



稀疏正则化: Gauss-Seidel 阈值迭代算法

曾锦山*, 何涛, 欧阳诗康

江西师范大学计算机信息工程学院, 南昌 330022

E-mail: jsh.zeng@gmail.com, htao1991@163.com, oyskang@163.com

收稿日期: 2017-01-17; 接受日期: 2017-07-25; 网络出版日期: 2018-06-08; * 通信作者

国家自然科学基金 (批准号: 61603162, 11501440, 61772246 和 61603163)、江西师范大学博士启动基金和江西省教育厅研究生创新 (批准号: YC2016-S171) 资助项目

摘要 本文考虑一类稀疏正则化问题, 该类问题在机器学习、信号处理和图像处理等众多领域中被广泛研究. 此类问题的一个典型特征是其诱导的阈值函数具有跳跃的不连续性. 本文提出一种基于 Gauss-Seidel 的迭代算法, 称作 Gauss-Seidel 跳跃阈值迭代算法 (Gauss-Seidel iterative jumping thresholding algorithm, GSIJT), 用以快速解决以上问题. 本文首先证明了由 GSIJT 所产生序列的支撑与符号的有限收敛性. 基于此收敛性质, 同时利用 restricted Kurdyka-Lojasiewicz (rKL) 性质给出 GSIJT 算法的全局收敛性. 此外给出了 GSIJT 的收敛率, 并且证明了任意的极限点都是驻点. 本文实施了一系列的数值实验来验证所提算法的有效性. 特别地, 通过与相关的阈值迭代算法进行比较, 表明所提算法不仅收敛更快, 同时可选择的步长范围更宽.

关键词 稀疏正则化 ℓ_q ($0 < q < 1$) 正则化 阈值迭代算法 Gauss-Seidel Kurdyka-Lojasiewicz 不等式

MSC (2010) 主题分类 90C26

1 引言

本文主要考虑如下稀疏正则化问题:

$$\min_{x \in \mathbb{R}^N} T_\lambda(x) := F(x) + \lambda \Phi(x), \quad (1.1)$$

其中 $\Phi(x)$ 是某种连续具有可分离的惩罚函数, 即 $\Phi(x) = \sum_{i=1}^N \phi(|x_i|)$, $\lambda > 0$ 是正则化参数. 问题 (1.1) 在科学研究和工程应用中都引起了广泛的关注, 一些典型的应用包括信号处理^[1]、图像处理^[2,3] 和合成孔径雷达成像^[4] 等.

求解问题 (1.1) 的算法主要有三类. 第一类是迭代重赋权算法, 包括迭代重赋权最小二乘 (iterative reweighted least squares, IRLS) 极小化^[5-7], 以及迭代重赋权 ℓ_1 极小化 (IRL1)^[8] 算法. 迭代重赋权算法的主要思想是通过解决一系列的加权最小二乘 (或者 ℓ_1 极小化) 问题, 来得到一个近似稀疏的结果.

英文引用格式: Zeng J S, He T, Ouyang S K. Sparse regularized optimization via Gauss-Seidel iterative thresholding (in Chinese). Sci Sin Math, 2018, 48: 969–986, doi: 10.1360/SCM-2017-0039

第二类是凸函数差分算法 (即 DC 规划 (difference of convex functions))^[9]. DC 规划的主要思想是通过目标函数进行适当的分解从而实现快速求解. 因此, 它只适用于那些能够分解成凸函数差的形式非凸罚范数. 最后一类是阈值迭代算法 (iterative thresholding algorithm, ITA)^[10–12], 它包含在前向-后向分裂算法^[13] 框架中. 与 IRLS 和 IRL1 等其他算法相比, ITA 在大规模问题^[4] 上通常有着更小的计算复杂度. 该优点使得阈值迭代算法在过去十年里得到了广泛研究 (如文献 [10–12, 14, 15]).

阈值迭代算法 (ITA) 可以通过多种方式推导得到. 第一种方式是从算子分裂的角度^[16] 来推导. 该类推导的主要思想是首先通过给出问题的最优条件, 然后利用算子分割分裂思想来得到一个不动点迭代格式. 另外一种方式是基于代理函数法^[11], 基本思想是, 在每次迭代中, 我们不去直接考虑原始问题的极小化问题, 而是转化为考虑一个更简单的所谓的代理函数的极小化问题, 并将其极小值再作下一步迭代. 在数学上, ITA 主要包括两个部分: 针对光滑函数部分采用梯度下降策略, 而针对不光滑的惩罚项采用相关的阈值算子. 具体地, 给定一个有限的初始值 x^0 ,

$$x^{n+1} = \mathcal{T}(x^n) \in \text{Prox}_{\mu, \lambda\Phi}(x^n - \mu \nabla F(x^n)), \quad \forall n \in \mathbb{N}, \quad (1.2)$$

其中 $\mu > 0$ 为步长, $\text{Prox}_{\mu, \lambda\Phi}$ 称为阈值函数 (或邻近算子), 定义如下:

$$\text{Prox}_{\mu, \lambda\Phi}(x) = \arg \min_{u \in \mathbb{R}^N} \left\{ \frac{\|x - u\|_2^2}{2\mu} + \lambda\Phi(u) \right\}. \quad (1.3)$$

从算法形式上来讲, ITA (1.2) 属于 Jacobi 迭代, 其中 x 的所有分量在每一次迭代时都同时更新. 另外一种重要的迭代格式是 Gauss-Seidel 迭代. 它是以循环的方式来更新每一个分量. 一般来讲, Gauss-Seidel 迭代的收敛速度比相应的 Jacobi 迭代要更快^[17], 并且可以允许更大的步长^[18], 主要因为其在每一步迭代中使用的是最近的更新. 本文旨在开发一个 Gauss-Seidel 版本的阈值迭代算法, 称为 Gauss-Seidel 迭代跳跃阈值 (GSIJT) 算法, 用于有效求解问题 (1.1). 给定一个有限的初始点 x^0 、当前步迭代 x^n 和步长 $\mu > 0$, 在下一步迭代中, 第 i 个分量按照如下周期性的方式进行更新:

$$i = \begin{cases} N, & 0 \equiv (n+1) \bmod N, \\ (n+1) \bmod N, & \text{其他}, \end{cases} \quad (1.4)$$

$$x_i^{n+1} = \mathcal{T}_i(x^n), \quad x_j^{n+1} = x_j^n, \quad \forall j \neq i, \quad (1.5)$$

其中 $\mathcal{T}_i(x^n) \in \text{Prox}_{\mu, \lambda\phi}(x_i^n - \mu[\nabla F(x^n)]_i)$, x_i^n 和 $[\nabla F(x^n)]_i$ 分别表示 x^n 的第 i 个分量和 x^n 处的 ∇F .

本文的主要目的是建立 GSIJT 的收敛性分析. 首先证明了所提算法具有支撑和符号的有限收敛性, 然后证明算法全局收敛¹⁾ 到一个驻点, 同时给出了 GSIJT 收敛速率的估计. 从证明技巧上来讲, 本文证明的创新之处主要包括, 首先利用阈值函数的跳跃不连续性给出算法的支撑和符号有限收敛性; 然后通过支撑和符号的有限收敛性给出整个序列的收敛性. 从数值实验的角度来讲, 本文提出的算法比经典的阈值迭代算法更快, 并且允许更大的收敛步长.

本文的剩余部分的结构如下: 第 2 节给出主要假设和主要结论; 第 3 节介绍相关的工作以及对比; 第 4 节给出我们进行的数值仿真结果; 第 5 节提供了证明; 第 6 节对本文进行总结.

下面给出本文的符号说明. 我们分别用 $\mathbb{R}$ 和 $\mathbb{N}$ 表示实数和自然数, 对于任意的向量 $x \in \mathbb{R}^N$, x_i 是其第 i 个分量, 对于一个给定的索引集 $I \subset I_N \triangleq \{1, 2, \dots, N\}$, x_I 表示包含指标集 I 中所有分量的子向量. I^c 表示 I 的补集, 即 $I^c = I_N \setminus I$. $\|x\|_2$ 表示向量 x 的欧式范数. $\text{Supp}(x)$ 表示 x 的支撑, 即

1) 本文所用的全局收敛是指对于算法产生的序列而言, 其收敛情况与初始点无关.

$\text{Supp}(x) = \{i : |x_i| > 0, i = 1, \dots, N\}$. 类似于向量的情形, 对于一个给定的指标集 I , A_I 表示 A 的子矩阵, 其包含了所有指标包含在 I 中的列. 对于任意的 $z \in \mathbb{R}$, $\text{sign}(z)$ 表示其符号函数, 即 $\text{sign}(z) = 1$ ($z > 0$)、 $\text{sign}(z) = -1$ ($z < 0$) 和 $\text{sign}(z) = 0$ ($z = 0$).

2 主要的假设和结论

2.1 主要的假设

我们给出如下的关于稀疏正则化优化问题 (1.1) 的基本假设.

假设 2.1 (逐坐标的 Lipschitz 可微性) $F : \mathbb{R}^N \rightarrow [0, \infty)$ 连续可微, 其梯度具有逐分量的 Lipschitz 连续性, 即对于任意 $i \in \{1, \dots, N\}$, 有 $F(u) \leq F(v) + [\nabla F(v)]_i(u - v_i) + \frac{L_i}{2}|u_i - v_i|^2$, 其中 $u, v \in \mathbb{R}^N$ 且满足 $u_i \neq v_i, u_j = v_j, [\nabla F(u)]_i$ 代表了 $\nabla F(u)$ 的第 i 个分量, $L_i > 0$ 是关于第 i 个坐标的 Lipschitz 常数. 令 $L_{\max} \triangleq \max_{1 \leq i \leq N} L_i$.

F 的另一个常用假设如下所示.

假设 2.2 (Lipschitz 连续性) F 是连续可微的, 且 ∇F 是 Lipschitz 连续的, 即对于任意 $u, v \in \mathbb{R}^N$, 有 $\|\nabla F(u) - \nabla F(v)\|_2 \leq L_F \|u - v\|_2$, 其中 $L_F > 0$ 是 Lipschitz 常数.

注意假设 2.2 一般来说要比假设 2.1 严格, 而 $L_{\max}$ 一般比 L_F 小. 例如, 在机器学习使用的 ℓ_2 和逻辑 (logistic) 损失函数是满足假设 2.1 和 2.2 的两个典型例子.

