

# 基于截断 $\ell_1$ 惩罚的可学习 Peaceman-Rachford 方法

严增平

浙江师范大学数学科学学院, 浙江 金华

收稿日期: 2026年4月15日; 录用日期: 2026年5月8日; 发布日期: 2026年5月19日

## 摘要

近年来, 模型与数据驱动的深度展开方法, 将基于凸或非凸正则化的迭代优化算法展开为深度神经网络, 在稀疏信号重建问题中取得了显著成效。本文围绕非凸截断  $\ell_1$  惩罚正则化下的稀疏重建问题, 提出了基于截断  $\ell_1$  惩罚的可学习 Peaceman-Rachford 方法。该模型具有良好的数学理论保障, 不仅在高概率下近似无假阳性, 而且可分析其线性收敛速率。数值实验表明, 本文提出的模型显著优于现有代表性模型。

## 关键词

稀疏重建, 深度展开, 截断  $\ell_1$  惩罚, Peaceman-Rachford 方法

# Learned Peaceman-Rachford Method Based on Capped $\ell_1$ Penalty

Zengping Yan

School of Mathematical Sciences, Zhejiang Normal University, Jinhua Zhejiang

Received: April 15, 2026; accepted: May 8, 2026; published: May 19, 2026

## Abstract

In recent years, model and data-driven deep unfolding methods have unfolded iterative optimization algorithms based on convex or nonconvex regularization into deep neural networks, achieving remarkable performance in sparse signal reconstruction problems. Focusing on the sparse reconstruction problem under nonconvex capped  $\ell_1$  penalty regularization, this paper proposes a learned Peaceman-Rachford method based on capped  $\ell_1$  penalty. The proposed model is supported

by sound mathematical theory: it not only yields approximate no-false-positives results with high probability, but also allows for the analysis of its linear convergence rate. Numerical experiments demonstrate that the proposed model significantly outperforms existing representative models.

## Keywords

Sparse Reconstruction, Deep Unfolding, Capped  $\ell_1$  Penalty, Peaceman-Rachford Method

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

### 1.1. 研究背景及意义

稀疏重建, 又称为稀疏恢复或稀疏信号重构, 其核心思想是基于“稀疏性先验”, 针对欠定线性系统, 通过引入正则化约束重建稀疏信号[1]。稀疏重建方法已在信号处理、机器学习、计算机视觉等多个领域中有着非常广泛的研究与应用[2]。

近年来, 得益于优化算法与机器学习技术的深度融合, 稀疏重建已在图像去噪、超分辨率重建、图像压缩以及缺失数据修复等方面取得显著进展[3], 其中在医学成像领域的应用价值尤为突出。现代医学影像模态(如磁共振成像 MRI、计算机断层扫描 CT、正电子发射断层成像 PET 等)虽能提供丰富的解剖与功能信息, 但往往面临扫描时间长、辐射剂量高、数据采集不完整等问题。例如, 在 MRI 中广泛采用的  $k$ -空间欠采样策略虽可显著缩短扫描时间, 却不可避免地引入混叠伪影; 在 CT 中则因患者移动或设备扫描范围的限制造成投影角度缺失, 导致扫描数据不完整。传统插值、滤波或解析重建方法在处理此类病态逆问题时, 往往难以兼顾细节保留与噪声抑制, 导致图像重建质量下降。相比之下, 稀疏重建方法充分利用医学图像在小波、曲波、全变分(Total Variation, TV)以及学习型字典等变换域中所呈现的高度结构化稀疏特性[4], 能够从不完整甚至含噪声的观测中稳健地重建高质量图像。例如, 在 MRI 中, 基于  $k$ -空间欠采样与 TV/小波联合稀疏约束的重建框架已被广泛验证; 在低剂量 CT 领域, 结合非局部先验与自适应正则化的算法显著提升了噪声鲁棒性与细节保留能力。这不仅为加速成像、降低辐射剂量提供了可行路径, 也为提升微小病灶的可视化能力提供了关键技术支撑。

进一步地, 稀疏重建在低剂量 CT、动态 MRI、功能性脑成像等前沿医疗场景中展现出更广阔的应用前景[5]: 一方面, 通过减少所需投影或相位编码步数, 显著缩短检查时间并降低患者不适; 另一方面, 借助稀疏先验增强图像对比度, 有助于增强病灶区域的可辨识度, 辅助医生进行更准确的诊断与治疗规划。然而, 现实医学图像常受复杂噪声(如 Rician 噪声、泊松噪声)、运动伪影及病灶形态异质性等因素干扰, 导致其稀疏结构发生偏移、退化甚至失效。同时, 基于凸松弛的重建方法虽具备坚实的理论支撑但其对稀疏性的诱导能力有限, 且迭代收敛速度较慢, 难以满足临床对高精度与实时性的双重需求[6]。这些因素严重挑战了现有稀疏重建方法的收敛效率和鲁棒性。

综上所述, 发展兼具强稀疏促进能力、快速收敛特性与抗噪和抗病态鲁棒性的重建算法, 已成为医学图像处理领域的重要研究课题。尤其是在深度学习与传统稀疏重建方法融合的背景下, 基于深度展开的稀疏重建方法应运而生, 为解决上述医学成像中的挑战提供了新的思路与技术路径[7] [8]。

## 1.2. 研究现状

稀疏重建旨在从欠采样或不完整观测中恢复具有稀疏结构的信号，其理论基础源于压缩感知。目标是从低维测量信号  $\mathbf{x} \in \mathbb{R}^m$  中恢复高维稀疏信号  $\mathbf{z} \in \mathbb{R}^n$ ，二者满足如下数学关系：

$$\mathbf{x} = \mathbf{D}\mathbf{z} + \mathbf{e},$$

其中， $\mathbf{D} \in \mathbb{R}^{m \times n}$  为感知矩阵(满足  $m \ll n$ )， $\mathbf{e}$  表示感知噪声。该问题本质上是在欠定系统中寻找正确解，但这是一个 NP-hard 问题。然而，在无穷多解中，通常仅关注具有稀疏性的解。因此，稀疏重建的数学模型可表述为：

$$\arg \min_{\mathbf{z} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_0, \quad (1)$$

