac{1}{\sqrt{j}} & l \in S^c \end{cases}$$

其中, j 为迭代步数。当 $j \rightarrow \infty$ 时, $w_j \odot a_i \rightarrow a_{i_S}$,因此原线性方程组退化成

$$b = Ax = A_S x_S$$

由于 $\kappa(A_S) \leq \kappa(A)$,因此稀疏随机Kaczmarz算法的收敛因子小于随机Kaczmarz算法。但是由于最小二乘解总是稠密的,该算法在求解超定线性方程组时,虽然在稀疏解零元对应位置上的元素很小,但仍然不等于零,因此在文献中提出稀疏Kaczmarz算法,文献[16]在稀疏Kaczmarz算法的基础上,提出了随机稀疏Kaczmarz算法(算法2)。

算法2 随机稀疏Kaczmarz算法^[16]。①输入 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$,最大迭代数 M 。②输出 x_k 。③初始化: $x_0 = x_0^* = 0$ 。④置 $k = 0$,当 $k \leq M - 1$ 时。⑤选择行向量 $a_{i_k}, i_k \in \{1, \dots, n\}$,每一行对应的概率为 $\frac{\|a_{i_k}\|_2^2}{\|A\|_F^2}$ 。⑥ $x_{k+1}^* = x_k^* - \frac{\langle a_{i_k}, x_k \rangle - b_{i_k}}{\|a_{i_k}\|_2^2} \cdot a_{i_k}$ 。⑦ $x_{k+1} = S_\lambda(x_{k+1}^*)$ 。⑧转步骤④。

算法2中 $\lambda > 0$, $S_\lambda(x) = \max\{|x| - \lambda, 0\} \cdot \text{sign}(x)$ 。结合稀疏随机Kaczmarz算法和随机稀疏Kaczmarz算法的思想,受贪婪算法的启发,本文提出稀疏贪婪随机Kaczmarz算法,在保证算法能够计算出稀疏解的同时,通过贪婪的概率准则来选择工作行从而达到加快算法收敛速度的目的。算法3给出稀疏贪婪随机Kaczmarz算法。

算法3 稀疏贪婪随机Kaczmarz算法。①输入 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$,最大迭代数 M 和估计的支持集的大小 \hat{k} 。②输出 x_k 。③初始化 $S = \{1, \dots, n\}$, $x_0 = x_0^* = 0$ 。④置 $k = 0$ 时,当 $k \leq M - 1$ 时。⑤计算

$$\epsilon_k = \frac{1}{2} \left(\frac{1}{\|b - Ax_k\|_2^2} \max_{1 \leq i_k \leq m} \left\{ \frac{|b_{i_k} - a_{i_k} x_k|^2}{\|a_{i_k}\|_2^2} \right\} + \frac{1}{\|A\|_F^2} \right) \quad (2)$$

⑥决定正整数指标集

$$\mathcal{U}_k = \left\{ i_k \mid |b_{i_k} - a_{i_k} x_k|^2 \geq \epsilon_k \|b - Ax_k\|_2^2 \|a_{i_k}\|_2^2 \right\} \quad (3)$$

⑦计算向量 \tilde{r}_k 的第 i 行 $\tilde{r}_k^{(i)}$

$$\tilde{r}_k^{(i)} = \begin{cases} b^{(i)} - A^{(i)}x_k & \text{若 } i \in U_k \\ 0 & \text{其他} \end{cases}$$

⑧以概率准则 $p(r_{\text{row}} = i_k) = \frac{|r_k^{(i)}|^2}{\|\tilde{r}_k\|_2^2}$ 选择 $i_k \in U_k$ 。⑨确

定估计的支持集 S , 如下: $S = \text{supp}\left(x_k \Big|_{\max\{\tilde{k}, n-k\}}\right)$ 。⑩生成权重值 w_{k+1}

$$w_{k+1}(l) = \begin{cases} 1 & l \in S \\ \frac{1}{\sqrt{k+1}} & l \in S^c \end{cases}$$

⑪计算

$$x_{k+1}^* = x_k^* + \frac{b_{i_k} - \langle w_{k+1} \odot a_{i_k}, x_k \rangle}{\|w_{k+1} \odot a_{i_k}\|_2} (w_{k+1} \odot a_{i_k})^T$$

⑫计算 $x_{k+1} = S_\lambda(x_{k+1}^*)$ 。⑬转步骤④。

2 稀疏贪婪随机 Kaczmarz 算法的收敛性分析

为了证明求解正则基追踪问题(1)稀疏贪婪随机 Kaczmarz 算法依期望线性收敛, 先回顾一些基本的概念和性质^[22]。

令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数, 用 $\partial f(x)$ 表示 f 在 $x \in \mathbb{R}^n$ 的次微分

$$\partial f(x) = \{x^* \in \mathbb{R}^n | f(y) \geq f(x) + \langle x^*, y - x \rangle \quad \forall y \in \mathbb{R}^n\}$$

其中, $\partial f(x)$ 是一个非空紧凸集。

定义 1 凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果存在 $\alpha > 0$, 使得对任意的 $x, y \in \mathbb{R}^n$ 且 $x^* \in \partial f(x)$, 有

$$f(y) \geq f(x) + \langle x^*, y - x \rangle + \frac{\alpha}{2} \cdot \|y - x\|_2^2$$

那么 f 是强凸的, 当和 α 的具体值相关时, 则称 f 是 α -强凸的。

定义 2^[22] 如果 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是 α -强凸, 那么它的共轭函数 $f^*(x^*) := \sup_{x \in \mathbb{R}^n} \langle x^*, x \rangle - f(x)$ 可微且有 $1/\alpha$ -Lipschitz 连续梯度, 例如

$$\langle \nabla f^*(x^*) - \nabla f^*(y^*) \rangle_2 \leq \frac{1}{\alpha} \cdot (x^* - y^*)_2, \quad \forall x^*, y^* \in \mathbb{R}^n$$