假设 2.3 (跳跃性质) $\phi : [0, \infty) \rightarrow [0, \infty)$ 是连续可微的并且满足如下条件:

- (i) ϕ 是强制 (coercive)、非降的且满足 $\phi(0) = 0$, 而当 $z > 0$ 时, $\phi(z) > 0$;
- (ii) 对于任意给定的 $\mu, \lambda > 0$, $\text{Prox}_{\mu, \lambda\phi}$ 定义良好, 并且具有跳跃的非连续性:

$$\text{Prox}_{\mu, \lambda\phi}(z) = \begin{cases} \text{sign}(z)\rho_\mu^{-1}(|z|), & \text{如果 } |z| \geq \tau_\mu, \\ 0, & \text{如果 } |z| \leq \tau_\mu, \end{cases} \quad (2.1)$$

$\rho_\mu^{-1}(\tau_\mu) > 0$, 其中 ρ_μ^{-1} 是 ρ_μ 的逆映射, 对于任意的 $u \in \mathbb{R}_+$, 以及 $\tau_\mu > 0$, 有 $\rho_\mu(u) = u + \lambda\mu\phi'(u)$, τ_μ 是一个阈值. 令 $\eta_\mu \triangleq \rho_\mu^{-1}(\tau_\mu)$, 那么 $\text{Prox}_{\mu, \lambda\phi}$ 的取值范围为 $\{0\} \cup [\eta_\mu, \infty)$.

满足假设 2.3 的惩罚函数有很多. 最典型的两个子类是文献 [15] 中研究的 $\phi(z) = z^q$ 和 $\phi(z) = \log(1 + z^q)$, 其中 $q \in (0, 1)$.

2.2 收敛性结果

定理 2.1 (支撑和符号有限收敛性) 令 $\{x^n\}$ 为一个由 GSJT 所生成的序列, $I^n = \text{Supp}(x^n)$. 在假设 2.1 和 2.3 的条件下, 如果 $0 < \mu < \frac{1}{L_{\max}}$, 那么,

- (1) $\{T_\lambda(x^n)\}$ 收敛到某个值 T_λ^* ;
- (2) $\{x^n\}$ 是有界的, 且有一个收敛子列;
- (3) 存在一个正整数 n_0 、一个指标集 I 和一个符号向量 S^* , 对于 $\forall n > n_0$, 如下成立:
 - (i) $I^n = I, \text{Supp}(x^*) = I$;
 - (ii) $\text{sign}(x^n) = S^*, \text{sign}(x^*) = S^*, \forall x^* \in \mathcal{L}$, 其中 $\mathcal{L}$ 是 $\{x^n\}$ 的极限点集.

定理 2.1 表示序列 $\{x^n\}$ 的支撑和符号在有限的迭代步数内收敛. 有限支撑收敛性在很多应用中都很重要, 包括但不限于稀疏信号恢复问题 [19] 和特征选择问题 [20]. 对于稀疏信号恢复问题, 支撑一

且收敛, 就可以通过求解最小二乘问题来重建向量的系数. 对于特征选择问题, 如果支撑已经保持不变, 那么可以从支撑中保留的这些特征中进一步提取主要特征.

定理 2.2 (全局收敛率) 考虑 GSIJT 生成的任意序列 $\{x^n\}$. 在假设 2.1 和 2.3 成立的条件下, 如果 $0 < \mu < \frac{1}{L_{\max}}$, 那么 $\{x^n\}$ 会收敛到一个极限点集 $\mathcal{L}$. 更进一步, 如果 F 满足假设 2.2, 并存在一个极限点 $x^* \in \mathcal{L}$ 使得 T_λ 满足有 restricted Kurdyka-Łojasiewicz (rKL) 性质 (见下文定义 5.2), 则整个序列收敛到 x^* .

此外, 如果 T_λ 在 x^* 处满足 rKL 不等式, 其中 $\varphi(s) = cs^{1-\theta}$ 对于某个常数 $c > 0$ 成立, 则有

- (1) 如果 $\theta = 0$, 则 x^n 在有限的多步迭代中收敛到 x^* ;
- (2) 如果 $\theta \in (0, \frac{1}{2}]$, 则对于所有 $n \geq n^*$, 对某个 $n^* > 0, C > 0, \tau \in [0, 1)$, 有 $\|x^n - x^*\|_2 \leq C\tau^n$;
- (3) 如果 $\theta \in (\frac{1}{2}, 1)$, 则对于所有 $n \geq n^*$, 对于某个 $n^* > 0, C > 0$, 有 $\|x^n - x^*\|_2 \leq Cn^{-(1-\theta)/(2\theta-1)}$.

定理 2.2(1)、2.2(2) 和 2.2(3) 分别对应有限收敛、渐近线性收敛²⁾ 和渐近次线性收敛. 对于非光滑非凸最优化问题, 可以从邻近算法^[21] 和块坐标下降算法^[22] 中得到相似的收敛率. 定理 2.2 建立了 GSIJT 的收敛性, 但我们仍需回答收敛到的极限点是什么. 接下来, 我们给出一个定理来说明极限点实际上是我们所考虑问题 (1.1) 的一个驻点. 本文中驻点是根据引理 5.6 来定义的, 因为问题 (1.1) 的最优性条件在引理 5.6 中进行了描述.

定理 2.3 (任意极限点是驻点) 令 $\{x^n\}$ 是一个 GSIJT 产生的迭代序列. 如果 $0 < \mu < L_{\max}^{-1}$, 则 $\mathcal{L} \subseteq \mathcal{F}_\Phi$, 其中 $\mathcal{F}_\Phi$ 是算子 $G_{\mu, \lambda\Phi}$ 的不动点集, $G_{\mu, \lambda\Phi}$ 定义如下:

$$G_{\mu, \lambda\Phi}(x) = \text{Prox}_{\mu, \lambda\Phi}(x - \mu \nabla F(x)), \quad \forall x \in \mathbb{R}^N. \quad (2.2)$$

事实上, 引理 5.6 中所说的 $\mathcal{F}_\Phi$ 也是最优化问题 (1.1) 的驻点集.

3 相关工作及比较

有许多现有的方法来解决稀疏正则化优化问题 (1.1). 一些普通的算法, 例如, 文献 [6, 8, 13, 15, 23–25] 以及它们的参考文献, 都不是通过使用 Gauss-Seidel 方法来更新迭代的. 文献 [15] 提出了迭代跳跃阈值 (iterative jumping thresholding, IJT) 算法, 用于解决一类具有 (1.1) 形式的稀疏正则化问题. 为了保证 IJT 和 GSIJT 算法收敛, 步长应分别满足 $0 < \mu < L_F^{-1}$ 和 $0 < \mu < L_{\max}^{-1}$, 其中 $L_{\max}$ 和 L_F 在假设 2.1 和 2.2 中分别进行了说明, 并且 $L_{\max} \leq L_f$. 这意味着 GSIJT 可以采用比 IJT 更大的步长. 除了 IJT 之外, 还有一些特定的阈值迭代算法 (如硬阈值^[10]、软阈值^[11] 和半阈值^[12, 14]) 迭代算法用来解决特定的 ℓ_q 正则化最小二乘问题. 这些算法的收敛条件为 $0 < \mu < \|A\|_2^{-2}$, 其中 A 是测量矩阵. 当应用于 ℓ_q 正则化最小二乘问题时, GSIJT 的收敛条件为 $0 < \mu < \frac{1}{\max_i \|A_i\|_2^2}$.

另一个紧密相关的算法是块坐标下降 (block coordinate descent, BCD) 算法. BCD 在许多应用中被大量使用. 它的原始形式, 块坐标最小化 (block coordinate minimization, BCM) 可以追溯到 20 世纪 50 年代^[26]. BCM 的主要思想是通过最小化某一块的原始目标来更新块坐标. BCD 在许多不同情形下的收敛性在文献 (如文献 [22, 27–29] 及其参考文献) 中有广泛研究. 在文献 [30] 中, BCD 的收敛速度是多块 (multi-block) 强凸假设下建立的, 在文献 [31] 中, 其收敛率是在两块 (two-block) 情形下基于凸性所建立的. 除了 BCM 之外, 块坐标梯度下降法 (block coordinate gradient descent, BCGD) 也得到广泛研究. 与 BCM 不同的是, BCGD 通过采取块梯度步更新一个块, 这等价于最小化目标的某一线性近似.

²⁾ 在许多其他论文中渐近线性收敛也称为最终或局部线性收敛.

它的全局收敛性在“局部 Lipschitz 误差约束”和目标函数不可微部分的凸性这两个假设下被证明^[32]. 然而, 根据假设 2.3, 问题 (1.1) 考虑的惩罚通常是非凸的、非光滑的, 甚至非 Lipschitz 连续性的.

BCD 的一个重要子类是循环坐标下降 (cyclic coordinate descent, CCD) 算法, 其中坐标是通过周期性的更新. 该工作^[22] 使用循环更新, 并且假定分块多凸 (block multi-convex). 特别地, 在文献 [22] 中所考虑的以下优化问题:

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}_1, \dots, \mathbf{x}_s) := g(\mathbf{x}_1, \dots, \mathbf{x}_s) + \sum_{i=1}^s r_i(\mathbf{x}_i), \quad (3.1)$$

其中, 变量 $\mathbf{x}$ 被分解为 s 块 $\mathbf{x}_1, \dots, \mathbf{x}_s$, $\mathbf{x}_i \in \mathbb{R}^{N_i}$, $i = 1, \dots, s$, $\sum_{i=1}^s N_i = N$, 函数 g 可微且具有分块多凸性³⁾ (r_i , $i = 1, \dots, s$) 是可扩展的 (extended) 凸函数. 从迭代形式的角度来看, GSIJT 与 BCD 算法在文献 [22] 中研究的线性近似更新类似. 当应用于问题 (1.1) 时, BCD 线性近似更新变成了

$$x_i^{n+1} \in \operatorname{argmin}_{v \in \mathbb{R}} \left\{ \langle [\nabla F(\hat{x}^n)]_i, v - \hat{x}_i^n \rangle + \frac{L_i}{2} |v - \hat{x}_i^n|^2 + \lambda \phi(|v|) \right\}, \quad (3.2)$$

其中

$$\hat{x}_j^n = \begin{cases} x_j^n + \omega_j^n (x_j^n - x_j^{n-1}), & \text{如果 } j = i, \\ x_j^n, & \text{其他} \end{cases}$$

表示一个外插点, $\omega_i^n \geq 0$ 为外插权重. 而 GSIJT 具有如下形式:

$$x_i^{n+1} \in \operatorname{argmin}_{v \in \mathbb{R}} \left\{ \langle [\nabla F(x^n)]_i, v - x_i^n \rangle + \frac{1}{2\mu} |v - x_i^n|^2 + \lambda \phi(|v|) \right\}, \quad (3.3)$$

其中 $\mu > 0$ 为步长. 从 (3.2) 和 (3.3) 可以看出, 如果外插权重 $\omega_i^n = 0$ 且 $\frac{1}{L_i}$ 为 (3.3) 中不同的坐标对 μ 进行取值, 那么这两个特殊的更新是相同的. 在某些条件下, 文献 [22] 证明了更新 (3.2) 的全局收敛性和收敛速度, 其主要收敛结果如下.

