非凸优化与数值算法专题

DOI:10.11721/cqnuj20260210

渐近奇异线性方程组的鲁棒 Kaczmarz 算法*

柯云营[1]，姜　文[1]，罗　浩[1,2,3]
(1. 重庆师范大学 重庆国家应用数学中心；2. 重庆师范大学 最优化技术与智能汽车重庆市高校重点实验室，重庆 401331；
3. 北京大学 重庆大数据研究院，重庆 401121)

摘要:针对经典 Kaczmarz 方法在求解渐近奇异线性方程组时收敛速度剧降的问题,提出具有一致稳定收敛率的鲁棒算法。将系数矩阵的行空间视为 Grassman 流形中的子空间,引入了一种具有行秩退化特点的渐近奇异矩阵新概念,并定义了渐近核空间与渐近对偶核空间。在此基础上,通过子空间校正框架,提出了鲁棒 Kaczmarz 方法,并利用 Xu-Zikatanov 恒等式完成了收敛性分析。最后,基于 Kaczmarz 和坐标下降法的等价性及渐近核与对偶核之间的关系,提出了鲁棒坐标下降算法,并证明了鲁棒 Kaczmarz 方法和鲁棒坐标下降算法之间的等价性。理论分析证明,所提出的鲁棒 Kaczmarz 方法和鲁棒坐标下降算法在求解渐近奇异线性方程组时具有一致稳定收敛速率。数值实验验证了算法的鲁棒性与有效性。

关键词:渐近奇异线性方程组;Kaczmarz 方法;坐标下降法;子空间校正法;Xu-Zikatanov 恒等式

中图分类号:O241.6　　**文献标志码**:A　　**文章编号**:1672-6693(2026)02-0116-13

1　研究背景

线性方程组广泛应用于各种领域,包括但不限于图像处理、信号处理、经济建模、控制系统等。在这些应用中,线性方程组的解往往代表着关键信息或决策的基础。因此,研究如何高效、准确地求解线性方程组具有重要意义。随着计算技术的不断发展,迭代方法因具有灵活性和高效性而受到越来越多的关注。特别是在处理大规模线性方程组时,迭代方法往往比直接方法更具优势[1]。

渐近奇异线性方程组是理论研究和实际应用中常见的问题,如 CT 重建[2]和不可压缩线性弹性问题的稳定离散化[3]等都是渐近奇异线性方程组的典型例子。如何设计出适用于渐近奇异线性方程组的鲁棒迭代算法,成为学者关注的问题。

Kaczmarz 方法[4]是求解线性方程组的重要迭代方法之一,该方法的核心思想是通过逐行处理线性方程组系数矩阵,逐步逼近方程的解。考虑一个相容的线性方程组:

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}, \tag{1}$$

其中:$\boldsymbol{A}\in\mathbf{R}^{m\times n}$,$\boldsymbol{b}\in\mathbf{R}^{m}$。从任意初始向量 $\boldsymbol{x}_0\in\mathbf{R}^n$ 开始,经典 Kaczmarz 方法如下所示:

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\frac{\boldsymbol{b}_i-\boldsymbol{A}_{(i)}\boldsymbol{x}_k}{\|\boldsymbol{A}_{(i)}\|^2}\boldsymbol{A}_{(i)}^{\mathrm{T}},\ i=k \bmod m。 \tag{2}$$

其中:$\boldsymbol{A}_{(i)}$ 表示 $\boldsymbol{A}$ 的第 i 行,$\boldsymbol{b}_i$ 是 $\boldsymbol{b}$ 的第 i 个分量。Strohmer 等人[5]提出了随机 Kaczmarz(randomized Kaczmarz,RK)方法,并证明了该方法的收敛速度为 ρ^k,其中收缩因子为 $\rho=1-1/\kappa^2(\boldsymbol{A})$,而 $\kappa(\boldsymbol{A})$ 和 $\sigma_{\min}^{+}(\boldsymbol{A})$ 分别代表 $\boldsymbol{A}$ 的比例条件数[6]和 $\boldsymbol{A}$ 的最小非零奇异值,其中 $\kappa(\boldsymbol{A}):=\|\boldsymbol{A}\|_F/\sigma_{\min}^{+}(\boldsymbol{A})$。在此之后,文献中涌现出许多 Kaczmarz 算法的变体[7-10]。

众所周知,Kaczmarz 方法求解式(1)等价于用坐标下降(coordinate descent, CD)方法求解如下无约束优化问题:

* **收稿日期**:2026-03-01　　**网络出版时间**:2026-05-09T09:07
资助项目:国家自然科学基金青年科学基金项目(No. 12401402);重庆市教育委员会科学技术研究计划项目(No. KJZDK202300505);重庆市自然科学基金面上项目(No. CSTB2024NSCQMSX0329);重庆师范大学基金项目资助(No. 22xwB020)
第一作者简介:柯云营,女,研究方向为数值线性代数,E-mail:2023110510025@stu.cqnu.edu.cn;**通信作者简介**:罗浩,男,副教授,博士,E-mail:luohao@cqnu.edu.cn
网络出版地址:https://link.cnki.net/urlid/50.1165.n.20260508.1546.002

$$\min_{\boldsymbol{y}\in\mathbf{R}^m} g(\boldsymbol{y})=\frac{1}{2}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}\|^2+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}。\tag{3}$$

对于这个无约束优化问题，CD 的迭代过程具体由下式给出：

$$\boldsymbol{y}_{k+1}=\boldsymbol{y}_k-\frac{1}{L_i}\boldsymbol{U}_i\boldsymbol{U}_i^{\mathrm{T}}\nabla g(\boldsymbol{y}_k),i=k \bmod m。\tag{4}$$

其中：$L_i>0$ 表示步长，$\boldsymbol{U}_i>0$ 表示 m 阶单位矩阵的第 i 列。若 $L_i=\|\boldsymbol{A}_{(i)}\|^{-2}$，将式(4)两边乘以 $-\boldsymbol{A}^{\mathrm{T}}$，就可得到经典的 Kaczmarz 迭代式，也即为式(2)。Nesterov 提出了式(4)的随机版本[11]，对于一般的凸目标函数，该迭代的收敛速度为 $O(1/k)$。同时，Nesterov 还提出了加速坐标下降法(accelerated coordinate descent method，ACDM)[11]，以实现更快的收敛速度 $O(1/k^2)$；对于强凸的目标函数，ACDM 具有线性收敛速率。将 ACDM 应用于式(1)的对偶问题，即可得到用于求解式(1)的加速随机 Kaczmarz(accelerated randomized Kaczmarz，ARK)方法[12]，若 $\lambda_{\min}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})>0$，ARK 的收敛速度为 $(1-m^{-1/2}/\kappa(\boldsymbol{A}))^k$。

此外，子空间校正(successive subspace correction，SSC)法是分析 Kaczmarz 算法极有力的工具之一。Xu 等人[13]提出了子空间校正方法的一般理论框架和著名的 Xu-Zikatanov 恒等式(以下简称 XZ 等式)。在 SSC 框架下，Oswald 等人[14]证明了循环 Kaczmarz 方法收敛到最小二乘解 $\boldsymbol{x}_{\mathrm{LS}}$，收缩因子为 $1-c/(\kappa^2(\boldsymbol{D}^{-1/2}\boldsymbol{A}))$，其中 $c=O(\ln m)$ 是一个正常数。

综上可见，Kaczmarz 方法的收敛性严重依赖于系数矩阵的条件数 $\kappa(\boldsymbol{A})$。对于渐近奇异线性方程组，由于病态条件数导致收缩因子接近 1，从而收敛速度变得极为缓慢。CD 法在求解线性方程组对偶问题时，与求解原线性方程组的 Kaczmarz 方法完全等价。基于此等价关系，CD 法在求解渐近奇异线性方程组时，必然有相同的缺陷，收敛性能将随线性方程组病态程度加剧而恶化。

考虑如下相容线性线性方程组：

$$\boldsymbol{A}(\varepsilon)\boldsymbol{x}=\boldsymbol{b},\tag{5}$$

其中：$\boldsymbol{A}(\varepsilon)\in\mathbf{R}^{m\times n}$，并依赖于一个小参数 $\varepsilon>0$，$\boldsymbol{b}\in\mathrm{Im}(\boldsymbol{A}(\varepsilon))$ 保证了相容性。这里，$\boldsymbol{b}\in\mathrm{Im}(\boldsymbol{A}(\varepsilon))$ 表示 $\boldsymbol{A}(\varepsilon)$ 的列空间。当 $\varepsilon\to0^+$ 时，$\boldsymbol{A}(\varepsilon)$ 变得病态，因为 $\sigma_{\min}^{+}(\boldsymbol{A}(\varepsilon))\to0$。考察一个简单例子：

$$\boldsymbol{A}(\varepsilon)=\begin{bmatrix}1 & -1\\ 1+\varepsilon & -1+\varepsilon\end{bmatrix}。\tag{6}$$

如表 1 所示，减小 ε 会急剧增大 $\kappa(\boldsymbol{A})$，并且 Kaczmarz 方法变得低效。这就引出了本文的主要动机：如何克服渐近奇异性对 Kaczmarz 方法带来的影响？

表 1　参数 ε 不同取值下的 $\kappa(A)$ 及 Kaczmarz 方法的迭代次数

Tab. 1　$\kappa(A)$ and the number of iterations for the Kaczmarz method under different values of ε

ε	迭代次数		ε	迭代次数	
	$\kappa(\boldsymbol{A})$	Kaczmarz 方法		$\kappa(\boldsymbol{A})$	Kaczmarz 方法
1/3	38	157	1/27	2 918	12 011
1/9	326	1 342	1/81	26 246	108 033

2　预备知识

本节将介绍一些相关的预备知识，包括基本符号、Grassman 流形以及 SSC 框架。

2.1　符号说明

设 $\boldsymbol{V}$ 为一个有限维 Hilbert 空间，内积为 $\sqrt{(\cdot,\cdot)}$，诱导范数为 $\|\cdot\|=\sqrt{(\cdot,\cdot)}$。用 $\mathcal{L}(\boldsymbol{V})$ 表示所有从 $\boldsymbol{V}$ 到 $\boldsymbol{V}$ 的线性映射的集合。恒等映射 $I:\boldsymbol{V}\to\boldsymbol{V}$ 定义为对所有 $\boldsymbol{v}\in\boldsymbol{V}$ 满足 $I\boldsymbol{v}=\boldsymbol{v}$。如果 $\boldsymbol{W}$ 是一个子空间，$\boldsymbol{W}\subset\boldsymbol{V}$，则定义包含算子 $\boldsymbol{\iota}_{\boldsymbol{W}\to\boldsymbol{V}}:\boldsymbol{W}\to\boldsymbol{V}$ 为恒等算子在上的限制，即 $\boldsymbol{\iota}_{\boldsymbol{W}\to\boldsymbol{V}}=\boldsymbol{I}_{\boldsymbol{W}}$。

给定任意 $\boldsymbol{A}\in\mathcal{L}(\boldsymbol{V})$，它的算子范数为：

$$\|\boldsymbol{A}\|_{\boldsymbol{V}\to\boldsymbol{V}}:=\sup_{\|\boldsymbol{v}\|=1}\|\boldsymbol{A}\boldsymbol{v}\|,$$

伴随算子 $\boldsymbol{A}':\boldsymbol{V}\to\boldsymbol{V}$ 由下式定义：

$$(\boldsymbol{A}'\boldsymbol{u},\boldsymbol{v})=(\boldsymbol{u},\boldsymbol{A}\boldsymbol{v}),\ \forall\,\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{V}。$$

