

基于交替方向乘子法的非线性动力学稀疏辨识框架^{*}

江锋^{1,2} 都琳^{1,3†} 袁铭⁴ 白建超^{1,3} 马琴^{1,3} 邓子辰^{1,5}

(1. 西北工业大学 复杂系统动力学与控制工信部重点实验室, 西安 710129)

(2. 西北工业大学 力学与交通运输工程学院, 西安 710129)

(3. 西北工业大学 数学与统计学院, 西安 710129)

(4. 上海交通大学 机械系统与振动国家重点实验室, 上海 200240)

(5. 西北工业大学 极端力学研究院, 西安 710072)

摘要 基于系统稀疏先验的非线性动力学稀疏辨识方法(简称稀疏建模)是数据驱动动力学建模的代表性技术,该方法通过将动力学建模转化为求解稀疏优化问题,有效实现了模型可解释性与应用效果的平衡.然而,传统基于特定损失函数和正则项的稀疏优化方法,往往难以兼顾模型的精度、稀疏性、鲁棒性与计算效率.为此,本文聚焦稀疏建模的核心优化问题,构建了一种基于交替方向乘子法(ADMM)的统一迭代求解框架,以灵活适配不同目标的稀疏优化问题.该框架中,损失函数涵盖平方损失、绝对偏差损失及 Huber 损失,正则项则包括常用的一范数及其变体与二分之一拟范数非凸正则项.通过综合引入非凸正则项、鲁棒损失函数及加速迭代技术,系统提升了模型在上述多维性能指标上的综合表现.最后,以六维 Lorenz96 混沌系统为例验证了所提方法的有效性.结果表明:(1)所提框架可统一求解包含多种损失函数与正则项组合的稀疏优化问题,涵盖非光滑/非凸优化问题;(2)引入非凸正则项可显著增强对系统稀疏特征的刻画能力;(3)采用加速迭代策略有效提升了稀疏解的求解效率.

关键词 数据驱动建模, 非线性动力学, 稀疏优化, 非线性动力学稀疏辨识, 交替方向乘子法

中图分类号:O313;O242.1;O415.6

文献标志码:A

ADMM-based Optimization Framework for Sparse Identification of Nonlinear Dynamics^{*}

Jiang Feng^{1,2} Du Lin^{1,3†} Yuan Ming⁴ Bai Jianchao^{1,3} Ma Qin^{1,3} Deng Zichen^{1,5}

(1. MIIT Key Laboratory of Dynamics and Control of Complex Systems, Northwestern Polytechnical University, Xi'an 710129, China)

(2. School of Mechanics and Transportation Engineering, Northwestern Polytechnical University, Xi'an 710129, China)

(3. School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China)

(4. State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China)

(5. Institute of Extreme Mechanics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract Nonlinear dynamics sparse identification methods based on system sparsity priors (hereafter referred to as sparse modeling) represent a quintessential technique in data-driven dynamic modeling. By reformulating dynamic modeling into sparse optimization problems, this method effectively achieves a balance between model interpretability and practical performance. However, traditional sparse optimiza-

2025-12-02 收到第 1 稿, 2025-12-24 收到修改稿.

^{*} 国家自然科学基金资助项目(12472357, 12232015, 12471298), 陕西省杰出青年科学基金资助项目((2024JC-JCQN-05), 陕西数理基础科学研究项目(23JSQ031); National Natural Science Foundation of China(12472357, 12232015, 12471298), Shaanxi Province Outstanding Youth Fund Project(2024JC-JCQN-05), Shaanxi Fundamental Science Research Project for Mathematics and Physics (23JSQ031).

[†] 通信作者 E-mail: lindu@nwpu.edu.cn

tion methods based on specific loss functions and regularizers often struggle to balance the accuracy, sparsity, robustness, and computational efficiency of the models. To address this, focusing on the core optimization problem of sparse modeling, this paper proposes a unified iterative solution framework based on the Alternating Direction Method of Multipliers (ADMM) to flexibly accommodate sparse optimization problems with diverse objectives. Within this framework, the loss functions encompass squared loss, absolute deviation loss, and Huber loss, while the regularization terms include the conventional one-norm and its variants, alongside the one-half quasi-norm non-convex regularizer. By comprehensively integrating non-convex regularizers, robust loss functions, and accelerated iterative techniques, the proposed method systematically enhances the overall performance of the model across the aforementioned multidimensional metrics. Finally, the effectiveness of the proposed method is validated using the six-dimensional Lorenz 96 chaotic system. The results indicate that: (1) the proposed framework can uniformly solve sparse optimization problems containing various combinations of loss functions and regularizers, encompassing non-smooth/non-convex optimization problems; (2) the introduction of non-convex regularizers significantly enhances the ability to characterize the sparse features of the system; (3) the adoption of an accelerated iterative strategy effectively improves the efficiency of solving for sparse solutions.

Key words data-driven modeling, nonlinear dynamics, sparse optimization, SINDy, ADMM

引言

在实际工程系统中,存在诸多动力学问题.传统动力学建模方法依赖牛顿力学、拉格朗日力学和哈密顿力学等物理原理及数学技巧来构建模型.然而,随着问题复杂度增加,基于物理原理的建模方法逐渐面临困难,例如经典动力学模型受限于假设条件,难以适应复杂的实际场景.此外,随着计算机仿真技术的发展,计算固体力学、计算流体力学及空气动力学等领域积累了海量的仿真数据.因此,充分挖掘并利用这些实验与仿真数据,进而实现数据驱动的动力学建模,具有极其重要的工程意义与实际应用价值.

近二十年来,数据驱动的动力学建模研究引起广泛关注,基于符号回归的动力学建模^[1]从数据中发现微分方程形式,极大地催生了可获取显式表达式的数据驱动建模研究.然而,符号回归方法获取动力学模型的计算成本较高.当系统存在先验结构时,可通过充分利用先验信息,更加简约的获取动力学模型.2011年,Wang等^[2]首次采用压缩感知技术,实现非线性动力学和复杂网络的模型重构,开启了基于压缩感知的动力学模型重构工作.Schaeffer等^[3]借助软阈值算子求解稀疏优化问题,实现偏微分方程(PDE)模型重构.Brunton等^[4]和Rudy等^[5]分别发展了序列阈值最小二乘

法(STLS)和序列阈值岭回归法(STRidge)实现超定采样情形下常微分方程(ODE)和PDE的重构工作.Boninsegna等^[6]将稀疏建模方法扩展到随机微分方程(SDE)领域.经典的稀疏优化问题形式与LASSO^[7]密切相关,其目标函数通常由刻画拟合精度的损失函数项与刻画稀疏性的正则项构成.其中,损失函数项多采用平方损失,因其光滑性而具有良好的优化性能;正则项采用 ℓ_1 正则项,作为 ℓ_0 正则项的最佳凸近似.当稀疏性表征的稳定性不足时,弹性网正则项^[8]能够获得更加稳定的优化结果,已广泛应用于稳定模型选择^[9-11]以及流体相关问题^[12-14];加权 ℓ_1 正则项^[15]可进一步增强稀疏性表征,相关代表工作见文献[16-20].当平方损失不足以刻画离群值特性时,引入绝对偏差损失^[21]或Huber损失^[22]可显著提升模型辨识的鲁棒性^[19,23,24].

在稀疏优化问题的求解算法方面,已有软阈值迭代法^[25]、硬阈值迭代法^[26]以及快速软阈值迭代法^[27]等基础算法,但这类算法主要针对损失函数项为平方损失的LASSO问题或类LASSO问题.Zhang等^[28]提出了稀疏松弛正则化回归的框架(SR3),Champion等^[29]将该框架应用于若干低维动力学模型的重构;然而,该方法的损失函数仍以平方损失为核心,尚未系统展示其他非凸正则项的应用价值.

