

一类稀疏优化问题的迭代阈值算法及收敛性研究

赵心雨

河北工业大学理学院, 天津

收稿日期: 2026年3月24日; 录用日期: 2026年4月18日; 发布日期: 2026年4月28日

摘要

Nesterov加速策略以其显著的收敛加速效果, 被广泛应用于各类梯度优化算法的设计中。本文将Nesterov加速技术与迭代有限收缩阈值算法相结合, 针对非凸稀疏优化问题提出了基于延拓技术的加速迭代有限收缩阈值算法(A-ILSTAC)和基于截断技术的加速迭代有限收缩阈值算法(A-ILSTAT)。在限制等距性质(RIP)条件下, 本文证明了所提算法能够收敛到近似真实稀疏解, 其误差由噪声水平和有限收缩幅度决定。初步数值实验表明, 本文算法能够找到近似真实稀疏解, 且效果显著优于原始迭代有限收缩阈值算法。

关键词

Nesterov加速策略, 非凸稀疏优化, 迭代有限收缩阈值算法, 限制等距性, 稀疏解

Research on Iterative Thresholding Algorithms and Convergence for a Class of Sparse Optimization Problems

Xinyu Zhao

School of Science, Hebei University of Technology, Tianjin

Received: March 24, 2026; accepted: April 18, 2026; published: April 28, 2026

Abstract

Nesterov's acceleration strategy is renowned in speeding up the convergence of gradient-based optimization algorithms. By merging this strategy and the iterative limited shrinkage thresholding algorithms, we propose an accelerated iterative limited shrinkage thresholding algorithm

with continuation (A-ILSTAC) and the one with truncation (A-ILSTAT) for nonconvex sparse optimization problems, respectively. Under the restricted isometry property (RIP) condition, we prove that the proposed algorithms converge to an approximate true sparse solution within a tolerance relevant to the noise level and the limited shrinkage magnitude. Preliminary numerical results show that our proposed algorithms can find approximate true sparse solutions that are much better than sparse solutions that are found by using original iterative limited shrinkage thresholding algorithms.

Keywords

Nesterov's Acceleration Strategy, Nonconvex Sparse Optimization, Iterative Limited Shrinkage Thresholding Algorithm, Restricted Isometry Property, Sparse Solution

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

考虑变量 x 的欠定线性系统:

$$b = Ax + \varepsilon, \quad (1)$$

其中 $A \in \mathbb{R}^{m \times n}$ ($m \ll n$) 是一个线性变换矩阵, $b \in \mathbb{R}^m$ 是带有未知噪声 $\varepsilon \in \mathbb{R}^m$ 的观测向量。向量 $x \in \mathbb{R}^n$ 的稀疏性定义为 x 的非零分量的个数, 记为 $\|x\|_0$ 。稀疏优化在压缩感知[1] [2]、图像科学[3] [4]、系统生物学[5] [6]和机器学习[7] [8]中有着广泛的应用, 它的目的是寻找这个欠定线性系统的稀疏解, 即只有极少数分量不为零。正则化方法是稀疏优化模型中一类常用的求解方法, 其通过求解如下复合优化模型来逼近真实解:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \lambda \varphi(x), \quad (2)$$

其中 $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$ 是稀疏促进惩罚函数, $\lambda > 0$ 为正则化参数。

当 $\varphi(x)$ 取凸的 ℓ_1 范数时, (2)称为 ℓ_1 正则化问题, 也就是著名的 Lasso 问题。得益于其凸性, 已有大量研究致力于其理论分析[9]-[11]和高效算法设计[3] [12]-[14]。但其精确恢复条件并不容易判断, 并且在许多情形下达不到真实解的稀疏性要求, 不满足无偏性和 oracle 性; 参见[8] [15]-[17]。

为克服这些局限, 研究者转向了能够产生更尖锐、更无偏估计的非凸惩罚函数。其中最具代表性的有三类:

- 平滑剪切绝对偏差(SCAD) [16]: 在原点附近具有与 ℓ_1 相似的惩罚以产生稀疏性, 但对大系数的惩罚趋于常数, 从而避免了过度收缩。
- 极小极大凹惩罚(MCP) [18]: 具有凹函数形式, 其导数在阈值后逐渐衰减至零, 实现了对大系数的近似无偏估计。
- ℓ_p 拟范数($0 < p < 1$) [19]: 惩罚函数为 $\varphi(x) = \sum_i |x_i|^p$ 。其非凸几何形态在原点附近比 ℓ_1 更陡峭, 从而对微小系数施加更强的压制力, 已被证明在理论上能以更少的测量次数恢复稀疏信号。

这些非凸惩罚函数在理论上的优越性已被广泛证实: 它们通常具有更强的稀疏促进能力, 且在一定条件下能产生具有 Oracle 性质的估计量。然而, 非凸性也带来了严峻的计算挑战: 问题(2)通常是一个多模

态的非凸优化问题,寻找其全局最优解是 NP-难的,参见[20]。因此,设计能够高效、稳定地找到高质量解的算法,并为之建立坚实的收敛性理论,成为了该领域的核心议题。

在算法层面,迭代阈值算法(ITA)因其形式简洁、计算复杂度低而成为求解此类问题的主流框架之一,参见[8][12][17][21]-[23]。ITA 本质上是应用于问题(2)的邻近梯度算法(PGA),其中阈值函数对应着惩罚函数 φ 的邻近算子。例如,软阈值、硬阈值和半阈值算子分别对应于 ℓ_0 、 ℓ_1 和 ℓ_1 惩罚。遗憾的是,对于非凸问题,标准 PGA 的收敛理论通常仅能保证迭代序列收敛到一个稳定点,且收敛速率通常为次线性的 $O(1/k)$,这在处理大规模问题时可能效率不足,参见[24][25]。

为了提升 ITA 的收敛速度与实用性能,研究者们主要发展了两种策略:延拓技术与截断技术。延拓技术通过从一个较大的正则化参数 λ 开始,沿一个递减序列进行求解,旨在提供更好的初始点并加速收敛,已在 ℓ_1 和 ℓ_0 问题中展现出理论优越性,参见[13][14][26][27]。截断技术则在迭代过程中主动保留最大的若干分量并强制其余分量为零,从而显式控制解的稀疏结构,其典型代表是迭代硬阈值算法[21][28][29]。然而,现有工作几乎完全集中于 ℓ_1 或 ℓ_0 惩罚,对于更一般的非凸惩罚(如 SCAD、MCP、 ℓ_p),如何系统地将这些加速策略与 ITA 结合,并建立严格的收敛性理论,仍然是一个未被充分探索的开放性课题。

最近, Hu 等人[30]通过引入有限收缩阈值算子这一统一框架,成功地将 SCAD、MCP 和 ℓ_p 惩罚的近端映射纳入同一分析范式,并提出了结合延拓技术与截断技术的迭代有限收缩阈值算法(ILSTAC/ILSTAT)。在限制等距性质(RIP)的假设下,他们首次为这类非凸 ITA 建立了逼近真实稀疏解的强收敛理论。这一开创性工作为非凸稀疏优化的加速算法设计奠定了重要基础。然而,其算法框架仍基于标准的(近端)梯度迭代,在收敛速率上存在固有瓶颈。

与此同时,在凸优化领域, Nesterov 加速技术[31]已被证明能将梯度方法的收敛速率从 $O(1/k)$ 提升至最优的 $O(1/k^2)$,其思想通过引入动量项来智能地结合历史梯度信息。经典的 FISTA 算法便是 Nesterov 加速与邻近算子结合的杰出代表,在 ℓ_1 正则化问题中取得了巨大成功。一个自然且极具吸引力的问题是:能否将 Nesterov 加速策略有机地融入针对非凸惩罚的统一 ITA 框架中,从而在保持理论保证的同时,显著提升算法的收敛速度?