其中，第一项为数据保真项，用于保证重建结果与测量信号的一致性；第二项为稀疏正则项，通过  $\ell_0$  范数约束解的稀疏性； $\lambda > 0$  为正则化参数，用于平衡二者之间的权衡。近年来，该领域经历了从传统稀疏重建方法，到数据驱动的深度学习方法，再到模型与数据融合的深度展开方法的发展历程。

### 1.2.1. 传统稀疏重建方法

由于问题(1)中的正则项为  $\ell_0$  范数，其本身具有非连续性与非凸性，难以直接求解。为此，传统稀疏重建方法主要基于凸松弛思想，将  $\ell_0$  范数近似为连续且凸的  $\ell_1$  范数，从而转化为可高效求解的凸优化问题——即最小绝对收缩与选择算子(Least Absolute Shrinkage and Selection Operator, LASSO)：

$$\arg \min_{\mathbf{z} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1.$$

在该凸优化框架下，研究者提出了多种经典的迭代算法，包括迭代收缩阈值算法(Iterative Shrinkage-Thresholding Algorithm, ISTA) [9]、迭代硬阈值算法(Iterative Hard Thresholding, IHT) [10]等。然而， $\ell_1$  范数作为  $\ell_0$  范数的凸松弛，在稀疏性逼近方面仍存在偏差，尤其在高维或强相关字典情形下易引入显著估计误差。为提升稀疏逼近精度，非凸正则化方法受到广泛关注。代表性工作包括光滑剪切绝对偏差(Smoothly Clipped Absolute Deviation, SCAD) [11]、极小极大凹惩罚(Minimax Concave Penalty, MCP) [12]，以及截断  $\ell_1$  惩罚(Capped  $\ell_1$  Penalty, Capped- $\ell_1$ ) [13]等。

尽管上述方法具有良好的理论保障，且算法效率不断提升，但其实际应用仍面临若干挑战：一方面，多数迭代算法收敛速度较慢，难以满足大规模高维场景下的实时性需求；另一方面，算法性能对超参数(如正则项系数  $\lambda$ )高度敏感。

### 1.2.2. 数据驱动的深度学习方法

上述传统方法的局限激发了以深度学习(Deep Neural Networks, DNNs)为核心的数据驱动新思路。DNNs 能够在几乎无需显式建模底层函数先验知识的前提下，通过端到端训练直接从数据中学习复杂的非线性映射关系。自 Goodfellow 等人系统阐述深度学习框架以来，DNNs 凭借其强大的表示能力，在计算机视觉、自然语言处理、医学成像等领域取得了突破性进展[14]。然而，此类纯数据驱动方法仍存在明显不足：第一，当训练样本有限时，DNNs 难以充分逼近真实映射函数，导致重建性能显著下降；第二，模型易发生过拟合，泛化能力受限；第三，训练所得网络通常缺乏明确的数学结构，被视为“黑箱”模型，难以提供理论保障与物理解释。

### 1.2.3. 模型与数据融合的深度展开方法

为兼顾重建效率与可解释性，模型与数据融合的深度展开方法应运而生。其核心思想是将传统稀疏重建方法(如迭代收缩阈值)的每一步展开为神经网络的一层，在保留原始算法结构的同时引入可学习参

数, 从而实现“具有理论可解释性的端到端学习”。其中, Gregor 等人[15]提出的可学习迭代收缩阈值算法(Learned Iterative Shrinkage Thresholding Algorithm, LISTA)是最早应用于稀疏重建的深度展开方法, 相较于原始 ISTA 算法, LISTA 显著加速了收敛过程, 并获得了更精确的稀疏解。进一步地, Chen 等人[16]对 LISTA 框架进行了系统性改进, 提出带部分耦合的 LISTA (LISTA with Partial Coupling, LISTA-CP)。作者不仅提升了重建性能, 还首次为该类深度展开网络建立了严格的理论保障。然而, 上述方法大多局限于  $\ell_1$  范数正则化问题。尽管  $\ell_1$  范数具有良好的凸性与计算便利性, 但其在诱导稀疏性方面不如非凸正则项有效。

为突破  $\ell_1$  正则化的局限, 研究者开始探索针对非凸正则化的深度展开方法。Kim 等人[17]提出了带自适应阈值 II 的 LISTA (LISTA with Adaptive Thresholds II, LISTA-AT2), 其核心在于利用凹函数的局部线性上界对非凸正则项进行逼近, 构建了具有理论保障的深度展开方法。

### 1.3. 研究内容及创新点

现有深度展开主要是基于近端梯度法, 但模型缺乏对噪声和病态系统的强鲁棒性。针对高感知噪声和病态感知矩阵的挑战, 我们基于截断  $\ell_1$  惩罚的 Peaceman-Rachford 方法, 提出可学习 Peaceman-Rachford 方法。我们证明了该模型不仅具有理论保障——高概率近似无假阳性及线性收敛速率, 还通过数值实验对所提模型进行了全面评估。实验结果表明在重建精度、抗噪能力及对高条件数的鲁棒性等关键指标上, 均显著优于现有代表性深度展开模型。