有式(4)不等式成立:

$$f^*(y^*) \leq f^*(x^*) + \langle \nabla f^*(x^*), y^* - x^* \rangle + \frac{1}{2\alpha} \|y^* - x^*\|_2^2 \quad \forall x^*, y^* \in \mathbb{R}^n \quad (4)$$

定义 3 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是强凸函数, 那么 $x, y \in \mathbb{R}^n$

之间的 Bregman 距离 $D_f^*(x, y)$ 由 f 和一个次梯度 $x^* \in \partial f(x)$ 定义

$$D_f^*(x, y) := f(y) - f(x) - \langle x^*, y - x \rangle = f^*(x^*) - \langle x^*, y \rangle + f(y)$$

如果 f 是可微的, 那么有 $\partial f(x) = \{\nabla f(x)\}$, 因为 $D_f(x, y) = f(y) - f(x) - \langle \nabla f(x), y - x \rangle$, 所以可以简化来写 $D_f(x, y) = D_f^*(x, y)$ 。

下面的引理都遵循强凸性的假设^[13], 给出了随机方法收敛性分析所需的 Bregman 距离的关键性质。

引理 1^[13] 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是 α 强凸函数, 对任意的 $x, y \in \mathbb{R}^n$ 和 $x^* \in \partial f(x), y^* \in \partial f(y)$ 有

$$\frac{\alpha}{2} \|x - y\|_2^2 \leq D_f^*(x, y) \leq \langle x^* - y^*, x - y \rangle \leq (x^* - y^*)_2 \cdot \|x - y\|_2$$

因此

$$D_f^*(x, y) = 0 \Leftrightarrow x = y$$

对于序列 x_k 和 $x_k^* \in \partial f(x_k), D_f^*(x_k, y)$ 的有界性意味着 x_k 和 x_k^* 有界。如果 f 有一个 L -Lipschitz 连续梯度, 那么也有 $D_f(x, y) \leq \frac{L}{2} \cdot \|x - y\|_2^2$ 。

定义 4 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是强凸函数, 且 $C \subset \mathbb{R}^n$ 是一个非空闭凸集, 则 x 到 C 上的 Bregman 投影是关于 f 和 $x^* \in \partial f(x)$ 满足下式的唯一点 $\Pi_C^*(x) \in C$:

$$D_f^*(x, \Pi_C^*(x)) = \min_{y \in C} D_f^*(x, y) =: \text{dist}_f^*(x, C)^2$$

对于可微的 f , 可以简化为 $\Pi_C(x)$ 和 $\text{dist}_f(x, C)$ 。

引理 2^[13] 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是强凸函数, x 到 C 上的 Bregman 投影点 $\hat{x} \in C$ 是与 f 和 $x^* \in \partial f(x)$ 相关, 当且仅当存在一个 $\hat{x}^* \in \partial f(\hat{x})$ 满足式(5)或式(6):

$$\langle \hat{x}^* - x^*, y - \hat{x} \rangle \geq 0, \quad \forall y \in C \quad (5)$$

$$D_f^*(\hat{x}, y) \leq D_f^*(x, y) - D_f^*(x, \hat{x}) \quad \forall y \in C \quad (6)$$

可以叫这样的点 \hat{x}^* 是 $\hat{x} = \Pi_C^*(x)$ 的可允次梯度。

引理 3 表明, Bregman 投影到仿射子空间和半空间可以通过求解可微共轭函数 f^* 的无约束优化问题来计算。

引理 3^[13] 令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是 α -强凸函数, $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, u \in \mathbb{R}^n$ 和 $\beta \in \mathbb{R}$ 。

(1) $x \in \mathbb{R}^n$ 到仿射空间 $L(A, b) := \{x \in \mathbb{R}^n | Ax = b\} \neq \emptyset$ 的 Bregman 投影是

$$\hat{x} := \Pi_{L(A, b)}^*(x) = \nabla f^*(x^* - A^T \hat{w})$$

其中 $\hat{w} \in \mathbb{R}^m$ 是下式的解:

$$\min_{w \in \mathbb{R}^m} f^*(x^* - A^T w) + \langle w, b \rangle$$

更重要的是,根据引理2, $\hat{x}^* := x^* - A^T \hat{w}$ 是 \hat{x} 的可允次梯度。如果 A 行满秩,那么对任意的 $y \in L(A, b)$ 有

$$D_f^{x^*}(z, y) \leq D_f^{x^*}(x, y) - \frac{\alpha}{2} \cdot \left((AA^T)^{-\frac{1}{2}} (Ax - b) \right)_2^2$$

(2) $x \in \mathbb{R}^n$ 到超平面 $H(u, \beta) := \{x \in \mathbb{R}^n | \langle u, x \rangle = \beta\}$ 满足 $u \neq 0$ 的Bregman投影是

$$\hat{x} := \Pi_{H(u, \beta)}^{x^*}(x) = \nabla f^*(x^* - \hat{t} \cdot u)$$

其中 $\hat{t} \in \mathbb{R}$ 是下式的解:

$$\min_{t \in \mathbb{R}} f^*(x^* - t \cdot u) + t \cdot \beta$$

更重要的是 $\hat{x}^* := x^* - \hat{t} \cdot u$ 是 \hat{x} 的可允次梯度,对任意的 $y \in H(u, \beta)$ 有

$$D_f^{\hat{x}^*}(\hat{x}, y) \leq D_f^{\hat{x}^*}(x, y) - \frac{\alpha}{2} \cdot \frac{(\langle u, x \rangle - \beta)^2}{\|u\|_2^2}$$

如果 x 不是在半空间 $H_{\leq}(u, \beta) := \{x \in \mathbb{R}^n | \langle u, x \rangle \leq \beta\}$, 那么有必要满足 $\hat{t} > 0$, $\Pi_{H_{\leq}(u, \beta)}^{x^*}(x) = \hat{x}$ 且上面的不等式对任意的 $y \in H_{\leq}(u, \beta)$ 都成立。