当 $\boldsymbol{A}$ 是对称正定(symmetric positive definite,SPD)的,即 $\boldsymbol{A}=\boldsymbol{A}'$,且对所有 $\boldsymbol{v}\in\boldsymbol{V}\backslash\{\boldsymbol{0}\}$ 有 $\langle\boldsymbol{A}\boldsymbol{v},\boldsymbol{v}\rangle>0$,还定义 $\boldsymbol{A}$-内积

$$(\boldsymbol{u},\boldsymbol{v})_{\boldsymbol{A}}:=(\boldsymbol{A}\boldsymbol{u},\boldsymbol{v}),\ \forall\,\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{V},$$

及它的诱导的 $\boldsymbol{A}$-范数 $\|\cdot\|_{\boldsymbol{A}}:=\sqrt{(\cdot,\cdot)_{\boldsymbol{A}}}$。对任意 $\boldsymbol{B}\in\mathcal{L}(\boldsymbol{V})$,它的 $\boldsymbol{A}$-范数定义为：

$$\|\boldsymbol{B}\|_{\boldsymbol{A}}:=\sup_{\|\boldsymbol{v}\|_{\boldsymbol{A}}=1}\|\boldsymbol{B}\boldsymbol{v}\|_{\boldsymbol{A}},$$

并用 $\boldsymbol{B}^{\mathrm{t}}:\boldsymbol{V}\to\boldsymbol{V}$ 表示关于 $\boldsymbol{A}$-内积的伴随算子：

$$(\boldsymbol{B}^{\mathrm{t}}\boldsymbol{u},\boldsymbol{v})_{\boldsymbol{A}}=(\boldsymbol{u},\boldsymbol{B}\boldsymbol{v})_{\boldsymbol{A}},\ \forall\,\boldsymbol{u},\boldsymbol{v}\in\boldsymbol{V}。$$

显然,对任意 $\boldsymbol{B}\in\mathcal{L}(\boldsymbol{V})$,有 $(\boldsymbol{B}\boldsymbol{A})^{\mathrm{t}}=\boldsymbol{B}'\boldsymbol{A}$。特别地,如果 $\boldsymbol{B}=\boldsymbol{B}'$ 是对称的,则 $(\boldsymbol{B}\boldsymbol{A})^{\mathrm{t}}=\boldsymbol{B}\boldsymbol{A}$,这意味着 $\boldsymbol{B}\boldsymbol{A}$ 关于 $\boldsymbol{A}$-内积是对称的。

令 $\|\cdot\|$ 为向量核矩阵的标准 2-范数,$\langle\cdot,\cdot\rangle$ 为向量间的 Euclid 内积。对于一个线性子空间 $\boldsymbol{S}(\boldsymbol{S}\subset\mathbf{R}^n)$,$\dim(\boldsymbol{S})$ 指该空间的维数。对于任意 $\boldsymbol{C}\in\mathbf{R}^{m\times n}$,用 $\boldsymbol{C}^{\mathrm{T}}$ 和 $\boldsymbol{C}^{\dagger}$ 分别表示 $\boldsymbol{C}$ 的转置和 Moore-Penrose 伪逆,并令 $\operatorname{rank}\boldsymbol{C}$、$\operatorname{Ker}(\boldsymbol{C})$ 和 $\operatorname{Im}(\boldsymbol{C})$ 分别表示 C 的秩、核空间和值域(或称列空间)。$\boldsymbol{C}^{\mathrm{T}}$ 的值域称为 $\boldsymbol{C}$ 的行空间,通常记为 $\operatorname{Row}(\boldsymbol{C})$。一个基本事实是,对所有 $\boldsymbol{C}\in\mathbf{R}^{m\times n}$,有 $\operatorname{Row}(\boldsymbol{C})=\operatorname{Row}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})$,这意味着 $\dim(\operatorname{Row}(\boldsymbol{C}))=\operatorname{rank}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}})$。对于一个对称矩阵 $\boldsymbol{C}\in\mathbf{R}^{m\times n}$,令 $\lambda_{\min}(\boldsymbol{C})$ 和 $\lambda_{\max}(\boldsymbol{C})$ 分别为 $\boldsymbol{C}$ 的最小、最大特征值。如果 $\boldsymbol{C}$ 是对称半正定的,则用 $\lambda_{\min}^{+}(\boldsymbol{C})$ 表示 $\boldsymbol{C}$ 的最小非零特征值,即：

$$\lambda_{\min}^{+}(\boldsymbol{C}):=\min_{\boldsymbol{v}\in\operatorname{Im}(\boldsymbol{C})/\{\boldsymbol{0}\}}\langle\boldsymbol{C}\boldsymbol{v},\boldsymbol{v}\rangle/\|\boldsymbol{v}\|^2。\tag{7}$$

2.2 Grassman 流形

设 $n,p\in\mathbf{N}_+$ 且满足 $p\leqslant n$。Grassman 流形 $\operatorname{Grass}(p,n)$ 由 $\mathbf{R}^n$ 中所有 p 维子空间构成。

定义 1($\operatorname{Grass}(p,n)$ 上的间隔度量) 对于任意 $U,W\in\operatorname{Grass}(p,n)$,间隔度量定义为：

$$\Delta(U,W):=\|P_U-P_W\|,$$

其中：$P_U:\mathbf{R}^n\to U$ 和 $P_W:\mathbf{R}^n\to W$ 均是正交投影算子。

众所周知,配备了间隔度量 $\Delta(\cdot,\cdot)$ 的 Grassmann 流形 $\operatorname{Grass}(p,n)$ 构成一个完备度量空间。

引理 1 设 $\boldsymbol{A}:[0,\delta]\to\mathbf{R}^{m\times n}$ 是一个逐分量连续的矩阵值函数。如果对于所有 $\varepsilon\in[0,\delta]$ 有 $\operatorname{Row}(\boldsymbol{A}(\varepsilon))\in\operatorname{Grass}(r,n)$,那么：

$$\lim_{\varepsilon\to0+}\Delta(\operatorname{Row}(\boldsymbol{A}(\varepsilon)),\operatorname{Row}(\boldsymbol{A}(0)))=0。$$

证明 到子空间 $\operatorname{Row}(\boldsymbol{A}(\varepsilon))$ 的正交投影算子 $P_{\boldsymbol{A}(\varepsilon)}:\mathbf{R}^n\to\operatorname{Row}(\boldsymbol{A}(\varepsilon))$ 由下式给出：

$$P_{\boldsymbol{A}(\varepsilon)}=\boldsymbol{A}^{\mathrm{T}}(\varepsilon)(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))^{\dagger}\boldsymbol{A}(\varepsilon),$$

其中：$(\cdot)^{\dagger}$ 表示 Moore-Penrose 伪逆。由于 $\boldsymbol{A}(\varepsilon)$ 逐分量连续且对所有 $\varepsilon\in[0,\delta]$ 有 $\operatorname{Row}(\boldsymbol{A}(\varepsilon))\in\operatorname{Grass}(r,n)$,可以断言 $\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)$ 也是连续的,且 $\operatorname{rank}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))=\dim\operatorname{Row}(\boldsymbol{A}(\varepsilon))=r$,并对所有 $\varepsilon\in[0,\delta]$ 成立。根据文献[15],可得 $(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))^{\dagger}$ 是连续的,因此 $P_{\boldsymbol{A}(\varepsilon)}$ 也是连续的。特别地,由定义 1 有：

$$\lim_{\varepsilon\to0^+}\Delta(\operatorname{Row}(\boldsymbol{A}(\varepsilon)),\operatorname{Row}(\boldsymbol{A}(0)))=\lim_{\varepsilon\to0^+}\|P_{\boldsymbol{A}(\varepsilon)}-P_{\boldsymbol{A}(0)}\|=0。$$

证毕

2.3 SSC 方法

设 $\boldsymbol{V}$ 是一个有限维 Hilbert 空间,带有内积 $(\cdot,\cdot)$ 及诱导范数 $\|\cdot\|$,$\|\cdot\|=\sqrt{(\cdot,\cdot)}$。给定一个 SPD 线性算子 $\boldsymbol{A}:\boldsymbol{V}\to\boldsymbol{V}$ 和 $\boldsymbol{f}\in\boldsymbol{V}$,旨在寻求 $\boldsymbol{u}\in\boldsymbol{V}$ 使得：

$$\boldsymbol{A}\boldsymbol{u}=\boldsymbol{f}。$$

下面简要重述一下求解式的逐次 SSC 的基本框架;全面介绍参考文献[13,16-17]。这样的框架对于分析 Kaczmarz 迭代以及发展求解渐近奇异线性方程组的鲁棒变体至关重要。

2.3.1 方法介绍

考虑一个有限维的辅助空间 $\{\boldsymbol{V}_j\}_{j=1}^{J}$ 的有限序列。每个 $\boldsymbol{V}_j$ 不一定是 $\boldsymbol{V}$ 的子空间,但通过线性算子 $\boldsymbol{R}_j:\boldsymbol{V}_j\to\boldsymbol{V}$ 与 $\boldsymbol{V}$ 相关联,该算子满足以下分解假设。

假设 1 $\boldsymbol{V}=\sum_{j=1}^{J}\boldsymbol{R}_j\boldsymbol{V}_j$ 成立。

根据假设 1，对于任意 $\boldsymbol{v}\in\boldsymbol{V}$，有 $\boldsymbol{v}=\sum_{j=1}^{J}\boldsymbol{R}_j\boldsymbol{v}_j$，其中 $\boldsymbol{v}_j\in\boldsymbol{V}_j$，$1\leqslant j\leqslant J$。这里不要求这样的分解是直和，即 $\boldsymbol{v}$ 的表示不必是唯一的。对于每个辅助空间，还提出以下假设。

假设 2　对于 $1\leqslant j\leqslant J$，$\boldsymbol{V}_j$ 是一个 Hilbert 空间，具有内积 $a_j(\cdot,\cdot)$。

设 $\boldsymbol{R}_j^*:\boldsymbol{V}\to\boldsymbol{V}_j$ 为 $\boldsymbol{R}_j$ 的伴随算子，即：

$$a_j(\boldsymbol{R}_j^*\boldsymbol{v},\boldsymbol{v}_j)=(\boldsymbol{v},\boldsymbol{R}_j\boldsymbol{v}_j),\ \forall\,\boldsymbol{v}_j\in\boldsymbol{V}_j,\boldsymbol{v}\in\boldsymbol{V}。$$

定义线性算子 $\boldsymbol{T}_j:\boldsymbol{V}\to\boldsymbol{V}_j$ 如下：

$$a_j(\boldsymbol{T}_j\boldsymbol{v},\boldsymbol{v}_j)=(\boldsymbol{v},\boldsymbol{R}_j\boldsymbol{v}_j)_{\boldsymbol{A}},\ \forall\,\boldsymbol{v}_j\in\boldsymbol{V}_j,\boldsymbol{v}\in\boldsymbol{V}。$$

显然，$\boldsymbol{T}_j$ 是 $\boldsymbol{R}_j$ 关于 $\boldsymbol{A}$-内积的伴随算子，且 $\boldsymbol{T}_j=\boldsymbol{R}_j^*\boldsymbol{A}$。注意 $\boldsymbol{R}_j\boldsymbol{R}_j^*:\boldsymbol{V}\to\boldsymbol{V}$ 是对称的，因此 $\boldsymbol{R}_j\boldsymbol{T}_j=\boldsymbol{R}_j\boldsymbol{R}_j^*\boldsymbol{A}$，则有 $\boldsymbol{R}_j\boldsymbol{T}_j$ 关于 $\boldsymbol{A}$-内积是对称的。