基于上述分析可知,传统的稀疏建模方法通常依赖于固定的损失函数与特定的正则化机制,缺乏

灵活且统一的求解框架. 然而, 面临实际问题提出的更高需求, 如强稀疏性、模型精度、鲁棒性及计算效率等, 现有方法的实现通常依赖多种优化算法或成熟的优化工具包. 基于此, 本文主要关注稀疏动力学建模中的核心优化问题, 通过构建基于交替方向乘子法 (ADMM) 的稀疏优化框架, 将不同形式的稀疏优化问题纳入该框架. 一方面, 可以考虑几类常见稀疏正则项, 同时引入非凸稀疏正则项, 以实现更优的稀疏优化性能; 另一方面, 传统平方损失函数难以刻画离群值特性, 因此可针对鲁棒性需求, 引入鲁棒损失函数及动量加速技术. 最终, 针对多种问题形式, 本文提出相应的基于 ADMM 的迭代求解策略, 从而在不同应用场景及同一场景的不同工况下, 通过合理调整优化问题并求解, 获得更加准确且简约的动力学模型.

1 数据驱动稀疏建模框架

一般而言, 考虑具有 n 个输出的未知系统

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad (1)$$

其中 $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$, $\mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})]^T$, 其系统初值 $\mathbf{x}(0) = [x_1(0), x_2(0), \dots, x_n(0)]^T$, 每个状态分量方程可表示为

$$\dot{x}_i = f_i(\mathbf{x}), \quad (2)$$

数据驱动建模主要目的在于从数据中推断出 $f_i(\mathbf{x})$ 的显式表达式. 稀疏动力学建模的基本假设为 $f_i(\mathbf{x})$ 可在特定的基函数下稀疏表示, 即少数关键特征基的系数非零, 多数冗余特征基系数为零. 记基函数库为 Θ , 稀疏系数向量为 ξ , 则式(2)函数表达式可形式化表示为

$$f_i(\mathbf{x}) = \Theta \xi. \quad (3)$$

常用的基函数包括多项式函数、三角函数以及依据先验经验人工构造的其他基函数等. 因此, 如何准确选取基函数以有效刻画稀疏性, 成为实现稀疏动力学建模的关键问题.

数据驱动的稀疏动力学建模以数据为出发点, 首先在有限时间 T 内对未知系统进行离散采样, 采样时刻记为 t_j ($j=0, 1, \dots, N$), 其中 $t_0=0, t_N=T$. 记录采样数据矩阵

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_N) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \cdots & x_n(t_N) \end{bmatrix}, \quad (4)$$

其中 $X_{i,j}$ 表示 t_j 时刻的状态变量 x_i . 相应地, 可通过有限差分法或先对函数进行光滑逼近后, 再进行数值微分, 以获得其导数数据矩阵

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^T(t_1) \\ \dot{\mathbf{x}}^T(t_2) \\ \vdots \\ \dot{\mathbf{x}}^T(t_N) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_N) & \dot{x}_2(t_N) & \cdots & \dot{x}_n(t_N) \end{bmatrix}, \quad (5)$$

其中 $\dot{X}_{i,j}$ 表示 t_j 时刻的状态变量 x_i 的导数. 值得注意的是, 该过程往往会引入较大误差, 尤其在数据存在显著噪声的情况下. 本文仅展示在最大阶数为 d 的多项式基下的稀疏动力学建模过程. 基函数库构建如下

$$\Theta(\mathbf{X}) = \begin{bmatrix} | & | & | & | & | \\ I(\mathbf{X}) & P_1(\mathbf{X}) & P_2(\mathbf{X}) & \cdots & P_d(\mathbf{X}) \\ | & | & | & | & | \end{bmatrix}, \quad (6)$$

其中 $I(\mathbf{X})_{i,j} = 1$ 表示 0 阶多项式, 即常数项; $P_2(\mathbf{X})$ 表示二阶多项式项, 具有如下形式

$$P_2(\mathbf{X}) = [x_1 \odot x_1, x_1 \odot x_2, \dots, x_i \odot x_j, \dots, x_n \odot x_n], \quad i \leq j, \quad (7)$$

其中, \odot 表示向量之间的哈达玛积 (Hadamard product), 即向量的对应分量逐点乘法; 其他更高阶多项式基满足类似构造法则. 此时, 在多项式最大阶数为 d 的情形, 函数库 $\Theta(\mathbf{X})$ 的行数 l 为采样数 N , 列数 p 为组合数 C_{n+d}^d .

对每一个子问题式(2)和式(3), 基于平方损失的通用稀疏优化问题可表述为带有正则项的结构优化问题

$$\xi_i = \arg \min_{\xi_i} \left\{ \frac{1}{2} \|\Theta(\mathbf{X})\xi_i' - \dot{\mathbf{X}}_{\cdot i}\|_2^2 + \mu R(\xi_i') \right\}, \quad i=1, 2, \dots, n, \quad (8)$$

其中 $R(\cdot)$ 为正则项, μ 为正则化系数, 用于平衡稀疏度与逼近精度. 当 $R(\xi) = \|\xi\|_0 = \#\{1 | \xi_i \neq 0\}$ 时, 该问题的求解与硬阈值算子高度相关, 例如压缩感知问题中的硬阈值迭代算法^[26] 和 SINDy 中的序列阈值最小二乘法^[4]. 当 $R(\xi) = \|\xi\|_1 = \sum_i |\xi_i|$ 时, 该问题的求解与软阈值算子高度相关, 例如 LASSO 问题的软阈值迭代算法^[25] 和快速软阈值迭代算法^[27]. 后文将讨论在引入其他正则项和损失函数项时, 如何对该类问题进行统一求解.

当各子问题均已求解时, 记其解集合的矩阵形

式为 $\Xi = [\xi_1, \xi_2, \dots, \xi_n] \in \mathbb{R}^{p \times n}$, 则可通过符号回归得到式(1)右端项的显式表达式

$$\dot{\mathbf{x}} = [\Theta(\mathbf{x}^T)\Xi]^T, \quad (9)$$

至此,即完成了从数据中推断得到动力学方程显式表达式的完整流程。

通常在数据获取阶段,需要考虑噪声干扰以及数据尺度不统一等问题,因此需进行适当的前处理。在优化问题求解过程中,单纯采用硬阈值或软阈值方法往往难以覆盖更广泛的应用场景,因而有必要引入其他损失函数项或正则项以实现更充分的表征。随后,需要对优化所得解进行去规范化处理,并提取相应的符号表达式。最后,通过模型重构对数据驱动辨识模型的有效性进行验证。上述流程可概括如图1所示,其中稀疏优化问题的构建与求解是本文的研究重点。

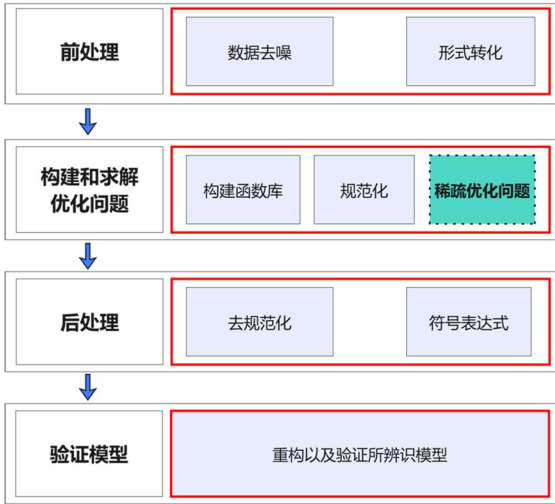


图1 动力学稀疏建模基本框架

Fig. 1 Fundamental framework of sparse dynamic modeling

2 基于ADMM的稀疏优化方法

在机器学习及数据驱动稀疏建模领域,出现了许多结构复杂,尤其是非凸且非光滑的稀疏优化问题。ADMM算法提供了一个适用范围广、易于理解与实现、性能可靠的综合解决方案^[30-32]。经典ADMM^[33]算法采用二分块形式,主要针对目标函数变量可分离且带有线性约束的优化问题(不局限于凸优化问题):

$$\begin{aligned} & \min_{\mathbf{x}, \mathbf{z}} f(\mathbf{x}) + g(\mathbf{z}) \\ & \text{s. t. } \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}, \end{aligned} \quad (10)$$

其中 $\mathbf{x} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{p \times m}, \mathbf{B} \in \mathbb{R}^{p \times n}$, 以及 $\mathbf{c} \in \mathbb{R}^p$ 。通过引入增广拉格朗日乘子

$$L_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \boldsymbol{\lambda}, \mathbf{Ax} +$$

$$\mathbf{Bz} - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}\|_2^2, \quad (11)$$