引理 4^[16] 对于任意的 $x \in \mathbb{R}^n$ 满足 $\partial f(x) \cap \mathcal{R}(A^T) \neq \emptyset$ 和任意的 $x^* = A^T y \in \partial f(x) \cap \mathcal{R}(A^T)$, 有

$$D_f^{x^*}(x, \hat{x}) \leq \frac{1}{\hat{\sigma}_{\min}^2(A)} \cdot \frac{|\hat{x}|_{\min} + 2\lambda}{|\hat{x}|_{\min}} \cdot \|Ax - b\|_2^2$$

定理 1 (无噪声情况) 令 $\bar{\kappa} = \frac{\|A\|_F}{\hat{\sigma}_{\min}(A)}$, $\epsilon_k =$

$$\frac{1}{2} \left(\frac{1}{(b - Ax_k)_2} \max_{1 \leq i \leq m} \left\{ \frac{|b_i - a_i x_k|^2}{\|a_i\|_2^2} \right\} + \frac{1}{\|A\|_F^2} \right)$$

稀疏贪婪随机Kaczmarz算法所产生的迭代序列 x_k 依期望线性收敛到唯一解 \hat{x} , 其收缩因子为

$$q = 1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda} \quad (7)$$

从而

$$E[D_f^{x_{k+1}^*}(x_{k+1}, \hat{x})] \leq q \cdot E[D_f^{x_k^*}(x_k, \hat{x})]$$

且

$$E[(x_k - \hat{x})_2] \leq q^{\frac{k}{2}} \cdot \sqrt{2\lambda} \|\hat{x}\|_1 + \|\hat{x}\|_2^2$$

证明 根据引理3的(2)中的估计,有式(8)成立:

$$D_f^{x_{k+1}^*}(x_{k+1}, \hat{x}) \leq D_f^{x_k^*}(x_k, \hat{x}) - \frac{1}{2} \cdot \frac{(\langle a_i, x_k \rangle - b_i)^2}{\|(a_i)_2\|} \quad (8)$$

对式(8)两边同时取条件期望有

$$E[D_f^{x_{k+1}^*}(x_{k+1}, \hat{x}) | i_0, \dots, i_{k-1}] \leq D_f^{x_k^*}(x_k, \hat{x}) - \sum_{i_k \in \mathcal{U}_k} \frac{|b_{i_k} - a_{i_k} x_k|^2}{\sum_{i \in \mathcal{U}_k} |b_i - a_i x_k|^2} \cdot \frac{1}{2} \cdot \frac{(\langle a_{i_k}, x_k \rangle - b_{i_k})^2}{(a_{i_k})_2^2} = D_f^{x_k^*}(x_k, \hat{x}) - \sum_{i_k \in \mathcal{U}_k} \frac{|b_{i_k} - a_{i_k} x_k|^2}{\sum_{i \in \mathcal{U}_k} |b_i - a_i x_k|^2} \cdot \frac{1}{2} \cdot \frac{|a_{i_k}(x_k - \hat{x})|^2}{\|a_{i_k}\|_2^2} \quad (9)$$

根据式(2)和式(3)的定义,有不等式(10)成立:

$$|b_{i_k} - a_{i_k} x_k|^2 \geq \epsilon_k \|b - Ax_k\|_2^2 \|a_{i_k}\|_2^2, \forall i_k \in \mathcal{U}_k \quad (10)$$

将式(10)代入到式(9)中,从而有

$$E[D_f^{x_{k+1}^*}(x_{k+1}, \hat{x}) | i_0, \dots, i_{k-1}] \leq D_f^{x_k^*}(x_k, \hat{x}) - \sum_{i_k \in \mathcal{U}_k} \frac{|b_{i_k} - a_{i_k} x_k|^2}{\sum_{i \in \mathcal{U}_k} |b_i - a_i x_k|^2} \cdot \frac{1}{2} \cdot \frac{|a_{i_k}(x_k - \hat{x})|^2}{\|a_{i_k}\|_2^2} \leq D_f^{x_k^*}(x_k, \hat{x}) - \frac{1}{2} \cdot \frac{\sum_{i_k \in \mathcal{U}_k} |a_{i_k}(x_k - \hat{x})|^2}{\sum_{i \in \mathcal{U}_k} |b_i - a_i x_k|^2} \cdot \epsilon_k \cdot \|b - Ax_k\|_2^2 = D_f^{x_k^*}(x_k, \hat{x}) - \frac{1}{2} \cdot \epsilon_k \cdot \|b - Ax_k\|_2^2 = D_f^{x_k^*}(x_k, \hat{x}) - \frac{1}{2} \cdot \epsilon_k \cdot \|b - Ax_k\|_2^2 \quad (11)$$

由式(11)的推导,可以得出

$$E[D_f^{x_{k+1}^*}(x_{k+1}, \hat{x}) | i_0, \dots, i_{k-1}] \leq D_f^{x_k^*}(x_k, \hat{x}) -$$

$$\frac{1}{2} \cdot \epsilon_k \cdot \|b - Ax_k\|_2^2$$

根据引理4有

$$E[D_f^{x_{k+1}^*}(x_{k+1}, \hat{x}) | i_0, \dots, i_{k-1}] \leq D_f^{x_k^*}(x_k, \hat{x}) - \frac{1}{2} \cdot \epsilon_k \cdot \frac{\hat{\sigma}_{\min}^2 |\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda} \cdot D_f^{x_k^*}(x_k, \hat{x}) \leq \left(1 - \frac{1}{2} \cdot \epsilon_k \cdot \frac{\hat{\sigma}_{\min}^2 |\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda} \right) \cdot D_f^{x_k^*}(x_k, \hat{x}) \quad (12)$$

由于 $\bar{\kappa} = \frac{\|A\|_F}{\bar{\sigma}_{\min}(A)}$, 从而

$$E[D_f^{x_k^*}(x_{k+1}, \hat{x}) | i_0, \dots, i_{k-1}] \leq (1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \cdot \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda}) \cdot D_f^{x_k^*}(x_k, \hat{x}) \quad (13)$$