本文旨在正面回应上述挑战,首次系统地将 Nesterov 加速策略融入非凸稀疏优化的统一 ITA 框架中。在有限收缩阈值算子的统一基础上,我们设计了两种新型 Nesterov 加速迭代算法: Nesterov 加速的迭代有限收缩阈值算法(带延拓技术)(A-ILSTAC)和 Nesterov 加速的迭代有限收缩阈值算法(带截断技术)(A-ILSTAT);参见算法 1 和算法 2。在 RIP 假设下,我们证明了 A-ILSTAC 的输出收敛到原问题的近似真稀疏解,其支撑集包含于真实支撑集之内,且误差界由噪声水平决定;并且 A-ILSTAT 生成的序列以几何速率收敛于(1)的近似真稀疏解。

数值实验表明, A-ILSTAC 和 A-ILSTAT 算法具有更强的稀疏恢复能力,相比于原始的 ILSTAC 和 ILSTAT 算法,收敛速度更快,精确度更高。

本文后续章节安排如下:第二节介绍必要的预备知识与有限收缩算子框架。第三节详细提出 A-ILSTAC 与 A-ILSTAT 算法。第四节致力于算法的收敛性理论分析。第五节通过数值实验展示算法的卓越性能。最后,第六节总结全文。

2. 预备知识

限制等距性质(RIP)是一个著名的正则性条件,用于度量感知矩阵在稀疏子空间上的子矩阵接近标准正交的程度。

定义 1 设 $A \in \mathbb{R}^{m \times n}$ 且 $s \in \mathbb{N}$ 。矩阵 A 的 s -限制等距常数 δ_s 定义为满足以下条件的最小非负数:

$$(1 - \delta_s) \|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_s) \|x\|^2, \forall x \in \mathbb{R}^n, \|x\|_0 \leq s.$$

如果 $\delta_s < 1$ ，则称矩阵 A 关于常数 δ_s 满足 s -限制等距性质(s -RIP)。

根据定义 1，常数 δ_s 关于 s 是非递减的，即当 $s \leq t$ 时 $\delta_s \leq \delta_t$ 。下述引理回顾了 s -RIP 的几个关键性质，这些性质将为本文后续分析算法的收敛性提供重要工具。

引理 1 设 $x \in \mathbb{R}^n$ ， $\varepsilon \in \mathbb{R}^m$ ， $\mathcal{M}, \mathcal{N} \subseteq [n]$ ，且 $v \in \left[0, \frac{1}{1 - \delta_s}\right]$ 。若矩阵 A 关于常数 δ_s 满足 s -RIP，则下列结论成立：

- (i) 如果 $\#(\mathcal{M} \cup \text{supp}(x)) \leq s$ ，那么 $\left\| \left((\mathbb{I} - vA^T A)x \right)_{\mathcal{M}} \right\| \leq (1 - v + v\delta_s) \|x\|$ 。
- (ii) 如果 $\#(\mathcal{M}) \leq s$ ，那么 $\|A_{\mathcal{M}}^T \varepsilon\| \leq \sqrt{1 + \delta_s} \|\varepsilon\|$ 。
- (iii) 如果 $\mathcal{M} \cap \mathcal{N} = \emptyset$ 且 $\#(\mathcal{M} \cup \mathcal{N}) \leq s$ ，那么 $\|A_{\mathcal{N}}^T A_{\mathcal{M}} x_{\mathcal{M}}\| \leq \delta_s \|x_{\mathcal{M}}\|$ 。

在介绍算法之前，本文先引入有限收缩阈值算子的定义。

定义 2 设 $\kappa: \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ 为一函数， $\lambda \in \mathbb{R}_+$ 为参数。

(i) 若算子 $\mathbb{T}_{\lambda}: \mathbb{R} \rightarrow \mathbb{R}$ 满足以下阈值条件：

$$\mathbb{T}_{\lambda}(t) = 0, \forall |t| \leq \kappa(\lambda), \tag{3}$$

则称 \mathbb{T}_{λ} 为关于 $\kappa(\lambda)$ 的阈值算子。

(ii) 若算子 $\mathbb{L}\mathbb{T}_{\lambda}: \mathbb{R} \rightarrow \mathbb{R}$ 满足上述(3)式，并进一步具有以下有限收缩特性：

$$|\mathbb{L}\mathbb{T}_{\lambda}(t) - t| \leq \kappa(\lambda), \forall t \in \mathbb{R}, \tag{4}$$

则称 $\mathbb{L}\mathbb{T}_{\lambda}$ 为关于 $\kappa(\lambda)$ 的有限收缩阈值算子。

全体满足上述条件的算子 $\mathbb{L}\mathbb{T}_{\lambda}$ 构成的集合记为 $\mathcal{LT}(\kappa; \lambda)$ ，称为有限收缩阈值算子族。在后续分析中，本文对该类算子提出如下结构性假设：

假设 1 设 $\alpha, \beta > 0$ ，且令 $\kappa(\lambda) = \alpha\lambda^{\beta}$ 。假设对任意 $\lambda > 0$ ，有 $\mathbb{L}\mathbb{T}_{\lambda} \in \mathcal{LT}(\kappa; \lambda)$ 。

有限收缩阈值算子为多种非凸稀疏惩罚的近端映射提供了统一的分析框架。特别地，SCAD、MCP 以及 ℓ_p ($0 < p < 1$) 等常用非凸惩罚的近端算子均可视为该算子的特例。

除此之外，在本文中，采用如下记法，设 \bar{x} 为问题(1)的一个 s -稀疏解，其支撑集记为 $\mathcal{S} := \text{supp}(\bar{x})$ 。

3. 一类新的加速迭代有限收缩阈值算法

Hu 等人[30]提出了迭代有限收缩阈值算法具有形式简单、计算复杂度低的特点，其一般形式可表示为

$$x^{k+1} := \mathbb{L}\mathbb{T}_{v\lambda} \left(x^k - vA^T (Ax^k - b) \right),$$