其中 $\boldsymbol{\lambda}$ 为拉格朗日乘子, $\langle \cdot, \cdot \rangle$ 表示向量的内积, ρ 为增广拉格朗日系数。借助式(11),可将式(10)转化为以下三个步骤进行迭代求解:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} L_\rho(\mathbf{x}, \mathbf{z}^k, \boldsymbol{\lambda}^k), \quad (12)$$

$$\mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} L_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k), \quad (13)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho(\mathbf{Ax}^{k+1} + \mathbf{Bz}^{k+1} - \mathbf{c}), \quad (14)$$

该过程包括两个单变量子优化问题及拉格朗日乘子的更新。由于子问题的求解顺序不影响最终结果,且子问题形式相似,本文仅讨论子问题式(12)的求解,子问题式(13)可采用相同策略。此时,子问题仅与变量 \mathbf{x} 相关, \mathbf{z}^k 和 $\boldsymbol{\lambda}^k$ 可被视为常数。在 $\arg \min$ 优化意义下,增加和减少常数项不影响结果,因此可通过配方法得如下形式的优化问题:

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} L_\rho(\mathbf{x}, \mathbf{z}^k, \boldsymbol{\lambda}^k) \\ &= \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz}^k - \mathbf{c} + \frac{1}{\rho} \boldsymbol{\lambda}^k\|_2^2 \right\}. \end{aligned} \quad (15)$$

该优化问题主要分为两种情形。

(1) $\mathbf{A} = \mathbf{I}$, 问题可表示为具有临近算子(参考附录)的标准形式:

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - (-\mathbf{Bz}^k + \mathbf{c} - \frac{1}{\rho} \boldsymbol{\lambda}^k)\|_2^2 \right\} \\ &= \text{Prox}_{1/\rho, f(\cdot)}(-\mathbf{Bz}^k + \mathbf{c} - \frac{1}{\rho} \boldsymbol{\lambda}^k), \end{aligned} \quad (16)$$

此时,可直接借助函数 f 的临近算子获得封闭形式解。例如,对于 $f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_i |x_i|$, 对应的标量形式软阈值算子 $S_t(x)$ 定义为

$$S_t(x) = \begin{cases} x - t, & x > t \\ 0, & x \in [-t, t], \\ x + t, & x < -t \end{cases} \quad (17)$$

其中 t 为非负系数。其向量形式可简单表示为 $S_t(\mathbf{x}) = \text{sgn}(\mathbf{x}) \odot \max\{|\mathbf{x}| - t, 0\}$, 其中 $\text{sgn}(\cdot)$ 表示符号函数,运算符 \odot 表示两个向量对应分量求积,此时迭代解可表示为 $S_{1/\rho}(-\mathbf{Bz}^k + \mathbf{c} - \boldsymbol{\lambda}^k/\rho)$ 。当 $f(\mathbf{x}) = \|\mathbf{x}\|_2^2/2 = \sum_i x_i^2/2$, 则可直接通过向量微分得到 $\mathbf{x}^{k+1} = (-\mathbf{Bz}^k + \mathbf{c} - \boldsymbol{\lambda}^k/\rho)\rho/(1+\rho)$ 。更多关于其他正则项临近算子的形式,可参考相关文献^[34,35]。

(2) $\mathbf{A} \neq \mathbf{I}$, 此时可在 \mathbf{x}^k 步对光滑项进行一次线性近似,从而得到类似式(16)的解形式。具体处

理如下:记式(15)右端的光滑项为

$$h(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} + \mathbf{Bz}^k - \mathbf{c} + \frac{1}{\rho} \boldsymbol{\lambda}^k\|_2^2, \quad (18)$$

对于 \mathbf{x}^{k+1} 的解,可在 $h(\mathbf{x}^k)$ 展开为

$$\begin{aligned} \tilde{h}(\mathbf{x}) &\approx h(\mathbf{x}^k) + \nabla h(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \frac{\tau_k}{2} \|\mathbf{x} - \mathbf{x}^k\|_2^2 \\ &= h(\mathbf{x}^k) + \frac{\tau_k}{2} \|\mathbf{x} - \mathbf{x}^k\|_2^2 + \mathbf{A}^T(\mathbf{Ax}^k + \\ &\quad \mathbf{Bz}^k - \mathbf{c} + \frac{1}{\rho} \boldsymbol{\lambda}^k)(\mathbf{x} - \mathbf{x}^k), \end{aligned} \quad (19)$$

其中 τ_k 为惩罚项参数,控制逼近范围.一般而言, $1/\tau_k$ 表示步长,因此 τ_k 通常选择范围为 $\tau_k \geq \|\mathbf{A}^T \mathbf{A}\|_2$. 在实际算法中,可直接取 $\tau_k = \|\mathbf{A}^T \mathbf{A}\|_2$. 此时,将 $\tilde{h}(\mathbf{x})$ 替换式(15)中的 $h(\mathbf{x})$, 并在 $\arg \min$ 优化意义下通过配方法,可得子问题的解式

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{\rho \tau_k}{2} \|\mathbf{x} - \mathbf{x}^k + \right. \\ &\quad \left. \frac{1}{\tau_k} \mathbf{A}^T(\mathbf{Ax}^k + \mathbf{Bz}^k - \mathbf{c} + \frac{1}{\rho} \boldsymbol{\lambda}^k)\|_2^2 \right\} \\ &= \text{Prox}_{1/(\rho \tau_k), f(\cdot)} \left[\mathbf{x}^k - \frac{1}{\tau_k} \mathbf{A}^T(\mathbf{Ax}^k + \right. \\ &\quad \left. \mathbf{Bz}^k - \mathbf{c} + \frac{1}{\rho} \boldsymbol{\lambda}^k) \right]. \end{aligned} \quad (20)$$

以下通过若干示例,详细阐述如何利用 ADMM 设计满足多种正则项及多种损失函数的稀疏优化问题求解算法.

2.1 基于平方损失稀疏优化问题的 ADMM 算法

稀疏优化问题式(8)可进一步统一表示为如下形式:

$$\arg \min_{\mathbf{x} \in \mathbb{R}^m} \{l(\mathbf{x}) + \mu R(\mathbf{x})\}, \quad (21)$$

其中, $l(\mathbf{x})$ 为损失函数项, $R(\mathbf{x})$ 为正则项, μ 为正则化系数,主要用于平衡模型精度与复杂度,通常可通过手动调整或交叉验证选取.

传统损失函数通常采用平方损失

$$l(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2, \quad (22)$$

其中 $\mathbf{A} \in \mathbb{R}^{p \times m}$, $\mathbf{b} \in \mathbb{R}^p$. 基于硬阈值算子的为 ℓ_0 正则项,其他常用的正则项包括:

(1) ℓ_1 正则项

$$R(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_i |x_i|, \quad (23)$$

该正则项为 ℓ_0 正则项的最佳凸近似,能够有效刻画稀疏性,理论基础完善,应用最为广泛.

(2) 加权 ℓ_1 正则项

$$R(\mathbf{x}) = \|\mathbf{Wx}\|_1 = \sum_i W_{i,i} |x_i|, \quad (24)$$

其中 \mathbf{W} 为对角线元素均为正值的对角矩阵,具有更强的稀疏性刻画能力.

(3) 弹性网正则项(elastic net)

$$R(\mathbf{x}) = \frac{1-\alpha}{2} \|\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_1, \quad (25)$$

其中 $\alpha \in (0, 1]$, 在刻画稀疏性的同时,使解更具稳定性.一般而言, α 可设置得非常接近 1, 例如 $\alpha = 0.999$.

基于增强刻画稀疏性的动机,本文初步引入 $\ell_{1/2}$ 拟范数非凸正则项,该正则项由徐宗本院士^[36] 于 2012 年提出. 研究表明,对于一般的 ℓ_q 非凸正则项, $q \in (0, 1)$, 其他正则项对比 $q=1/2$ 不存在显著优势. 因此,本文仅以 $\ell_{1/2}$ 正则项初步阐述 ADMM 算法在非凸稀疏优化问题中的应用,以及非凸正则项所带来的更强稀疏表征能力.