令 $q = 1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \cdot \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda}$, 则

$$E[D_f^{x_k^*}(x_{k+1}, \hat{x})] \leq q \cdot E[D_f^{x_k^*}(x_k, \hat{x})]$$

根据引理 1, 令 $\alpha=1$ 可得

$$E[\|x_k - \hat{x}\|_2] \leq q^{\frac{k}{2}} \cdot \sqrt{2\lambda \| \hat{x} \|_1 + \| \hat{x} \|_2^2}$$

证毕。

定理 2 (有噪声情况) 已知右段噪声项 $b^\delta \in \mathbb{R}^m$ 满足 $\|b - b^\delta\|_2 \leq \delta$ 。如果稀疏贪婪随机 Kaczmarz 算法的迭代值 x_k 由 b^δ 替代 b 来计算, 那么有噪声干扰和无噪声干扰有相同的收缩因子

$$q = 1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda} \text{ 且}$$

$$E[(x_k - \hat{x})_2] \leq q^{\frac{k}{2}} \cdot \sqrt{2\lambda \| \hat{x} \|_1 + \| \hat{x} \|_2^2} +$$

$$\sqrt{\frac{2|\hat{x}|_{\min} + 4\lambda}{\epsilon_k \cdot |\hat{x}|_{\min}}} \cdot \frac{\delta}{\|A\|_F \cdot \bar{\sigma}_{\min}(A)}$$

证明 令

$$x_k^\delta := \hat{x} + \frac{b_{i_k}^\delta - b_{i_k}}{\|a_{i_k}\|_2} \cdot a_{i_k} \in H(a_{i_k}, b_{i_k}^\delta)$$

和式(8)相类似有

$$D_f^{x_k^\delta}(x_{k+1}, x_k^\delta) \leq D_f^{x_k^\delta}(x_k, x_k^\delta) - \frac{1}{2} \cdot \frac{(\langle a_{i_k}, x_k \rangle - b_{i_k}^\delta)^2}{(a_{i_k})_2^2} \quad (14)$$

将式(14)进行简单的变换等价于式(15):

$$D_f^{x_k^\delta}(x_{k+1}, \hat{x}) \leq D_f^{x_k^\delta}(x_k, \hat{x}) - \frac{1}{2} \cdot \frac{(\langle a_{i_k}, x_k \rangle - b_{i_k}^\delta)^2}{(a_{i_k})_2^2} + \langle x_{k+1}^* - x_k^*, x_k^\delta - \hat{x} \rangle \quad (15)$$

在稀疏贪婪随机 Kaczmarz 算法中有

$$x_{k+1}^* - x_k^* = -\frac{\langle a_{i_k}, x_k \rangle - b_{i_k}^\delta}{\|a_{i_k}\|_2} \cdot a_{i_k} \text{ 因此}$$

$$\langle x_{k+1}^* - x_k^*, x_k^\delta - \hat{x} \rangle = \frac{b_{i_k}^\delta - b_{i_k}}{\|a_{i_k}\|_2} \cdot \langle x_{k+1}^* - x_k^*, a_{i_k} \rangle =$$

$$\frac{(b_{i_k}^\delta - b_{i_k})^2}{\|a_{i_k}\|_2^2} - \frac{(b_{i_k}^\delta - b_{i_k}) \cdot (\langle a_{i_k}, x_k \rangle - b_{i_k})}{\|a_{i_k}\|_2^2}$$

而

$$-\frac{1}{2} \cdot \frac{(\langle a_{i_k}, x_k \rangle - b_{i_k}^\delta)^2}{\|a_{i_k}\|_2^2} = -\frac{1}{2} \cdot$$

$$\frac{(\langle a_{i_k}, x_k \rangle - b_{i_k})^2}{\|a_{i_k}\|_2^2} +$$

$$\frac{(b_{i_k}^\delta - b_{i_k}) \cdot (\langle a_{i_k}, x_k \rangle - b_{i_k})}{\|a_{i_k}\|_2^2} - \frac{1}{2} \cdot \frac{(b_{i_k}^\delta - b_{i_k})^2}{\|a_{i_k}\|_2^2}$$

所以

$$D_f^{x_k^\delta}(x_{k+1}, \hat{x}) \leq D_f^{x_k^\delta}(x_k, \hat{x}) - \frac{1}{2} \cdot$$

$$\frac{(\langle a_{i_k}, x_k \rangle - b_{i_k})^2}{\|a_{i_k}\|_2^2} + \frac{1}{2} \cdot \frac{(b_{i_k}^\delta - b_{i_k})^2}{\|a_{i_k}\|_2^2}$$

类似于无噪声情况的证明方法可以得到

$$E[D_f^{x_k^\delta}(x_{k+1}, \hat{x})] \leq q \cdot E[D_f^{x_k^\delta}(x_k, \hat{x})] + \frac{1}{2} \cdot \frac{(b^\delta - b)_2^2}{\|A\|_F^2}$$

通过归纳得出

$$E[D_f^{x_k^\delta}(x_k, \hat{x})] \leq q^k \cdot \left(\lambda \| \hat{x} \|_1 + \frac{1}{2} \| \hat{x} \|_2^2 \right) + \frac{1}{1-q} \cdot \frac{1}{2} \cdot \frac{(b^\delta - b)_2^2}{\|A\|_F^2}$$

由引理 1, 令 $\alpha=1$, 且 $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$, 有

$$E[(x_k - \hat{x})_2] \leq q^{\frac{k}{2}} \cdot \sqrt{2\lambda \| \hat{x} \|_1 + \| \hat{x} \|_2^2} + \sqrt{\frac{1}{1-q} \cdot \frac{\|b^\delta - b\|_2^2}{\|A\|_F^2}} \quad (16)$$