其中 $\mathbb{L}\mathbb{T}_{v\lambda}$ 是关于 $\kappa(v\lambda)$ 的有限收缩阈值算子。然而，其算法框架仍基于标准的(近端)梯度迭代，在收敛速率上存在固有瓶颈。与此同时，在凸优化领域，Nesterov 加速技术[31]已被证明能将梯度方法的收敛速率从 $O(1/k)$ 提升至最优的 $O(1/k^2)$ ，其思想通过引入动量项来智能地结合历史梯度信息。经典的 FISTA 算法便是 Nesterov 加速与邻近算子结合的杰出代表，在 ℓ_1 正则化问题中取得了巨大成功。受 FISTA 算法的启发，我们考虑将 Nesterov 加速策略有机地融入针对非凸惩罚的统一 ITA 框架中，来提升算法的收敛速度。基于此，本文在 ILSTAC 算法和 ILSTAT 算法的基础上加入 Nesterov 动量项，得到新的算法 A-ILSTAC 和 A-ILSTAT，改进后的算法分别如下：

算法1 A-ILSTAC 算法

- 1: 初始化正则化参数 $\lambda_0 > 0$, $\lambda > 0$ 和延拓参数 $\lambda \in (0,1)$, 初始点 $x^0 := x^{-1} := 0$, 步长 $v > 0$ 和 $k = 0$.
- 2: if $\lambda_k < \lambda$ then
- 3: break
- 4: else
- 5: $y^{k+1} := x^k + \frac{k-1}{k+2}(x^k - x^{k-1})$
- 6: $x^{k+1} := \mathbb{L}T_{v\lambda_k}(y^{k+1} - vA^T(Ay^{k+1} - b))$
- 7: $\lambda_k := \gamma\lambda_k$
- 8: $k = k + 1$
- 9: end if
- 10: 输出 $x^* := x^k$

和

算法2 A-ILSTAT 算法

- 1: 初始化: 正则化参数 $\lambda > 0$, 截断参数 $s \in \mathbb{N}$, 随机初始点 $x^0 \in \mathbb{R}^n$, $x^{-1} \in \mathbb{R}^n$, 步长 $v > 0$ 和 $k = 0$.
- 2: 迭代: 对每个 $k \in \mathbb{N}$, 已知当前迭代点 x^k , 按以下公式更新 x^{k+1}

$$y^{k+1} := x^k + \frac{k-1}{k+2}(x^k - x^{k-1}),$$

$$x^{k+1} = \mathbb{H}_s \circ LT_{v\lambda}(y^{k+1} - vA^T(Ay^{k+1} - b)).$$

4. 收敛性分析

在本节中, 分析 A-ILSTAC 和 A-ILSTAT 算法的收敛性。证明了在 RIP 假设下, 算法 1 在与噪声水平成比例的容差范围内收敛到问题(1)的近似真稀疏解, 且误差界由噪声水平决定。此外, 算法 1 输出解的支撑集不会产生伪预测, 并且一定是真稀疏解支撑集的子集。

定理 1 设假设 1 成立, 矩阵 A 满足 s -RIP 且

$$v\delta_{s+1}(\sqrt{s}+1)(2v-v\delta_s) < 3v-2-2v\delta_s, \quad v\delta_{s+1}(\sqrt{s}+1) < 2v-v\delta_s, \quad (5)$$

取

$$\eta \in (0, 1 - \max\{A, B\}), \quad (6)$$

其中 $A = \frac{v\delta_{s+1}(\sqrt{s}+1)(2v-v\delta_s)}{3v-2-2v\delta_s}$, $B = \frac{v\delta_{s+1}(\sqrt{s}+1)}{2v-v\delta_s}$ 。步长设为 $\frac{2}{3-2\delta_s} < v < \frac{1}{1-\delta_s}$, 正则化参数设为

$$\lambda_0 \geq \frac{1}{v} \left(\frac{v\delta_{s+1} \|\bar{x}\|}{\alpha \cdot \max\{A, B\}} \right)^{\frac{1}{\beta}}, \quad \lambda = \frac{1}{v} \left(\frac{v\sqrt{1+\delta_s} \|\varepsilon\|}{\alpha\eta} \right)^{\frac{1}{\beta}}, \quad (7)$$

延拓参数设为

$$\gamma \in \left[C^{\frac{1}{\beta}}, 1 \right), \quad (8)$$

其中 $C = \frac{2(1-\nu) + \frac{\nu\delta_{s+1}}{1-\eta}(\sqrt{s}+1) + 2\sqrt{\left[2(1-\nu) + \frac{\nu\delta_{s+1}}{1-\eta}(\sqrt{s}+1) + \nu\delta_s\right]^2 + 4(1-\nu + \nu\delta_s)}}{2}$ 。若算法 1 基于上述参数输出 x^* ，则有

$$\text{supp}(x^*) \subseteq S, \|x^* - \bar{x}\| \leq \frac{(1-\eta)\sqrt{1+\delta_s}}{\eta\delta_{s+1}} \|\varepsilon\|. \tag{9}$$

证明：根据假设(5)， $1 - \max\{A, B\} > 0$ ，因此(6)式中 η 是定义良好的。由(6)式可知， $C > 1$ 。因此(8)式中的 γ 和使用这些参数的算法 1 是定义良好的。为了进行证明，让算法 1 生成有限序列 $\{x^k\}_{k=0}^K$ 并输出 $x^* = x^K$ ，记

$$\rho = \frac{1-\eta}{\delta_s}, \tag{10}$$

并对每个 $k = 0, \dots, K$ 定义

$$\begin{aligned} y^{k+1} &= x^k + \frac{k-1}{k+2}(x^k - x^{k-1}), \\ z^k &= y^{k+1} - \nu A^T(Ay^{k+1} - b), \\ \mathcal{S}_k &= \text{supp}(x^k), \\ r^k &= x^k - \bar{x}, \\ x^{k+1} &= \mathbb{L}\mathbb{T}_{\nu\lambda_k}(z^k). \end{aligned} \tag{11}$$

由假设 1 知，当取 $\kappa(\lambda) = \alpha\lambda^\beta$ 时，(3)式与(4)式均成立。下面通过归纳法证明：对每个 $k = 0, \dots, K$ ，有

$$\mathcal{S}_k \subseteq S, \|r^k\| \leq \frac{\rho}{\nu} \kappa(\nu\lambda_k) = \rho\alpha\nu^{\beta-1}\lambda_k^\beta.$$

取初始点 $x^0 = x^{-1} = 0$ ，则 $\mathcal{S}_0 = \emptyset \subseteq S$ 。根据(10)式中 ρ 的定义及假设(6)，并结合(7)中对 λ_0 的选取，可得

$$\rho\alpha\nu^{\beta-1}\lambda_0^\beta > \frac{\max\{A, B\} \cdot \alpha\nu^{\beta-1}\lambda_0^\beta}{\delta_{s+1}} \geq \|\bar{x}\| = \|r^0\|. \tag{12}$$