## 2. 预备知识

本文中, 粗体大写字母表示矩阵, 粗体小写字母表示向量。 $\mathbf{D}$  的转置记作  $\mathbf{D}^\top$ , 且  $\tau$  为  $\mathbf{D}^\top \mathbf{D}$  的最大特征值。稀疏目标信号  $\mathbf{z}^*$  的支撑集记为  $S = \text{supp}(\mathbf{z}^*)$ , 即  $S = \{i | z_i^* \neq 0\}$ , 其补集为  $S^c = \{1, 2, \dots, n\} \setminus S$ 。对于任意矩阵  $\mathbf{B}$ , 记  $\mathbf{B}_{S,S}$  为其由索引集  $S$  所确定的主子矩阵,  $\mathbf{B}_{:,S}$  与  $\mathbf{B}_{S,:}$  分别表示  $\mathbf{B}$  中由  $S$  指定的列子矩阵与行子矩阵。类似地, 对任意向量  $\mathbf{z}$ ,  $\mathbf{z}_S$  表示其在支撑集  $S$  上的投影子向量, 而  $z_i$  表示  $\mathbf{z}$  的第  $i$  个分量。在算法迭代过程(或深度展开)中的第  $k$  步(或第  $k$  层展开)以右上标  $k$  标记, 如  $\mathbf{z}^k$ 。此外, 定义可行参数集  $\mathcal{Z}(L, s, \epsilon) = \{(\mathbf{z}^*, \mathbf{e}) | 0 < |z_i^*| < L, \forall i \in S; \|\mathbf{z}^*\|_0 \leq s; \|\mathbf{e}\|_2 \leq \epsilon\}$ , 其中  $L > 0$ ,  $s > 0$  为给定常数, 分别控制非零元幅值的上下界与稀疏度上限,  $\epsilon \geq 0$  表示观测噪声的容许范围。

### 2.1. 近端梯度法

**定义 2.1** 设  $F$  是适当闭函数, 且具有有限的下界, 即满足  $\inf_{\mathbf{z} \in \mathbb{R}^n} F(\mathbf{z}) > -\infty$ , 则  $F$  的近端算子为

$$\text{prox}_F(\mathbf{u}) := \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ F(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{u}\|_2^2 \right\}.$$

根据最优性条件, 有  $\mathbf{0} \in \partial F(\mathbf{z}) + \mathbf{z} - \mathbf{u}$ 。进而可解得  $\mathbf{z}$ , 即  $\text{prox}_F(\mathbf{u})$ 。直观上, 近端算子的目的是求解一个接近  $\mathbf{u}$  的点, 并使函数值  $F(\mathbf{z})$  相对较小。近端梯度法先是做梯度下降, 再做近端算子投影, 求解 LASSO 问题的迭代格式如下:

$$\begin{cases} \mathbf{u}^k = \mathbf{z}^k - \alpha^k \nabla G(\mathbf{z}^k), \\ \mathbf{z}^{k+1} = \text{prox}_{\alpha^k \lambda \|\cdot\|}(\mathbf{u}^k), \end{cases} \quad (2)$$

其中梯度  $\nabla G(\mathbf{z}^k) = -\mathbf{D}^\top (\mathbf{x} - \mathbf{D}\mathbf{z}^k)$ ,  $\alpha^k > 0$  为步长, 通常取  $\alpha^k \leq 1/\tau$ 。

## 2.2. 深度展开方法

LISTA-CP 将(2)中每一步迭代映射为深度神经网络的一层，从而构建一个可端到端训练的深度展开网络。其第  $k$  层展开如下：

$$\begin{cases} \mathbf{y}^k = \mathbf{x} - \mathbf{D}\mathbf{z}^k, \\ \mathbf{z}^{k+1} = P_{\theta^k \|\cdot\|}(\mathbf{z}^k + \mathbf{B}^k \mathbf{y}^k), \end{cases}$$

其中  $\mathbf{B}^k = \alpha^k \mathbf{D}^\top$  为待学习的网络参数集，非线性激活函数  $P_{\theta^k \|\cdot\|}$  通常取为正则项  $F$  对应的近端算子。特别地，如果令  $\mathbf{B}^k = \alpha^k \mathbf{D}^\top$ ， $\theta^k = \alpha^k \lambda$ ，则该模型退化为近端梯度法。

在训练策略上，深度神经网络通常存在两种学习方式。对于无监督学习，深度神经网络以最小化损失函数  $\mathbf{G}(\mathbf{z}^K) + \lambda F(\mathbf{z}^K)$  训练，其中  $K$  表示网络层数。对于有监督学习，深度神经网络以最小化损失函数  $\|\mathbf{z}^K - \mathbf{z}^*\|_2^2$  训练。深度展开网络一般采用有监督学习。

在训练过程中，给定训练数据集  $\{(\mathbf{z}_i^*, \mathbf{x}_i)\}_{i=1}^Q$ ，其中  $\mathbf{z}_i^*$  为对应于测量信号  $\mathbf{x}_i$  的稀疏目标信号，网络参数通过求解如下经验风险最小化问题进行学习：

$$\Theta^* = \arg \min_{\Theta} \sum_{i=1}^Q \|\mathbf{z}^K(\Theta, \mathbf{x}_i, \mathbf{z}^0) - \mathbf{z}_i^*\|_2^2,$$

其中  $\mathbf{z}^K(\Theta, \mathbf{x}_i, \mathbf{z}^0)$  表示以  $\mathbf{x}_i$  为输入、初始状态  $\mathbf{z}^0$  (通常取零向量) 经  $K$  层网络传播后所得的输出。训练完成后，可获得最优参数集  $\Theta^* = \{(\mathbf{W}^k)^*, (\mathbf{B}^k)^*, (\theta^k)^*\}$ 。最后，对于任意新的测量信号  $\mathbf{x}$ ，应用已训练好的深度展开网络(2)，即可高效完成稀疏信号的重建。