将 $q = 1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda}$ 代入式(16)

中就可以得到

$$E[(x_k - \hat{x})_2] \leq q^{\frac{k}{2}} \cdot \sqrt{2\lambda \| \hat{x} \|_1 + \| \hat{x} \|_2^2} + \sqrt{\frac{2|\hat{x}|_{\min} + 4\lambda}{\epsilon_k \cdot |\hat{x}|_{\min}}} \cdot \frac{\delta}{\|A\|_F \cdot \bar{\sigma}_{\min}(A)}$$

证毕。

注意到稀疏贪婪随机 Kaczmarz 算法的收缩因子为

$$q = 1 - \frac{1}{\bar{\kappa}^2} \cdot \frac{\epsilon_k \|A\|_F^2}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda}$$

文献[16]定理 3.2 中随机稀疏 Kaczmarz 算法的收缩因子为

$$\tilde{q} = 1 - \frac{1}{\tilde{\kappa}^2} \cdot \frac{1}{2} \cdot \frac{|\hat{x}|_{\min}}{|\hat{x}|_{\min} + 2\lambda}$$

因

$$\max_{1 \leq i_k \leq m} \left(\frac{|b_{i_k} - a_{i_k} x_k|^2}{\|a_{i_k}\|_2^2} \right) \geq \sum_{i_k=1}^m \frac{\|a_{i_k}\|_2^2}{\|A\|_F^2} \cdot \frac{|b_{i_k} - a_{i_k} x_k|^2}{\|a_{i_k}\|_2^2} = \frac{\|b - Ax_k\|_2^2}{\|A\|_F^2}$$

从而 $\epsilon_k \|A\|_F^2 \geq 1$, 因此

$$q \leq \tilde{q}$$

由此可以得出稀疏贪婪随机Kaczmarz算法的收缩因子小于随机稀疏Kaczmarz算法的收缩因子。一般而言,收缩因子越小,收敛速度越快,因此理论分析表明稀疏贪婪随机Kaczmarz算法收敛速度会比原有的随机稀疏Kaczmarz算法的收敛速度更快。

3 数值实验

通过随机生成的高斯矩阵和矩阵市场中的矩阵2个数值实验算例来比较稀疏贪婪随机Kaczmarz(SGRK)算法、随机稀疏Kaczmarz(RaSK)算法和稀疏随机Kaczmarz(SRK)算法的收敛速度。为了能够使三者收敛到相同解上进行比较,在本次数值实验中,在稀疏随机Kaczmarz算法最后一步加上了软阈值函数 $S_\lambda(x)$ 参与迭代以保证求得稀疏解。对相同参数的同一个线性方程组 $Ax = b$ 做了20次试验,得到迭代步数(IT)和CPU计算时间的平均值。解的相对误差(RSE)定义为

$$r_{\text{RSE}} = \frac{\|x_k - \hat{x}\|_2^2}{\|\hat{x}\|_2^2}$$

例1 利用Matlab软件中的randn函数随机生成系数矩阵 A 和 $x_{n \times 1}$, 其中有 $k \times n$ 个非零元,之后利用 $b = A \times x$ 得到相容线性方程组,初始向量 $x_0 = 0$ 。取 $\tilde{k} = 2 \times k$, 通过 \tilde{k} 来设置解的稀疏度,其中 \tilde{k} 是对稀疏度的初始估计值。

当解稀疏度为0.2、0.4、0.6以及0.8时,表1和表2分别给出了系数矩阵 A 的维数为 1000×150 和 4000×300 时的SGRK算法、RaSK算法和SRK算法的迭代步数和计算时间。

当稀疏度为0.2和0.8时,图1和图2分别给出系数矩阵 A 的维数为 1000×150 和 4000×300 时3种算法近似解的相对误差随迭代步数变化的曲线。

通过表1和表2中的实验数据以及图1和图2中的变化曲线可以看出,随着稀疏解非零元个数的不断增加,SRK算法的收敛曲线不断向RaSK算法的

表1 $m = 1000, n = 150$ 时SGRK、RaSK和SRK的迭代步数和计算时间

Tab. 1 Iteration steps and computation time of SGRK, RaSK and SRK ($m = 1000, n = 150$)

稀疏度	SGRK		RaSK		SRK	
	迭代步数	计算时间	迭代步数	计算时间	迭代步数	计算时间
0.2 n	158.80	0.014 6	2 496.50	0.053 0	1 082.10	0.038 1
0.4 n	273.35	0.023 5	2 491.20	0.053 3	1 972.00	0.064 9
0.6 n	360.65	0.026 3	2 552.00	0.055 2	2 559.40	0.082 7
0.8 n	363.05	0.025 9	2 597.50	0.054 7	2 557.30	0.080 4

表2 $m = 4000, n = 300$ 时SGRK、RaSK和SRK的迭代步数和计算时间

Tab. 2 Iteration steps and computation time of SGRK, RaSK and SRK ($m = 4000, n = 300$)

稀疏度	SGRK		RaSK		SRK	
	迭代步数	计算时间	迭代步数	计算时间	迭代步数	计算时间
0.2 n	259.95	0.094 0	4 224.40	0.226 4	2 057.00	0.152 6
0.4 n	418.45	0.146 8	4 302.50	0.229 9	3 736.00	0.306 7
0.6 n	522.90	0.188 2	4 344.60	0.245 1	4 372.70	0.330 1
0.8 n	523.30	0.190 9	4 435.90	0.237 3	4 442.40	0.321 0

收敛曲线靠拢,这是因为SRK算法在对应零元位置上乘以权重,加速减小零元位置上的元素,因此在零元较多的情况下,SRK算法的收敛速度更快。但是,由于SRK算法的计算过程较为复杂,因此在计算时间上一开始低于RaSK算法,后来在迭代步数不明显的情况下,计算时间要多于RaSK算法。SGRK算法在2种算法的基础之上,每次以贪婪的概率准则随机选取行指标进行迭代,因此SGRK算法的计算时间和迭代步数都要优于其他2种算法。