因此，(12)式对 $k = 0$ 成立。

现假设(12)式对第 k ($< K$)次迭代成立。由(11)式及(1)式可得

$$\begin{aligned} z^k &= y^{k+1} - \nu A^T(Ay^{k+1} - b) \\ &= x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) - \nu A^T\left(A\left(x^k + \frac{k-1}{k+2}(x^k - x^{k-1})\right) - b\right) \\ &= x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) - \frac{k-1}{k+2}\nu A^T A(x^k - x^{k-1}) - \nu A^T Ar^k + \nu A^T \varepsilon \\ &= x^k + \frac{k-1}{k+2}(\mathbb{I} - \nu A^T A)(x^k - x^{k-1}) - \nu A^T Ar^k + \nu A^T \varepsilon \\ &= x^k + \frac{k-1}{k+2}(\mathbb{I} - \nu A^T A)(x^k - x^{k-1}) - \nu A^T A_S r_S^k + \nu A^T \varepsilon, \end{aligned} \tag{13}$$

其中最后一个等式利用了归纳假设(12)中的 $\mathcal{S}_k \subseteq S$ 。任取 $i \in S^c$ ，由归纳假设 $\mathcal{S}_k \subseteq S$ 可知 $x_i^k = 0$ ， $x_i^{k-1} = 0$ ，于是(13)式化为

$$|z_i^k| \leq v |A_i^T A_S r_S^k| + v |A_i^T \varepsilon|. \quad (14)$$

由于 $\{i\} \cap S = \emptyset$ ，根据引理 1 (iii)和(ii)可得

$$|A_i^T A_S r_S^k| \leq \delta_{s+1} \|r_S^k\| = \delta_{s+1} \|r^k\|, \quad |A_i^T \varepsilon| \leq \sqrt{1+\delta_1} \|\varepsilon\| \leq \sqrt{1+\delta_s} \|\varepsilon\|.$$

其中最后一个不等式利用了 $\delta_1 \leq \delta_s$ 的单调性。结合(14)式，得到

$$|z_i^k| \leq v \delta_{s+1} \|r^k\| + v \sqrt{1+\delta_s} \|\varepsilon\|. \quad (15)$$

由算法的停止准则 $\lambda_k < \lambda$ 及(7)式中 λ 的定义，有 $\lambda_k \geq \lambda = \frac{1}{v} \left(\frac{v\sqrt{1+\delta_s}}{\alpha\eta} \|\varepsilon\| \right)^\beta$ ，即 $\|\varepsilon\| \leq \frac{\alpha\eta}{v\sqrt{1+\delta_s}} (v\lambda_k)^\beta$ 。结合

假设(12)，将上述估计代入(15)得

$$|z_i^k| \leq v \delta_{s+1} \|r^k\| + v \sqrt{1+\delta_s} \|\varepsilon\| \leq \delta_{s+1} \rho \alpha (v\lambda_k)^\beta + \alpha \eta (v\lambda_k)^\beta = \alpha (v\lambda_k)^\beta, \quad (16)$$

其中等式利用了(10)中 ρ 的定义。根据阈值条件(3)可知 $x_i^{k+1} = 0$ ，从而 $i \in S_{k+1}^c$ 。上述推导对任意 $i \in S^c$ 均成立，因此有 $S^c \subseteq S_{k+1}^c$ ，即 $S_k \subseteq S$ 。

另一方面，由包含关系 $S_k \subseteq S$ 可得

$$\|x^{k+1} - \bar{x}\| = \|x_S^{k+1} - \bar{x}_S\| \leq \|x_S^{k+1} - z_S^k\| + \|z_S^k - \bar{x}\|. \quad (17)$$

根据(4)式和算法 1 中 $x^{k+1} = \mathbb{L}\mathbb{T}_{v\lambda_k}(z^k)$ ，有

$$\|x_S^{k+1} - z_S^k\| \leq \sqrt{s} \|x^{k+1} - z^k\|_\infty = \sqrt{s} \|\mathbb{L}\mathbb{T}_{v\lambda_k}(z^k) - z^k\|_\infty \leq \sqrt{s} \kappa (v\lambda_k), \quad (18)$$

根据(11)和(13)式，可得

$$\begin{aligned} \|z_S^k - \bar{x}_S\| &= \left\| \left(x^k + \frac{k-1}{k+2} (\mathbb{I} - vA^T A) (x^k - x^{k-1}) - vA^T A_S r_S^k + vA^T \varepsilon \right)_S - \bar{x}_S \right\| \\ &= \left\| \frac{k-1}{k+2} ((\mathbb{I} - vA^T A) (x^k - x^{k-1}))_S + ((\mathbb{I} - vA^T A) r^k)_S + vA_S^T \varepsilon \right\|. \end{aligned} \quad (19)$$

然后推出

$$\|z_S^k - \bar{x}_S\| \leq \frac{k-1}{k+2} \left\| ((\mathbb{I} - vA^T A) (x^k - x^{k-1}))_S \right\| + \left\| ((\mathbb{I} - vA^T A) r^k)_S \right\| + \|vA_S^T \varepsilon\|.$$

根据引理 1 (i)和(iii)有

$$\begin{aligned} \left\| ((\mathbb{I} - vA^T A) (x^k - x^{k-1}))_S \right\| &\leq (1-v+v\delta_s) \|x^k - x^{k-1}\|, \\ \left\| ((\mathbb{I} - vA^T A) r^k)_S \right\| &\leq (1-v+v\delta_s) \|r^k\|, \\ \|vA_S^T \varepsilon\| &\leq v\sqrt{1+\delta_s} \|\varepsilon\|. \end{aligned}$$

从而有

$$\begin{aligned} \|z_S^k - \bar{x}_S\| &\leq \frac{k-1}{k+2} (1-v+v\delta_s) \|x^k - x^{k-1}\| + (1-v+v\delta_s) \|r^k\| + v\sqrt{1+\delta_s} \|\varepsilon\| \\ &\leq (1-v+v\delta_s) \|x^k - x^{k-1}\| + (1-v+v\delta_s) \|r^k\| + v\sqrt{1+\delta_s} \|\varepsilon\|. \end{aligned} \quad (20)$$

根据 $\delta_s \leq \delta_{s+1}$ 及(16)式，我们得到

$$v\delta_s \|r^k\| + v\sqrt{1+\delta_s} \|\varepsilon\| \leq v\delta_{s+1} \|r^k\| + v\sqrt{1+\delta_s} \|\varepsilon\| \leq \kappa(v\lambda_k).$$

因此(20)式可以化为

$$\|z_s^k - \bar{x}_s\| \leq (1-v+v\delta_s) \|x^k - x^{k-1}\| + (1-v) \|r^k\| + \kappa(v\lambda_k). \quad (21)$$

根据 $\lambda_{k+1} = \gamma\lambda_k$ 及(12)式, 有

$$\begin{aligned} \|r^k\| &\leq \frac{\rho}{v} \kappa(v\lambda_k) = \rho\alpha v^{\beta-1} \lambda_k^\beta, \\ \|r^{k-1}\| &\leq \frac{\rho}{v} \kappa(v\lambda_{k-1}) = \rho\alpha v^{\beta-1} \lambda_{k-1}^\beta = \frac{\rho\alpha v^{\beta-1} \lambda_k^\beta}{\gamma^\beta}. \end{aligned}$$