(4) $\ell_{1/2}$ 正则项

$$R(\mathbf{x}) = \|\mathbf{x}\|_{1/2} = \sum_i |x_i|^{1/2}. \quad (26)$$

通过引入辅助变量 \mathbf{z} , 可将式(21)的无约束优化问题转化为约束优化问题

$$\arg \min_{\mathbf{x} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^m} \{l(\mathbf{x}) + \mu R(\mathbf{z}) \mid \text{s.t. } \mathbf{x} - \mathbf{z} = \mathbf{0}\}. \quad (27)$$

此时,采用 ADMM 算法分别求解式(12)和式(13),得到 x 子问题的显式解为

$$\mathbf{x}^{k+1} = (\mathbf{A}^T \mathbf{A} + \rho \mathbf{I}_{m \times m})^{-1} (\mathbf{A}^T \mathbf{b} + \rho \mathbf{z}^k - \boldsymbol{\lambda}^k), \quad (28)$$

z 子问题的解由式(16)给出,即

$$\mathbf{z}^{k+1} = \text{Prox}_{\mu/\rho, R(\cdot)}(\mathbf{x}^{k+1} + \frac{1}{\rho} \boldsymbol{\lambda}^k). \quad (29)$$

2.2 基于绝对偏差损失稀疏优化问题的 ADMM 算法

目前,大多数稀疏优化问题采用平方损失作为损失函数项.其优点在于平方损失为光滑函数,可直接进行线性近似,或如上一节所述得到子问题的精确解.然而,平方损失在刻画离群值方面存在显著缺陷,对于较大离群值,平方损失往往失去其有效度量功能.因此,可考虑使用绝对偏差损失替代式(22),此时损失函数项为

$$l(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_1, \quad (30)$$

正则项仍然考虑稀疏建模算法中的 ℓ_0 正则项以及上一节中的四种正则项,用 $R(\mathbf{x})$ 表示. 通过引入辅助变量 $\mathbf{z} = \mathbf{Ax} - \mathbf{b}$, 可得到对应的约束优化问题

$$\arg \min_{\substack{\mathbf{x} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^p}} \{ \|\mathbf{z}\|_1 + \mu R(\mathbf{x}) \mid \text{s. t. } \mathbf{Ax} - \mathbf{z} - \mathbf{b} = \mathbf{0} \}, \quad (31)$$

采用 ADMM 算法求解式(12)和式(13),其中 x 子问题的解满足式(15) $\mathbf{A} \neq \mathbf{I}$ 的条件,解由式(20)给出,即

$$\mathbf{x}^{k+1} = \text{Prox}_{\mu/(\rho\tau_k), R(\cdot)} \left(\mathbf{x}^k - \frac{1}{\tau_k} \mathbf{c} \right), \quad (32)$$

其中 $\mathbf{c} = \mathbf{A}^\top (\mathbf{Ax}^k - \mathbf{z}^k - \mathbf{b} + \boldsymbol{\lambda}^k / \rho)$. \mathbf{z} 子问题的解可由软阈值算子式(17)表示,即

$$\mathbf{z}^{k+1} = S_{1/\rho} (\mathbf{Ax}^{k+1} - \mathbf{b} + \boldsymbol{\lambda}^k). \quad (33)$$

2.3 基于 Huber 损失稀疏优化问题的 ADMM 算法

前文提到的平方损失与绝对偏差损失在不同应用场景下各有优势,然而,绝对偏差损失函数在零点处非光滑,对于高维信号恢复等实际问题,其性质并不理想. 因此,有必要引入光滑的鲁棒损失函数. Huber 损失函数综合了平方损失的光滑性与绝对偏差损失的鲁棒性,其定义如下

$$l^{\text{hub}}(\mathbf{x}) = \sum_i \text{Huber}(\mathbf{Ax} - \mathbf{b})_i, \quad (34)$$

$$\text{Huber}(v) = \begin{cases} \frac{v^2}{2}, & |v| \leq 1, \\ |v| - \frac{1}{2}, & |v| > 1. \end{cases}$$

对该问题的求解,同样可引入辅助变量 $\mathbf{z} = \mathbf{Ax} - \mathbf{b}$, 从而得到相应的约束优化问题

$$\arg \min_{\substack{\mathbf{x} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^p}} \{ l^{\text{hub}}(\mathbf{z}) + \mu R(\mathbf{x}) \mid \text{s. t. } \mathbf{Ax} - \mathbf{z} - \mathbf{b} = \mathbf{0} \}, \quad (35)$$

采用 ADMM 算法求解式(12)和式(13),并参考绝对偏差损失情形下子问题的求解及文献[33]中关于 Huber 函数临近算子的定义,可知 x 子问题的解与式(32)相同, \mathbf{z} 子问题的解为

$$\mathbf{z}^{k+1} = \frac{\rho}{1+\rho} (\mathbf{Ax}^{k+1} - \mathbf{b} + \boldsymbol{\lambda}^k) + \frac{1}{1+\rho} S_{1+\rho} (\mathbf{Ax}^{k+1} - \mathbf{b} + \boldsymbol{\lambda}^k), \quad (36)$$

由子问题解的表达形式可知,该解同时兼具平方损失与绝对偏差损失的特性.

此外,其他损失函数同样可以在 ADMM 算法

框架下进行统一表示,因篇幅所限,此处不再赘述.

3 数值算例

本节以高维 Lorenz 96 混沌系统^[37,38]为例,验证基于 ADMM 的稀疏优化算法的有效性. 由于高维系统重构面临维度灾难问题,本文采用 Jiang 等^[39]提出的基于字典学习的快速非线性动力学稀疏辨识方法(FSINDy)以实现仿真加速. Lorenz 96 系统表示如下:

$$\begin{aligned} \dot{x}_i &= (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i = 1, 2, \dots, n, \\ x_{i-n} &= x_{i+n} = x_i, \end{aligned} \quad (37)$$

其中, F 表示外部激励,当设定为固定值 $F = 8$, 系统会产生混沌运动; n 表示系统的维度 ($n \geq 4$). 图 2 展示了 Lorenz 96 系统的混沌特性.

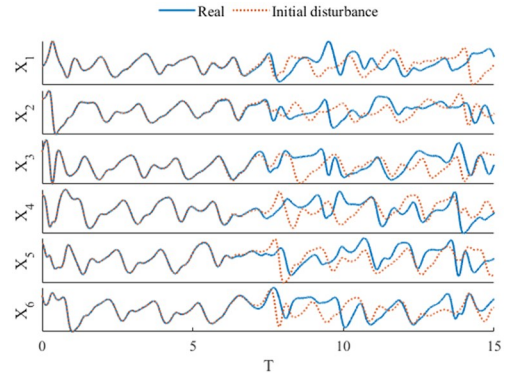


图2 具有混沌动力学特性的 Lorenz 96 模型
Fig. 2 Lorenz 96 model with chaotic dynamics

仿真参数设置如下:系统状态变量数 $n = 6$, 数据采样区间 $T = [0, 15]$, 时间间隔 $\Delta t = 0.001$; 为模拟环境扰动,对仿真系统添加 1% 的高斯白噪声 (见图 3); 多项式函数库阶数 $d = 2$, 正则项系数 $\mu \in [10^{-4}, 20]$, 并于该区间以对数间隔均匀选取 100 个正则化系数序列,以进行充分对比.

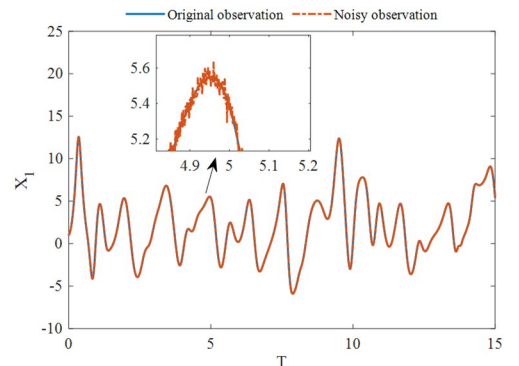


图3 原始观测数据以及噪声观测数据
Fig. 3 Original observation data and noisy observation data

类似于第 1 节的处理流程,该系统的稀疏优化问题可表示为

$$\xi_i = \arg \min_{\xi_i} \{l[\Theta(\mathbf{X})\xi_i' - \dot{\mathbf{X}}_{:,i}] + \mu R(\xi_i')\},$$

$$i = 1, 2, \dots, n, \quad (38)$$

其通用求解方案采用 ADMM 算法. 为便于与第 2 节算法部分对应, 每个子问题可表示为: $\mathbf{x} = \xi_i$, $\mathbf{A} = \Theta(\mathbf{X})$, $\mathbf{x}' = \xi_i'$, $\mathbf{b} = \dot{\mathbf{X}}_{:,i}$.