## 3. 基于截断惩罚的可学习 Peaceman-Rachford 方法

本文旨在求解如下稀疏重建问题：

$$\arg \min_{\mathbf{z} \in \mathbb{R}^n} G(\mathbf{z}) + \lambda F(\mathbf{z}), \tag{3}$$

其中数据保真项  $G(\mathbf{z}) = \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2$ ，正则项  $F(\mathbf{z}) = \sum_{i=1}^n \min(|z_i|, \sigma)$  为带有截断参数  $\sigma > 0$  的截断  $\ell_1$  惩罚函数， $\lambda > 0$  为正则项参数。截断  $\ell_1$  惩罚在促进稀疏性的同时，通过截断参数  $\sigma$  避免对幅值较大的系数施加过度惩罚，从而在保持稀疏结构的同时更好地保留目标信号的主要结构信息。应用 Peaceman-Rachford 方法求解问题(3)，其迭代格式如下：

$$\begin{cases} \mathbf{u}^k = \mathbf{z}^k - \alpha^k \nabla G(\mathbf{z}^k), \\ \mathbf{z}^{k+\frac{1}{2}} = \text{prox}_{\alpha^k \lambda F}(\mathbf{u}^k), \\ \mathbf{z}^{k+1} = \text{prox}_{\alpha^k G}\left(2\mathbf{z}^{k+\frac{1}{2}} - \mathbf{z}^k + \alpha^k \nabla G(\mathbf{z}^k)\right), \end{cases} \tag{4}$$

其中梯度  $\nabla G(\mathbf{z}^k) = -\mathbf{D}^\top(\mathbf{x} - \mathbf{D}\mathbf{z}^k)$ ， $\alpha^k > 0$  为步长。类似 LISTA-CP 的思想，本文将(4)展开为深度网络，提出基于截断  $\ell_1$  惩罚的可学习 Peaceman-Rachford 方法(Learned Peaceman-Rachford Method, LPRM)，其第  $k$  层展开为：

$$\begin{cases} \mathbf{y}^k = \mathbf{x} - \mathbf{D}\mathbf{z}^k, \\ \mathbf{z}^{k+\frac{1}{2}} = P_{\theta^k, F}(\mathbf{z}^k + \mathbf{B}^k \mathbf{y}^k), \\ \mathbf{z}^{k+1} = \text{prox}_{\alpha^k G} \left( 2\mathbf{z}^{k+\frac{1}{2}} - \mathbf{z}^k + \alpha^k \nabla G(\mathbf{z}^k) \right), \end{cases} \quad (5)$$

其中  $\mathbf{B}^k = \alpha^k \mathbf{D}^\top$  为可学习参数,  $\alpha^k$  为超参数, 激活函数  $P_{\theta^k, F}$  采用如下形式:

$$\left( P_{\theta^k, F}(\mathbf{u}^k) \right)_i = \begin{cases} 0, |u_i^k| \leq \theta^k, \\ \text{sign}(u_i^k) (|u_i^k| - \theta^k), \theta^k < |u_i^k| < \sigma + \frac{1}{2}\theta^k, \\ u_i^k, |u_i^k| \geq \sigma + \frac{1}{2}\theta^k, \end{cases} \quad (6)$$

特别地, 如果令  $\mathbf{B}^k = \alpha^k \mathbf{D}^\top$ ,  $\theta^k = \alpha^k \lambda$ , 则该模型退化为经典的 Peaceman-Rachford 方法。

### 解析型线性速率分析

**引理 3.1** 设  $\{\mathbf{z}^k\}$  是由(5)生成的序列, 初始值  $\mathbf{z}^0 = \mathbf{0}$ , 测量信号  $\mathbf{x} = \mathbf{D}\mathbf{z}^* + \mathbf{e}$ , 超参数  $\alpha^k > 0$ 。其中,  $\mathbf{z}^*$  是一个任意  $s$ -稀疏目标信号, 满足  $\|\mathbf{z}^*\|_2 \leq \mu_z$ 。假设可学习参数序列  $\{\mathbf{B}^k, \theta^k\}$  满足  $\|\mathbf{B}^k\|_2 \leq \mu_B$  一致有界, 并且  $\theta^k \geq \mu_S \cdot \sup_{(\mathbf{z}^*, \mathbf{e}) \in \mathcal{Z}(L, s, \epsilon)} \|\mathbf{z}^k - \mathbf{z}^*\|_2 + \mu_B \epsilon$ , 其中  $\mu_S = \sup_k \max_{i \in S^c} \|\mathbf{B}_{i,:}^k \mathbf{D}_{:,S}\|_2$ , 则 LPRM 高概率近似无假阳性。

**证明.** 假设感知矩阵  $\mathbf{D} \in \mathbb{R}^{m \times n}$  的列  $\ell_2$  范数归一化, 元素满足独立同分布且为子高斯随机变量, 由子高斯集中不等式, 有

$$\mu = \max_{i \neq j} |d_i^\top d_j| \leq \mathcal{O} \left( \sqrt{\frac{\log n}{m}} \right).$$

令  $\mathbf{D}^\top \mathbf{D} = \mathbf{I} + \mathbf{E}$ , 其中对角线元素  $E_{ii} = 0$ , 非对角线元素  $|E_{ij}| \leq \mu$ , 则

$$(\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})^{-1} = ((1 + \alpha^k) \mathbf{I} + \alpha^k \mathbf{E})^{-1} = \frac{1}{1 + \alpha^k} \left( \mathbf{I} + \frac{\alpha^k}{1 + \alpha^k} \mathbf{E} \right)^{-1}.$$

由于  $\mu$  极小, 利用 Neumann 级数进行一阶近似:

$$(\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})^{-1} \approx \frac{1}{1 + \alpha^k} \left( \mathbf{I} - \frac{\alpha^k}{1 + \alpha^k} \mathbf{E} \right). \quad (7)$$