带松弛的 SSC 方法如下所示[17]。给定 $\boldsymbol{u}_k\in\boldsymbol{V}$，令 $\boldsymbol{v}_{k,1}=\boldsymbol{u}_k$，并通过下式更新 $\boldsymbol{u}_{k+1}=\boldsymbol{v}_{k,J+1}$：

$$\boldsymbol{v}_{k,j+1}=\boldsymbol{v}_{k,j}+\omega\boldsymbol{R}_j\boldsymbol{R}_j^*(\boldsymbol{f}-\boldsymbol{A}\boldsymbol{v}_{k,j}),j=1,\cdots,J,$$

其中：$\omega>0$ 为松弛参数。上述格式可改写为乘法 Schwarz 迭代：

$$\boldsymbol{u}_{k+1}=\boldsymbol{u}_k+\boldsymbol{B}_{\mathrm{ssc}}(\omega)(\boldsymbol{f}-\boldsymbol{A}\boldsymbol{u}_k),\tag{8}$$

其中：迭代算子 $\boldsymbol{B}_{\mathrm{ssc}}(\omega):\boldsymbol{V}\to\boldsymbol{V}$ 满足 $\boldsymbol{I}-\boldsymbol{B}_{\mathrm{ssc}}(\omega)\boldsymbol{A}=(\boldsymbol{I}-\omega\boldsymbol{R}_J\boldsymbol{T}_J)\cdot\cdots\cdot(\boldsymbol{I}-\omega\boldsymbol{R}_1\boldsymbol{T}_1)$。

2.3.2　收敛理论

基于假设 1，引入一个乘积型的辅助空间 $\widetilde{\boldsymbol{V}}:=\boldsymbol{V}_1\times\boldsymbol{V}_2\times\cdots\times\boldsymbol{V}_J$ 以及相应的内积 $(\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}:=\sum_{j=1}^{J}a_j(\boldsymbol{u}_j,\boldsymbol{v}_j)$，对所有 $\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{v}}\in\widetilde{\boldsymbol{V}}$ 成立，其中 $\widetilde{\boldsymbol{u}}_j=\boldsymbol{u}_j,\widetilde{\boldsymbol{v}}_j=\boldsymbol{v}_j$。诱导范数为 $\|\cdot\|_{\widetilde{\boldsymbol{V}}}:=\sqrt{(\cdot,\cdot)_{\widetilde{\boldsymbol{V}}}}$。定义线性算子 $\mathcal{R}:\widetilde{\boldsymbol{V}}\to\boldsymbol{V}$ 为：

$$\mathcal{R}\widetilde{\boldsymbol{u}}=\sum_{j=1}^{J}\boldsymbol{R}_j\widetilde{\boldsymbol{u}}_j=\sum_{j=1}^{J}\boldsymbol{R}_j\boldsymbol{u}_j。$$

可知 $\mathcal{R}:\widetilde{\boldsymbol{V}}\to\boldsymbol{V}$ 是满射，即 $\mathcal{R}$ 是满射。用 $\mathcal{R}^\star:\boldsymbol{V}\to\widetilde{\boldsymbol{V}}$ 表示 $\mathcal{R}$ 的伴随算子：

$$(\mathcal{R}^\star\boldsymbol{u},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}=(\boldsymbol{u},\mathcal{R}\widetilde{\boldsymbol{v}})。$$

还定义 $\mathcal{R}$ 关于 $\boldsymbol{A}$-内积的伴随算子：

$$(\mathcal{R}^{\mathrm{a}}\boldsymbol{u},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}=(\boldsymbol{u},\mathcal{R}\widetilde{\boldsymbol{v}})_{\boldsymbol{A}}。$$

显然 $\mathcal{R}^{\mathrm{a}}=(\boldsymbol{A}\mathcal{R})^\star=\mathcal{R}^\star\boldsymbol{A}$。

令 $\widetilde{\boldsymbol{A}}=\mathcal{R}^\star\boldsymbol{A}\mathcal{R}$ 为扩展算子。若将每个 $\widetilde{\boldsymbol{v}}\in\widetilde{\boldsymbol{V}}$ 写成一个“列向量”，则按一致的块形式有：

$$\mathcal{R}=(\boldsymbol{R}_1,\cdots,\boldsymbol{R}_J)\text{且}\ \widetilde{\boldsymbol{A}}=\begin{bmatrix}\boldsymbol{T}_1\boldsymbol{R}_1 & \boldsymbol{T}_1\boldsymbol{R}_2 & \cdots & \boldsymbol{T}_1\boldsymbol{R}_J\\ \boldsymbol{T}_2\boldsymbol{R}_1 & \boldsymbol{T}_2\boldsymbol{R}_2 & \cdots & \boldsymbol{T}_2\boldsymbol{R}_J\\ \vdots & \vdots & & \vdots\\ \boldsymbol{T}_J\boldsymbol{R}_1 & \boldsymbol{T}_J\boldsymbol{R}_2 & \cdots & \boldsymbol{T}_J\boldsymbol{R}_J\end{bmatrix},\boldsymbol{R}^*=\begin{bmatrix}\boldsymbol{R}_1^\star\\ \boldsymbol{R}_2^\star\\ \vdots\\ \boldsymbol{R}_J^\star\end{bmatrix},\widetilde{\boldsymbol{v}}=\begin{bmatrix}\boldsymbol{v}_1\\ \boldsymbol{v}_2\\ \vdots\\ \boldsymbol{v}_J\end{bmatrix}。$$

现在考虑块分裂 $\widetilde{\boldsymbol{A}}=\widetilde{\boldsymbol{D}}+\widetilde{\boldsymbol{L}}+\widetilde{\boldsymbol{U}}$，其中 $\widetilde{\boldsymbol{D}}=\mathrm{diag}(\widetilde{\boldsymbol{A}})$ 表示块对角部分，$\widetilde{\boldsymbol{U}}=\mathrm{triu}(\widetilde{\boldsymbol{A}})$ 是严格上块三角部分。不难验证 $\widetilde{\boldsymbol{L}}'=\widetilde{\boldsymbol{U}}$ 和 $\widetilde{\boldsymbol{D}}'=\widetilde{\boldsymbol{D}}$。

在以下等价于 $2\widetilde{\boldsymbol{I}}-\omega\widetilde{\boldsymbol{D}}$ 是 SPD 的假设下，可以建立式(8)的收敛性结果。该结果被称为著名的 XZ 等式[13]，为本文分析 Kaczmarz 迭代及变体提供了有力工具。

假设 3　$2\widetilde{\boldsymbol{I}}-\omega\widetilde{\boldsymbol{D}}$ 是 SPD 的。或者等价地，$\omega\in(0,2/\|\widetilde{\boldsymbol{D}}\|_{\widetilde{\boldsymbol{V}}\to\widetilde{\boldsymbol{V}}})$。

引理 2[13]（XZ 等式）　在假设 1、假设 2 和假设 3 下，Schwarz 迭代式收敛，也即式(8)收敛，且：

$$\|\boldsymbol{I}-\boldsymbol{B}_{\mathrm{ssc}}(\omega)\boldsymbol{A}\|_{\boldsymbol{A}}^2=1-\frac{1}{c_0(\omega)},$$

上式中的有限正常数 $c_0(\omega)$ 定义为：

$$c_0(\omega):=\frac{1}{\omega}\sup_{\|\boldsymbol{v}\|_{\boldsymbol{A}}=1}\inf_{\mathcal{R}\widetilde{\boldsymbol{v}}=\boldsymbol{v}}((2\widetilde{\boldsymbol{I}}-\omega\widetilde{\boldsymbol{D}})^{-1}(\widetilde{\boldsymbol{I}}+\omega\widetilde{\boldsymbol{U}})\widetilde{\boldsymbol{v}},(\widetilde{\boldsymbol{I}}+\omega\widetilde{\boldsymbol{U}})\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}。$$

3 渐近奇异性

Lee 等人[18]考虑了一种特殊的渐近奇异情形,即 $\boldsymbol{A}(\varepsilon)=\varepsilon\boldsymbol{I}+\boldsymbol{A}_0$,其中 $\boldsymbol{A}_0$ 为对称半正定矩阵。显然,对于当 $\varepsilon\to 0^+$ 时,$\boldsymbol{A}(\varepsilon)$的最小特征值逐渐趋于 0。本节旨在将此类渐近奇异性质推广至更一般的非方阵情形。

3.1 渐近奇异性的定义

众所周知,Kaczmarz 迭代本质上是一种行作用法[4],并且也可以重新表述到 SSC 框架[13-14]中,只需对 $\boldsymbol{A}(\varepsilon)$的行空间进行适当的子空间分解,详见后文第 4 节。受此启发,现在不再关注最小特征值或奇异值的退化,而是关注当 ε 趋近于 0 时 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$的渐近行为。更精确地说,本文要求矩阵满足以下假设。

假设 4 矩阵值函数 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 是逐分量连续的,且具有双块结构:

$$\boldsymbol{A}(\varepsilon)=\begin{bmatrix}\boldsymbol{A}_0(\varepsilon)\\ \boldsymbol{A}_1(\varepsilon)\end{bmatrix},\ \forall\, 0\leqslant\varepsilon\leqslant\bar{\varepsilon},$$

其中:$\bar{\varepsilon}>0$,$\boldsymbol{A}_0(\varepsilon)\in\mathbf{R}^{m_0\times n}$,$\boldsymbol{A}_1(\varepsilon)\in\mathbf{R}^{m_1\times n}$,$m_0+m_1=m$。此外,对角部分

$$\boldsymbol{D}_\varepsilon:=\mathrm{diag}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)):[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times m}$$

非退化,即:

$$\eta_0\boldsymbol{I}\prec\mathrm{diag}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))\prec\eta_1\boldsymbol{I},\ \forall\, 0\leqslant\varepsilon\leqslant\bar{\varepsilon},\tag{9}$$

其中:$0<\eta_0\leqslant\eta_1<\infty$,且 η_0 和 η_1 均与 ε 无关。

从 SSC 的角度来看,行空间的稳定性起着关键作用。受此启发,引入以下渐近奇异性质的定义,该性质会导致不稳定的子空间分解。

定义 2 设 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 是一个满足假设 4 的矩阵值函数。如果 $\boldsymbol{A}(\varepsilon)$满足以下条件,则称它在 $\varepsilon=0$ 处是渐近奇异的:

1) 对所有 $0<\varepsilon\leqslant\bar{\varepsilon}$,有 $\dim\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))<\dim\mathrm{Row}(\boldsymbol{A}(\varepsilon))$;

2) 存在 $r_0\in\mathbf{N}_+$,使得对所有 $0\leqslant\varepsilon\leqslant\bar{\varepsilon}$,有 $\dim\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))=r_0$;

3) 存在 $\boldsymbol{\Pi}\in\mathbf{R}^{m_1\times m_0}$,使得 $\boldsymbol{A}_1(0)=\boldsymbol{\Pi}\boldsymbol{A}_0(0)$。换句话说,有 $\mathrm{Row}(\boldsymbol{A}(0))\subset\mathrm{Row}(\boldsymbol{A}_0(0))$。

下面简要讨论一下上述定义。第 1 项表明 $\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$确实是 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$的一个真子空间;第 2 项意味着 $\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$的维数在 $\varepsilon\in[0,\bar{\varepsilon}]$上保持不变,并且在 $\varepsilon=0$ 处是连续的。值得注意的是,当 $\varepsilon\to 0^+$ 时,在间隔度量意义下,$\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$收敛到 $\mathrm{Row}(\boldsymbol{A}_0(0))$。