本实验中, ADMM 算法的通用系数设置如下: 最大迭代次数 $N = 2000$, 增广拉格朗日系数基于经验原则选取并设定为固定值 $\rho = 0.9$, 部分讨论可见第 4.2 节. 记真实稀疏系数矩阵为 $\Xi = [\xi_1, \xi_2, \dots, \xi_n]$, 其非零元素位置的指标集为 \mathcal{T} , 其零元素的指标集为 \mathcal{F} , 稀疏优化后稀疏系数矩阵为 $\hat{\Xi} = [\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_n]$. 同时定义三种指标综合考察辨识准确度:

(1) 成功辨识率

$$S = \frac{\#\{\hat{\Xi}(\mathcal{T})\} + \#\{\hat{\Xi}(\mathcal{F})\}}{\#\{\hat{\Xi}\}} \times 100\%, \quad (39)$$

(2) 最大系数误差

$$E_{\text{coef}} = \max \left(\frac{\|\hat{\Xi} - \Xi\|_2}{\|\Xi\|_2} \right), \quad (40)$$

(3) 最大真实系数相对误差

$$E_{\text{rel_coef}} = \max \left[\frac{\|\hat{\Xi}(\mathcal{T}) - \Xi(\mathcal{T})\|_2}{\|\Xi(\mathcal{T})\|_2} \times 100\% \right]. \quad (41)$$

针对不同问题, 迭代初值的设定以及停止条件略有差异.

3.1 损失函数为平方损失

针对损失函数为平方损失的情形, 优化问题由式(27)给出. 为了增强迭代结果的稳定性, 可将迭代初值 \mathbf{x}^0 设置为岭回归问题的解, 即

$$\mathbf{x}^0 = \arg \min_x \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \epsilon \|\mathbf{x}\|_2^2 \right\}, \quad (42)$$

此时, 迭代初值为 $\mathbf{x}^0 = (\mathbf{A}^T \mathbf{A} + \epsilon \mathbf{I}_{p \times p})^{-1} (\mathbf{A}^T \mathbf{b})$, 其中 $\epsilon = 10^{-6}$, \mathbf{I} 表示单位矩阵. 根据等式约束关系, 可令迭代初值 $\mathbf{z}^0 = \mathbf{x}^0$, 以及 $\boldsymbol{\lambda}^0 = \mathbf{I}^{p \times 1}$, 其中 \mathbf{I} 表示全 1 常数向量. 另外该问题中, x 子问题具有显式解, 且迭代过程中 $(\mathbf{A}^T \mathbf{A} + \rho \mathbf{I}_{m \times m})$ 为对称正定矩阵, 可提前通过乔里斯基(Cholesky)分解进行矩阵缓存, 以避免后续频繁的大规模矩阵求逆运算. 停止条件由迭代相对误差 (IRE) 刻画, 定义为

$$\epsilon_{\text{IRE}} = \frac{\max(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2, \|\mathbf{z}^{k+1} - \mathbf{z}^k\|_2, \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|_2)}{\max(\|\mathbf{x}^k\|_2, \|\mathbf{z}^k\|_2, \|\boldsymbol{\lambda}^k\|_2, 1)}, \quad (43)$$

并设定迭代相对误差限为 ϵ_{lim} . 停止条件为当 $\epsilon_{\text{IRE}} < \epsilon_{\text{lim}}$ 或达到预设最大迭代次数 N 时. 本实验中设定迭代相对误差限 $\epsilon_{\text{lim}} = 10^{-12}$. 图 4 综合对比了五种正则项在平方损失函数场景下的数值实验结果.

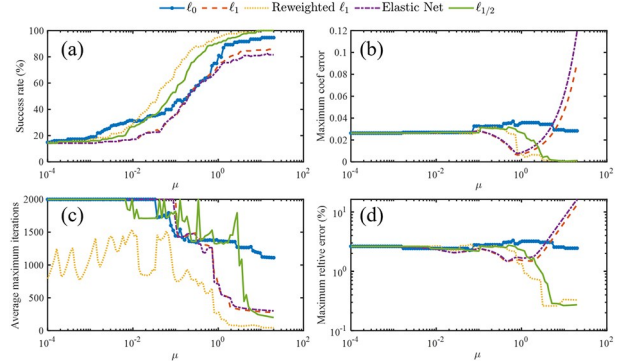


图 4 基于平方损失的五种正则项对比: (a) 成功辨识率; (b) 最大系数误差; (c) 平均最大迭代次数; (d) 最大真实系数相对误差
Fig. 4 Comparison of five regularization terms based on square loss: (a) Success rate; (b) Maximum coefficient error; (c) Average maximum iterations; (d) Maximum relative error of true coefficients

值得注意的是, 加权 ℓ_1 正则化问题的求解具有特殊性, 其权重矩阵需动态更新, 因此需要外层迭代. 该过程主要包含以下五个步骤.

步骤 1: 初始化 $k = 0$, $\mathbf{W}^0 = \mathbf{I}_{p \times p}$, 解式(42)得 \mathbf{x}^0 ;

步骤 2: 求解问题

$$\mathbf{x}^k = \arg \min_x \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{x}' - \mathbf{b}\|_2^2 + \mu \|\mathbf{W}^k \mathbf{x}'\|_1 \right\};$$

步骤 3: 更新权重矩阵 $\mathbf{W}_j^{k+1} = \frac{1}{|x_j^k|^q + \epsilon}$;

步骤 4: 下一次迭代 $k = k + 1$;

步骤 5: 判断 $k \geq k_{\text{max}}$. 满足条件, 退出; 不满足条件, 回到步骤 2.

初次迭代时, 将权重矩阵设置为单位矩阵, 此时该问题等同于 ℓ_1 正则化问题. 本文中, 设置外层迭代次数 $k_{\text{max}} = 10$, $q = 1$, $\epsilon = 10^{-4}$, 其他关于加权 ℓ_1 正则项问题细节可参考文献[17]. 出于可视化对比的目的, 仅展示平均迭代次数. 结果表明, 加权 ℓ_1 正则化能够带来更高的正确率, 但代价是远高于其他正则项的迭代次数(图 4 中的平均迭代次数需乘以 k_{max}). 作为对比, 其他正则项仅需处理步骤 2, 因为无需更新权重矩阵 \mathbf{W} . 后续研究可考虑将权重矩阵引入其他正则项, 以探究能否实现更快的收敛速度.

由上述结果可知, 加权 ℓ_1 正则项、 $\ell_{1/2}$ 正则项

具有最高的成功辨识率,其次为 ℓ_0 正则项,而 ℓ_1 正则项以及弹性网正则项的成功辨识率最低. 然而从最大迭代次数的统计结果可以看出, ℓ_1 正则项与弹性网正则项作为凸正则化方法,表现出更好的数值稳定性,其迭代次数波动较小,且整体收敛速度更快. 进一步比较最大系数误差及最大真实系数相对误差可发现,加权 ℓ_1 正则项和 $\ell_{1/2}$ 正则项在两项误差指标上均表现更优,表明其在系数辨识方面具有更高的估计精度. 尽管 ℓ_1 正则项以及弹性网正则项的误差在达到较低水平后出现一定程度的回升,但在可接受误差范围内,其辨识结果仍具有一定参考价值.

一般而言,在含噪数据条件下,稀疏优化方法能够获得具有稀疏结构的解,但往往难以实现充分的稀疏化,部分冗余特征对应的系数仍保留较小的非零值. 针对这一问题,可通过额外引入筛选阈值,对尚未充分优化的参数进行过滤. 在本算例中,若对 ℓ_0 正则项、 ℓ_1 正则项以及弹性网正则项的辨识结果施加上述后处理,可实现 100% 的成功辨识率. 需要指出的是,该处理方式存在两方面需要说明:一方面,从理论角度看,额外引入筛选阈值(即硬阈值操作)相当于在原优化问题求解完成后,引入一个基于稀疏先验的附加优化过程;另一方面,从工程实践角度出发,有必要引入合理的评价指标,以评估删除小系数对最终辨识结果是否产生显著影响.