例2 选取了矩阵市场中的一些矩阵进行数值实验。为了得到超定的系数矩阵,将所有的欠定矩阵取转置之后再做数值实验,取稀疏度估计为 $\hat{k} = 2k$, 稀疏度 $k = 0.2n$ 。表3列出的是测试矩阵的相关信息。

对于测试的矩阵,表4给出了SGRK算法、RaSK算法和SRK算法的迭代步数和运行时间。

图3对应于表4,给出了3种算法近似解的相对误差随着迭代步数的变化曲线。通过对比可以发现,在无噪声干扰的情况下,SGRK算法的迭代步数和CPU运行时间都优于原来的RaSK算法和SRK算法。

图4和图5描述SGRK算法、SRK算法和RaSK算法在有噪声干扰的系统的一些数值实验结果。用Matlab软件中的randn随机生成高斯矩阵和范数为0.02的独立高斯噪声进行数值实验。

图4和图5中水平的实线是SGRK算法的相对

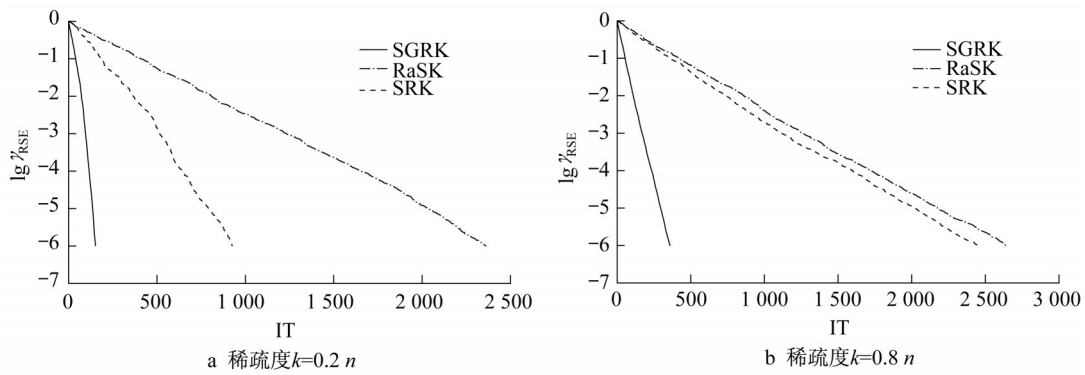


图 1 当矩阵 A 的维数为 1 000 × 150 时 SGRK、RaSK 和 SRK 算法的收敛曲线

Fig. 1 Convergence curves of SGRK, RaSK and SRK methods for Example 1 at a matrix A of 1 000 × 150

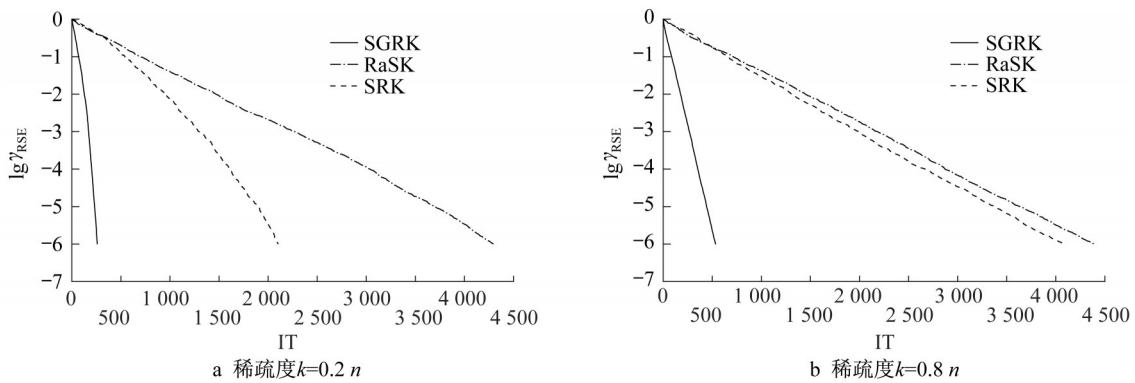


图 2 当矩阵 A 的维数为 4 000 × 300 时 SGRK、RaSK 和 SRK 算法的收敛曲线

Fig. 2 Convergence curves of SGRK, RaSK and SRK methods for Example 1 at a matrix A of 4 000 × 300

表 3 矩阵信息

Tab. 3 Matrix information for Example 2

矩阵名称	阶数	稠密度/%	秩	条件数
WorldCities	315 × 100	23.87	100	66.00
bibd_13_6	78 × 1 716	19.23	78	6.27
Crew1	135 × 6 469	5.38	135	18.2
Trec8	23 × 84	28.42	23	26.89

误差阈值,水平的虚线是 RaSK 算法的相对误差阈值,图 4 是在有噪声干扰的情况下,解的稀疏度为 0.2 时 3 种算法的近似解的相对误差随迭代步数的变化曲线。其中图 4a 描述系数矩阵 A 的维数为 1 000 × 150 时的收敛情况,图 4b 描述的是系数矩阵 A 的维数为 4 000 × 300 时的收敛情况。

表 4 SGRK、RaSK 和 SRK 的迭代步数和计算时间

Tab. 4 Iteration steps and computation time of SGRK, RaSK and SRK

矩阵名称	SGRK		RaSK		SRK	
	迭代步数	计算时间	迭代步数	计算时间	迭代步数	计算时间
bibd_13_6	134.40	0.015 0	1 681.80	0.078 9	753.15	0.039 4
Trec8	53.85	0.002 4	10 697.00	0.217 8	580.80	0.015 1
crew1	433.35	0.118 0	13 222.00	1.468 4	6 050.30	0.747 1
WorldCities	482.30	0.032 1	19 706.00	0.654 5	2 909.10	0.116 6