定理 3.1 (参见文献 [22, 定理 2.8 和 2.9]) 令 $\{x^n\}$ 为 BCD 算法产生的序列, 其参数 ω_i^n 满足一定的条件. 假定

- (i) ∇g 在任意有界集上 Lipschitz 连续;
- (ii) f 在 $\{x^n\}$ 的一个有限的极限点 x^* 处满足 Kurdyka-Łojasiewicz (KL) 性质 (见下文定义 5.1);
- (iii) x^0 足够接近 x^* , 对于 $n > 0$, 目标序列在 $n > 0$ 时满足 $f(x^n) > f(x^*)$,

则 $\{x^n\}$ 收敛到临界点 x^* .

此外, 如果当 $\varphi(s) = cs^{1-\theta}$, $c > 0$ 并且 $\theta \in [0, 1)$, f 满足 KL 不等式, 则以下成立:

- (1) 如果 $\theta = 0$, 那么 $\{x^n\}$ 在有限步迭代内收敛到 x^* ;
- (2) 如果 $\theta \in (0, 1/2]$, 则对于所有的 $n \geq n_0$, 对于某个 $n_0 > 0$, $C > 0$, $\rho \in [0, 1)$, 有 $\|x^n - x^*\| \leq C\rho^n$;
- (3) 如果 $\theta \in (1/2, 1)$, 那么对于所有的 $n \geq n_0$, 对于某个 $n_0 > 0$, $C > 0$, 有

$$\|x^n - x^*\| \leq Cn^{-(1-\theta)/(2\theta-1)}.$$

直观上比较, 文献 [22] 所研究的 BCD 的主要收敛结果与 GSIJT 的定理 2.2 类似. 然而, 文献 [22] 与本文的结果仍然存在一些差异. 首先, 假设 (3.1) 中的正则化项 r_i 是凸的, 用于保证每一个子问

3) 当其他坐标块固定时, 如果对于每一个 i , g 都是 $\mathbf{x}_i$ 的凸函数, 则称 g 是分块多凸的.

题有唯一解, 而本文所研究的函数 ϕ 是非凸的, 从而导致子问题通常有多个解. 其次, 本文除了对全局收敛性和收敛率进行了研究, 还对 GSIJT 的有限的支撑和符号收敛进行了分析. 最后一点是本文与文献 [22] 的证明思路有着很大的不同. 我们的全局收敛的证明有两个关键之处, 即通过利用阈值函数的“跳跃”性质得到支撑的有限步收敛性, 以及基于支撑和符号的有限收敛性得到的全局收敛. 关于这一点我们将在第 5 节中进行详细说明.

需要指出的是 GSIJT 与 Bolte 等 [33] 提出的邻近交替线性极小化 (proximal alternating linearized minimization, PALM) 算法非常相似. 文献 [33] 基于 KL 性质证明了 PALM 的全局收敛. 然而, 满足假设 2.1 和 2.3 的函数未必满足 KL 性质 (文献 [15] 给出了一个一维的反例). 本文在一个比 KL 性质稍弱的 rKL 性质下给出 GSIJT 的全局收敛, 同时还证明了 GSIJT 的支撑和符号有限步收敛性.

4 数值实验

本节通过提供了一系列的实验并与 ITA 进行比较来说明 GSIJT 的有效性. 在这些实验中, 我们主要考虑的是 ℓ_q 正则化最小二乘问题中两个特定的 q , 即 $q = \frac{1}{2}$ (参见文献 [12]) 和 $q = \frac{2}{3}$ (参见文献 [2]), 因为它们的阈值函数 $\text{Prox}_{\mu, \lambda \cdot \|\cdot\|_q^q}$ 有闭形式, 从而算法更容易实现. 特别地, 非凸 ℓ_q ($0 < q < 1$) 正则化最小二乘问题描述如下: $\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_q^q$, 其中 $A \in \mathbb{R}^{m \times N}$ 通常是一个扁矩阵, 即 $m \leq N$, $y \in \mathbb{R}^m$ 通常称为测量向量, $x \in \mathbb{R}^N$, $\|x\|_q = (\sum_{i=1}^N |x_i|^q)^{1/q}$ 称为 ℓ_q 拟范数. $\ell_{1/2}$ 拟范数的阈值函数表示如下 (参见文献 [12]):

$$\text{Prox}_{\mu, \lambda \cdot |\cdot|^{1/2}}(u) = \begin{cases} \frac{2}{3} u \left(1 + \cos \left(\frac{2\pi}{3} - \frac{2}{3} \arccos \left(\frac{\lambda \mu}{4} \left(\frac{|u|}{3} \right)^{-3/2} \right) \right) \right), & |u| > \frac{3}{2} (\lambda \mu)^{2/3}, \\ 0, & \text{其他}, \end{cases} \quad (4.1)$$

$\ell_{2/3}$ 拟范数 (参见文献 [2]) 的阈值函数表示如下:

$$\text{Prox}_{\mu, \lambda \cdot |\cdot|^{2/3}}(u) = \begin{cases} \text{sign}(u) \left(\frac{\phi_\tau(u) + \sqrt{\frac{2|u|}{\phi_\tau(u)} - \phi_\tau(u)^2}}{2} \right)^3, & |u| > \tau, \\ 0, & \text{其他}, \end{cases} \quad (4.2)$$

其中 $\text{sign}(u)$ 表示为 u 的符号函数, $\tau = 2(\frac{4}{3}\lambda\mu)^{3/4}$, $\phi_\tau(u) = \frac{2^{13/16}}{4\sqrt{3}}\tau^{3/16}(\cosh(\frac{\theta_\tau(u)}{3}))^{1/2}$,

$$\theta_\tau(u) = \text{arccosh} \left(\frac{3\sqrt{3}u^2}{2^{7/4}(2\tau)^{9/8}} \right).$$

对于其他的 $q \in (0, 1)$, 我们可以使用在文献 [34] 中提出的近似求解方案, 得到阈值函数的近似值.

4.1 收敛更快

我们通过一系列的实验并与 ITA [15] 进行比较来说明 GSIJT 的收敛速度更快. 给定一个维度 $N = 500$ 的稀疏信号 x_{tr} 和稀疏度 $k^* = 15$, 如图 1(c) 所示, 我们通过观察 $y = Ax_{\text{tr}}$ 来恢复稀疏信号 x_{tr} , 其中 A 是测量矩阵. 原始稀疏信号 x_{tr} 通常是根据 Gauss 分布随机生成的. A 的维度 $m \times N = 250 \times 500$, 根据 $\mathcal{N}(0, 1/250)$ 独立同分布产生, 通过列标准化进行预处理, 即对于任意 i , 有 $\|A_i\|_2 = 1$. 然后我们将 GSIJT 和 ITA 分别应用到 ℓ_q 正则化问题, 其中 q 取两个不同的值, 即 $q = 1/2$ 和 $q = 2/3$. 在这两种情形下, 正则化系数 λ 取值为 0.0012, GSIJT 的步长 $\mu = \frac{0.95}{\max_i \|A_i\|_2^2}$ ($= 0.95$), ITA 的步长为

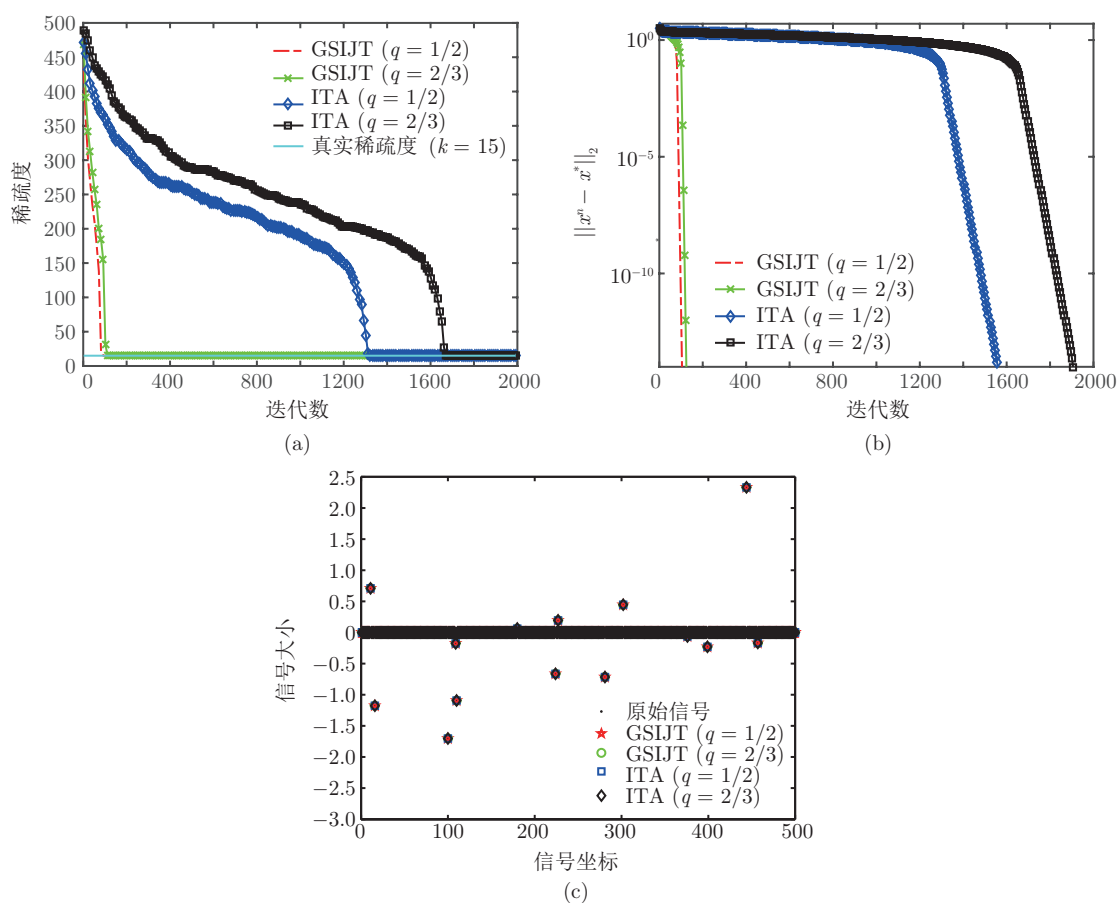


图 1 更快收敛的实验. (a) 稀疏度变化曲线; (b) 收敛率, 即 $\|x^n - x^*\|_2$ 的趋势, 其中 $x^* = x^{n_0}$, 根据极限点的近似值取 $n_0 = 5000$; (c) 恢复信号. 四种情形 GSIJT ($q = 1/2$)、GSIJT ($q = 2/3$)、ITA ($q = 1/2$) 和 ITA ($q = 2/3$) 的恢复精度分别为 2.06×10^{-8} 、 5.14×10^{-9} 、 2.12×10^{-8} 和 5.28×10^{-9}