引理 3 设 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 满足假设 4。若 $\boldsymbol{A}(\varepsilon)$是渐近奇异的,则对所有 $\varepsilon\in[0,\bar{\varepsilon}]$,有 $\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))\in\mathrm{Grass}(r_0,n)$,且 $\lim\limits_{\varepsilon\to 0^+}\Delta(\mathrm{Row}(\boldsymbol{A}_0(\varepsilon)),\mathrm{Row}(\boldsymbol{A}_0(0)))=0$,其中 $\Delta(\cdot,\cdot)$表示间隔度量(参见定义 1)。

证明 将引理 1 应用于 $\boldsymbol{A}_0:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$,即得证。 证毕

然而,对于整个行空间 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$,由于定义 2 中第 3 项的包含关系,在 $\varepsilon=0$ 处发生了维数缩减。如果考虑通过行向量对 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$进行分解,那么当 $\varepsilon\to 0^+$ 时,这种分解是不稳定的。例如,设 $\boldsymbol{A}(\varepsilon)$由式(6)给出,它对所有 $\varepsilon>0$ 都有 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))=\mathbf{R}^2$。取 $\boldsymbol{A}_0(\varepsilon)=[1,-1]$和 $\boldsymbol{A}_1(\varepsilon)=[1+\varepsilon,-1+\varepsilon]$,显然对所有 $\varepsilon\geqslant 0$ 有 $\dim\mathrm{Row}(\boldsymbol{A}(\varepsilon))=1$。因此,根据定义 2 知,$\boldsymbol{A}(\varepsilon)$是渐近奇异的。

3.2 渐近核与渐近对偶核

接着引入渐近奇异矩阵 $\boldsymbol{A}(\varepsilon)$的渐近核,这对从 SSC 框架发展鲁棒迭代方法至关重要。

定义 3 设 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 是一个满足假设 4 的矩阵值函数 $\boldsymbol{A}(\varepsilon)$的渐近核定义为:

$$\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon)):=\mathrm{Ker}(\boldsymbol{A}_0(\varepsilon))\cap\mathrm{Row}(\boldsymbol{A}(\varepsilon))=\{\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}:\boldsymbol{A}_0(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}=\mathbf{0},\boldsymbol{y}\in\mathbf{R}^m\}。$$

$\boldsymbol{A}(\varepsilon)$的渐近对偶核定义为:

$$\widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon)):=\mathrm{Ker}(\boldsymbol{A}_0(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))=\{\boldsymbol{y}\in\mathbf{R}^m:\boldsymbol{A}_0(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}=\mathbf{0}\}。$$

注 1 易证得,渐近 $\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))$实际上是 $\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$在 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$中的正交补。此外 $\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))$是渐近对偶核 $\widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon))$在映射 $\boldsymbol{A}^{\mathrm{T}}(\varepsilon):\mathbf{R}^m\to\mathbf{R}^n$ 下的像。由此可得 $\dim\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))=\dim\widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon))$。当 $\boldsymbol{A}(\varepsilon)$

满行秩时，$\dim \mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))=\dim \widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon))$等号成立。

注 2　设 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 是一个满足假设 4 的矩阵值函数。若 $\boldsymbol{A}(\varepsilon)$是渐近奇异的，则有正交直和：

$$\mathrm{Row}(\boldsymbol{A}(\varepsilon))=\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))\oplus \mathrm{Row}(\boldsymbol{A}_0(\varepsilon)),$$

且存在 $\varepsilon_1\in(0,\bar{\varepsilon}]$使得：

$$\sigma_0(\varepsilon)=\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0))\cos^2\theta_1(\varepsilon)-\|\boldsymbol{A}_0(0)-\boldsymbol{A}_0(\varepsilon)\|^2>0,\ \forall\varepsilon\in[0,\varepsilon_1],$$

其中：$\theta_1(\varepsilon)\in[0,\pi/2]$表示 $\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$与 $\mathrm{Row}(\boldsymbol{A}_0(0))$之间的最大典范角。此外，还可以证明：

$$\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(\varepsilon)\boldsymbol{A}_0(\varepsilon))\geqslant\sigma_0(\varepsilon)>0$$

以及

$$\lim_{\varepsilon\to0^+}\sigma_0(\varepsilon)=\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0))。$$

以上结论可以表明 $\lambda_{\min}^{+}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{A}(\varepsilon))$被 $\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0))$的一个小扰动所界定，而 $\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0))$与 ε 无关。这对证明本文提出的方法的鲁棒收敛速度至关重要。

最后提供一个例子加以说明。

例 1　考虑：

$$\boldsymbol{A}(\varepsilon)=\begin{bmatrix}1 & -2\\ 1+\varepsilon & -2+\varepsilon\\ 2 & -4\end{bmatrix},\text{其中 }\boldsymbol{A}_0(\varepsilon)=[1,-2],\boldsymbol{A}_1(\varepsilon)=\begin{bmatrix}1+\varepsilon & -2+\varepsilon\\ 2 & -4\end{bmatrix}。$$

显然 $\boldsymbol{A}(\varepsilon)$是渐近奇异的。直接计算得：

$$\widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon))=\mathrm{span}\{[\varepsilon-5,5,0]^{\mathrm{T}},[-2,0,1]^{\mathrm{T}}\},\mathrm{Ker}_\varepsilon\boldsymbol{A}(\varepsilon)=\mathrm{span}\{[2,1]^{\mathrm{T}}\}。$$

则由式(7)可得：

$$\lambda_{\min}^{+}(\boldsymbol{A}_0^{\mathrm{T}}(\varepsilon)\boldsymbol{A}_0(\varepsilon))=2,\lambda_{\min}^{+}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{A}(\varepsilon))=O(\varepsilon^2)。$$

注意，$\boldsymbol{A}(\varepsilon)$不满行秩，且 $\dim \mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))<\dim \widehat{\mathrm{Ker}}_\varepsilon(\boldsymbol{A}(\varepsilon))$。

4　鲁棒 Kaczmarz 算法

本节关注渐近奇异线性方程组式(5)，其中 $\boldsymbol{b}\in\mathrm{Im}(\boldsymbol{A}(\varepsilon))$且 $\boldsymbol{A}:[0,\bar{\varepsilon}]\to\mathbf{R}^{m\times n}$ 满足假设 4 和定义 2。求解式(5)的经典 Kaczmarz 方法为[19]：

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\omega\frac{\boldsymbol{b}_i-\boldsymbol{A}_{(i)}(\varepsilon)\boldsymbol{x}_k}{\|\boldsymbol{A}_{(i)}(\varepsilon)\|^2}\boldsymbol{A}_{(i)}^{\mathrm{T}}(\varepsilon),\tag{10}$$

其中：$\omega>0$ 为松弛参数。由 $b\in\mathrm{Im}(\boldsymbol{A}(\varepsilon))$可得式(5)是相容的。此时通解为 $\boldsymbol{x}^*=\boldsymbol{x}_{\mathrm{LS}}+\hat{\boldsymbol{x}}$，这里 $\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^+\boldsymbol{b}$ 且 $\hat{\boldsymbol{x}}\in\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))$。已知，若 $x_0\in\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))$，则经典 Kaczmarz 方法收敛到 $\boldsymbol{x}_{\mathrm{LS}}$[14]。为了克服渐近奇异性，受文献[14,18]的启发，本节将渐近核 $\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))$(参见定义 3)添加到式(10)的 SSC 设定中：

$$\begin{cases}\boldsymbol{V}:=\mathrm{Row}(\boldsymbol{A}(\varepsilon)),\text{配备 Euclid 内积 }(\boldsymbol{x},\boldsymbol{y}):=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y};\\ \boldsymbol{A}:=\boldsymbol{I}:\boldsymbol{V}\to\boldsymbol{V},\text{是恒等算子，且 }\boldsymbol{f}:=\boldsymbol{x}_{\mathrm{LS}};\\ \boldsymbol{V}=\sum_{i=1}^{m+1}\boldsymbol{V}_i,\text{其中 }\boldsymbol{V}_i:=\mathbf{R}(1\leqslant i\leqslant m),\boldsymbol{V}_{m+1}:=\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon));\\ \boldsymbol{R}_i:=\boldsymbol{A}_{(i)}^{\mathrm{T}}(\varepsilon):\boldsymbol{V}_i\to\boldsymbol{V}(1\leqslant i\leqslant m),\boldsymbol{R}_{m+1}=\boldsymbol{\iota}:\boldsymbol{V}_{m+1}\to\boldsymbol{V};\\ a_i(x_i,y_i)=d_ix_iy_i,\text{其中 }d_i=\|\boldsymbol{A}_{(i)}(\varepsilon)\|^2(1\leqslant i\leqslant m),a_{m+1}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}。\end{cases}\tag{11}$$

基于如上设定，很容易验证假设 1 和假设 2 成立。另外，若去掉上式的最后一行以及 $\boldsymbol{R}_{m+1}$ 的设定，则可以得到经典 Kaczmarz 算法的 SSC 表达。容易发现：

$$\boldsymbol{T}_i=\boldsymbol{R}_i^*=\boldsymbol{A}_{(i)}(\varepsilon)/d_i(1\leqslant i\leqslant m)。$$

同时，还可以得到正交投影：

$$\boldsymbol{T}_{m+1}=\boldsymbol{R}_{m+1}^*:=\boldsymbol{R}:\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Ker}_\varepsilon(\boldsymbol{A}(\varepsilon))。$$

此时,对所有 $1\leqslant i\leqslant m+1$,都有 $\frac{2}{\omega}\boldsymbol{I}-\boldsymbol{T}_i\boldsymbol{R}_i=\frac{2}{\omega}-1>0$,从而有假设 3 成立。这导出如下 SSC 表述:给定 $\boldsymbol{x}_k\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))$,令 $\boldsymbol{v}_{k,1}=\boldsymbol{x}_k$,通过下式更新 $\boldsymbol{x}_{k+1}=\boldsymbol{v}_{k,m+2}$:

$$\begin{cases}\boldsymbol{v}_{k,i+1}=\boldsymbol{v}_{k,i}+\omega\dfrac{\boldsymbol{b}_i-\boldsymbol{A}_{(i)}(\varepsilon)\boldsymbol{v}_{k,i}}{\|\boldsymbol{A}_{(i)}(\varepsilon)\|^2}\boldsymbol{A}_{(i)}^{\mathrm{T}}(\varepsilon),1\leqslant i\leqslant m,\\ \boldsymbol{v}_{k,m+2}=\boldsymbol{v}_{k,m+1}+\omega\boldsymbol{R}(\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{v}_{k,m+1}),\end{cases}\tag{12}$$

其中:$\omega\in(0,2)$。作为 Kaczmarz 方法的推广,上述方法被称为鲁棒 Kaczmarz 方法。它等价于定常迭代:

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{B}_1(\omega)(\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{x}_k),$$

不加核的经典 Kaczmarz 方法显然等价于定常迭代形式:

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{B}_0(\omega)(\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{x}_k),\tag{13}$$

其中,迭代算子 $\boldsymbol{B}_0(\omega):\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon)$满足:

$$\boldsymbol{I}-\boldsymbol{B}_0(\omega)\boldsymbol{A}=(\boldsymbol{I}-\omega\boldsymbol{R}_J\boldsymbol{T}_J)\cdot\cdots\cdot(\boldsymbol{I}-\omega R_1T_1);$$

迭代算子 $\boldsymbol{B}_1(\omega):\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon))$满足:

$$\boldsymbol{I}-\boldsymbol{B}_1(\omega)=(\boldsymbol{I}-\omega\boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*)(\boldsymbol{I}-\boldsymbol{B}_0(\omega))。$$