从而有

$$\|x^k - x^{k-1}\| = \|r^k - r^{k-1}\| \leq \|r^k\| + \|r^{k-1}\| \leq \rho\alpha v^{\beta-1} \lambda_k^\beta + \frac{\rho\alpha v^{\beta-1} \lambda_k^\beta}{\gamma^\beta} = \left(1 + \frac{1}{\gamma^\beta}\right) \rho\alpha v^{\beta-1} \lambda_k^\beta.$$

由(17)式, (18)式及(21)式, 可得

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq \|x_s^{k+1} - z_s^k\| + \|z_s^k - \bar{x}\| \\ &\leq \sqrt{s} \kappa(v\lambda_k) + (1-v+v\delta_s) \|x^k - x^{k-1}\| + (1-v) \|r^k\| + \kappa(v\lambda_k) \\ &\leq \sqrt{s} \kappa(v\lambda_k) + (1-v+v\delta_s) \left(1 + \frac{1}{\gamma^\beta}\right) \rho\alpha v^{\beta-1} \lambda_k^\beta + (1-v) \rho\alpha v^{\beta-1} \lambda_k^\beta + \kappa(v\lambda_k) \\ &\leq \sqrt{s} \kappa(v\lambda_k) + (1-v+v\delta_s) \frac{(1+\gamma^\beta)\rho}{v\gamma^\beta} \kappa(v\lambda_k) + \frac{(1-v)\rho}{v} \kappa(v\lambda_k) + \kappa(v\lambda_k). \end{aligned}$$

注意到

$$\begin{aligned} &\sqrt{s} \kappa(v\lambda_k) + (1-v+v\delta_s) \frac{(1+\gamma^\beta)\rho}{v\gamma^\beta} \kappa(v\lambda_k) + \frac{(1-v)\rho}{v} \kappa(v\lambda_k) + \kappa(v\lambda_k) \\ &= \left[\sqrt{s} + (1-v+v\delta_s) \frac{(1+\gamma^\beta)\rho}{v\gamma^\beta} + \frac{(1-v)\rho}{v} + 1 \right] \kappa(v\lambda_k) \\ &= \frac{\rho}{v} \left[\frac{v}{\rho} (\sqrt{s} + 1) + (1-v+v\delta_s) \left(1 + \frac{1}{\gamma^\beta}\right) + 1 - v \right] \kappa(v\lambda_k). \end{aligned}$$

因此我们得到

$$\|x^{k+1} - \bar{x}\| \leq \frac{\rho}{v} \left[\frac{v}{\rho} (\sqrt{s} + 1) + (1-v+v\delta_s) \left(1 + \frac{1}{\gamma^\beta}\right) + 1 - v \right] \kappa(v\lambda_k). \quad (22)$$

根据 ρ 和 γ 的定义, 有

$$\frac{v}{\rho} (\sqrt{s} + 1) + (1-v+v\delta_s) \left(1 + \frac{1}{\gamma^\beta}\right) + 1 - v = \frac{v\delta_{s+1}}{1-\eta} (\sqrt{s} + 1) + (1-v+v\delta_s) \left(1 + \frac{1}{\gamma^\beta}\right) + 1 - v \leq \gamma^\beta.$$

结合 $\lambda_k = \gamma\lambda_{k-1}$, (22)式可以化为

$$\|x^{k+1} - \bar{x}\| \leq \frac{\rho}{v} \kappa(v\lambda_{k+1}).$$

由此, 结合 $\mathcal{S}_{k+1} \subseteq \mathcal{S}$ 可知, (12)式对每个迭代步 $k = 0, \dots, K$ 均成立. 进而由(12)式可推出 $\text{supp}(x^*) \subseteq \mathcal{S}$, 且

$$\|x^* - \bar{x}\| \leq \rho \alpha v^{\beta-1} \lambda_K^\beta < \rho \alpha v^{\beta-1} \lambda^\beta = \frac{(1-\eta)\sqrt{1+\delta_s}}{\eta\delta_{s+1}} \|\varepsilon\|,$$

其中最后的等式利用了(7)中 λ 的定义及(10)中 ρ 的定义。因此(9)得证。证明完成。

在定理 2 中, 我们证明了在 RIP 假设下, 算法 2 以几何速率收敛到问题(1)的近似真稀疏解。

定理 2 设假设 1 成立, 且 A 满足 $3s$ -RIP。令 $\{x^k\}$ 为算法 2 生成的序列, 步长 v 满足

$$\frac{5}{6(1-\delta_{3s})} < v \leq \frac{1}{1-\delta_{3s}}. \quad (23)$$

那么 $\{x^k\}$ 以几何速率近似地收敛于 \bar{x} ; 具体有

$$\|x^k - \bar{x}\| \leq \rho^k \|x^0 - \bar{x}\| + \frac{2}{1-\rho} (v\sqrt{1+\delta_{2s}} \|\varepsilon\| + \sqrt{2s}\kappa(v\lambda)), \quad (24)$$

其中 $\rho = 2(1-v+v\delta_{3s}) \in \left(0, \frac{1}{3}\right)$ 。

证明: 由假设 1 知, (3)和(4)式在 $\kappa(\lambda) = \alpha\lambda^\beta$ 时成立。为了进行收敛性分析, 将算法 2 的过程改写为以下四步:

$$\begin{aligned} y^{k+1} &= x^k + \frac{k-1}{k+2} (x^k - x^{k-1}), \\ z^k &= y^{k+1} - vA^T (Ay^{k+1} - b), \\ t^k &= \mathbb{L}\mathbb{T}_{v\lambda}(z^k), \\ x^{k+1} &= \mathbb{H}_s(t^k). \end{aligned} \quad (25)$$

为简化符号, 记

$$r^k = x^k - \bar{x}, \quad \mathcal{S}_k = \text{supp}(x^k), \quad \mathcal{I}^k = \mathcal{S} \cup \mathcal{S}_k; \quad (26)$$

从而由定义可得

$$\#(\mathcal{S}) = s, \quad \#(\mathcal{S}_k) \leq s, \quad \#(\mathcal{I}^k) \leq 2s \quad \forall k \in \mathbb{N}. \quad (27)$$

注意到

$$\text{supp}(\bar{x}) = \mathcal{S} \subseteq \mathcal{I}_{k+1}, \quad \text{supp}(x^{k+1}) = \mathcal{S}_{k+1} \subseteq \mathcal{I}_{k+1}, \quad (28)$$

根据(26)式, 得到

$$\|r^{k+1}\| = \|x_{\mathcal{I}_{k+1}}^{k+1} - \bar{x}_{\mathcal{I}_{k+1}}\| \leq \|x_{\mathcal{I}_{k+1}}^{k+1} - t_{\mathcal{I}_{k+1}}^k\| + \|t_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\|. \quad (29)$$