图 5 是在有噪声干扰的情况下,解的稀疏度为 0.6 时 3 种算法的近似解的相对误差随迭代步数的变化曲线。其中图 5a 和图 5b 分别画出的是系数矩阵 A 的维数为 1 000 × 150 和 4 000 × 300 时 3 种算法的收敛曲线。

由图 4 和图 5 可以看出,在有噪声干扰的情况下,SGRK 算法优于其他 2 种算法最先达到稳定阈值。另外,在图 4 和图 5 中分别用水平的实线和虚线画出了 SGRK 算法的相对误差阈值和 RaSK 算法的相对误差阈值,通过数值实验表明,由定理 2 中

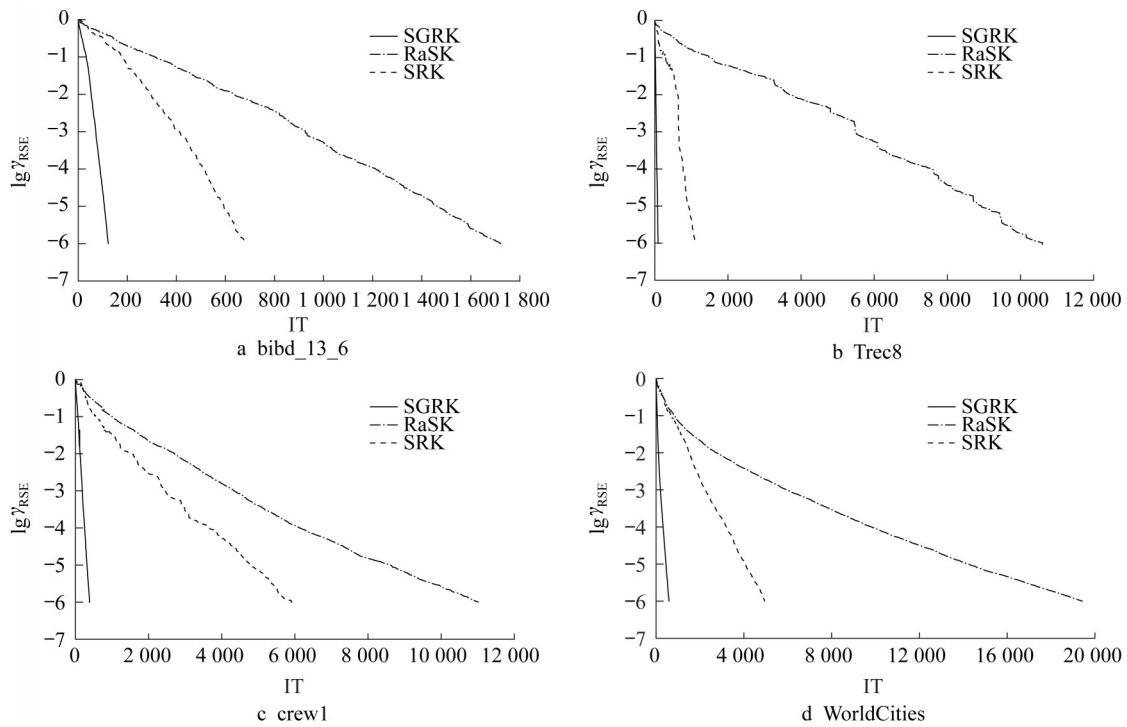


图3 SGRK、RaSK 和 SRK 算法在稀疏度 $k = 0.2 n$ 时的收敛曲线

Fig. 3 Convergence curves of SGRK, RaSK and SRK methods convergence curve for Examples 2 at a sparsity $k = 0.2 n$

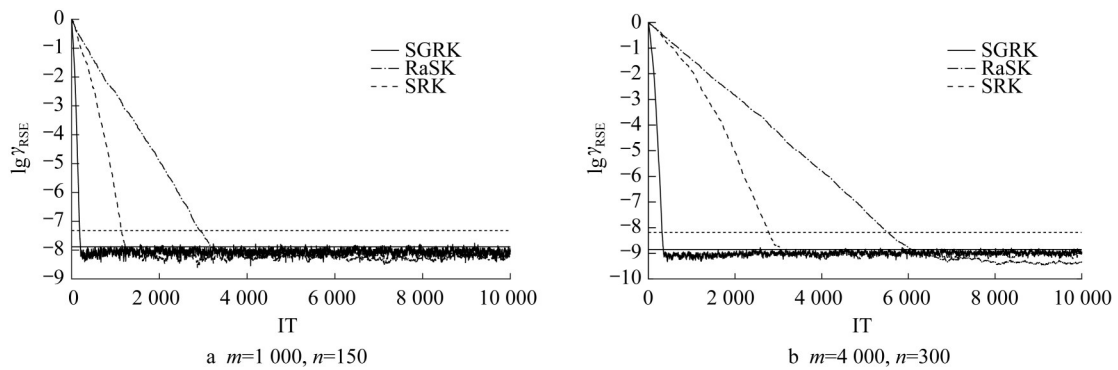


图4 SGRK、RaSK 和 SRK 算法在稀疏度 $k = 0.2 n$ 时误差阈值和实际误差比较

Fig. 4 Comparison of error threshold and actual error of SGRK, RaSK and SRK methods at a sparsity $k = 0.2 n$