基于引理 2,得到一个为证明鲁棒 Kaczmarz 方法的收敛速度奠定了基础的重要估计。在设定式(11)下,乘积空间 $\widetilde{\boldsymbol{V}}=\mathbf{R}^m$ 具有内积$(\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}:=(\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{D}}_\varepsilon}$。给出如下定义:

$$\delta_{\max}(\boldsymbol{A}(\varepsilon)):=\sup_{\substack{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\\ \|\boldsymbol{v}\|=1}}(\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\mathcal{R}^{\mathrm{a}}\boldsymbol{v})_{\widetilde{\boldsymbol{V}}},\delta_{\min}(\boldsymbol{A}(\varepsilon)):=\inf_{\substack{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\\ \|\boldsymbol{v}\|=1}}(\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\mathcal{R}^{\mathrm{a}}\boldsymbol{v})_{\widetilde{\boldsymbol{V}}}。$$

于是有[14]:

$$\delta_{\min}(\boldsymbol{A}(\varepsilon))=\lambda_{\min}^{+}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)),\delta_{\max}(\boldsymbol{A}(\varepsilon))=\lambda_{\max}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)),$$

其中:$\boldsymbol{D}_\varepsilon=\mathrm{diag}(d_1,\cdots,d_m)=\mathrm{diag}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))$,显然有:

$$\delta_{\min}(\boldsymbol{A}(\varepsilon))=\lambda_{\min}^{+}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon))\leqslant\frac{1}{\eta_0}\lambda_{\min}^{+}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{A}(\varepsilon))。$$

引理 4 设定式(11)下,若 $\omega\in(0,2/(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))$,则对于式(12)有:

$$\|\boldsymbol{I}-\boldsymbol{B}_1(\omega)\|_{\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon))}^2\leqslant1-C_0(\varepsilon,\omega)\inf_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\frac{(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\boldsymbol{v})}{(\boldsymbol{v},\boldsymbol{v})},\tag{14}$$

其中:

$$C_0(\varepsilon,\omega):=\frac{4\omega(2-\omega(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))}{(2+\omega\lfloor\log_2(2m)\rfloor\delta_{\max}(\boldsymbol{A}(\varepsilon))+2\omega\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\|)^2}。\tag{15}$$

证明 显然有:

$$\mathcal{R}=[\boldsymbol{A}^{\mathrm{T}}(\varepsilon),\boldsymbol{I}],\mathcal{R}^{\mathrm{a}}=\mathcal{R}^{\bigstar}=\begin{bmatrix}\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\\ \boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\end{bmatrix},\widetilde{\boldsymbol{A}}=\begin{bmatrix}\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon) & \boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\\ \boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\boldsymbol{A}^{\mathrm{T}}(\varepsilon) & \boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\end{bmatrix}。\tag{16}$$

令$\widetilde{\boldsymbol{D}}_0:=\mathrm{diag}(\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))$,$\widetilde{\boldsymbol{U}}_0:=\mathrm{triu}(\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))$。随后,对$\widetilde{\boldsymbol{D}}$ 和 $\boldsymbol{U}$ 进行如下分解:

$$\widetilde{\boldsymbol{D}}=\begin{bmatrix}\widetilde{\boldsymbol{D}}_0 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\end{bmatrix},\widetilde{\boldsymbol{U}}=\begin{bmatrix}\widetilde{\boldsymbol{U}}_0 & \boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}。$$

令$\widetilde{\boldsymbol{V}}_0=\mathbf{R}^m$,对于任意 $\widetilde{\boldsymbol{u}}=(\widetilde{\boldsymbol{u}}_0,\boldsymbol{u})\in\widetilde{\boldsymbol{V}}$,$\widetilde{\boldsymbol{v}}=(\widetilde{\boldsymbol{v}}_0,\boldsymbol{v})\in\widetilde{\boldsymbol{V}}$,乘积空间$\widetilde{\boldsymbol{V}}=\widetilde{\boldsymbol{V}}_0\times\boldsymbol{V}_{m+1}$ 具有内积$(\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{v}})_{\widetilde{\boldsymbol{V}}}:=(\widetilde{\boldsymbol{u}}_0,\widetilde{\boldsymbol{v}}_0)_{\boldsymbol{D}_\varepsilon}+(\boldsymbol{u},\boldsymbol{v})$。之前已经提到过:$\boldsymbol{R}_{m+1}=\boldsymbol{\iota}:\boldsymbol{V}_{m+1}\to\boldsymbol{V}$,且对所有 $\boldsymbol{v}\in\boldsymbol{V}_{m+1}$ 有 $\boldsymbol{T}_{m+1}\boldsymbol{R}_{m+1}\boldsymbol{v}=\boldsymbol{T}_{m+1}\boldsymbol{v}$。

由文献[14]知$\|\widetilde{\boldsymbol{D}}_0\|_{\widetilde{\boldsymbol{V}}_0\to\widetilde{\boldsymbol{V}}_0}\leqslant\delta_{\max}(\boldsymbol{A}(\varepsilon))$。因此:

$$(\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{v}})_{\boldsymbol{V}}=(\widetilde{\boldsymbol{D}}_0\widetilde{\boldsymbol{v}}_0,\boldsymbol{D}_0\widetilde{\boldsymbol{v}}_0)_{\boldsymbol{V}_0}+a_{m+1}(\boldsymbol{T}_{m+1}\boldsymbol{v},\boldsymbol{T}_{m+1}\boldsymbol{v})\leqslant\delta_{\max}^2(\boldsymbol{A}(\varepsilon))\|\widetilde{\boldsymbol{v}}_0\|_{\widetilde{\boldsymbol{V}}_0}^2+\|\boldsymbol{T}_{m+1}\boldsymbol{v}\|^2\leqslant$$
$$(1+\delta_{\max}^2(\boldsymbol{A}(\varepsilon)))(\|\widetilde{\boldsymbol{v}}_0\|_{\widetilde{\boldsymbol{V}}_0}^2+\|\boldsymbol{R}_{m+1}^*\boldsymbol{v}\|^2)=(1+\delta_{\max}^2(\boldsymbol{A}(\varepsilon)))\|\widetilde{\boldsymbol{v}}\|_{\widetilde{\boldsymbol{V}}}^2,$$

这表明$\|\widetilde{\boldsymbol{D}}\|_{\widetilde{\boldsymbol{V}}\to\widetilde{\boldsymbol{V}}}\leqslant1+\delta_{\max}(\boldsymbol{A}(\varepsilon))$。由于 $\omega\in(0,2/(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))$,$2\widetilde{\boldsymbol{I}}-\omega\widetilde{\boldsymbol{D}}$ 是 SPD 的,并且由引理 2 可得:

$$\|\boldsymbol{I}-\boldsymbol{B}_1(\omega)\|^2_{\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon))}=1-\frac{1}{c_0(\omega)},$$

其中：

$$c_0(\omega)=\frac{1}{\omega}\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\inf_{\mathcal{R}\tilde{\boldsymbol{v}}=\boldsymbol{v}}\frac{((2\tilde{\boldsymbol{I}}-\omega\tilde{\boldsymbol{D}})^{-1}(\tilde{\boldsymbol{I}}+\omega\tilde{\boldsymbol{U}})\tilde{\boldsymbol{v}},(\tilde{\boldsymbol{I}}+\omega\tilde{\boldsymbol{U}})\tilde{\boldsymbol{v}})_{\tilde{\boldsymbol{V}}}}{(\boldsymbol{v},\boldsymbol{v})}\leqslant$$

$$\frac{1}{\omega(2-\omega\|\tilde{\boldsymbol{D}}\|_{\tilde{\boldsymbol{V}}\to\tilde{\boldsymbol{V}}})}\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\inf_{\mathcal{R}\tilde{\boldsymbol{v}}=\boldsymbol{v}}\frac{((\tilde{\boldsymbol{I}}+\omega\tilde{\boldsymbol{U}})\tilde{\boldsymbol{v}},(\tilde{\boldsymbol{I}}+\omega\tilde{\boldsymbol{U}})\tilde{\boldsymbol{v}})_{\tilde{\boldsymbol{V}}}}{(\boldsymbol{v},\boldsymbol{v})}。\tag{17}$$

此外，注意到$\tilde{\boldsymbol{U}}_0'=\tilde{\boldsymbol{L}}_0'$ 以及$\|\tilde{\boldsymbol{L}}_0\|_{\tilde{\boldsymbol{V}}_0\to\tilde{\boldsymbol{V}}_0}\leqslant\lfloor\log_2(2m)\rfloor/2\delta_{\max}(\boldsymbol{A}(\varepsilon))$[14]，于是有：

$$\|\tilde{\boldsymbol{U}}\|^2_{\tilde{\boldsymbol{V}}\to\tilde{\boldsymbol{V}}}=\sup_{\substack{\tilde{\boldsymbol{v}}=(\tilde{\boldsymbol{v}}_0,\boldsymbol{v})\in\boldsymbol{V}\\ \|\tilde{\boldsymbol{v}}\|_{\tilde{\boldsymbol{V}}}=1}}\|\tilde{\boldsymbol{U}}\tilde{\boldsymbol{v}}\|^2_{\tilde{\boldsymbol{V}}}=\sup_{\substack{\tilde{\boldsymbol{v}}=(\tilde{\boldsymbol{v}}_0,\boldsymbol{v})\in\boldsymbol{V}\\ \|\tilde{\boldsymbol{v}}\|_{\tilde{\boldsymbol{V}}}=1}}(\|\tilde{\boldsymbol{U}}_0\tilde{\boldsymbol{v}}_0\|^2_{\tilde{\boldsymbol{V}}_0}+\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}\|^2)\leqslant$$

$$\sup_{\substack{\tilde{\boldsymbol{v}}=(\tilde{\boldsymbol{v}}_0,\boldsymbol{v})\in\boldsymbol{V}\\ \|\tilde{\boldsymbol{v}}\|_{\tilde{\boldsymbol{V}}}=1}}\left(\frac{\lfloor\log_2(2m)\rfloor^2}{4}\delta^2_{\max}(\boldsymbol{A}(\varepsilon))\|\tilde{\boldsymbol{v}}_0\|^2_{\tilde{\boldsymbol{V}}_0}+\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\|^2\|\boldsymbol{v}\|^2\right)\leqslant$$

$$\frac{\lfloor\log_2(2m)\rfloor^2}{4}\delta^2_{\max}(\boldsymbol{A}(\varepsilon))+\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\|^2,$$

从而可以推出：

$$\|\tilde{\boldsymbol{U}}\|_{\tilde{\boldsymbol{V}}\to\tilde{\boldsymbol{V}}}\leqslant\frac{1}{2}\lfloor\log_2(2m)\rfloor\delta_{\max}(\boldsymbol{A}(\varepsilon))+\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\|。$$

接下来，将上述估计代入式(17)得：

$$c_0(\omega)\leqslant\frac{(1+\omega\|\tilde{\boldsymbol{U}}\|_{\tilde{\boldsymbol{V}}\to\tilde{\boldsymbol{V}}})^2}{\omega(2-\omega\|\tilde{\boldsymbol{D}}\|_{\tilde{\boldsymbol{V}}\to\tilde{\boldsymbol{V}}})}\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\inf_{\mathcal{R}\tilde{\boldsymbol{v}}=\boldsymbol{v}}\frac{(\tilde{\boldsymbol{v}},\tilde{\boldsymbol{v}})_{\tilde{\boldsymbol{V}}}}{(\boldsymbol{v},\boldsymbol{v})}\leqslant\frac{1}{C_0(\varepsilon,\omega)}\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\inf_{\mathcal{R}\tilde{\boldsymbol{v}}=\boldsymbol{v}}\frac{(\tilde{\boldsymbol{v}},\tilde{\boldsymbol{v}})_{\tilde{\boldsymbol{V}}}}{(\boldsymbol{v},\boldsymbol{v})}。$$