注意到由(25)式有 $x^{k+1} = \mathbb{H}_s(t^k) = \arg \min_{\|x\|_0 \leq s} \|x - t^k\|$, 由于 $\|\bar{x}\|_0 = s$, 得到 $\|t^k - x^{k+1}\| \leq \|t^k - \bar{x}\|$ 。结合(28)式, 推出 $\|t_{\mathcal{I}_{k+1}}^k - x_{\mathcal{I}_{k+1}}^{k+1}\| \leq \|t_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\|$ 。因此(29)可化简为

$$\|r^{k+1}\| \leq 2\|t_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\| \leq 2(\|t_{\mathcal{I}_{k+1}}^k - z_{\mathcal{I}_{k+1}}^k\| + \|z_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\|). \quad (30)$$

由(25)式有 $t^k = \mathbb{L}\mathbb{T}_{v\lambda}(z^k)$, 根据(4), 有 $\|t^k - z^k\|_\infty \leq \kappa(v\lambda)$ 。结合(27)式, 得到

$$\|t_{\mathcal{I}_{k+1}}^k - z_{\mathcal{I}_{k+1}}^k\| \leq \sqrt{\#(\mathcal{I}_{k+1})} \|t^k - z^k\|_\infty \leq \sqrt{2s}\kappa(v\lambda). \quad (31)$$

另一方面, 由(25)式和(1)式, 得到

$$\begin{aligned}
 z^k &= y^{k+1} - vA^T (Ay^{k+1} - b) \\
 &= x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) - vA^T \left(A \left[x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) \right] - b \right) \\
 &= x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) - vA^T A \left(x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) \right) + vA^T (A\bar{x} + \varepsilon) \\
 &= x^k - vA^T Ax^k + vA^T A\bar{x} + vA^T \varepsilon + \frac{k-1}{k+2}(x^k - x^{k-1}) - vA^T A \frac{k-1}{k+2}(x^k - x^{k-1}) \\
 &= x^k - vA^T Ar^k + vA^T \varepsilon + \frac{k-1}{k+2}(\mathbb{I} - vA^T A)(x^k - x^{k-1});
 \end{aligned} \tag{32}$$

因此,

$$\begin{aligned}
 \|z_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\| &= \left\| \left(x^k - \bar{x} - vA^T Ar^k + vA^T \varepsilon + \frac{k-1}{k+2}(\mathbb{I} - vA^T A)(x^k - x^{k-1}) \right)_{\mathcal{I}_{k+1}} \right\| \\
 &= \left\| \left((\mathbb{I} - vA^T A)r^k + vA^T \varepsilon + \frac{k-1}{k+2}(\mathbb{I} - vA^T A)(x^k - x^{k-1}) \right)_{\mathcal{I}_{k+1}} \right\| \\
 &= \left\| \left((\mathbb{I} - vA^T A)r^k + vA^T \varepsilon + \frac{k-1}{k+2}(\mathbb{I} - vA^T A)(r^k - r^{k-1}) \right)_{\mathcal{I}_{k+1}} \right\| \\
 &= \left\| \left((\mathbb{I} - vA^T A)r^k + \frac{k-1}{k+2}(\mathbb{I} - vA^T A)r^k + vA^T \varepsilon - \frac{k-1}{k+2}(\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\| \\
 &= \left\| \frac{2k+1}{k+2} \left((\mathbb{I} - vA^T A)r^k \right)_{\mathcal{I}_{k+1}} + vA_{\mathcal{I}_{k+1}}^T \varepsilon - \frac{k-1}{k+2} \left((\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\| \\
 &\leq \frac{2k+1}{k+2} \left\| \left((\mathbb{I} - vA^T A)r^k \right)_{\mathcal{I}_{k+1}} \right\| + \|vA_{\mathcal{I}_{k+1}}^T \varepsilon\| + \frac{k-1}{k+2} \left\| \left((\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\|.
 \end{aligned}$$

由(27)式和(28)式, 可以推出 $\#(\mathcal{I}_{k+1}) \leq 2s$ 和 $\#(\mathcal{I}_{k+1} \cup \text{supp}(r^k)) = \#(\mathcal{I}_{k+1} \cup \mathcal{S}_k) \leq 3s$ 。根据 A 满足 $3s$ -RIP 并应用引理 1 (i)和(ii), 有

$$\begin{aligned}
 \left\| \left((\mathbb{I} - vA^T A)r^k \right)_{\mathcal{I}_{k+1}} \right\| &\leq (1 - v + v\delta_{3s}) \|r^k\|, \\
 \left\| \left((\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\| &\leq (1 - v + v\delta_{3s}) \|r^{k-1}\|, \\
 \|vA_{\mathcal{I}_{k+1}}^T \varepsilon\| &\leq v\sqrt{1 + \delta_{2s}} \|\varepsilon\|.
 \end{aligned}$$

根据上述三个不等式, 我们推出

$$\begin{aligned}
 \|z_{\mathcal{I}_{k+1}}^k - \bar{x}_{\mathcal{I}_{k+1}}\| &\leq \frac{2k+1}{k+2} \left\| \left((\mathbb{I} - vA^T A)r^k \right)_{\mathcal{I}_{k+1}} \right\| + \|vA_{\mathcal{I}_{k+1}}^T \varepsilon\| + \frac{k-1}{k+2} \left\| \left((\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\| \\
 &\leq 2 \left\| \left((\mathbb{I} - vA^T A)r^k \right)_{\mathcal{I}_{k+1}} \right\| + \|vA_{\mathcal{I}_{k+1}}^T \varepsilon\| + \left\| \left((\mathbb{I} - vA^T A)r^{k-1} \right)_{\mathcal{I}_{k+1}} \right\| \\
 &\leq 2(1 - v + v\delta_{3s}) \|r^k\| + v\sqrt{1 + \delta_{2s}} \|\varepsilon\| + (1 - v + v\delta_{3s}) \|r^{k-1}\|.
 \end{aligned}$$

结合(30)式与(31)式, 得到递推关系式

$$\|r^{k+1}\| \leq 2\sqrt{2s}\kappa(v\lambda) + 4(1 - v + v\delta_{3s}) \|r^k\| + 2v\sqrt{1 + \delta_{2s}} \|\varepsilon\| + 2(1 - v + v\delta_{3s}) \|r^{k-1}\|.$$

容易推出

$$\|x^k - \bar{x}\| \leq (\rho + \sqrt{\rho^2 + \rho})^k (\|x^0 - \bar{x}\| + \|x^{-1} - \bar{x}\|) + \frac{2\sqrt{2s}\kappa(v\lambda) + 2v\sqrt{1 + \delta_{2s}}\|\varepsilon\|}{1 - (\rho + \sqrt{\rho^2 + \rho})},$$