$\mu = 0.99\|A\|_2^{-2}$ ($= 0.1676$). 此外, 在所有情形下, 初始值都为 0. 为了更好比较, 我们将 GSIJT 的每 N 步迭代作为一个阶段, 因为在这种情形下, 所有的坐标都只更新一次, 每步迭代的计算复杂度与 ITA 相同. 实验结果如图 1 所示.

由图 1(a) 可看出, 在有限步迭代内, GSIJT 和 ITA 都收敛到真正的稀疏水平, 但识别 GSIJT 支撑所需要的迭代次数明显小于 ITA. 具体地, 如图 1(a) 所示, GSIJT 需要的迭代次数在两种情形下都是 150, 而 ITA 需要更多迭代 (例如, 对于 $q = 1/2$ 和 $q = 2/3$, 大约分别需要 1,500 和 1,700 次). 图 1(b) 展示了 GSIJT 和 ITA 最终的线性收敛性. 我们可以观察到 GSIJT 的收敛速度比 ITA 快得多. 从图 1(a) 和 1(b) 可以观察到 GSIJT 和 ITA 都呈线性衰减直到支撑收敛. 此外, 图 1(c) 显示原始的稀疏信号可以被 GSIJT 和 ITA 高精度恢复. 实验结果表明 GSIJT 具有更快的收敛性和渐近线性收敛率等.

4.2 更弱的条件

我们实施一系列实验来说明 GSIJT 可以采用比 ITA 更大的步长. 实验设置与上一节相同. 然后, 我们将 GSIJT 和 ITA 应用于 $q = 1/2$ 的 ℓ_q 正则化最小二乘问题. 在这样的设置下, ITA 的收敛条件为 $\mu \in (0, 0.1759)$, 而 GSIJT 的相关条件为 $\mu \in (0, 1)$. 对于 GSIJT 和 ITA 我们都使用不同的 μ , 即 $\mu = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$. 目标序列的图像见图 2.

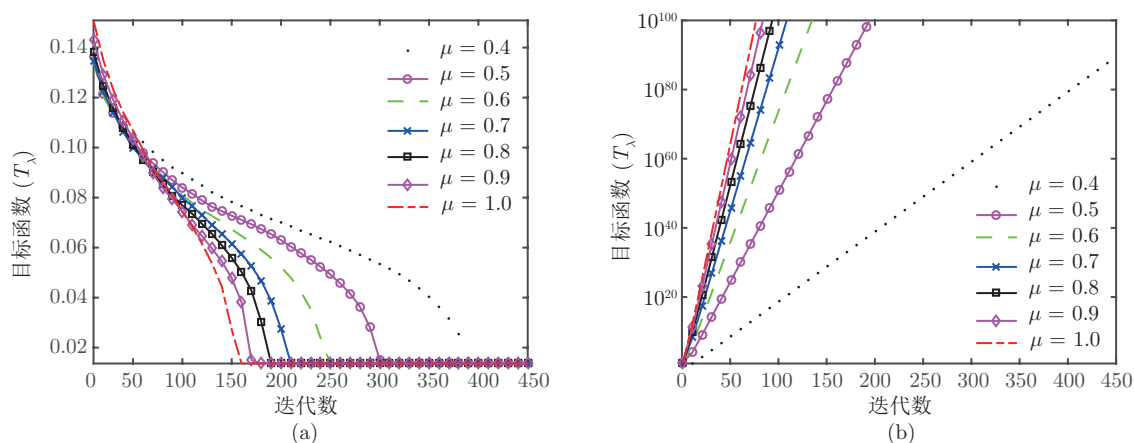


图 2 所允许步长范围的实验. (a) GSIJT 的收敛性, μ 取不同值时, 目标序列, 即 GSIJT 中 $\{T_\lambda(x^n)\}$ 的趋势; (b) ITA 发散, μ 取不同值时, ITA 的目标序列的趋势. 在所有情形中, 正则化系数 λ 设置为 0.0015

从图 2 可知, 对于所有的 μ , ITA 的目标函数序列都会发散, 而 GSIJT 收敛. 由图 2(a) 可知, GSIJT 的目标函数收敛很快 (所有情形下, 大约 400 次迭代就足以收敛); 而如图 2(b) 所示, ITA 的序列会迅速发散.

4.3 稀疏信号恢复性能

我们进行一系列的数值实验, 比较了 GSIJT 和 ITA 在 $q = 1/2$ 和 $q = 2/3$ 两种情形下的 ℓ_q 正则化问题的稀疏恢复性能. 考虑如下的信号恢复问题: 给定测量向量 $y = Ax_{\text{tr}} + \varepsilon$, 其中测量矩阵 $A \in \mathbb{R}^{250 \times 500}$, 原始稀疏信号 $x_{\text{tr}} \in \mathbb{R}^{500}$, k 非零元素是通过与第 4.1 小节中相同方式产生的, ε 是 Gauss 噪声. 测量值 y 增加了 40 分贝的噪声. 令 k 为 0 到 100 之间的等分点. 对于每个 k , 我们独立重复 50 次实验, 并记录了它们的恢复误差, 即 $\frac{\|\hat{x} - x_{\text{tr}}\|_\infty}{\|x_{\text{tr}}\|_\infty}$, 其中 $\hat{x}$ 是恢复信号. 然后我们将这 50 次独立实验的平均恢复误差作为这个例子的恢复误差. 此外, 我们还记录了恢复误差的成功概率, 从某种意义上来说, 如果恢复误差 $\frac{\|\hat{x} - x_{\text{tr}}\|_\infty}{\|x_{\text{tr}}\|_\infty} < 10^{-2}$, 那么就计一次成功. 恢复概率是指成功的次数与实验总数 (即 50 次) 之间的比值. 实验结果如图 3 所示.

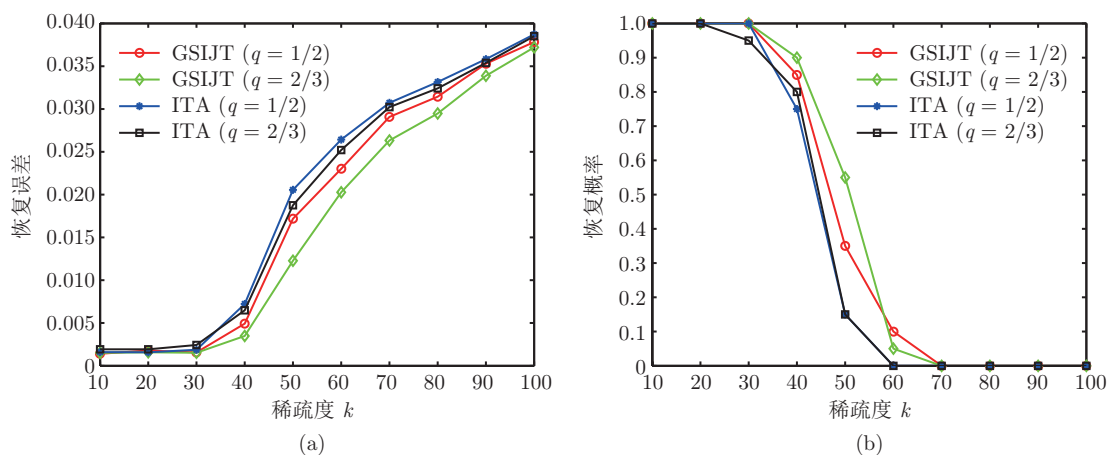


图 3 将 GSIJT 与 ITA 的恢复性能进行对比实验. (a) 恢复误差的趋势; (b) 恢复概率的趋势

通过图 3 可以看出, 在恢复误差和恢复概率上 GSIJT 的性能要略优于 ITA. 由图 3 可知, 当稀疏度 k 在 40 以下时, GSIJT 和 ITA 都能够高概率恢复稀疏信号; 一旦 k 上升到 50, GSIJT 和 ITA 的恢复概率就会急剧衰减到一个较小的值; 当 k 大于 60 时, 两者都无法恢复稀疏信号.

5 证明

5.1 支撑和符号有限步收敛

引理 5.1 令 $\{x^n\}$ 为 GSIJT 产生的一个迭代序列. 在假设 2.1 和 2.3 成立的情形下, 如果 i 是通过 (1.4) 来决定的, 那么, x_i^{n+1} 满足

(1) $x_i^{n+1} = 0$, 或者,

(2) $|x_i^{n+1}| \geq \eta_\mu$, $[\nabla F(x^{n+1})]_i + \lambda \text{sign}(x_i^{n+1}) \phi'(|x_i^{n+1}|) = \frac{1}{\mu}(x_i^n - x_i^{n+1}) - ([\nabla F(x^n)]_i - [\nabla F(x^{n+1})]_i)$, 即 $\nabla_i T_\lambda(x^{n+1}) = \frac{1}{\mu}(x_i^n - x_i^{n+1}) - ([\nabla F(x^n)]_i - [\nabla F(x^{n+1})]_i)$, 其中 $\nabla_i T_\lambda(x^{n+1})$ 表示 T_λ 在 x^{n+1} 处的第 i 个坐标的梯度.