由(5)可得,

$$\mathbf{z}^{k+1} = \mathbf{z}^k + 2(\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})^{-1} \left( \mathbf{z}^{k+\frac{1}{2}} - \mathbf{z}^k \right). \quad (8)$$

结合(7)与(8)有

$$\mathbf{z}^{k+1} \approx \mathbf{z}^k + \frac{2}{1 + \alpha^k} \left( \mathbf{I} - \frac{\alpha^k}{1 + \alpha^k} \mathbf{E} \right) \left( \mathbf{z}^{k+\frac{1}{2}} - \mathbf{z}^k \right).$$

又因为  $\theta^k \geq \mu_S \cdot \sup_{(\mathbf{z}^*, \mathbf{e}) \in \mathcal{Z}(L, s, \epsilon)} \|\mathbf{z}^k - \mathbf{z}^*\|_2 + \mu_B \epsilon \geq \|\mathbf{B}_{i,:}^k \mathbf{D}_{:,S}\|_2 \cdot \|\mathbf{z}^k - \mathbf{z}^*\|_2 + \|\mathbf{B}^k\|_2 \cdot \|\mathbf{e}\|_2 \geq |u_i^k|$ , 于是

$$\mathbf{z}_{S^c}^{k+1} \approx \mathbf{0} - \frac{\alpha^k}{(1+\alpha^k)^2} \mathbf{E}_{S^c, S} \left( \mathbf{z}_S^{k+\frac{1}{2}} - \mathbf{z}_S^k \right).$$

由于  $|\mathbf{E}_{ij}| \leq \mu$ ，即高概率近似无假阳性。

**定义 3.2** 给定字典  $\mathbf{D} \in \mathbb{R}^{m \times n}$  (每列均已归一化)，其广义互相干性为：

$$\tilde{\mu} = \inf_B \left\{ \max_{\substack{i \neq j \\ 1 \leq i, j \leq n}} \mathbf{B}_{i,:} \mathbf{D}_{:,j} \right\}, \quad \text{s.t. } \mathbf{B}_{i,:} \mathbf{D}_{:,i} = 1, \forall i \in \{1, \dots, n\}.$$

**定理 3.3** 设  $\{z^k\}$  是由(5)生成的序列，初始值  $z^0 = \mathbf{0}$ ，测量信号  $\mathbf{x} = \mathbf{D}\mathbf{z}^* + \mathbf{e}$ ，超参数  $\alpha^k > 0$ 。其中， $\mathbf{z}^*$  是一个任意  $s$ -稀疏目标信号，满足  $\|\mathbf{z}^*\|_2 \leq \mu_z$ 。记  $\left\| (\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})_{S,S}^{-1} \right\|_2 = a^k$ ， $\left\| (\alpha^k \mathbf{D}^\top \mathbf{D})_{S,S} \right\|_2 = b^k$ 。假设可学习参数序列  $\{\theta^k, \gamma^k\}$  满足  $\theta^k = \gamma^k \sqrt{s} \tilde{\mu} \cdot \sup_{(\mathbf{z}^*, \mathbf{e}) \in \mathcal{Z}(L, s, \epsilon)} \|\mathbf{z}^k - \mathbf{z}^*\|_2 + \gamma^k \|\mathbf{B}\|_2 \epsilon$ ，

只要  $c = \sup_k a^k (b^k + |1 - 2\gamma^k| + \gamma^k \tilde{\mu} (2s - 1)) < 1$ ，则

$$\|\mathbf{z}^{k+1} - \mathbf{z}^*\|_2 \leq c^{k+1} \mu_z + \frac{2a\gamma\epsilon(\sqrt{s} + 1)\|\mathbf{B}\|_2}{1 - c},$$

其中  $a\gamma = \sup_k (a^k \gamma^k)$ 。

**证明.** 注意到，若  $\mathbf{B}^k = \gamma^k \mathbf{B}$ ，且阈值参数满足  $\theta^k \geq \gamma^k \sqrt{s} \tilde{\mu} \cdot \sup_{(\mathbf{z}^*, \mathbf{e}) \in \mathcal{Z}(L, s, \epsilon)} \|\mathbf{z}^k - \mathbf{z}^*\|_2 + \|\gamma^k \mathbf{B}\|_2 \epsilon$ ，则引理 3.1

仍然成立。

对任意的  $j \in S$ ，根据近端算子最优性条件，有

$$2z_j^{k+\frac{1}{2}} - z_j^k - z_j^* = (z_j^k - z_j^*) + 2 \left( \gamma^k \mathbf{B}_{j,:} \mathbf{D}_{:,S} (\mathbf{z}_S^* - \mathbf{z}_S^k) + \gamma^k \mathbf{B}_{j,:} \mathbf{e} - \theta^k \nabla F \left( z_j^{k+\frac{1}{2}} \right) \right). \quad (9)$$

由(5)可得，

$$(\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})_{j,:} (\mathbf{z}_S^{k+1} - \mathbf{z}_S^*) = \left( 2z_j^{k+\frac{1}{2}} - z_j^k - z_j^* \right) + (\alpha^k \mathbf{D}^\top \mathbf{D})_{j,:} (\mathbf{z}_S^k - \mathbf{z}_S^*). \quad (10)$$

结合(9)与(10)，可得

$$\mathbf{z}_S^{k+1} - \mathbf{z}_S^* = (\mathbf{I} + \alpha^k \mathbf{D}^\top \mathbf{D})_{S,S}^{-1} \left( \left( (\alpha^k \mathbf{D}^\top \mathbf{D})_{S,S} + (1 - 2\gamma^k) \right) (\mathbf{z}_S^k - \mathbf{z}_S^*) + \mathbf{A}^k \tilde{\mathbf{z}} + 2\gamma^k \mathbf{B}_{S,:} \mathbf{e} - 2\theta^k \nabla F \left( \mathbf{z}_S^{k+\frac{1}{2}} \right) \right),$$