其中 $\rho = 2(1 - v + v\delta_{3s}) \in (0, \frac{1}{3})$ 。证明完成。

5. 数值实验

本节通过数值实验展示提出的 A-ILSTAC 和 A-ILSTAT 算法在求解问题(2)时的性能表现, 惩罚函数分别取 SCAD、MCP 和 ℓ_p 惩罚($p = \frac{1}{2}$), 并分别与原始的 ILSTAC 和 ILSTAT 算法进行比较。

实验数据通过压缩感知的标准流程生成。具体而言, 随机生成独立同分布的高斯矩阵 $A \in \mathbb{R}^{m \times n}$, 满足 $AA^T = I$ 。真实的 s -稀疏解 $x \in \mathbb{R}^n$ 通过随机选取 s 个非零位置生成, 其非零元素服从独立同分布的高斯分布。观测向量 b 由下式生成:

$$b = A\bar{x} + \sigma\varepsilon_1,$$

其中 $\sigma \in \mathbb{R}$ 为噪声水平, $\varepsilon_1 \sim \mathcal{N}(0, 1)$ 为标准高斯噪声, 在数值实验中, 设定问题的规模为 $m = 256$ 和 $n = 1024$, 噪声水平为 $\sigma = 0.1\%$ 。

非凸惩罚函数参数设置如下: SCAD 惩罚和 MCP 惩罚中设置参数 $a = 16$, ℓ_p 惩罚中取 $p = \frac{1}{2}$ 。在算法实现中, 我们设置初始点 $x^0 = x^{-1} = 0$, 最大迭代次数为 1000。正则化参数设置如下: 在 A-ILSTAC 和 ILSTAC 算法中, 设置初始参数 $\lambda_0 := \frac{\|\bar{x}\|}{\sqrt{s} + 1}$, 终止阈值 $\lambda = 10^{-4}$; 在 A-ILSTAT 和 ILSTAT 算法中, 通过交叉验证在区间 $[10^{-4}, 1]$ 内选择正则化参数 λ 。

算法的停止准则设置如下:

(i) A-ILSTAC 和 ILSTAC: 满足以下任一条件即停止:

- 迭代次数大于 1000。
- 正则化参数满足 $\lambda_k < \lambda$ 。

(ii) A-ILSTAT 和 ILSTAT: 满足以下任一条件即停止:

- 迭代次数大于 1000。
- 相邻迭代点满足 $\|x^k - x^{k-1}\| \leq 10^{-6}$ 。

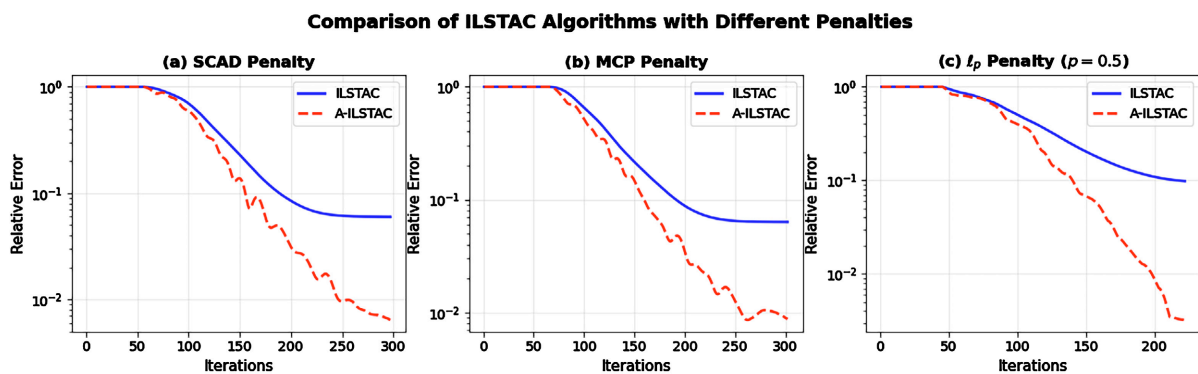


Figure 1. Performance comparison of A-ILSTAC and ILSTAC algorithms under different penalty functions

图 1. 不同惩罚函数下 A-ILSTAC 与 ILSTAC 算法性能比较

第一个实验旨在比较 A-ILSTAC 和 ILSTAC 算法的数值性能。图 1 显示, A-ILSTAC 算法比原始 ILSTAC 收敛更快, 并能获得更精确的解。这验证了在稀疏优化中, Nesterov 加速技术在 ILSTAC 框架下具有加速能力, 并能收敛到近似真实解。

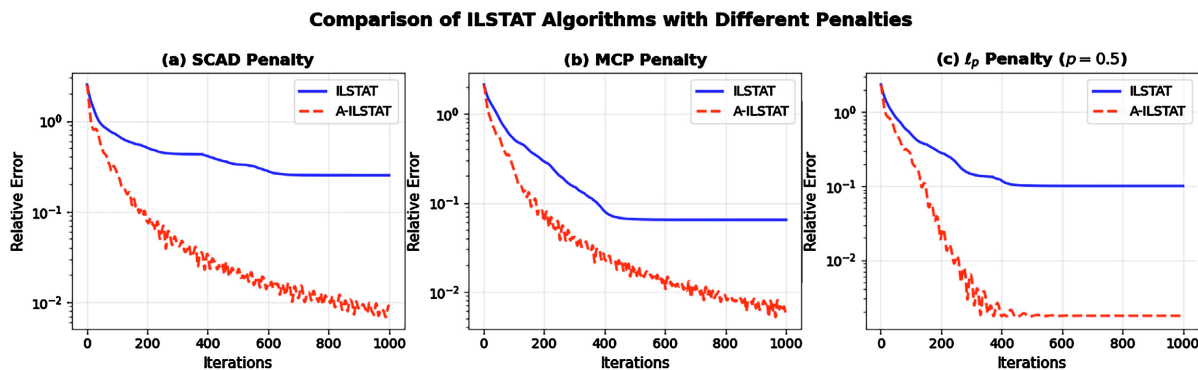


Figure 2. Performance comparison of A-ILSTAT and ILSTAT algorithms under different penalty functions
图 2. 不同惩罚函数下 A-ILSTAT 与 ILSTAT 算法性能比较

第二个实验旨在比较 A-ILSTAT 和 ILSTAT 算法的数值性能。图 2 显示, 相比与 ILSTAT, A-ILSTAT 精确度更高, 收敛速率更快。这表明 Nesterov 加速不仅提高了收敛速度, 还增强了算法的稀疏恢复能力。

6. 总结

本文将 Nesterov 加速策略融入非凸稀疏优化的统一 ITA 框架中。在有限收缩阈值算子的统一基础上, 本文设计了两种新型 Nesterov 加速迭代算法: Nesterov 加速的迭代有限收缩阈值算法(带延拓技术)(A-ILSTAC)和 Nesterov 加速的迭代有限收缩阈值算法(带截断技术)(A-ILSTAT)。在 RIP 假设下, 证明了 A-ILSTAC 的输出在与噪声水平成比例的容差范围内逼近于(1)的近似真稀疏解, 并且 A-ILSTAT 生成的序列以几何速率收敛于(1)的近似真稀疏解。数值结果表明, Nesterov 加速技术能有效加速算法收敛, 提升精确度, 使算法具有更强的稀疏恢复能力。