图 5 给出了基于 ℓ_1 正则项的辨识结果与真实系数的对比,可以看到,尽管所有真实特征均已被成功识别,但仍存在稀疏化不足的问题:一方面,对幅值较大的系数尚未获得足够精确的估计;另一方面,优化结果中仍保留了大量接近于零的小系数,从而导致成功辨识率未能达到 100%.

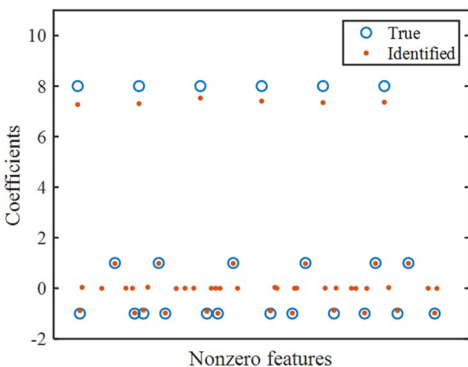


图 5 基于 ℓ_1 正则项的辨识结果与真实系数的对比
Fig. 5 Comparison of identified coefficients based on ℓ_1 regularization and true coefficients

3.2 损失函数为绝对偏差损失

针对损失函数为绝对偏差损失的情形,优化问题由式(31)给出. 参考平方损失场景下初值的选择与设置策略,在例中设定迭代相对误差限 $\epsilon_{lim} = 10^{-6}$. 图 6 综合对比了五种正则项在绝对偏差损失函数场景下的数值实验结果.

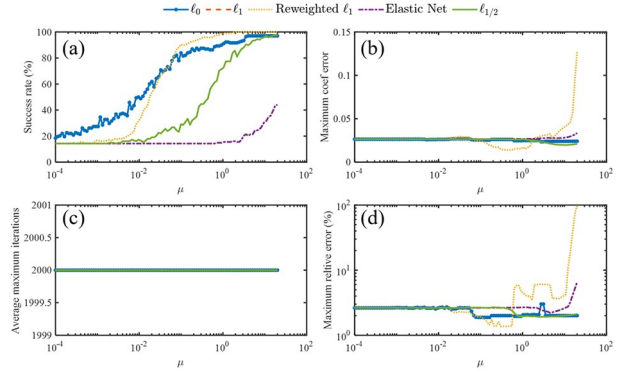


图 6 基于绝对偏差损失的五种正则项对比:(a)成功辨识率;(b)最大系数误差;(c)平均最大迭代次数;(d)最大真实系数相对误差
Fig. 6 Comparison of five regularization terms based on absolute deviation loss:(a) Success rate;(b) Maximum coefficient error;(c) Average maximum iterations;(d) Maximum relative error of true coefficients

由图 6 可见,当选用绝对偏差损失时,加权 ℓ_1 正则项、 $\ell_{1/2}$ 正则项和 ℓ_0 正则项仍具有较高的成功辨识率,但随着正则化系数的增加,加权 ℓ_1 正则项会导致误差显著增加. ℓ_1 正则项及弹性网正则项成功辨识率较低,但在误差允许范围内仍可使用,同时需通过额外设置小阈值以过滤掉尚未充分优化的小系数. 然而,在无离群值场景下,绝对偏差损失的优化性能表现不如平方损失,表现为所有优化器均达到预设最大迭代次数上限. 但在存在离群值的场景下,绝对偏差损失通常表现更佳,正如笔者前期工作^[24]以及其他相关研究^[19,23]所示. 本文在此处主要展示 ADMM 算法的可行性.

3.3 损失函数为 Huber 损失

针对损失函数为 Huber 损失的情形,优化问题由式(35)给出. 针对绝对偏差损失场景下尚未充分优化的问题,本文初步引入动量加速技术,此时仅需对 x 子问题迭代求解公式做适当修正为:

$$\begin{cases} \tilde{x}^k = x^k + \frac{k}{k+3}(x^k - x^{k-1}), \\ x^{k+1} = \text{Prox}_{\mu/(\rho\tau_k), R(\cdot)}\left(\tilde{x}^k - \frac{1}{\tau_k}c\right), \end{cases} \quad (44)$$

其中, $c = A^T(A\tilde{x}^k - z^k - b + \lambda^k/\rho)$, \tilde{x}^k 利用前两

次迭代的解,实现动量加速作用^[35].参考平方损失场景下初值的选择与设置策略,本例中迭代相对误差限设为 $\epsilon_{\text{lim}} = 10^{-6}$.图7综合对比了五种正则项在 Huber 损失函数场景下的数值实验结果.

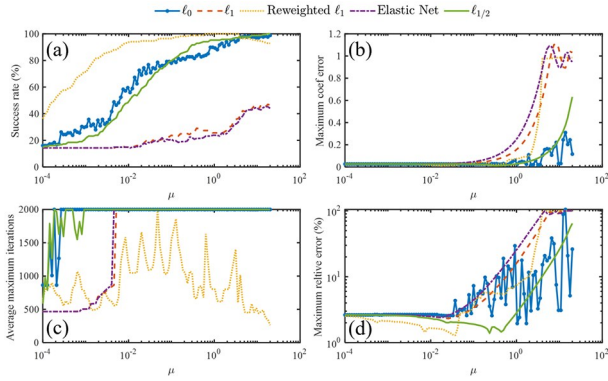


图7 基于 Huber 损失的五种正则项对比:(a)成功辨识率;(b)最大系数误差;(c)平均最大迭代次数;(d)最大真实系数相对误差
Fig. 7 Comparison of five regularization terms based on Huber loss:
(a) Success rate; (b) Maximum coefficient error; (c) Average maximum iterations; (d) Maximum relative error of true coefficients

由图7可知,当正则化系数较小时,动量加速效果显著,表现为较低的迭代次数即可达到收敛条件;然而,随着正则化系数增大,会出现过度优化现象,即成功辨识率从100%降低,且误差呈增大趋势.对于实际优化问题而言,合理引入加速技术可显著降低计算时间成本.

4 讨论

4.1 两种稀疏优化范式及其应用场景

目前,稀疏优化领域主要存在两种范式:其一为对系统进行欠定采样,构建欠定稀疏优化问题,主要利用压缩感知技术求解;其二为对系统进行超定采样,构建超定稀疏优化问题,主要利用 STLS 或 STRidge 方法求解.两者在一定程度上均可转化为 LASSO^[7]形式,即

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu \|\mathbf{x}\|_1 \right\}, \quad (45)$$

其中 μ 为正则化系数. ADMM 优化算法广泛应用于压缩感知问题的求解,其中普遍采用欠定采样策略构建稀疏优化问题.然而,本文采用超定采样策略,即构建超定系统的优化问题,同时拓展 LASSO 问题的形式,引入不同损失函数项和稀疏正则项.

4.1.1 整体架构

首先,在平方损失下,为刻画稀疏性引入多种

稀疏正则项,并在统一范式下进行求解,将问题转化为变量可分离的约束优化问题,并采用 ADMM 算法求解.其次,考虑传统平方损失函数在距离度量上的局限性,额外引入绝对偏差损失和 Huber 损失两类鲁棒损失函数,并给出了基于 ADMM 的迭代求解策略.同时,该框架可结合动量加速技术,以提升稀疏解的求解效率.综上,在实际应用中,可根据需要调整损失函数、正则项及加速策略,从而实现模型辨识的鲁棒性、稳定性与稀疏性目标.