其中

$$\mathbf{A}^k = \gamma^k \begin{pmatrix} \mathbf{B}_{j_1,:} \mathbf{D}_{:,S \setminus j_1} & & & \\ & \mathbf{B}_{j_2,:} \mathbf{D}_{:,S \setminus j_2} & & \\ & & \ddots & \\ & & & \mathbf{B}_{j_s,:} \mathbf{D}_{:,S \setminus j_s} \end{pmatrix}, \quad \tilde{\mathbf{z}} = \begin{pmatrix} \mathbf{z}_{S \setminus j_1}^* - \mathbf{z}_{S \setminus j_1}^k \\ \mathbf{z}_{S \setminus j_2}^* - \mathbf{z}_{S \setminus j_2}^k \\ \vdots \\ \mathbf{z}_{S \setminus j_s}^* - \mathbf{z}_{S \setminus j_s}^k \end{pmatrix},$$

且  $j_l \in S$  对所有  $l=1, 2, \dots, s$  成立。一方面，

$$\begin{aligned} \|\mathbf{A}^k\|_2^2 &= \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}^k \mathbf{v}\|_2^2 \leq (\gamma^k)^2 \sup_{\|\mathbf{v}\|_2=1} \sum_{l=1}^s \|\mathbf{B}_{j_l,:} \mathbf{D}_{:,S \setminus j_l}\|_2^2 \|\mathbf{v}_l\|_2^2 \\ &\leq (s-1) (\gamma^k \tilde{\mu})^2 \sup_{\|\mathbf{v}\|_2=1} \sum_{l=1}^s \|\mathbf{v}_l\|_2^2 = (s-1) (\gamma^k \tilde{\mu})^2, \end{aligned}$$

其中  $\mathbf{v} = (\mathbf{v}_1^\top, \mathbf{v}_2^\top, \dots, \mathbf{v}_s^\top)^\top$ ,  $\mathbf{v}_l \in \mathbb{R}^{s-1}$ 。另一方面,  $\|\tilde{\mathbf{z}}\|_2^2 = (s-1)\|\mathbf{z}_s^k - \mathbf{z}_s^*\|_2^2$ 。

因此

$$\|\mathbf{z}_s^{k+1} - \mathbf{z}_s^*\|_2 \leq a^k \left( (b^k + |1 - 2\gamma^k| + \gamma^k \tilde{\mu}(s-1)) \|\mathbf{z}_s^k - \mathbf{z}_s^*\|_2 \right) + 2\sqrt{s}a^k\theta^k + 2a^k\gamma^k \|\mathbf{B}\|_2 \epsilon.$$

对上式在可行参数集  $(\mathbf{z}^*, \epsilon) \in \mathcal{Z}(L, s, \epsilon)$  上取上确界, 并将阈值参数  $\theta^k$  代入可得

$$\begin{aligned} \sup_{(\mathbf{z}^*, \epsilon)} \|\mathbf{z}_s^{k+1} - \mathbf{z}_s^*\|_2 &\leq a^k (b^k + |1 - 2\gamma^k| + \gamma^k \tilde{\mu}(s-1)) \sup_{(\mathbf{z}^*, \epsilon)} \|\mathbf{z}_s^k - \mathbf{z}_s^*\|_2 + 2\sqrt{s}a^k\theta^k + 2a^k\gamma^k \|\mathbf{B}\|_2 \epsilon \\ &\leq c^{k+1} \sup_{(\mathbf{z}^*, \epsilon)} \|\mathbf{z}_s^0 - \mathbf{z}_s^*\|_2 + \frac{2 \sup_{(\mathbf{z}^*, \epsilon)} (a^k \gamma^k) (\sqrt{s} + 1) \|\mathbf{B}\|_2 \epsilon}{1 - c}, \end{aligned}$$

因此,

$$\|\mathbf{z}^{k+1} - \mathbf{z}^*\|_2 \leq \sup_{(\mathbf{z}^*, \epsilon)} \|\mathbf{z}^{k+1} - \mathbf{z}^*\|_2 = \sup_{(\mathbf{z}^*, \epsilon)} \|\mathbf{z}_s^{k+1} - \mathbf{z}_s^*\|_2 \leq c^{k+1} \mu_z + \frac{2a\gamma\epsilon(\sqrt{s} + 1)\|\mathbf{B}\|_2}{1 - c}.$$

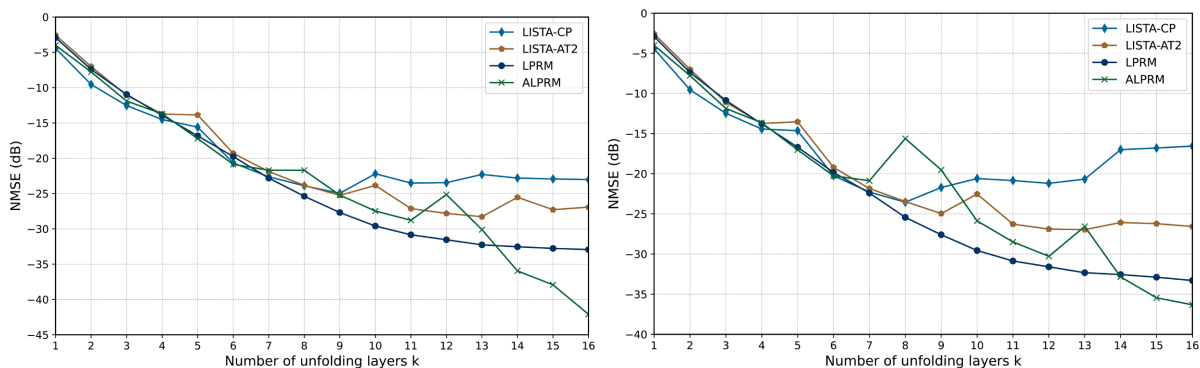
定理表明, 迭代误差的衰减率不仅受限于字典的广义互相关性  $\tilde{\mu}$  的影响, 还受到特定矩阵的谱范数的调制作用。本文将满足此定理的模型称为解析型 LPRM (ALPRM)。