根据 $\text{Prox}_{\mu, \lambda \Phi}$ 的定义、 ϕ 的跳跃假设 (见假设 2.3) 以及 GSIJT 的 (1.4) 和 (1.5) 容易证明该引理成立.

引理 5.2 (充分下降性) 令 $\{x^n\}$ 为 GSIJT 产生的迭代序列. 在假设 2.1 和 2.3 成立的情形下, 如果 $0 < \mu < L_{\max}^{-1}$, 则

$$T_\lambda(x^n) - T_\lambda(x^{n+1}) \geq a \|x^n - x^{n+1}\|_2^2, \quad \forall n \in \mathbb{N},$$

其中 $a \triangleq \frac{1}{2}(\frac{1}{\mu} - L_{\max})$.

证明 给定当前迭代 x^n , 令 i 为根据 (1.4) 决定的激活坐标, 那么通过 (1.2)、(1.3) 和 (1.5), 可得

$$x_i^{n+1} \in \arg \min_{v \in \mathbb{R}} \left\{ \frac{|z_i^n - v|^2}{2\mu} + \lambda \phi(|v|) \right\},$$

其中 $z_i^n = x_i^n - \mu[\nabla F(x^n)]_i$. 上式意味着

$$\frac{1}{2}|\mu[\nabla F(x^n)]_i|^2 + \lambda \mu \phi(|x_i^n|) \geq \frac{1}{2}|(x_i^{n+1} - x_i^n) + \mu[\nabla F(x^n)]_i|^2 + \lambda \mu \phi(|x_i^{n+1}|).$$

化简得

$$\lambda \phi(|x_i^n|) - \lambda \phi(|x_i^{n+1}|) \geq \frac{|x_i^{n+1} - x_i^n|^2}{2\mu} + [\nabla F(x^n)]_i(x_i^{n+1} - x_i^n).$$

此外, 由于对任意 $j \neq i$, 有 $x_j^{n+1} = x_j^n$, 从而上式变为

$$\lambda \Phi(x^n) - \lambda \Phi(x^{n+1}) \geq \frac{|x_i^{n+1} - x_i^n|^2}{2\mu} + [\nabla F(x^n)]_i(x_i^{n+1} - x_i^n).$$

在上面不等式的两边加上 $F(x^n) - F(x^{n+1})$, 就有

$$\begin{aligned} T_\lambda(x^n) - T_\lambda(x^{n+1}) &\geq \frac{|x_i^{n+1} - x_i^n|^2}{2\mu} - [F(x^{n+1}) - F(x^n) - [\nabla F(x^n)]_i(x_i^{n+1} - x_i^n)] \\ &\geq \frac{1}{2} \left(\frac{1}{\mu} - L_i \right) |x_i^n - x_i^{n+1}|^2 \geq \frac{1}{2} \left(\frac{1}{\mu} - L_{\max} \right) \|x^n - x^{n+1}\|_2^2, \end{aligned} \quad (5.1)$$

其中第二个不等式是通过 ∇F 逐坐标的 Lipschitz 连续性 (即假设 2.1) 所得到的. $\square$

定理 2.1 的证明 通过引理 5.2 可知, $T_\lambda(x^n)$ 是单调递减的. 因为其下界也为 0, 所以, $\{T_\lambda(x^n)\}$ 收敛. 从而表明 $T_\lambda^* = \lim_{n \rightarrow \infty} T_\lambda(x^n)$.

通过引理 5.2 知, 对于任意 $n \in \mathbb{N}$, 有 $T_\lambda(x^n) \leq T_\lambda(x^0) < \infty$. 由 $T_\lambda(x)$ 的强制性知, $\{x^n\}$ 有界并有收敛子列. 结合引理 5.2 和 $\{T_\lambda(x^n)\}$ 的收敛性, 可以给出

$$\sum_{n=0}^{\infty} \|x^{n+1} - x^n\|_2^2 \leq \frac{T_\lambda(x^0)}{a}, \quad (5.2)$$

$$\|x^n - x^{n+1}\|_2 \rightarrow 0, \quad \text{当 } n \rightarrow +\infty, \quad (5.3)$$

其中 (5.3) 表明, 存在一个足够大的正整数 n_0 使得当 $n > n_0$ 时, $\|x^n - x^{n+1}\|_2 < \eta_\mu$.

(1) $\{I^n\}$ 的收敛性. 首先通过反证法来证明

$$I^{n+1} = I^n, \quad \forall n > n_0. \quad (5.4)$$

假设结论不成立, 即对于某些 $n_1 > n_0$, 有 $I^{n_1+1} \neq I^{n_1}$. 令 $i = N - ((n_1 + 1) \bmod N)$, 也就是说, 存在一个 $j_0 \in \mathbb{N}$ 使得 $n_1 + 1 = j_0 N + i$. 一方面, 通过 (1.5), 我们有

$$x_j^{n_1+1} \neq x_j^{n_1}, \quad \forall j \neq i.$$

那么 $I^{n_1+1} \neq I^{n_1}$ 意味着 $x_i^{n_1+1}$ 和 $x_i^{n_1}$ 其中之一为 0, 而另一个不为 0. 不失一般性, 假设 $x_i^{n_1} = 0$, $x_i^{n_1+1} \neq 0$. 通过引理 5.1, 我们有 $|x_i^{n_1+1}| \geq \eta_\mu$. 因此,

$$\|x^{n_1+1} - x^{n_1}\|_2 = |x_i^{n_1+1}| \geq \eta_\mu,$$

这与 $\|x^n - x^{n+1}\|_2 < \eta_\mu$ 矛盾. 因此, 假设的序列 $\{I^n\}$ 收敛.

(2) $\{\text{sign}(x^n)\}$ 的收敛性. 与证明 (1) 类似, 我们将通过反证法来证明, 对于任意的 $i \in I$, 有 $\text{sign}(x_i^{n_1+1}) = \text{sign}(x_i^{n_1})$. 假设结论不成立, 那么存在一个 $i^* \in I$ 使得 $\text{sign}(x_{i^*}^{n_1+1}) \neq \text{sign}(x_{i^*}^{n_1})$, 因此,

$$\text{sign}(x_{i^*}^{n_1+1})\text{sign}(x_{i^*}^{n_1}) = -1.$$

通过引理 5.1, 很容易验证 $\|x^{n_1+1} - x^{n_1}\|_2 \geq |x_{i^*}^{n_1+1} - x_{i^*}^{n_1}| = |x_{i^*}^{n_1+1}| + |x_{i^*}^{n_1}| \geq \min_{i \in I} \{|x_i^{n_1+1}| + |x_i^{n_1}|\} \geq 2\eta_\mu$, 这又与 $\|x^{n_1+1} - x^{n_1}\|_2 < \eta_\mu$ 矛盾. 这一矛盾说明了当 $n > n_0$ 时, $\text{sign}(x^{n+1}) = \text{sign}(x^n)$. 符号序列 $\{\text{sign}(x^n)\}$ 收敛到 S^* , 其中 S^* 为符号序列 $\{\text{sign}(x^n)\}$ 的极限. 对于任意 $x^* \in \mathcal{L}$, 有 $\text{sign}(x^*) = S^*$, 该结论可通过类似的方式证明. $\square$

5.2 全局收敛及收敛率

通过定理 2.1, 算法的支撑和符号序列会在有限步迭代内收敛. 然后可以构建一个新的与 $\{x^n\}$ 具有相同收敛行为的序列.

引理 5.3 ($\{u^n\}$ 的构造) 在定理 2.1 的条件下, 存在一个序列 $\{u^n\}$ 与 $\{x^n\}$ 有相同的收敛行为.

证明 通过定理 2.1, 对于足够大的 $n > n_0$, $\{x^n\}$ 的支撑和符号相同. 对于一些正整数 j_0 , 令 $n^* = j_0 N > n_0$, 那么可以定义一个新的序列 $\{\hat{x}^n\}$, 对于 $n \in \mathbb{N}$, 有 $\hat{x}^n = x^{n^*+n}$. 根据定理 2.1, $\{\hat{x}^n\}$ 的所有支撑和符号都相同, 令 K 为 I 的基数. 不失一般性, 假设 $1 \leq I(1) < I(2) < \dots < I(K) \leq N$. 根据 GSIJT 中 (1.4) 和 (1.5) 的更新, 我们可以观察到 $\{\hat{x}^n\}$ 的许多相邻的迭代都是相同的. 因此, 我们可

以将这些连续的迭代合并为单步迭代. 另外, 索引的更新是循环的, 其具有周期性. 因此, 序列合并过程可以周期性地重复. 具体地, 我们认为这是一个周期子序列, 其长度 N 为 $\{\hat{x}^n\}$ 的长度. 对于 $j \in \mathbb{N}$, 考虑如下的一个周期的子列:

$$\{\hat{x}^{jN+I(1)}, \hat{x}^{jN+I(1)+1}, \dots, \hat{x}^{jN+I(1)+N-1}\},$$

按照如下的规则将上述长度为 N 的序列合并成长度为 K 的新序列 $\{\bar{x}^{jK+1}, \bar{x}^{jK+2}, \dots, \bar{x}^{jK+K}\}$, 即

$$\{\hat{x}^{jN+I(i)}, \dots, \hat{x}^{jN+I(i+1)-1}\} \mapsto \bar{x}^{jK+i},$$

其中 $\bar{x}^{jK+i} = \hat{x}^{jN+I(i)}$, $i = 1, 2, \dots, K$, 因为对于 $k = 1, \dots, I(i+1) - I(i) - 1$, 有 $\hat{x}^{jN+I(i)+k} = \hat{x}^{jN+I(i)}$. 此外, 我们将 $\{\hat{x}^n\}$ 的第一个 $I(1)$ 迭代合并为 $\bar{x}^0$, 即

$$\{\hat{x}^0, \dots, \hat{x}^{I(1)-1}\} \mapsto \bar{x}^0,$$

其中 $\bar{x}^0 = \hat{x}^0$, 因为这些迭代保持不变并且等于 $\hat{x}^0$. 在这个过程之后, 我们得到一个新的序列 $\{\bar{x}^n\}$, 其中 $n = jK + i$, $i = 0, \dots, K-1$, $j \in \mathbb{N}$. 可以看到, 这样一个合并过程保持了 $\{\bar{x}^n\}$ 的收敛行为, 同样也保持了 $\{\hat{x}^n\}$ 和 $\{x^n\}$ 的收敛行为. 此外, 定义 $\{u^n\}$ 如下:

$$u^n = \bar{x}_I^n, \quad \forall n \in \mathbb{N}.$$

可以观察到 u^n 去掉了 $\bar{x}^n$ 的所有 0 元素, 而保留了 $\bar{x}^n$ 的所有非 0 元素, 因此, $\{u^n\}$ 具有与 $\{x^n\}$ 一样的收敛行为. $\square$