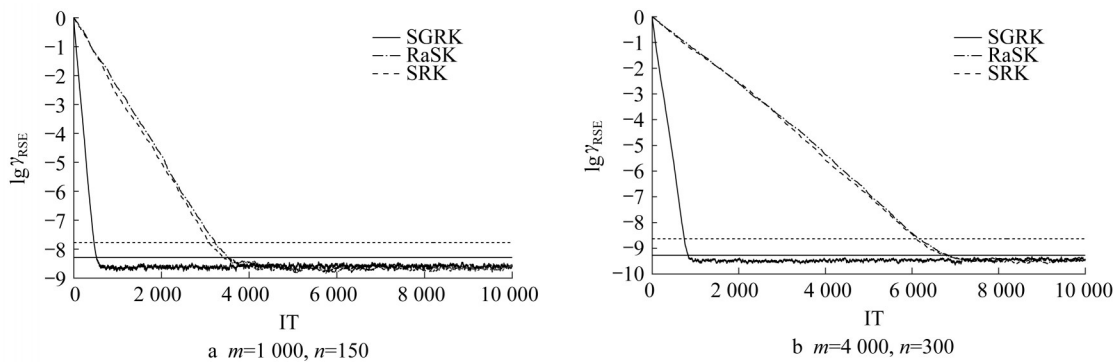


图5 SGRK、RaSK 和 SRK 算法在稀疏度 $k = 0.6 n$ 时误差阈值和实际误差比较

Fig. 5 Comparison of error threshold and actual error of SGRK, RaSK and SRK methods at a sparsity $k = 0.6 n$

SGRK算法推导出来的阈值更接近实际情况。

4 结论

在随机稀疏 Kaczmarz 算法和稀疏随机 Kaczmarz 算法的基础上,受贪婪算法的启发,提出稀疏贪婪随机 Kaczmarz 算法,给出了新算法的收敛性分析,且在有噪声干扰和无噪声干扰的情况下,通过理论证明了新算法的收缩因子小于随机稀疏 Kaczmarz 算法的收缩因子。数值实验表明所提出的新算法在迭代步数和计算时间上均优于传统的随机稀疏 Kaczmarz 算法。

作者贡献声明:

王 泽:算法设计者和算法研究的执行人,构造新的算法,给出收敛性证明,完成数值实验和数据分析、论文初稿的写作。

殷俊锋:研究的构思者及负责人,指导实验设计,数据分析,论文写作与修改。

参考文献:

- [1] TSAIG Y, DONOHO D L. Extensions of compressed sensing [J]. *Signal Processing*, 2005, 86(3):549.
- [2] CANDÈS E, ROMBERG J, TAO T. Stable signal recovery from incomplete and inaccurate measurements [J]. *Communications on Pure and Applied Mathematics*, 2006, 59(8):1207.
- [3] ELAD M. Sparse and redundant representations: From theory to applications in signal and image processing [M]. Berlin: Springer, 2010.
- [4] BYRD R H, CHIN G M, WU Y, *et al.* Sample size selection in optimization models for machine learning [J]. *Mathematical Programming*, 2012, 134(1):127.
- [5] CHEN S S, DONOHO D L, SAUNDERS M A. Atomic decomposition by basis pursuit [J]. *SIAM Review*, 2001, 43(1): 129.
- [6] YIN W, OSHER S, GOLDFARB D, *et al.* Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing [J]. *SIAM Journal on Imaging Sciences*, 2008, 1(1): 143.
- [7] SCHOPFER F. Linear convergence of descent methods for the unconstrained minimization of restricted strongly convex functions [J]. *SIAM Journal on Optimization*, 2016, 26(3): 1883.
- [8] STROHMER T, VERSHYNIN R. A randomized Kaczmarz algorithm with exponential convergence [J]. *Journal of Fourier Analysis and Applications*, 2009, 15(2): 262.
- [9] NEEDELL D, TROPP J A. Paved with good intentions: Analysis of a randomized block Kaczmarz method [J]. *Linear Algebra and its Applications*, 2014, 441(1): 199.
- [10] AGASKAR A, WANG C, LU Y M. Randomized Kaczmarz algorithms: Exact MSE analysis and optimal sampling probabilities [C]// 2014 IEEE Global Conference on Signal and Information Processing (Global SIP). [s.l.]: IEEE, 2015: 389-393.
- [11] LIU J, WRIGHT S. An accelerated randomized Kaczmarz algorithm [J]. *Mathematics of Computation*, 2016, 85(297): 153.
- [12] LORENZ D A, SCHOPFER F, WENGER S. A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing [C]//2014 IEEE International Conference on Image Processing (ICIP). [s.l.]:IEEE, 2015:1347-1351.
- [13] LORENZ D A, SCHOPFER F, WENGER S. The linearized Bregman method via split feasibility problems: Analysis and generalizations [J]. *SIAM Journal on Imaging Sciences*, 2014, 7(2):1237.
- [14] ZOUZIAS A, FRERIS N M. Randomized extended Kaczmarz for solving least squares[J]. *SIAM Journal on Matrix Analysis and Applications*, 2013, 34(2): 773.
- [15] NEEDELL D. Randomized Kaczmarz solver for noisy linear system [J]. *BIT Numerical Mathematics*, 2010, 50(2): 395.
- [16] SCHOPFER F, LORENZ D A. Linear convergence of the randomized sparse Kaczmarz method [J]. *Mathematical Programming*, 2019, 173(1/2): 509.
- [17] 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: A592.
- [18] 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.
- [19] 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算法[J]. *同济大学学报(自然科学版)*, 2020, 48(8): 1224. DOI: 10.11908/j. issn. 0253-374x. 20041. 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. DOI: 10.11908/j. issn. 0253-374x. 20041.
- [20] 荆燕飞, 李彩霞, 胡少亮. 求解大型稀疏线性系统的贪婪双空间随机Kaczmarz方法 [J]. *同济大学学报(自然科学版)*, 2021, 49(10): 1473. DOI: 10.11908/j. issn. 0253-374x. 21054. JING Yanfei, LI Caixia, HU Shaoliang. A greedy two-subspace randomized Kaczmarz method for solving large sparse linear systems [J]. *Journal of Tongji University (Natural Science)*, 2021, 49(10): 1473. DOI: 10.11908/j. issn. 0253-374x. 21054.
- [21] MANSOUR H, YILMAZ O. A fast randomized Kaczmarz algorithm for sparse solutions of consistent linear system [J]. *arXiv*, 2013, 1305.3803v1.
- [22] ROCKAFELLAR R T, WETS R, *et al.* *Variational analysis* [M]. Berlin: Springer, 2009.