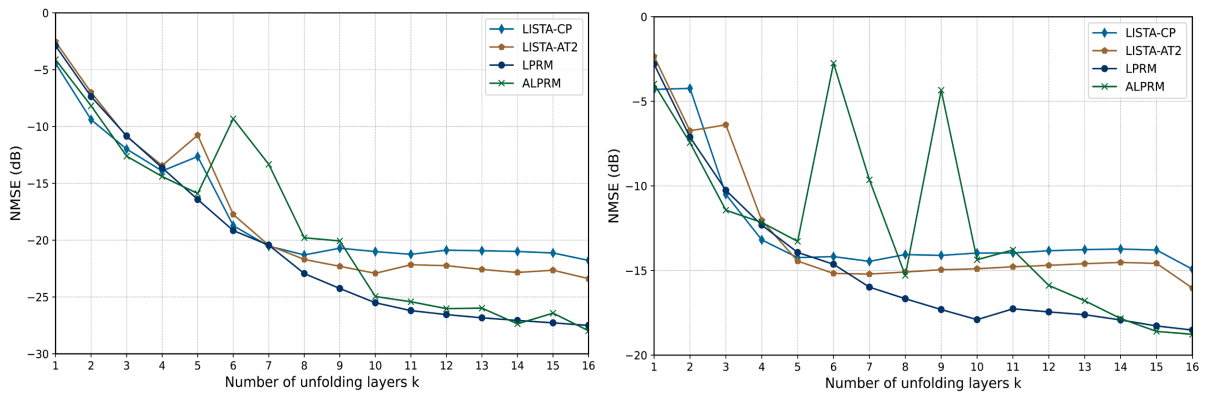
### 4. 数值实验

所有实验均基于同一合成数据集进行, 包含 51,200 个样本用于训练, 2048 个样本用于验证, 以及 2048 个样本用于测试。其中, 感知矩阵  $\mathbf{D}$  的维度为  $m \times n = 250 \times 500$ , 其元素独立同分布(i.i.d.)地从高斯分布  $\mathcal{N}(0, 1/m)$  中采样, 并对  $\mathbf{D}$  进行列向量的  $\ell_2$  范数归一化。稀疏目标信号  $\mathbf{z}^*$  由分布  $\mathcal{N}(0, 1) \cdot \text{Ber}(0.1)$  生成。网络默认包含  $K = 16$  层, 所有可学习参数的优化均采用 Adam (Adaptive Moment Estimation) 优化算法求解。

#### 4.1. 重建精度对比

图 1 展示了多种深度展开网络模型在不同信噪比(依次是 SNR =  $\infty$ 、40、30、20 dB)条件下, 归一化均方误差(NMSE)随展开层数的变化趋势。本文提出的 LPRM 与 ALPRM 模型在各类 SNR 水平下均展现出优于 LISTA-CP、LISTA-AT2 的稀疏信号重构性能。值得注意的是, 随着感知噪声水平的增加, 各算法的稳定性差异逐渐显现。这一现象表明: 基于 Peaceman-Rachford 框架的算法在噪声环境下具备更强的鲁棒性, 能够有效抑制迭代过程中的数值振荡。