给定一个指标集 $J \subset I_N$, 定义一个映射 $P_J: \mathbb{R}^N \rightarrow \mathbb{R}^{|J|}$, $P_J x = x_J$, $\forall x \in \mathbb{R}^N$, 其中 $|J|$ 代表它的基数. 我们还用 P_J^T 表示 P_J 的转置,

$$P_J^T: \mathbb{R}^{|J|} \rightarrow \mathbb{R}^N, \quad (P_J^T z)_J = z \quad \text{和} \quad (P_J^T z)_{J^c} = 0 \quad (5.5)$$

对于任意 $z \in \mathbb{R}^{|J|}$ 成立, 其中 $J^c = I_N \setminus J$, $(P_J^T z)_J$ 和 $(P_J^T z)_{J^c}$ 分别代表限制在 J 和 J^c 上 $P_J^T z$ 的子向量. 基于上述定义, 我们定义了两个新函数 T 和 f , 具体如下:

$$\begin{aligned} T: \mathbb{R}^K &\rightarrow \mathbb{R}, \quad T(u) = T_\lambda(P_I^T u), \quad \forall u \in \mathbb{R}^K, \\ f: \mathbb{R}^K &\rightarrow \mathbb{R}, \quad f(u) = F(P_I^T u), \quad \forall u \in \mathbb{R}^K, \end{aligned} \quad (5.6)$$

其中 I 在定理 2.1 中说明, $K = |I|$, 是 I 的基数. 令 $\{u^n\}$ 为引理 5.3 构造的序列, 那么 $\{u^n\}$ 具有以下性质.

引理 5.4 序列 $\{u^n\}$ 具有以下性质:

(1) 考虑到当前的迭代 u^n , u^{n+1} 通过如下迭代产生:

$$u_i^{n+1} = \hat{\mathcal{T}}_i(u^n) \quad \text{和} \quad u_j^{n+1} = u_j^n, \quad j \neq i, \quad (5.7)$$

其中

$$i = \begin{cases} K, & \text{如果 } 0 \equiv (n+1) \bmod K, \\ (n+1) \bmod K, & \text{其他,} \end{cases} \quad (5.8)$$

$$\hat{T}_i(x^n) \in \text{Prox}_{\mu, \lambda\phi}(u_i^n - \mu[\nabla f(u^n)]_i).$$

(2) 根据 (5.7), 对于 $n \geq K$, 存在两个整数 $1 \leq i_0 \leq K$ 和 $j_0 \geq 1$ 满足 $n = j_0 K + i_0$ 且

$$u_j^n = \begin{cases} u_j^{n-(i_0-j)}, & \text{如果 } 1 \leq j \leq i_0, \\ u_j^{n-K-(i_0-j)}, & \text{如果 } i_0 + 1 \leq j \leq K. \end{cases} \quad (5.9)$$

(3) $u^n \in \mathbb{R}_{\eta_\mu}^K$, 其中 $\mathbb{R}_{\eta_\mu} \triangleq \mathbb{R} \setminus (-\eta_\mu, \eta_\mu)$.

(4) 给定 u^n , 假定 i 由 (1.4) 决定, 则

$$\nabla_i T(u^{n+1}) = \frac{1}{\mu}(u_i^n - u_i^{n+1}) - ([\nabla f(u^n)]_i - [\nabla f(u^{n+1})]_i),$$

其中 $\nabla_i T(u^{n+1})$ 代表 u^{n+1} 处第 i 个坐标 T 的梯度.

(5) $T(u^n) - T(u^{n+1}) \geq a\|u^n - u^{n+1}\|_2^2, \forall n \in \mathbb{N}$.

证明 引理 5.4 中列的 $\{u^n\}$ 的性质是 $\{x^n\}$ 性质的直接推广. 更具体地, 引理 5.4(1) 可以由 (1.4)、(1.5) 和构造过程得到. 引理 5.4(2) 由循环更新规则直接得到. 引理 5.4(3) 和 5.4(4) 可以由引理 5.1(2) 和 5.4(1) 推导得出. 引理 5.4(5) 可由引理 5.2 和 T 的定义 (5.6) 得到. $\square$

通过引理 5.4(5) 可知, $\{u^n\}$ 满足充分下降性, 根据引理 5.4(3), $\{u^n\}$ 处于一个特殊的子空间, 它不包含零点和坐标轴上的点. 因此, 函数 T 在这种特殊的子空间里足够光滑, 可通过采用光滑分析中的一些证明技术来分析 $\{u^n\}$ 的收敛性. 因此, 这种构造极大地降低了证明 $\{x^n\}$ 全局收敛的难度. 除了引理 5.4, 接下来的引理表明, $\{u^n\}$ 同样满足相对误差 (relative error) 的性质.

引理 5.5 (相对误差) 在定理 2.1 的条件下, 假定 F 也满足假设 2.2. 对于任意 $n \geq K - 1$, 有

$$\|\nabla T(u^{n+1})\|_2 \leq b\|u^{n+1} - u^n\|_2,$$

其中 $b \triangleq (\frac{1}{\mu} + KL_F)\sqrt{K}$.

证明 假设对于一些正整数 $j^* \geq 1$ 和 $1 \leq i^* \leq K$, 满足 $n + 1 = j^* K + i^*$. 为了简单起见, 令 $i^* = K$. 否则, 由于迭代规则 (5.8) 是循环的从而使其具有周期性, 我们可以将坐标重新编号使得 $i^* = K$ 成立, 且序列 $\{u^n\}$ 保持不变. 这样的操作可以描述如下: 对于任意 $n \geq K$, 通过引理 5.4(2), 我们知道 u^n 只与 $K - 1$ 步迭代有关. 因此, 考虑如下一段迭代序列, 即 $\{i^* + 1, \dots, K, 1, \dots, i^*\}$, 然后将上面的坐标更新顺序重新编号为 $\{1', \dots, (K - i^*)', (K - i^* + 1)', \dots, K'\}$, 其中

$$j' = \begin{cases} i^* + j, & \text{如果 } 1 \leq j \leq K - i^*, \\ j - (K - i^*), & \text{如果 } K - i^* < j \leq K. \end{cases}$$

接下来, 我们通过递归方式来计算 $\nabla_i T(u^{n+1})$. 对于 $i = K$, 通过引理 5.4(4), 可以得到

$$\nabla_K T(u^{n+1}) = \frac{1}{\mu}(u_K^n - u_K^{n+1}) - ([\nabla f(u^n)]_K - [\nabla f(u^{n+1})]_K). \quad (5.10)$$

对于任意 $i = K - 1, K - 2, \dots, 1$, $\nabla_i T(u^{n+1}) = [\nabla f(u^{n+1})]_i + \lambda \text{sign}(u_i^{n+1})\phi'(|u_i^{n+1}|)$, 且 $u_i^{n+1} = u_i^n$. 因此, 对于 $i = K - 1, K - 2, \dots, 1$, 都有 $\nabla_i T(u^{n+1}) = \nabla_i T(u^n) - ([\nabla f(u^n)]_i - [\nabla f(u^{n+1})]_i)$. 对于任意 $j = 1, \dots, K - 1$, 通过递归的方式, 有

$$\nabla_{K-j} T(u^{n+1}) = \nabla_{K-j} T(u^n) - ([\nabla f(u^n)]_{K-j} - [\nabla f(u^{n+1})]_{K-j})$$

$$\begin{aligned}
&= \nabla_{K-j} T(u^{n-1}) - \sum_{i=0}^1 ([\nabla f(u^{n-i})]_{K-j} - [\nabla f(u^{n+1-i})]_{K-j}) \\
&= \nabla_{K-j} T(u^{n-j+1}) - \sum_{i=0}^{j-1} ([\nabla f(u^{n-i})]_{K-j} - [\nabla f(u^{n+1-i})]_{K-j}). \quad (5.11)
\end{aligned}$$

此外, 引理 5.4(4) 给出

$$\nabla_{K-j} T(u^{n-j+1}) = \frac{1}{\mu}(u_{K-j}^{n-j} - u_{K-j}^{n-j+1}) - ([\nabla f(u^{n-j})]_{K-j} - [\nabla f(u^{n-j+1})]_{K-j}).$$

将其代入 (5.11), 得到

$$\nabla_{K-j} T(u^{n+1}) = \frac{1}{\mu}(u_{K-j}^{n-j} - u_{K-j}^{n-j+1}) - \sum_{i=0}^j ([\nabla f(u^{n-i})]_{K-j} - [\nabla f(u^{n+1-i})]_{K-j}),$$

上式意味着

$$|\nabla_{K-j} T(u^{n+1})| \leq \frac{1}{\mu}|u_{K-j}^{n-j} - u_{K-j}^{n-j+1}| + L_F \sum_{i=0}^j |u_{K-i}^{n-i} - u_{K-i}^{n+1-i}|$$

对于 $j = 0, 1, \dots, K-1$ 成立, 其中不等式是由 ∇F 的 Lipschitz 连续性进而 ∇f 的连续性所得到的. 此外, 由引理 5.4(2), 对于任意 $0 \leq i \leq K-1$, 有 $u_{K-i}^{n+1-i} = u_{K-i}^{n+1}$, $u_{K-i}^{n-i} = u_{K-i}^n$. 因此, 对于 $j = 0, 1, \dots, K-1$, 有下式成立:

$$|\nabla_{K-j} T(u^{n+1})| \leq \frac{1}{\mu}|u_{K-j}^n - u_{K-j}^{n+1}| + L_F \sum_{i=0}^j |u_{K-i}^n - u_{K-i}^{n+1}| \leq \frac{1}{\mu}|u_{K-j}^n - u_{K-j}^{n+1}| + L_F \|u^{n+1} - u^n\|_1.$$

对 $|\nabla_{K-j} T(u^{n+1})|$ 关于 j 进行累加, 可以得到

$$\|\nabla T(u^{n+1})\|_1 \leq \frac{1}{\mu}\|u^{n+1} - u^n\|_1 + L_F K \|u^{n+1} - u^n\|_1 \leq \left(\frac{1}{\mu} + KL_F\right) \sqrt{K} \|u^{n+1} - u^n\|_2, \quad (5.12)$$

其中第二个不等式由如下不等式给出:

$$\|u\|_2 \leq \|u\|_1 \leq \sqrt{K} \|u\|_2, \quad \forall u \in \mathbb{R}^K. \quad (5.13)$$

此外, 由 (5.13) 的左边, (5.12) 推出

$$\|\nabla T(u^{n+1})\|_2 \leq \left(\frac{1}{\mu} + KL_F\right) \sqrt{K} \|u^{n+1} - u^n\|_2.$$

从而, 引理 5.5 得证. $\square$

为了证明定理 2.2, 我们还需要文献 [15] 中介绍的 T_λ 的 rKL 性质. 在描述 rKL 性质之前, 首先给出 KL 性质的定义.