由于$\mathcal{R}:\tilde{\boldsymbol{V}}\to\boldsymbol{V}$ 是满射且$\mathcal{R}^\star=\mathcal{R}^{\mathrm{a}}$。由文献[13]可得：

$$\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\inf_{\mathcal{R}\tilde{\boldsymbol{v}}=\boldsymbol{v}}\frac{(\tilde{\boldsymbol{v}},\tilde{\boldsymbol{v}})_{\tilde{\boldsymbol{V}}}}{(\boldsymbol{v},\boldsymbol{v})}=\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\frac{((\mathcal{R}\mathcal{R}^{\mathrm{a}})^{-1}\boldsymbol{v},\boldsymbol{v})}{(\boldsymbol{v},\boldsymbol{v})}=$$

$$\left(\inf_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\frac{(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\boldsymbol{v})}{(\boldsymbol{v},\boldsymbol{v})}\right)^{-1}=\left(\inf_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\frac{(\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\mathcal{R}\boldsymbol{v})_{\boldsymbol{V}}}{(\boldsymbol{v},\boldsymbol{v})}\right)^{-1}。$$

结合前一个不等式即得式(14)。　　证毕

鲁棒 Kaczmarz 方法的最终收敛结果由以下定理给出，它表明对小 ε 具有一致压缩常系数；更多讨论见注 3。

定理 1　设 $\omega\in(0,2/(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))$，并定义收缩因子：

$$\rho(\varepsilon,\omega):=C_0(\varepsilon,\omega)(\min(1,\sigma_0(\varepsilon)/\eta_1)-C_1(\varepsilon)/\eta_0),\tag{18}$$

其中：$C_0(\varepsilon,\omega)$由式(15)定义，而：

$$C_1(\varepsilon):=\|\boldsymbol{E}_1\|^2+2\|\boldsymbol{A}_1(0)\|\|\boldsymbol{E}_1\|+\|\boldsymbol{E}_0\|\|\boldsymbol{A}_1(0)\|\|\boldsymbol{\Pi}\|,$$

这里 $\boldsymbol{E}_0:=\boldsymbol{A}_0(\varepsilon)-\boldsymbol{A}_0(0)$，$\boldsymbol{E}_1:=\boldsymbol{A}_1(\varepsilon)-\boldsymbol{A}_1(0)$。对于式(12)有：

$$\|\boldsymbol{I}-\boldsymbol{B}_1(\omega)\|^2_{\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon))}\leqslant1-\rho(\varepsilon,\omega)。$$

证明　由式(14)可知，现只需证明：

$$\inf_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))\setminus\{\boldsymbol{0}\}}\frac{(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\boldsymbol{v})}{(\boldsymbol{v},\boldsymbol{v})}\geqslant\min(1,\sigma_0(\varepsilon)/\eta_1)-C_1(\varepsilon)/\eta_0\tag{19}$$

根据注 1，$\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))$中 $\mathrm{Row}(\boldsymbol{A}(\varepsilon))$的正交补即为渐近核 $\mathrm{Ker}(\boldsymbol{A}(\varepsilon))$。因此对任意 $\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))$有正交分解 $\boldsymbol{v}=\boldsymbol{v}_n+\boldsymbol{v}_r$，其中 $\boldsymbol{v}_n\in\mathrm{Ker}(\boldsymbol{A}(\varepsilon))$，$\boldsymbol{v}_r\in\mathrm{Ker}(\boldsymbol{A}(\varepsilon))$且$(\boldsymbol{v}_r,\boldsymbol{v}_n)=0$。于是$\|\boldsymbol{v}\|^2=\|\boldsymbol{v}_r\|^2+\|\boldsymbol{v}_n\|^2$ 且：

$$(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\boldsymbol{v})=(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_r)+2(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)+(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n,\boldsymbol{v}_n)=$$

$$(\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r)_{\tilde{\boldsymbol{V}}}+2(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)+(\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n,\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n)_{\tilde{\boldsymbol{V}}}=\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r\|^2_{\tilde{\boldsymbol{V}}}+2(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)+\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n\|^2_{\tilde{\boldsymbol{V}}}。$$

由式(9)、注 2 和式(16)得:

$$\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r\|_{\widetilde{\mathbf{V}}}^2=\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_r\|_{\boldsymbol{D}_\varepsilon}^2+\|\boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\boldsymbol{v}_r\|^2=\boldsymbol{v}_r^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_r\geqslant\frac{1}{\eta_1}\boldsymbol{v}_r^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{A}(\varepsilon)\boldsymbol{v}_r=$$

$$\frac{1}{\eta_1}\boldsymbol{v}_r^{\mathrm{T}}(\boldsymbol{A}_0^{\mathrm{T}}(\varepsilon)\boldsymbol{A}_0(\varepsilon)+\boldsymbol{A}_1^{\mathrm{T}}(\varepsilon)\boldsymbol{A}_1(\varepsilon))\boldsymbol{v}_r\geqslant\frac{\sigma_0(\varepsilon)}{\eta_1}\|\boldsymbol{v}_r\|^2。\tag{20}$$

这里用到 $\boldsymbol{v}_r\in\mathrm{Row}(\boldsymbol{A}_0(\varepsilon))\subset\mathrm{Row}(\boldsymbol{A}(\varepsilon))$,故 $\boldsymbol{R}_{m+1}^*\boldsymbol{v}_r=0$。类似地,

$$\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n\|_{\widetilde{\mathbf{V}}}^2=\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_n\|_{\boldsymbol{D}_\varepsilon}^2+\|\boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\boldsymbol{v}_n\|_{\boldsymbol{D}_\varepsilon}^2=\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_n\|_{\boldsymbol{D}_\varepsilon}^2+\|\boldsymbol{v}_n\|^2\geqslant\|\boldsymbol{v}_n\|^2。\tag{21}$$

交叉项为:

$$(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)=\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_r+(\boldsymbol{R}_{m+1}\boldsymbol{R}_{m+1}^*\boldsymbol{v}_r,\boldsymbol{v}_n)=\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\boldsymbol{v}_r。$$

考虑对 $\boldsymbol{D}_\varepsilon$ 的分块:$\boldsymbol{D}_\varepsilon=\mathrm{diag}(\boldsymbol{D}_0,\boldsymbol{D}_1)$,其中 $\boldsymbol{D}_0=\mathrm{diag}(d_1,\cdots,d_{m_0})$,$\boldsymbol{D}_1=\mathrm{diag}(d_{m_0+1},\cdots,d_m)$。由等式 $\boldsymbol{A}_0(\varepsilon)\boldsymbol{v}_n=0$ 可得:

$$(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)=\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(\varepsilon)\boldsymbol{v}_r=\boldsymbol{v}_n^{\mathrm{T}}(\boldsymbol{E}_1+\boldsymbol{A}_1(0))^{\mathrm{T}}\boldsymbol{D}_1^{-1}(\boldsymbol{E}_1+\boldsymbol{A}_1(0))\boldsymbol{v}_r=$$

$$\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}(0)\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(0)\boldsymbol{v}_r+\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{E}_1\boldsymbol{v}_r+\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(0)\boldsymbol{v}_r+\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}(0)\boldsymbol{D}_1^{-1}\boldsymbol{E}_1\boldsymbol{v}_r,$$

其中:$\boldsymbol{E}_1=\boldsymbol{A}_1(\varepsilon)-\boldsymbol{A}_1(0)$。注意到 $|\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{v}_r|\leqslant\frac{1}{2}(\|\boldsymbol{v}_r\|^2+\|\boldsymbol{v}_n\|^2)=\frac{1}{2}\|\boldsymbol{v}\|^2$,且:

$$\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{E}_1\boldsymbol{v}_r+\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(0)\boldsymbol{v}_r+\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}(0)\boldsymbol{D}_1^{-1}\boldsymbol{E}_1\boldsymbol{v}_r\geqslant$$

$$-\frac{|\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{v}_r|}{\eta_0}(\|\boldsymbol{E}_1\|^2+2\|\boldsymbol{E}_1\|\|\boldsymbol{A}_1(0)\|)\geqslant-\frac{\|\boldsymbol{v}\|^2}{2\eta_0}(\|\boldsymbol{E}_1\|^2+2\|\boldsymbol{E}_1\|\|\boldsymbol{A}_1(0)\|)。$$

另一方面,由定义 2 知 $\boldsymbol{A}_1(0)=\boldsymbol{\Pi}\boldsymbol{A}_0(0)=\boldsymbol{\Pi}(\boldsymbol{E}_0+\boldsymbol{A}_0(\varepsilon))$,其中 $\boldsymbol{E}_0=\boldsymbol{A}_0(0)-\boldsymbol{A}_0(\varepsilon)$。则:

$$\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}(0)\boldsymbol{D}_1^{-1}A_1(0)\boldsymbol{v}_r=\boldsymbol{v}_n^{\mathrm{T}}(\boldsymbol{E}_0+\boldsymbol{A}_0(\varepsilon))^{\mathrm{T}}\boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(0)\boldsymbol{v}_r=\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{E}_0^{\mathrm{T}}\boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_1^{-1}\boldsymbol{A}_1(0)\boldsymbol{v}_r\geqslant$$

$$-\frac{|\boldsymbol{v}_n^{\mathrm{T}}\boldsymbol{v}_r|}{\eta_0}\|\boldsymbol{E}_0\|\|\boldsymbol{A}_1(0)\|\|\boldsymbol{\Pi}\|\geqslant-\frac{\|\boldsymbol{v}\|^2}{2\eta_0}\|\boldsymbol{E}_0\|\|\boldsymbol{A}_1(0)\|\|\boldsymbol{\Pi}\|。$$

综上可得 $(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)\geqslant-\frac{C_1(\varepsilon)}{2\eta_0}\|\boldsymbol{v}\|^2$,结合式(20)、(21)得:

$$(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v},\boldsymbol{v})=\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r\|_{\mathbf{V}}^2+2(\mathcal{R}\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_r,\boldsymbol{v}_n)+\|\mathcal{R}^{\mathrm{a}}\boldsymbol{v}_n\|_{\mathbf{V}}^2\geqslant$$

$$\frac{1}{\eta_1}\min(\eta_1,\sigma_0(\varepsilon))\|\boldsymbol{v}\|^2-\frac{C_1(\varepsilon)}{\eta_0}\|\boldsymbol{v}\|^2=(\min(1,\sigma_0(\varepsilon)/\eta_1)-C_1(\varepsilon)/\eta_0)\|\boldsymbol{v}\|^2,$$

对任意 $\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))$ 成立。代入式(19)即得所需估计。 证毕

注 3 下面将详细解释定理 2 中给出的鲁棒估计。由假设 4、注 1 和注 2,有:

1) 当 $\varepsilon\to0^+$ 时 $C_1(\varepsilon)\to0$;

2) 对所有 $\varepsilon\in[0,\bar{\varepsilon}]$,有:

$\|\boldsymbol{A}(\varepsilon)\|\leqslant C_2<\infty$,$\|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon)\|\leqslant\|\boldsymbol{A}(\varepsilon)\|/\eta_0$,$\delta_{\max}(\boldsymbol{A}(\varepsilon))=\lambda_{\max}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{A}(\varepsilon))\leqslant\|\boldsymbol{A}(\varepsilon)\|^2/\eta_0$。

3) 对所有 $\varepsilon\in[0,\varepsilon_1]$,$\sigma_0(\varepsilon)>0$,且 $\lim\limits_{\varepsilon\to0^+}\sigma_0(\varepsilon)=\lambda_{\min}^+(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0))>0$。