4.1.2 正则项的效果

针对本文结果,首先可以看出 ℓ_1 正则项具有稀疏优化能力,但准确率并未达到100%,原因在于优化结果中仍存在许多小量,通常量级较小.通过额外引入阈值过滤,这些小量可以被去除,同时仍保持较好的性能.然而, ℓ_1 正则项的优势并不在此,其主要体现在两个层面:第一,在误差允许的范围内,优化结果具有稳定性.这是因为 ℓ_1 正则项是对原始 ℓ_0 正则项的最佳凸近似,而凸优化问题天然具有稳定性;第二,该方法具有良好的拓展性.例如,弹性网正则项和加权 ℓ_1 正则项的求解均与 ℓ_1 正则化问题的解——软阈值算子密切相关.同时,其他非凸正则项在刻画稀疏性时,也要求在原点处左右导数为-1和+1^[40],即 ℓ_1 正则项可以作为其他非凸正则项在原点附近的最佳凸近似,从而使非凸正则项的解同样与软阈值算子密切相关,如 SCAD^[41]和 MCP^[42]正则项.然而,本文未考虑这两类正则项,因为它们额外引入了控制曲线开口的参数,并且与正则化系数耦合,缺乏良好的可分离特性.

其次,加权 ℓ_1 正则项能够更好地刻画稀疏性,即在相同的正则化参数下,其准确率通常高于其他方法.然而,该方法涉及外层循环,需要引入额外参数以控制循环次数并避免分母除零错误,从而增加了迭代开销.基于 ℓ_0 正则项的结果表现出较高的准确率,这得益于采用了多量数据模式,构建了超定系统的稀疏优化问题.在此情形下,基于最小二乘法的方法如 STLS 和 STRidge 也能取得较好表现.然而,对于欠定系统,该方法效果明显下降(最小二乘失效),难以与基于 ℓ_1 正则项的结果相媲美.

同时,本文考虑了一种特殊的 ℓ_q ($q=1/2$) 非凸正则项,该正则项具有更强稀疏刻画能力.实验结果表明,相较于 ℓ_1 正则项, $\ell_{1/2}$ 正则项在稀疏表

征方面表现更优,同时算法的迭代次数和优化误差也具有明显优势.然而,由于其过强的稀疏性,在增大正则化系数时会出现过度优化现象,从而导致精度下降.在本文的示例中,基于 ℓ_0 正则项的结果尚未出现过度优化,但随着正则化系数继续增大,同样会出现类似现象.在极端情形下,当系数超过某一阈值时,所有变量都被压缩为零.因此,正则化系数的选择应控制在合理区间内.

更进一步,从本质上讲,加权 ℓ_1 正则项方法属于非凸优化方法,与对数惩罚项密切相关^[15].本文展示的一般结果表明,基于非凸正则项(加权 ℓ_1 正则项、 ℓ_0 正则项以及 $\ell_{1/2}$ 正则项)的优化问题在特定情形下具有更优的稀疏优化性能,但稳定性略逊,例如,图中显示随着正则化系数的微小变化,最大迭代次数可能出现较大波动.因此,后续工作的重点之一是探索其他非凸正则项,使其在保证更高精度的同时,实现更少的迭代次数和更稳定的迭代结果.

4.1.3 正则化系数的选取

仍以 LASSO 问题式(45)为例,其损失函数项可表示为 $l(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2$,正则项可表示为 $R(\mathbf{x}) = \|\mathbf{x}\|_1$,此时优化问题满足一般形式

$$\mathbf{x}^* = \arg \min_x \{\omega_1 l(\mathbf{x}) + \omega_2 R(\mathbf{x})\}, \quad (46)$$

其中 $\omega_1 = \frac{1}{1+\mu}$, $\omega_2 = \frac{\mu}{1+\mu}$, $\omega_1 + \omega_2 = 1$.

针对欠定采样,构建压缩感知问题时,系数矩阵 $\mathbf{A}_{m \times n}$ 中的 m 往往远小于 n ,方程 $\mathbf{Ax} = \mathbf{b}$ 具有无穷多个解.通过引入稀疏正则项,可将解约束到满足特定稀疏性条件,即从无穷多解限制到一个解.此时损失函数 $l(\mathbf{x})$ 往往具有非常小的误差,可通过 ϵ 来限制,即 $\sup[l(\mathbf{x})] \leq \epsilon$.然而正则项的量级通常远大于损失函数的上界,例如通过 M 来刻画,即 $\sup[R(\mathbf{x})] \leq M$.因此,需要通过正则化系数来平衡损失项和正则项的权重,其选取通常与 $\|\mathbf{A}^T \mathbf{b}\|_\infty$ 相关,例如可选择 $\mu = 10^{-6} \times \|\mathbf{A}^T \mathbf{b}\|_\infty$.

针对超定采样,在构建带稀疏约束的最小二乘问题时,系数矩阵 $\mathbf{A}_{m \times n}$ 中的 m 往往远大于 n ,根据线性代数基本知识,方程 $\mathbf{Ax} = \mathbf{b}$ 无解,但在数值计算上具有最小二乘解.通过引入稀疏正则项,可将最小二乘解约束为具有稀疏特性的解,因此迭代初值可设置为最小二乘解,或根据稳定性需求设置为岭回归解.此时,损失函数 $l(\mathbf{x})$ 通常具有误差下

界,可通过 M_1 来表示,即 $\inf[l(\mathbf{x})] \geq M_1$.正则项可通过 M_2 来刻画,即 $\sup[R(\mathbf{x})] \leq M_2$.随着采样数量的增加,最小二乘解的误差下界也会提高,因此需要较大的正则化系数以平衡损失项与正则项的权重.如本文结果图4所示,较大的正则化系数可提升成功辨识率,但过大的正则化系数突出了稀疏性的权重,可能导致过度稀疏,从而降低成功辨识率.因此,在其他问题中,可根据优化问题形式和采样数量合理预设正则化系数,以快速获得稀疏解,而无需通过大范围遍历来确定合适的正则化系数取值范围.

4.2 ADMM 优化算法及稀疏建模问题的局限性

本文主要针对 ADMM 所求解问题的形式,构造了一系列相应问题的迭代求解格式.参考算法收敛性分析^[33],原始 ADMM 算法具有 $O(1/k)$ 的收敛速度,通过引入超松弛技术、动量加速技术和变步长技术等,可显著提升 ADMM 算法的收敛性能,例如实现 $O(1/k^2)$ 或更快的收敛速度.然而,这些技术的引入通常会引入额外的超参数,这些超参数需预先设定,或在运行中动态调整.理论收敛结果往往依赖于特定假设,因此如何有效调整超参数涉及理论与经验的结合.例如,本文所用的一个重要超参数——增广拉格朗日系数 ρ ,在所有实验中均设定为固定值,用于平衡原始残差与对偶残差.对于该参数,也可采用动态调整方法^[43,44],但对于某些复杂问题,尤其是非凸问题,迭代解易陷入局部极值点,因此超参数调整仍是方法的一个局限.

本文所提出的方法采用二分块形式的 ADMM 算法.当稀疏性不是唯一约束时,通常需要引入其他正则项,从而增加可分离变量的数量,这时可采用多分块形式的 ADMM 算法.多分块 ADMM 提供了一种可行的求解途径,但已有研究表明,其理论保障仍不足,例如针对简单三分块问题的反例^[45].尽管如此,在实际应用中,多分块 ADMM 仍可作为探索性求解方法进行尝试.

此外,本文尚未涉及动力学稀疏建模的其他潜在难题.如图1所示,稀疏动力学建模包含多个步骤,每一步均可能存在潜在难题,同时已有部分研究为相关问题提供了可能的解决方案.例如:前处理中的降噪问题可能导致数值微分困难^[46,47];形式变换问题可通过积分形式或偏微分方程弱形式

加以解决^[48,49];优化过程中基函数库的多重共线性问题可采用其他稀疏优化技术处理^[50,51];反问题解的非唯一性则可通过正则化方法加以缓解^[52-54].完整考虑所有潜在问题超出了本文研究范围,相关问题可留待后续研究.

5 结论

本文从 ADMM 优化算法的角度提供了一种统一的稀疏优化框架,用于从数据中构建动力学模型.该框架从损失函数与正则项两个方面整合了平方损失、绝对偏差损失以及 Huber 损失,同时给出了基于 ℓ_0 、 ℓ_1 、弹性网、加权 ℓ_1 以及 $\ell_{1/2}$ 正则项的 ADMM 迭代求解策略.在六维 Lorenz 96 混沌系统模型上验证了多种算法的实际性能,结果表明:非凸正则项通常能更好地刻画稀疏性;基于 ℓ_1 正则项的稀疏优化是许多凸/非凸正则项的基础,其结果具有更高稳定性,显示出强大的应用潜力.