定义 5.1 (KL 性质^[35,36]) 一个下半连续函数 $h: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ 称为在 $x^* \in \text{dom}(\partial h)$ 处具有 Kurdyka-Łojasiewicz (KL) 性质是指, 如果存在 $\eta \in (0, +\infty]$ 、 x^* 的一个领域 U 和一个连续的凹函数满足

- (i) $\varphi(0) = 0$, φ 在 $(0, \eta)$ 上连续可微;
- (ii) 对于所有 $s \in (0, \eta)$, 有 $\varphi'(s) > 0$;

(iii) 对于所有 $U \cap \{x : h(x^*) < h(x) < h(x^*) + \eta\}$ 中的 x , 以下 KL 不等式成立:

$$\varphi'(h(x) - h(x^*)) \cdot \text{dist}(0, \partial h(x)) \geq 1. \quad (5.14)$$

在 h 定义域 $\text{dom}(\partial h)$ 内都满足 KL 不等式的下半连续函数称为 KL 函数.

KL 函数包括实解析函数、次解析函数、半代数函数和局部强凸函数. 更多相关函数可以参见文献 [22, 第 2.2 小节] 及其参考文献. 然而, 正如文献 [15] 中命题 1 的证明部分所显示的, 有一类函数满足假设 2.1–2.3, 但是不满足 KL 性质. 归咎于此, 我们利用文献 [15] 中介绍的 rKL 性质而不是用 KL 性质来得到 GSIJT 的全局收敛.

定义 5.2 (rKL 性质^[15]) 称函数 $h : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ 在 $x^* \in \text{dom}(\partial h)$ 满足 restricted Kurdyka-Lojasiewicz (rKL) 性质是指, 如果 $g : \mathbb{R}^{|J|} \rightarrow \mathbb{R} \cup \{+\infty\}$, $g(z) = h(P_J^T z)$ 在 $z^* = x_J^*$ 处, $J = \text{Supp}(x^*)$ 满足 KL 性质.

通过定义 5.2 可知, rKL 性质仅要求 h 的次梯度在具有非 0 点处可以通过一些特定的凹变换变得更加的陡峭, 而原始的 KL 性质需要假定函数的次梯度在该点周围所有的元素都具有非常好的性质. 由文献 [15, 命题 1] 知, rKL 性质是 KL 性质的更一般形式, 并且比 KL 性质更弱.

定理 2.2 的证明 由定理 2.1 知, $\{x^n\}$ 的支撑和符号在有限步迭代内收敛. 基于此, 我们可以构造一个新的序列 $\{u^n\}$ 保持 $\{x^n\}$ 在引理 5.3 中的性质. 因此, 为了证明 $\{x^n\}$ 的全局收敛, 我们只需证明 $\{u^n\}$ 的收敛性. 通过引理 5.4 和 5.5, $\{u^n\}$ 满足所谓的充分下降性、相对误差和连续性条件. 与文献 [13, 定理 2.9] 类似, 我们可以给出 $\{u^n\}$ 在 T_λ 满足 rKL 条件下的全局收敛性, 从而得到 $\{x^n\}$ 的全局收敛性.

为了给出 $\{x^n\}$ 的收敛率, 我们可以通过估计 $\{u^n\}$ 的收敛率. 通过引理 5.4 和 5.5, $\{u^n\}$ 满足文献 [13] 中列出的充分下降性、相对误差和连续性条件. 在这三个条件下, 以及 T 在 u^* 处满足 KL 性质 (其中 $\varphi(s) = cs^{1-\theta}$), 然后根据文献 [13, 引理 2.6], 对于一些 $n^* > 0$ 和所有 $n \geq n^*$, 有

$$2\|u^{n+1} - u^n\|_2 \leq \|u^n - u^{n-1}\|_2 + \frac{cb}{a}((T(u^n) - T(u^*))^{1-\theta} - (T(u^{n+1}) - T(u^*))^{1-\theta}),$$

经过简单推导之后,

$$\sum_{k=n_0}^n \|u^{k+1} - u^k\|_2 \leq \|u^{n_0} - u^{n_0-1}\|_2 + \frac{cb}{a}((T(u^{n_0}) - T(u^*))^{1-\theta} - (T(u^{n+1}) - T(u^*))^{1-\theta}).$$

此外, 根据文献 [21, 定理 5] 中相似的推导证明, 我们可以估计 $\{u^n\}$ 在 θ 取不同值时的收敛率. 因此, 可以估计 $\{x^n\}$ 的收敛率. $\square$

推论 5.1 在定理 2.1 的假设下, 假定 x^* 是 $\{x^n\}$ 的一个极限点. 令 $I = \text{Supp}(x^*)$, $u^* = P_I x^*$, $T(u) = T_\lambda(P_I^T u)$. 此外, 假设 F 和 ϕ 在 x^* 的邻域内有二阶连续可微, 并且 $\nabla^2 T(u^*)$ 是非奇异的, 那么 $\{x^n\}$ 以渐近线性收敛率收敛到 x^* , 即

$$\|x^n - x^*\|_2 \leq C\tau^n$$

对于所有 $n \geq n^*$, 及某些 $n^* > 0$, $C > 0$, $\tau \in [0, 1)$ 成立.

该推论可直接由文献 [15, 引理 2] 和定理 2.2 得到.

5.3 收敛到一个驻点

根据 $\text{Prox}_{\mu, \lambda \Phi}$ 的定义, 稀疏正则化问题的某种最优性条件可以表示如下:

引理 5.6 给定一个点 x^* , 当且仅当以下三个条件成立, 则 $x^* \in \mathcal{F}_\Phi$.

- (1) 对于 $i \in \text{Supp}(x^*)$, 有 $|x_i^*| \geq \eta_\mu$;
- (2) 对于 $i \in \text{Supp}(x^*)$, 有 $[\nabla F(x^*)]_i + \lambda \text{sign}(x_i^*) \phi'(|x_i^*|) = 0$;
- (3) 对于 $i \in \text{Supp}(x^*)^c$, 有 $|[\nabla F(x^*)]_i| \leq \tau_\mu / \mu$.

根据 $\text{Prox}_{\mu, \lambda \Phi}$ 的定义, 该引理的证明是显然的. 因此省略. 如果 x^* 满足引理 5.6 的最优性条件, 那么在下文中称它为稀疏正则化问题的一个驻点.

对于任意 $n \in \mathbb{N}$, 定义

$$\Delta(x^n, x^{n+1}) \triangleq T_\lambda(x^n) - T_\lambda(x^{n+1}). \quad (5.15)$$

注意 x^n 和 x^{n+1} 只有在它们第 i 个系数中才会不同, 其中 i 由 (1.4) 决定. 从现在起, 如果没有特别说明, 假设 x_i^{n+1} 由 (1.5) 给出, i 由 (1.4) 给出.

引理 5.7 令 $\{x^n\}$ 是一个由 GSIJT 产生的迭代序列. 假设 $0 < \mu < L_{\max}^{-1}$, 则有

$$\Delta(x^n, x^{n+1}) = 0 \text{ 当且仅当 } x_i^{n+1} = x_i^n.$$

此引理的证明是显然的. 一方面, 如果 $x_i^{n+1} = x_i^n$, 那么 $x^{n+1} = x^n$, 因此, $\Delta(x^n, x^{n+1}) = 0$. 另一方面, 如果 $\Delta(x^n, x^{n+1}) = 0$, 那么引理 5.2 可推出 $x^{n+1} = x^n$, 所以, $x_i^{n+1} = x_i^n$.

$$\text{令 } z_i^n = x_i^n - \mu[\nabla F(x^n)]_i,$$

$$\mathcal{T}(z_i^n, x_i^n) = \text{Prox}_{\lambda, \mu\phi}(z_i^n)$$

对于任意 $n \in \mathbb{N}$ 和 $i \in \{1, 2, \dots, N\}$ 成立. 类似于文献 [34, 定理 19], 我们可以称 $\mathcal{T}(\cdot, \cdot)$ 是一个闭映射, 表示如下.

引理 5.8 $\mathcal{T}(\cdot, \cdot)$ 是一个闭的映射, 即假设

- (1) $x_i^n \rightarrow x_i^*$, $n \rightarrow \infty$;
- (2) $x_i^{n+1} \rightarrow x_i^{**}$, $n \rightarrow \infty$, 其中 $x_i^{n+1} = \mathcal{T}(z_i^n, x_i^n)$,

那么 $x_i^{**} = \mathcal{T}(z_i^*, x_i^*)$, 其中 $z_i^* = x_i^* - \mu[\nabla F(x^*)]_i$.

该引理的证明与文献 [34, 定理 10] 的证明本质上是一样的. 我们在此省略. 基于这些引理, 我们可以证明定理 2.3.