因此,存在某个 $\varepsilon_2\in(0,\bar{\varepsilon}]$,使得对所有 $\varepsilon\in[0,\varepsilon_2]$ 有:

$$\sigma_0(\varepsilon)\geqslant\frac{1}{2}\lambda_{\min}^+(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0)),C_1(\varepsilon)\leqslant\frac{\eta_0}{2}\min(1,\sigma_0(\varepsilon)/\eta_1)。$$

这意味着收缩因子式(18)有一个一致的正下界:对所有 $\varepsilon\in[0,\varepsilon_2]$,有 $0<\rho_0(\omega)\leqslant\rho(\varepsilon,\omega)\leqslant1$,其中:

$$\rho_0(\omega)=\frac{\omega(2-\omega(1+C_2^2/\eta_0))\min(2\eta_1,\lambda_{\min}^+(\boldsymbol{A}_0^{\mathrm{T}}(0)\boldsymbol{A}_0(0)))}{\eta_1(2+\omega\lfloor\log_2(2m)\rfloor C_2^2/\eta_0+2\omega C_2/\sqrt{\eta_0})^2}。$$

5 鲁棒 CD 法

前一节展示了鲁棒 Kaczmarz 算法对渐近奇异方程组的有效性。然而,该方法需要先验信息 $\boldsymbol{x}_{\mathrm{LS}}$。为了克服这一局限,本文将 Kaczmarz 算法求解原来的渐近奇异线性方程组转化为用 CD 法求解对偶问题式(3)。该问题等价于 $\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}=-\boldsymbol{b}$。前面已经提到过,用 Kaczmarz 求解式(5)等价于用 CD 法求解式(3)。又注意到

Kaczmarz 方法本身可以由 SSC 表述，由此，本文尝试通过用 SSC 框架来分析相应的鲁棒坐标下降方法。

基于 Kaczmarz 和 CD 方法的等价性，遵循 SSC 法的思想，考虑对 CD 法的 SSC 设定中直接加入渐近对偶核空间，并由此得到：

$$\begin{cases} \boldsymbol{V} := \mathbf{R}^m，配备\ \text{Euclid}\ 内积\ (\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}; \\ \boldsymbol{A} := \boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon): \boldsymbol{V} \to \boldsymbol{V}, \boldsymbol{f} := -\boldsymbol{b}; \\ \boldsymbol{V} = \sum_{i=1}^{m+1} \boldsymbol{V}_i，其中\ \boldsymbol{V}_i := \mathbf{R},\ 1 \leqslant i \leqslant m, \boldsymbol{V}_{m+1} := \widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon)); \\ \boldsymbol{R}_i := \boldsymbol{U}_i: \boldsymbol{V}_i \to \boldsymbol{V}(1 \leqslant i \leqslant m), \boldsymbol{R}_{m+1} := \boldsymbol{\iota}: \boldsymbol{V}_{m+1} \to \boldsymbol{V}; \\ a_i(x_i, y_i) = d_i x_i y_i, d_i = \|\boldsymbol{A}_{(i)}(\varepsilon)\|^2 > 0, 1 \leqslant i \leqslant m; \\ \boldsymbol{V}_{m+1}\ 上的内积定义为\ a_{m+1}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}。 \end{cases} \tag{22}$$

由此容易得到：

$$\boldsymbol{T}_i = \boldsymbol{R}_i^* \boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon) = \boldsymbol{U}_i \boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)/\boldsymbol{d}_i, 1 \leqslant i \leqslant m,$$

而 $\boldsymbol{R}_{m+1}^* := \hat{\boldsymbol{R}}: \mathbf{R}^m \to \widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 为正交投影，这就导出另一种鲁棒 SSC 形式：给定 $\boldsymbol{y}_k \in \mathbf{R}^m$，$\boldsymbol{u}_{k,1} = \boldsymbol{y}_k$，通过下式更新 $\boldsymbol{y}_{k+1} = \boldsymbol{u}_{k,m+2}$：

$$\begin{cases} \boldsymbol{u}_{k,i+1} = \boldsymbol{u}_{k,i} - \omega \dfrac{\boldsymbol{U}_i \boldsymbol{U}_i^{\mathrm{T}}}{\|\boldsymbol{A}_{(i)}(\varepsilon)\|^2}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,i} + \boldsymbol{b}), 1 \leqslant i \leqslant m; \\ \boldsymbol{u}_{k,m+2} = \boldsymbol{u}_{k,m+1} - \omega \hat{\boldsymbol{R}}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,i} + \boldsymbol{b})。 \end{cases} \tag{23}$$

这被称为鲁棒 CD 方法。注意，有定常迭代形式：

$$\boldsymbol{y}_{k+1} = \boldsymbol{y}_k - \boldsymbol{B}_2(\omega)(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}_k + \boldsymbol{b}) \tag{24}$$

其中 $\boldsymbol{B}_2(\omega): \mathbf{R}^m \to \mathbf{R}^m$ 满足：

$$\boldsymbol{I} - \boldsymbol{B}_2(\omega) = (\boldsymbol{I} - \omega \boldsymbol{R}_{m+1}\boldsymbol{T}_{m+1})(\boldsymbol{I} - \omega \boldsymbol{R}_m \boldsymbol{T}_m) \cdot \cdots \cdot (\boldsymbol{I} - \omega \boldsymbol{R}_1 \boldsymbol{T}_1)。$$

下面给出鲁棒 CD 方法的具体内容。

算法 1　用于求解式(3)的鲁棒 CD 算法

步骤 0，设置松弛参数：$\omega \in (0, 2/(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))$ 和 $\boldsymbol{y}_k \in \mathbf{R}^m$，正交投影：$\hat{\boldsymbol{R}}: \mathbf{R}^m \to \widehat{\mathrm{Ker}}(\boldsymbol{A}(\varepsilon))$。

步骤 1，令 $\boldsymbol{v} = \boldsymbol{y}_k$。

步骤 2，计算 $\boldsymbol{v} = \boldsymbol{v} - \omega \dfrac{\boldsymbol{U}_i \boldsymbol{U}_i^{\mathrm{T}}}{d_i}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{v} + \boldsymbol{b})$，$i = 1, \cdots, m$。

步骤 3，计算 $\boldsymbol{v} = \boldsymbol{v} - \omega \hat{\boldsymbol{R}}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{v} + \boldsymbol{b})$。

步骤 4，令 $\boldsymbol{y}_{k+1} = \boldsymbol{v}$。

定理 2　设 $\boldsymbol{A}(\varepsilon)$ 满行秩，松弛参数满足 $\omega \in (0, 2/(1+\delta_{\max}(\boldsymbol{A}(\varepsilon))))$。设 $\{x_k\}_{k \geqslant 0}$ 由鲁棒 Kaczmarz 方法生成，$\{y_k\}_{k \geqslant 0}$ 由鲁棒 CD 方法生成，且 $x_0 = -\boldsymbol{A}^{\mathrm{T}}(\varepsilon)y_0$。则有如下结论：

1) 在设定式(22)下的鲁棒 CD 方法等价于在设定式(11)下的鲁棒 Kaczmarz 方法，即对所有：

$$k \geqslant 0\ 有\ \boldsymbol{x}_k = -\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}_k。$$

2)
$$\boldsymbol{B}_1(\omega) = \boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon),$$

$$\|\boldsymbol{I} - \boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\|^2_{\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)} = \|\boldsymbol{I} - \boldsymbol{B}_1(\omega)\|^2_{\mathrm{Row}(\boldsymbol{A}(\varepsilon)) \to \mathrm{Row}(\boldsymbol{A}(\varepsilon))} \leqslant 1 - \rho(\varepsilon, \omega)。 \tag{25}$$

其中：$\rho(\varepsilon, \omega)$ 由式(18)定义。

证明　首先证明关系式 $\boldsymbol{R} = \boldsymbol{A}^{\mathrm{T}}(\varepsilon)\hat{\boldsymbol{R}}\boldsymbol{A}(\varepsilon)$，其中 $\boldsymbol{R}: \mathrm{Row}(\boldsymbol{A}(\varepsilon)) \to \mathrm{Ker}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 和 $\hat{\boldsymbol{R}}: \mathbf{R}^m \to \widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 均为正交投影。由定义 2 知，$\mathrm{Ker}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 是 $\widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 在线性映射 $\boldsymbol{A}^{\mathrm{T}}(\varepsilon): \mathbf{R}^m \to \mathbf{R}^n$ 下的像。由于 $\boldsymbol{A}(\varepsilon)$ 满行秩，可得：

$$\dim \mathrm{Ker}_{\varepsilon}(\boldsymbol{A}(\varepsilon)) = \dim \widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon)) = r_1。$$

设 $\boldsymbol{S} \in \mathbf{R}^{m \times r_1}$ 的列构成 $\widehat{\mathrm{Ker}}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$ 的一组基，则：

$$\hat{\boldsymbol{R}}=\boldsymbol{S}(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}。$$

注意 $\boldsymbol{W}=\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{S}\in\mathbf{R}^{n\times r_1}$ 的列构成 $\mathrm{Ker}_{\varepsilon}(\boldsymbol{A}(\varepsilon))$的一组基,因此有:

$$\boldsymbol{R}=\boldsymbol{W}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})^{-1}\boldsymbol{W}^{\mathrm{T}}=\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{S}(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}(\varepsilon)=\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\hat{\boldsymbol{R}}\boldsymbol{A}(\varepsilon)。$$

下面逐一验证以上结论。

假设 $\boldsymbol{x}_k=-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}_k$,将式(23)与鲁棒 Kaczmarz 方法比较,立即得到对所有 $2\leqslant i\leqslant m+1$ 有 $\boldsymbol{v}_{k,i}=-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,i}$。因此,再次利用式(23),可得(注意 $\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{b}$):

$$-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,m+2}=-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,m+1}+\omega\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\hat{\boldsymbol{R}}(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,m+1}+\boldsymbol{b})=$$

$$\boldsymbol{v}_{k,m+1}+\omega\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\hat{\boldsymbol{R}}\boldsymbol{A}(\varepsilon)(\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{v}_{k,m+1})=\boldsymbol{v}_{k,m+1}+\omega(\boldsymbol{R}\boldsymbol{x}_{\mathrm{LS}}-\boldsymbol{v}_{k,m+1}),$$

这表明 $\boldsymbol{x}_{k+1}=\boldsymbol{v}_{k,m+2}=-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}_{k,m+2}=-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}_{k+1}$。这就验证了结论 1)。

根据式(13)和(24),很容易推导出 $\boldsymbol{B}_1(\omega)=\boldsymbol{A}^{\mathrm{T}}(\omega)\boldsymbol{B}_2(\omega)\boldsymbol{A}(\omega)$。根据引理 2 可得:

$$\|\boldsymbol{I}-\boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\|^2_{\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)}=\sup_{\boldsymbol{u}\in\mathbf{R}^m}\frac{\|(\boldsymbol{I}-\boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon))\boldsymbol{u}\|^2_{\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)}}{\|\boldsymbol{u}\|^2_{\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)}}=$$

$$\sup_{\boldsymbol{u}\in\mathbf{R}^m}\frac{\|(\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon))\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}\|^2}{\|\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{u}\|^2}=\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))}\frac{\|(\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{B}_2(\omega)\boldsymbol{A}(\varepsilon))\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}=$$

$$\sup_{\boldsymbol{v}\in\mathrm{Row}(\boldsymbol{A}(\varepsilon))}\frac{\|(\boldsymbol{I}-\boldsymbol{B}_1(\omega))\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}=\|\boldsymbol{I}-\boldsymbol{B}_1(\omega)\|^2_{\mathrm{Row}(\boldsymbol{A}(\varepsilon))\to\mathrm{Row}(\boldsymbol{A}(\varepsilon))}\leqslant 1-\rho(\varepsilon,\omega),$$