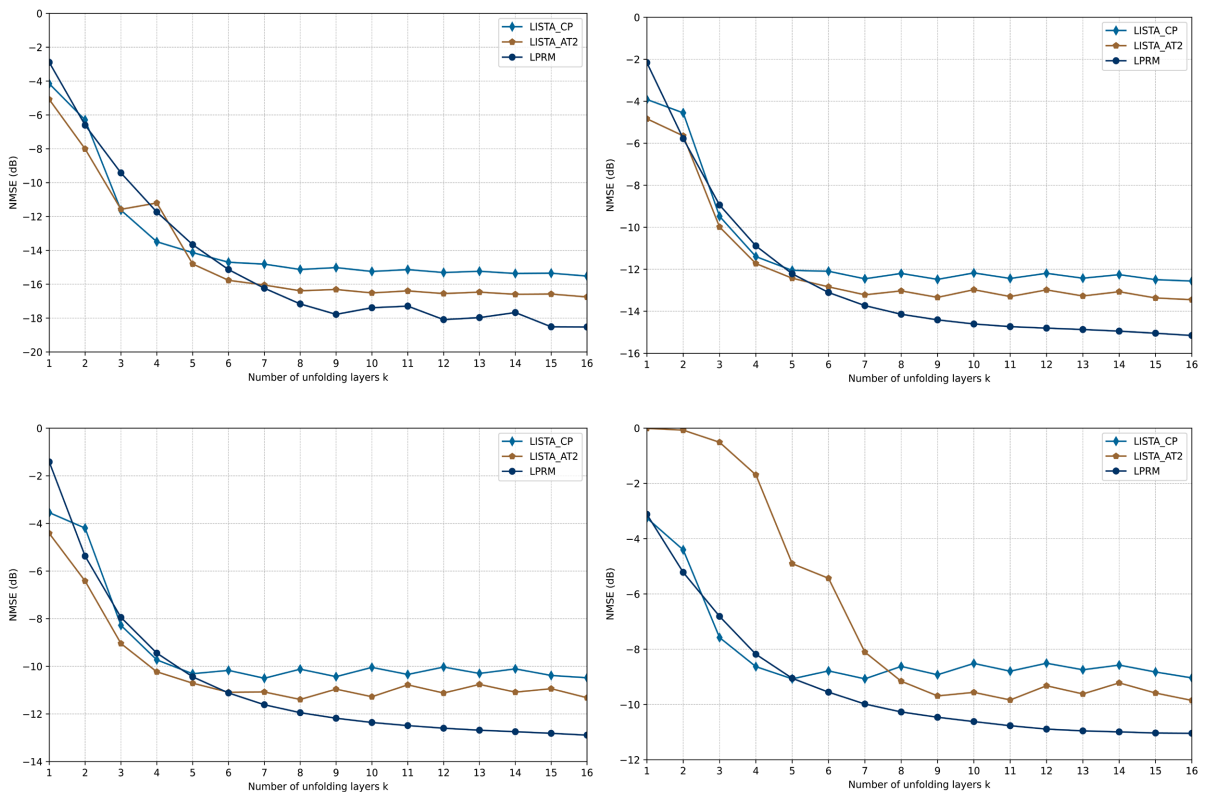




**Figure 1.** NMSE versus the number of unfolding layers for different models  
**图 1.** 不同模型下 NMSE 随展开层数的变化曲线

### 4.2. 噪声与条件数鲁棒性对比

进一步地，本文验证了 LPRM 模型在病态感知矩阵下显著的鲁棒性。实验设置高噪声环境(信噪比  $SNR = 20\text{ dB}$ )，并在条件数分别为 5, 15, 30, 50 的一系列病态感知矩阵  $D$  上对 LISTA-CP、LISTA-AT2 与 LPRM 进行训练与测试。如图 2 所示，即使在高噪声环境下( $SNR = 20\text{ dB}$ )，当条件数为 5, 15, 30 时，LPRM 的重建性能均显著优于 LISTA-CP 与 LISTA-AT2，且随着条件数增大，其优势更加明显。这一结果说明：基于 Peaceman-Rachford 框架的 LPRM 算法在处理高条件数字典时具备更好的病态鲁棒性。



**Figure 2.** NMSE versus the number of unfolding layers for different condition numbers  
**图 2.** 不同条件数下 NMSE 随展开层数的变化曲线

## 5. 总结与展望

本文围绕非凸截断  $\ell_1$  惩罚正则化下的稀疏信号重建问题, 构建了一类兼具理论严谨性与实用效能的深度展开模型。通过将 Peaceman-Rachford 方法进行深度展开, 并引入可学习参数, 所提出的方法不仅显著提升了重建性能, 还建立了坚实的理论保障。

未来的工作可以将本文所提模型推广至其他非凸正则化函数(如 MCP、SCAD 等), 以构建更具普适性的理论分析框架, 并探索将所提模型应用于更复杂的实际成像场景, 如动态 MRI、低剂量 CT 等, 以检验其在真实世界中的泛化能力与实用价值。

## 参考文献

- [1] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. <https://doi.org/10.1109/tit.2006.871582>
- [2] 李彤, 肖锋, 张文娟, 等. 基于深度学习的压缩感知图像重构算法综述[J]. *微电子学与计算机*, 2022, 39(12): 1-12.
- [3] Elad, M. (2010) *Sparse and Redundant Representations*. Springer.
- [4] Chen, C. and Huang, J. (2012) Compressive Sensing MRI with Wavelet Tree Sparsity. *26th Annual Conference on Neural Information Processing Systems 2012*, Lake Tahoe, 3-6 December 2012, 1106-1114.
- [5] Chen, H., Li, Q., Zhou, L. and Li, F. (2024) Deep Learning-Based Algorithms for Low-Dose CT Imaging: A Review. *European Journal of Radiology*, **172**, Article ID: 111355. <https://doi.org/10.1016/j.ejrad.2024.111355>
- [6] Selesnick, I. (2017) Sparse Regularization via Convex Analysis. *IEEE Transactions on Signal Processing*, **65**, 4481-4494. <https://doi.org/10.1109/tsp.2017.2711501>
- [7] Zhang, H., Wang, Q., Shi, J., Ying, S. and Wen, Z. (2025) Deep Unfolding Network with Spatial Alignment for Multi-Modal MRI Reconstruction. *Medical Image Analysis*, **99**, Article ID: 103331. <https://doi.org/10.1016/j.media.2024.103331>
- [8] You, D., Xie, J. and Zhang, J. (2021) ISTA-NET<sup>++</sup>: Flexible Deep Unfolding Network for Compressive Sensing. *2021 IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, 5-9 July 2021, 1-6. <https://doi.org/10.1109/icme51207.2021.9428249>
- [9] 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>
- [10] 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>
- [11] 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>
- [12] 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>
- [13] Gong, P., Ye, J. and Zhang, C. (2012) Multi-Stage Multi-Task Feature Learning. *26th Annual Conference on Neural Information Processing Systems 2012*, Lake Tahoe, 3-6 December 2012, 1997-2005.
- [14] Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. MIT Press.
- [15] Gregor, K. and Lecun, Y. (2010) Learning Fast Approximations of Sparse Coding. *Proceedings of the 27th International Conference on Machine Learning*, Haifa, 21-24 June 2010, 399-406.
- [16] Chen, X., Liu, J., Wang, Z., et al. (2018) Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montreal, 3-8 December 2018, 9061-9071.
- [17] Kim, D. and Park, D. (2020) Element-Wise Adaptive Thresholds for Learned Iterative Shrinkage Thresholding Algorithms. *IEEE Access*, **8**, 45874-45886. <https://doi.org/10.1109/access.2020.2978237>