定理 2.3 的证明 由于序列 $\{x^n\}$ 有界, 因此, 它就有极限点. 令 $x^* \in \mathcal{L}$. 现在我们关注序列 x^n 的第 i 个系数, 其中 $n = n(i) = jN + i - 1$, $i = 1, 2, \dots, N$, $j = 0, 1, \dots$. 然而, 这里简单地使用 n 来表示 $n(i)$. 不妨设, 有一个子序列 $\{x^{n_1}, x^{n_2}, \dots\}$ 满足

$$\{x^{n_1}, x^{n_2}, \dots\} \rightarrow x^* \quad \text{以及} \quad \{x_i^{n_1}, x_i^{n_2}, \dots\} \rightarrow x_i^*. \quad (5.16)$$

此外, 由于序列 $\{x^{n_1+1}, x^{n_2+1}, \dots\}$ 也是有界的, 那么其也有极限点. 将其中一个极限点表示为 x^{**} , 那么存在一个子序列 $\{x^{l_1+1}, x^{l_2+1}, \dots\}$ 满足

$$\{x^{l_1+1}, x^{l_2+1}, \dots\} \rightarrow x^{**} \quad \text{以及} \quad \{x_i^{l_1+1}, x_i^{l_2+1}, \dots\} \rightarrow x_i^{**}, \quad (5.17)$$

其中 $\{l_1, l_2, \dots\} \subset \{n_1, n_2, \dots\}$. 在这种情形, 可得

$$\{x^{l_1}, x^{l_2}, \dots\} \rightarrow x^* \quad \text{以及} \quad \{x_i^{l_1}, x_i^{l_2}, \dots\} \rightarrow x_i^*, \quad (5.18)$$

因为这是 (5.16) 的一个子序列. 令 $z_i^n = x_i^n - \mu[\nabla F(x^n)]_i, \forall n \in \mathbb{N}$. 从 (5.18), 有

$$z_i^{l_j} \rightarrow z_i^*, \quad j \rightarrow \infty.$$

因此, 通过引理 5.8, 可得

$$x_i^{**} = \mathcal{T}(z_i^*, x_i^*). \quad (5.19)$$

此外, 由 (5.17)、(5.18) 和 (1.5) 可知, 如下式子成立:

$$x_j^* = x_j^{**}, \quad j \neq i. \quad (5.20)$$

接下来, 通过 $T_\lambda(\cdot)$ 的连续性可知 $\Delta(\cdot, \cdot)$ 的连续性, 满足

$$\Delta(x^{l_j}, x^{l_j+1}) \rightarrow \Delta(x^*, x^{**}).$$

此外, 由于序列 $\{T_\lambda(x^n)\}$ 收敛, 因此,

$$\Delta(x^{l_j}, x^{l_j+1}) = T_\lambda(x^{l_j}) - T_\lambda(x^{l_j+1}) \rightarrow 0, \quad j \rightarrow \infty,$$

这意味着

$$\Delta(x^*, x^{**}) = 0.$$

此外, 通过引理 5.7、(5.19) 和 (5.20), 可得

$$x_i^{**} = x_i^*. \quad (5.21)$$

合并 (5.19) 和 (5.21), 得到

$$x_i^* = \mathcal{T}(z_i^*, x_i^*). \quad (5.22)$$

由于 i 是任意的, 因此, 对于所有 $i \in \{1, \dots, N\}$, 有 (5.22) 成立. 这意味着 x^* 是 $G_{\mu, \lambda\Phi}$ 的一个不动点, 也就是说, $x^* \in \mathcal{F}_\Phi$. 类似地, 因为 $x^* \in \mathcal{L}$ 也是任意的, 所以, $\mathcal{L} \subset \mathcal{F}_\Phi$. 因此, 我们完成了该定理的证明. $\square$

6 结论

本文主要致力于将 Gauss-Seidel 迭代格式用于求解一大类稀疏正则化问题的阈值迭代算法, 并提出了一个新的算法称作 Gauss-Seidel 跳跃迭代阈值 (GSIJT) 算法. 本文的主要贡献是建立所提出算法的收敛性. 概括来说,

(i) 当步长 $0 < \mu < 1/L_{\max}$ 时, 所提出的算法具有支撑和符号有限收敛性. 这意味着 GSIJT 产生的支撑和符号的序列能够收敛并在有限多迭代步内保持不变.

(ii) 结合 rKL 性质, 本文证明了 GSIJT 的全局收敛性并给出了收敛速率的估计. 与 ITA 算法如文献 [15] 中的 IJT 算法相比, IJT 对步长的收敛条件为 $0 < \mu < \frac{1}{L_F}$, 而 GSIJT 为 $0 < \mu < \frac{1}{L_{\max}}$, 其中 L_F 和 $L_{\max}$ 分别在假设 2.2 和 2.1 中说明 $L_{\max} \leq L_F$.

(iii) 一系列的数值实验证明了 GSIJT 的有效性, 特别是通过与 ITA 进行比较, 证明了预期中更快的收敛速度和所需较弱的收敛条件.

致谢 感谢三位审稿人提出的宝贵意见. 基于他们的意见, 本文的质量得以极大的提升. 同时也感谢编委在处理本文时所付出的努力.

参考文献

- Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process Lett*, 2007, 14: 707–710
- Cao W, Sun J, Xu Z. Fast image deconvolution using closed-form thresholding formulas of L_q ($q = \frac{1}{2}, \frac{2}{3}$) regularization. *J Vis Commun Image Represent*, 2013, 24: 31–41
- Chen X, Han Z, Wang Y, et al. Nonconvex plus quadratic penalized low-rank and sparse decomposition for noisy image alignment. *Sci China Inf Sci*, 2016, 59: 052107
- Zeng J, Fang J, Xu Z. Sparse SAR imaging based on $L_{1/2}$ regularization. *Sci China Inf Sci*, 2012, 55: 1755–1775
- Chartrand R, Yin W. Iterative reweighted algorithms for compressed sensing. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Las Vegas: IEEE, 2008, 3869–3872
- Daubechies I, DeVore R, Fornasier M, et al. Iteratively reweighted least squares minimization for sparse recovery. *Comm Pure Appl Math*, 2010, 63: 1–38
- Lu Z. Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming. *Math Program*, 2014, 147: 277–307
- Candés E, Wakin M, Boyd S. Enhancing sparsity by reweighted l_1 minimization. *J Fourier Anal Appl*, 2008, 14: 877–905
- Gasso G, Rakotomamonjy A, Canu S. Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Trans Signal Process*, 2009, 57: 4686–4698
- Blumensath T, Davies M. Iterative thresholding for sparse approximations. *J Fourier Anal Appl*, 2008, 14: 629–654
- Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm Pure Appl Math*, 2004, 57: 1413–1457
- Xu Z, Chang X, Xu F, et al. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans Neural Netw Learn Syst*, 2012, 23: 1013–1027
- Attouch H, Bolte J, Svaiter B. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math Program*, 2013, 137: 91–129
- Zeng J, Lin S, Wang Y, et al. $L_{1/2}$ regularization: Convergence of iterative half thresholding algorithm. *IEEE Trans Signal Process*, 2014, 62: 2317–2329
- Zeng J, Lin S, Xu Z. Sparse regularization: Convergence of iterative jumping thresholding algorithm. *IEEE Trans Signal Process*, 2016, 64: 5106–5118
- Combettes P, Wajs V. Signal recovery by proximal forward-backward splitting. *Multiscale Model Simul*, 2005, 4: 1168–1200
- Tritsiklis J. A comparison of Jacobi and Gauss-Seidel parallel iterations. *Appl Math Lett*, 1989, 2: 167–170
- Peng Z, Wu T, Xu Y, et al. Coordinate-friendly structures, algorithms and applications. *Ann Math Sci Appl*, 2016, 1: 57–119
- Wang Y, Zeng J, Peng Z, et al. Linear convergence of adaptively iterative thresholding algorithms for compressed sensing. *IEEE Trans Signal Process*, 2015, 63: 2957–2971
- Zhu L, Li L, Li R, et al. Model-free feature screening for ultrahigh dimensional data. *J Amer Statist Assoc*, 2011, 106: 1464–1475
- Attouch H, Bolte J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math Program*, 2009, 116: 5–16
- Xu Y, Yin W. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J Imag Sci*, 2013, 6: 1758–1789
- Bagirov A, Jin L, Karimisa N, et al. Subgradient method for nonconvex nonsmooth optimization. *J Optim Theory Appl*, 2013, 157: 416–435

- 24 Burke J, Lewis A, Overton M. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J Optim*, 2005, 15: 751–779
- 25 Chen X. Smoothing methods for nonsmooth, nonconvex minimization. *Math Program*, 2012, 134: 71–99
- 26 Hildreth C. A quadratic programming procedure. *Naval Res Logist*, 1957, 4: 79–85
- 27 Grippo L, Sciandrone M. Globally convergent block-coordinate techniques for unconstrained optimization. *Optim Methods Softw*, 1999, 10: 587–637
- 28 Razaviyayn M, Hong M, Luo Z. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J Optim*, 2013, 23: 1126–1153
- 29 Tseng P. Convergence of a block coordinate descent method for nondifferentiable minimization. *J Optim Theory Appl*, 2001, 109: 475–494
- 30 Luo Z, Tseng P. On the convergence of the coordinate descent method for convex differentiable minimization. *J Optim Theory Appl*, 1992, 72: 7–35
- 31 Beck A, Tetruashvili L. On the convergence of block coordinate descent type methods. *SIAM J Optim*, 2013, 23: 2037–2060
- 32 Tseng P, Yun S. A coordinate gradient descent method for nonsmooth separable minimization. *Math Program*, 2009, 117: 387–423
- 33 Bolte J, Sabach S, Teboulle M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math Program*, 2014, 146: 459–494
- 34 Marjanovic G, Solo V. l_q sparsity penalized linear regression with cyclic descent. *IEEE Trans Signal Process*, 2014, 62: 1464–1475
- 35 Łojasiewicz S. Sur la géométrie semi- et sous- analytique. *Ann Inst Fourier (Grenoble)*, 1993, 43: 1575–1595
- 36 Bolte J, Daniilidis A, Lewis A. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J Optim*, 2007, 17: 1205–1223

Sparse regularized optimization via Gauss-Seidel iterative thresholding

Jinshan Zeng, Tao He & Shikang Ouyang

Abstract This paper considers a class of sparse regularized optimization problems, which are widely studied in many applications such as machine learning, signal processing and image processing. A typical feature of such a class of sparse regularizers is that their induced thresholding functions are discontinuous with jump discontinuities. We develop a Gauss-Seidel based algorithm, called Gauss-Seidel iterative jumping thresholding algorithm (GSIJT) for fast solution to these problems. We first justify the finite convergence of the support and sign of any sequence generated by GSIJT. The convergence property together with the restricted Kurdyka-Łojasiewicz (rKL) property yields the global convergence of GSIJT. Moreover, we give the rate of convergence of GSIJT, and show that any limit point is a stationary point. A series of numerical experiments are provided to show the effectiveness of GSIJT. Particularly, compared with the associated iterative thresholding algorithms, the proposed algorithm converges much faster and can adopt a larger step size.

Keywords sparse regularization, ℓ_q ($0 < q < 1$) regularization, iterative thresholding algorithm, Gauss-Seidel, Kurdyka-Łojasiewicz inequality

MSC(2010) 90C26

doi: 10.1360/SCM-2017-0039