这证明了式(25)并验证了结论 2)。 证毕

6 数值结果

本节提供了一些数值测试来验证鲁棒 CD 方法的性能。由于鲁棒 Kaczmarz 方法需要 $\boldsymbol{x}_{\mathrm{LS}}$,所以该方法不实用。但鲁棒 CD 方法作为等价形式是可实现的,因二者具有等价性,故不在此展现鲁棒 Kaczmarz 的实验结果。

在所有情况下,矩阵 $\boldsymbol{A}(\varepsilon)$ 相对于小参数 ε 是渐近奇异的。松弛参数选择如下:对于 CD 方法,$\omega=\dfrac{0.9\times 2}{\delta_{\max}(\boldsymbol{A}(\varepsilon))}$;对于鲁棒 CD 方法,$\omega=\dfrac{0.9\times 2}{1+\delta_{\max}(\boldsymbol{A}(\varepsilon))}$。对于 ACDM 和 ARK,凸性参数相同,为:

$$\delta_{\min}(\boldsymbol{A}(\varepsilon))=\lambda^{+}_{\min}(\boldsymbol{A}^{\mathrm{T}}(\varepsilon)D_{\varepsilon}^{-1}\boldsymbol{A}(\varepsilon))。$$

对于所有方法,初始值为零向量,相对停止准则为:坐标下降型方法满足 $\|\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{T}}(\varepsilon)\boldsymbol{y}_k+\boldsymbol{b}\|/\|\boldsymbol{b}\|<10^{-6}$,Kaczmarz 类型方法满足 $\|\boldsymbol{A}(\varepsilon)\boldsymbol{x}_k-\boldsymbol{b}\|/\|\boldsymbol{b}\|<10^{-6}$。

例 2 考虑如下的稠密系数矩阵:

$$\boldsymbol{A}(\varepsilon)=\begin{bmatrix}1.271\,1 & \varepsilon & \varepsilon & \varepsilon\\ 1.5 & 1 & 0 & 0\\ 1.512\,4 & \varepsilon & \varepsilon & 0\end{bmatrix},$$

其中第 1 列为随机生成。当 $\varepsilon\to 0$ 时,$\boldsymbol{A}(\varepsilon)$的行秩趋近于 2。采用 $\varepsilon=1/3^k$ 进行测试,其中 $k=1,2,3,4$。表 2 呈现了 CD 和鲁棒 CD 方法对于这个渐近奇异问题的表现结果。这无疑验证了后者的鲁棒性。

表 2 CD 方法和鲁棒 CD 方法的比较结果

Tab. 2 Comparison of results between CD and robust CD

方法	ε			
	1/3	1/9	1/27	1/81
CD	1 396	5 486	29 429	143 003
鲁棒 CD	44	40	29	39

例 3 从 suite sparse 矩阵集合[20]中选择一个样本矩阵。从一个 453×453 的矩阵开始,提取一个满行秩子矩阵 $\boldsymbol{A}\in\mathbf{R}^{40\times 453}$。为了生成具有可调秩亏的测试问题,通过 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ 的奇异值分解定义参数化矩阵 $\boldsymbol{A}(\varepsilon)$:

$$\boldsymbol{A}(\varepsilon)=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}},\text{其中 }\boldsymbol{\Sigma}(\varepsilon)_{ii}=\begin{cases}\varepsilon\cdot\boldsymbol{\Sigma}_{ii},\text{对于 5 个最小的奇异值；}\\ \boldsymbol{\Sigma}_{ii},\text{其他。}\end{cases}$$

设置 $\varepsilon=1/2^k$，$k=1,2,\cdots,5$。当 ε 趋于 0 时，$\boldsymbol{A}(\varepsilon)$的秩从 40 突然降至 35。为了详细比较，选择了以下 4 种算法：CD 法、ACDM[11]、RRK 方法[21]和 ARK 方法[12]，比较结果见表 3。

表 3　CD、鲁棒 CD、ARK、ACDM 和 RRK 的比较结果

Tab. 3　Comparison of results among CD, the robust CD, ARK, ACDM and RRK

ε	CD	鲁棒 CD	ARK	ACDM	RRK
1/2	2 350	**263**	610	833	1 196
1/4	8 714	**246**	1 235	2 274	4 083
1/8	18 849	**221**	1 882	5 621	15 945
1/16	63 405	**220**	3 965	9 421	50 629
1/32	135 834	**243**	8 455	13 032	232 809

表 3 展示了数值输出。由于 ARK 方法和 ACDM 是随机算法，它们的迭代次数已除以 40，即 $\boldsymbol{A}(\varepsilon)$的行数。但 RRK 方法本身计算的就是循环次数，所以不需要除以行数。结果表明：RRK、ARK、CD 和 ACDM 这 4 种标准方法对小奇异值引起的病态性高度敏感。相比之下，鲁棒 CD 方法表现出很强的鲁棒性。

7　结论

本文提出了用于求解渐近奇异线性方程组的鲁棒 Kaczmarz 方法。即使最小非零奇异值趋近于 0，该方法的收敛速度仍能保持一致稳定。本文利用著名的 XZ 等式证明了所提算法的一致收敛速率。基于 Kaczmarz 和 CD 之间的等价性，本文还提出了与鲁棒 Kaczmarz 方法等价的鲁棒 CD 方法。通过数值结果验证了所提方法相较于现有 Kaczmarz 方法和 CD 方法的性能。此外，算法的计算开销以及在大规模问题上的求解能力同样是衡量该算法性能的重要指标，限于篇幅未能在本文中充分展开，相关内容将在后续工作中进一步探讨。

参考文献：

[1]　A Ferreira I，A Acebrón J，Monteiro J. Survey of a class of iterative row-action methods：the Kaczmarz method[J]. Numerical Algorithms，2025，99 (4)：2097-2135.

[2]　Beister M，Kolditz D，Kalender W A. Iterative reconstruction methods in X-ray CT[J]. Physica Medica，2012，28 (2)：94-108.

[3]　Yi S Y. A new nonconforming mixed finite element method for linear elasticity[J]. Mathematical Models and Methods in Applied Sciences，2006，16 (7)：979-999.

[4]　KACZMARZ S. Angenäherte Auflösung von Systemen linearer Gleichungen[J]. Bulletin International de l' Académie Polonaise des Sciences et des Lettres，1937，35：355-357.

[5]　Strohmer T，Vershynin R. A randomized kaczmarz algorithm with exponential convergence[J]. Journal of Fourier Analysis and Applications，2009，15 (2)：262-278.

[6]　Demmel J W. The probability that a numerical analysis problem is difficult[J]. Mathematics of Computation，1988，50 (182)：449-480.

[7]　Zeng Y，Han D R，Su Y S，et al. Randomized Kaczmarz method with adaptive stepsizes for inconsistent linear systems[J]. Numerical Algorithms，2023，94 (3)：1403-1420.

[8]　Su Y S，Han D R，Zeng Y，et al. On greedy multi-step inertial randomized Kaczmarz method for solving linear systems [J]. Calcolo，2024，61 (4)：68.

[9]　郭璇，李蕊，殷俊锋. 基于二维搜索的加速替代超平面 Kaczmarz 方法[J]. 计算数学，2025，47 (3)：490-501.
Guo Xuan，Li Rui，Yin Junfeng. An accelerated surrogate hyperplane kaczmarz method based on two-dimensional search [J]. Mathematica Numerica Sinica，2025，47 (3)：490-501.

[10]　王泽，殷俊锋. 求解线性方程组稀疏解的稀疏贪婪随机 Kaczmarz 算法[J]. 同济大学学报（自然科学版），2021，49

(11): 封 2.

Wang Ze, Yin Junfeng. Sparse greedy randomized kaczmarz method for sparse solutions to linear equations[J]. Journal of Tongji University (Natural Science), 2021, 49 (11): inside front cover.

[11] Nesterov Y. Efficiency of coordinate descent methods on huge-scale optimization problems[J]. SIAM Journal on Optimization, 2012, 22 (2): 341-362.

[12] Liu J, Wright S. An accelerated randomized Kaczmarz algorithm[J]. Mathematics of Computation, 2016, 85 (297): 153-178.

[13] Xu J C, Zikatanov L. The method of alternating projections and the method of subspace corrections in Hilbert space[J]. Journal of the American Mathematical Society, 2002, 15 (3): 573-597.

[14] Oswald P, Zhou W Q. Convergence analysis for Kaczmarz-type methods in a Hilbert space framework[J]. Linear Algebra and Its Applications, 2015, 478: 131-161.

[15] Stewart G W, Sun J. Matrix Perturbation Theory[M]. Boston: Academic Press, 1990.

[16] Xu J C. The method of subspace corrections[J]. Journal of Computational and Applied Mathematics, 2001, 128 (1/2): 335-362.

[17] Xu J C. Iterative methods by space decomposition and subspace correction[J]. SIAM Review, 1992, 34 (4): 581-613.

[18] Lee Y T, Sidford A. Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems[C]// 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. Piscataway: IEEE, 2013: 147-156.

[19] Moorman J D, Tu T K, Molitor D, et al. Randomized kaczmarz with averaging[J]. BIT Numerical Mathematics, 2021, 61 (1): 337-359.

[20] Kolodziej S, Aznaveh M, Bullock M, et al. The SuiteSparse matrix collection website interface[J]. Journal of Open Source Software, 2019, 4 (35): 1244.

[21] Han D R, Xie J X. A simple linear convergence analysis of the randomized reshuffling Kaczmarz method [PP/OL]. V3. arXiv (2025-08-07) [2026-03-01]. https://doi.org/10.48550/arXiv.2410.01140.

Special Section on Non-Convex Optimization and Numerical Algorithms

Robust Kaczmarz Algorithm for Asymptotically Singular Linear Systems of Equations

KE Yunying[1], JIANG Wen[1], LUO Hao[1,2,3]

(1. National Center for Applied Mathematics in Chongqing, Chongqing Normal University; 2. Chongqing Municipal University Key Lab of Opeimization Techniques and Intelligent Vehide, Chongqing Normal University, Chongqing 401331; 3. Chongqing Research Institute of Big Data, Peking University, Chongqing 401121, China)

Abstract: To address the issue of a sharp decline in convergence speed of the classical Kaczmarz method when solving asymptotically singular linear equations, a robust algorithm with a consistently stable convergence rate is proposed. By regarding the row space of the coefficient matrix as a subspace in the Grassmann manifold, a new concept of asymptotically singular matrices with row rank degeneracy is introduced, and the asymptotic null space and asymptotic dual null space are defined. Based on this, a robust Kaczmarz method is proposed through the subspace correction framework, and the convergence analysis is completed using the Xu-Zikatanov identity. Finally, based on the equivalence between the Kaczmarz method and the coordinate descent method and the relationship between the asymptotic null space and the dual null space, a robust coordinate descent algorithm is proposed, and the equivalence between the robust Kaczmarz method and the robust coordinate descent algorithm is proved. Theoretical analysis shows that the proposed robust Kaczmarz method and robust coordinate descent algorithm have a consistently stable convergence rate when solving asymptotically singular linear equations. Numerical experiments verify the robustness and effectiveness of the algorithms.

Keywords: asymptotically singular linear equations; Kaczmarz method; coordinate descent method; subspace correction method; Xu-Zikatanov identity

(责任编辑 黄 颖)