参考文献

- [1] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1998) Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, **20**, 33-61. <https://doi.org/10.1137/s1064827596304010>
- [2] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. <https://doi.org/10.1109/tit.2006.871582>
- [3] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. <https://doi.org/10.1137/080716542>
- [4] Elad, M. (2010) Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer Science & Business Media.
- [5] Qin, J., Hu, Y., Xu, F., Yalamanchili, H.K. and Wang, J. (2014) Inferring Gene Regulatory Networks by Integrating ChIP-seq/chip and Transcriptome Data via LASSO-Type Regularization Methods. *Methods*, **67**, 294-303. <https://doi.org/10.1016/j.ymeth.2014.03.006>
- [6] Wang, J., Hu, Y., Li, C. and Yao, J. (2017) Linear Convergence of CQ Algorithms and Applications in Gene Regulatory Network Inference. *Inverse Problems*, **33**, Article ID: 055017. <https://doi.org/10.1088/1361-6420/aa6699>
- [7] Bach, F., Jenatton, R., Mairal, J. and Obozinski, G. (2012) Structured Sparsity through Convex Optimization. *Statistical Science*, **27**, 450-468. <https://doi.org/10.1214/12-sts394>
- [8] Hu, Y., Li, C., Meng, K., Qin, J. and Yang, X. (2017) Group Sparse Optimization via $\ell_{p,q}$ Regularization. *Journal of Machine Learning Research*, **18**, 1-52.

- [9] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009) Simultaneous Analysis of Lasso and Dantzig Selector. *The Annals of Statistics*, **37**, 1705-1732. <https://doi.org/10.1214/08-aos620>
- [10] Cai, T.T., Xu, G. and Zhang, J. (2009) On Recovery of Sparse Signals via ℓ_1 Minimization. *IEEE Transactions on Information Theory*, **55**, 3388-3397. <https://doi.org/10.1109/tit.2009.2021377>
- [11] Candes, E.J. and Tao, T. (2005) Decoding by Linear Programming. *IEEE Transactions on Information Theory*, **51**, 4203-4215. <https://doi.org/10.1109/tit.2005.858979>
- [12] Daubechies, I., Defrise, M. and De Mol, C. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, **57**, 1413-1457. <https://doi.org/10.1002/cpa.20042>
- [13] Hale, E.T., Yin, W. and Zhang, Y. (2008) Fixed-Point Continuation for ℓ_1 -Minimization: Methodology and Convergence. *SIAM Journal on Optimization*, **19**, 1107-1130. <https://doi.org/10.1137/070698920>
- [14] Xiao, L. and Zhang, T. (2013) A Proximal-Gradient Homotopy Method for the Sparse Least-Squares Problem. *SIAM Journal on Optimization*, **23**, 1062-1091. <https://doi.org/10.1137/120869997>
- [15] Chartrand, R. and Staneva, V. (2008) Restricted Isometry Properties and Nonconvex Compressive Sensing. *Inverse Problems*, **24**, Article ID: 035020. <https://doi.org/10.1088/0266-5611/24/3/035020>
- [16] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [17] Xu, Z., Chang, X., Xu, F. and Zhang, H. (2012) L_{ν_2} Regularization: A Thresholding Representation Theory and a Fast Solver. *IEEE Transactions on Neural Networks and Learning Systems*, **23**, 1013-1027. <https://doi.org/10.1109/tnnls.2012.2197412>
- [18] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-aos729>
- [19] Chen, X., Xu, F. and Ye, Y. (2010) Lower Bound Theory of Nonzero Entries in Solutions of $\ell_2 - \ell_p$ Minimization. *SIAM Journal on Scientific Computing*, **32**, 2832-2852. <https://doi.org/10.1137/090761471>
- [20] Locatelli, M. and Schoen, F. (2013) Global Optimization: Theory, Algorithms, and Applications. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611972672>
- [21] Blumensath, T. and Davies, M.E. (2009) Iterative Hard Thresholding for Compressed Sensing. *Applied and Computational Harmonic Analysis*, **27**, 265-274. <https://doi.org/10.1016/j.acha.2009.04.002>
- [22] Foucart, S. and Rauhut, H. (2013) A Mathematical Introduction to Compressive Sensing. Springer.
- [23] Lu, Z. (2014) Iterative Reweighted Minimization Methods for ℓ_p Regularized Unconstrained Nonlinear Programming. *Mathematical Programming*, **147**, 277-307. <https://doi.org/10.1007/s10107-013-0722-4>
- [24] Attouch, H., Bolte, J. and Svaiter, B.F. (2013) Convergence of Descent Methods for Semi-Algebraic and Tame Problems: Proximal Algorithms, Forward-Backward Splitting, and Regularized Gauss-Seidel Methods. *Mathematical Programming*, **137**, 91-129. <https://doi.org/10.1007/s10107-011-0484-9>
- [25] Boţ, R.I., Csetnek, E.R. and Nguyen, D. (2019) A Proximal Minimization Algorithm for Structured Nonconvex and Nonsmooth Problems. *SIAM Journal on Optimization*, **29**, 1300-1328. <https://doi.org/10.1137/18m1190689>
- [26] Wen, Z., Yin, W., Goldfarb, D. and Zhang, Y. (2010) A Fast Algorithm for Sparse Reconstruction Based on Shrinkage, Subspace Optimization, and Continuation. *SIAM Journal on Scientific Computing*, **32**, 1832-1857. <https://doi.org/10.1137/090747695>
- [27] Jiao, Y., Jin, B. and Lu, X. (2017) Iterative Soft/Hard Thresholding with Homotopy Continuation for Sparse Recovery. *IEEE Signal Processing Letters*, **24**, 784-788. <https://doi.org/10.1109/lsp.2017.2693406>
- [28] Blumensath, T. and Davies, M.E. (2010) Normalized Iterative Hard Thresholding: Guaranteed Stability and Performance. *IEEE Journal of Selected Topics in Signal Processing*, **4**, 298-309. <https://doi.org/10.1109/jstsp.2010.2042411>
- [29] Zhao, Y. (2020) Optimal k -Thresholding Algorithms for Sparse Optimization Problems. *SIAM Journal on Optimization*, **30**, 31-55. <https://doi.org/10.1137/18m1219187>
- [30] Hu, Y., Hu, X. and Yang, X. (2025) On Convergence of Iterative Thresholding Algorithms to Approximate Sparse Solution for Composite Nonconvex Optimization. *Mathematical Programming*, **211**, 181-206. <https://doi.org/10.1007/s10107-024-02068-1>
- [31] Nesterov, Y. (1983) A Method of Solving a Convex Programming Problem with Convergence Rate $O\left(\frac{1}{k^2}\right)$. *Proceedings of the USSR Academy of Sciences*, **269**, 543-547.