在后续工作中,可从以下三个方面进行拓展.第一,探索具有稳定性、显式解以及更强稀疏表征能力的非凸正则项;第二,在 ADMM 框架基础上引入加速收敛技术,以便更好地应用于对时效要求高的工业问题;第三,从 ADMM 优化框架出发,关注数据驱动动力学模型辨识中的其他潜在难题,如噪声问题、多重共线性问题等,并将 ADMM 模块嵌入其他问题求解中.

参考文献

- [1] SCHMIDT M, LIPSON H. Distilling free-form natural laws from experimental data [J]. *Science*, 2009, 324(5923): 81–85.
- [2] WANG W X, YANG R, LAI Y C, et al. Predicting catastrophes in nonlinear dynamical systems by compressive sensing [J]. *Physical Review Letters*, 2011, 106(15): 154101.
- [3] SCHAEFFER H, CAFLISCH R, HAUCK C D, et al. Sparse dynamics for partial differential equations [J]. *Proceedings of the National Academy of Sciences*, 2013, 110(17): 6634–6639.
- [4] BRUNTON S L, PROCTOR J L, KUTZ J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems [J]. *Proceedings of the National Academy of Sciences*, 2016, 113(15): 3932–3937.
- [5] RUDY S H, BRUNTON S L, PROCTOR J L, et al. Data-driven discovery of partial differential equations [J]. *Science Advances*, 2017, 3(4): e1602614.
- [6] BONINSEGNA L, NÜSKE F, CLEMENTI C. Sparse learning of stochastic dynamical equations [J]. *The Journal of Chemical Physics*, 2018, 148(24): 241723.
- [7] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58(1): 267–288.
- [8] ZOU H, HASTIE T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 2005, 67(2): 301–320.
- [9] QUADE M, ABEL M, SHAFI K, et al. Prediction of dynamical systems by symbolic regression [J]. *Physical Review E*, 2016, 94(1): 012214.
- [10] SUN W K, BRAATZ R D. ALVEN: Algebraic learning via elastic net for static and dynamic nonlinear model identification [J]. *Computers & Chemical Engineering*, 2020, 143: 107103.
- [11] LEJARZA F, BALDEA M. Data-driven discovery of the governing equations of dynamical systems via moving horizon optimization [J]. *Scientific Reports*, 2022, 12(1): 11836.
- [12] VADDIREDDY H, SAN O. Equation discovery using fast function extraction: a deterministic symbolic regression approach [J]. *Fluids*, 2019, 4(2): 111.
- [13] SCHMELZER M, DWIGHT R P, CINNELLA P. Discovery of algebraic Reynolds-stress models using sparse symbolic regression [J]. *Flow, Turbulence and Combustion*, 2020, 104(2/3): 579–603.
- [14] DANIELE S, DARIO B, MATTEO D, et al. Modified formulation of laminar kinetic energy transition models by means of elastic-net of a big experimental database of separated flows [J]. *Flow, Turbulence and Combustion*, 2020, 105(3): 671–697.
- [15] CANDÈS E J, WAKIN M B, BOYD S P. Enhancing sparsity by reweighted l1 minimization [J]. *Journal of Fourier Analysis and Applications*, 2008, 14(5/6): 877–905.
- [16] BAI J C, ZHANG H C, LI J C. A parameterized proximal point algorithm for separable convex optimization [J]. *Optimization Letters*, 2018, 12(7):

- 1589–1608.
- [17] CORTIELLA A, PARK K C, DOOSTAN A. Sparse identification of nonlinear dynamical systems via reweighted l_1 -regularized least squares [J]. *Computer Methods in Applied Mechanics and Engineering*, 2021, 376: 113620.
- [18] YUAN Y, LI X T, LI L, et al. Machine discovery of partial differential equations from spatiotemporal data: a sparse Bayesian learning framework [J]. *Chaos*, 2023, 33(11): 113122.
- [19] HE X, SUN Z K. Sparse identification of dynamical systems by reweighted l_1 -regularized least absolute deviation regression [J]. *Communications in Nonlinear Science and Numerical Simulation*, 2024, 131: 107813.
- [20] LU X F, MA H M, ZHANG L N. Model selection via reweighted partial sparse recovery [J]. *Journal of Computational and Applied Mathematics*, 2024, 438: 115553.
- [21] WANG H S, LI G D, JIANG G H. Robust regression shrinkage and consistent variable selection through the LAD-lasso [J]. *Journal of Business & Economic Statistics*, 2007, 25(3): 347–355.
- [22] LAMBERT-LACROIX S, ZWALD L. Robust regression through the Huber's criterion and adaptive lasso penalty [J]. *Electronic Journal of Statistics*, 2011, 5: 1015–1053.
- [23] TRAN G, WARD R. Exact recovery of chaotic systems from highly corrupted data [J]. *Multiscale Modeling & Simulation*, 2017, 15(3): 1108–1129.
- [24] JIANG F, DU L, YANG F, et al. Regularized least absolute deviation-based sparse identification of dynamical systems [J]. *Chaos*, 2023, 33(1): 013103.
- [25] DAUBECHIES I, DEFRISE M, DE MOL C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint [J]. *Communications on Pure and Applied Mathematics*, 2004, 57(11): 1413–1457.
- [26] BLUMENSATH T, DAVIES M E. Iterative thresholding for sparse approximations [J]. *Journal of Fourier Analysis and Applications*, 2008, 14(5): 629–654.
- [27] BECK A, TEOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems [J]. *SIAM Journal on Imaging Sciences*, 2009, 2(1): 183–202.
- [28] ZHENG P, ASKHAM T, BRUNTON S L, et al. A unified framework for sparse relaxed regularized regression: SR3 [J]. *IEEE Access*, 2019, 7: 1404–1423.
- [29] CHAMPION K, ZHENG P, ARAVKIN A Y, et al. A unified sparse optimization framework to learn parsimonious physics-informed models from data [J]. *IEEE Access*, 2020, 8: 169259–169271.
- [30] YUAN M, BAI J C, JIANG F, et al. A systematic DNN weight pruning framework based on symmetric accelerated stochastic ADMM [J]. *Neurocomputing*, 2024, 575: 127327.
- [31] YUAN M, DU L, JIANG F, et al. Reweighted alternating direction method of multipliers for DNN weight pruning [J]. *Neural Networks*, 2024, 179: 106534.
- [32] BAI J C, ZHANG M, ZHANG H C. An inexact ADMM for separable nonconvex and nonsmooth optimization [J]. *Computational Optimization and Applications*, 2025, 90(2): 445–479.
- [33] BOYD S, PARIKH N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends[®] in Machine Learning*, 2010, 3(1): 1–122.
- [34] BACH F, JENATTON R, MAIRAL J, et al. Optimization with sparsity-inducing penalties [J]. *Foundations and Trends[®] in Machine Learning*, 2011, 4(1): 1–106.
- [35] PARIKH N, BOYD S. Proximal algorithms [J]. *Foundations and Trends in Optimization*, 2014, 1(3): 127–239.
- [36] XU Z B, CHANG X Y, XU F M, et al. $L_{1/2}$ regularization: a thresholding representation theory and a fast solver [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(7): 1013–1027.
- [37] LORENZ E N. Predictability: a problem partly solved [C]. *Shinfield Park, Reading; ECMWF*, 1995: 1–18.
- [38] LORENZ E N, EMANUEL K A. Optimal sites for supplementary weather observations: simulation with a small model [J]. *Journal of the Atmospheric Sciences*, 1998, 55(3): 399–414.
- [39] JIANG F, DU L, XUE Q, et al. Adaptive backward stepwise selection of fast sparse identification of nonlinear dynamics [J]. *Applied Mathematics and Mechanics*, 2025, 46(12): 2361–2384.

- [40] SELESNICK I, FARSHCHIAN M. Sparse signal approximation via nonseparable regularization [J]. IEEE Transactions on Signal Processing, 2017, 65(10): 2561–2575.
- [41] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. Journal of the American Statistical Association, 2001, 96(456): 1348–1360.
- [42] ZHANG C H. Nearly unbiased variable selection under minimax concave penalty [J]. The Annals of Statistics, 2010, 38(2): 894–942.
- [43] HE B S, YANG H, WANG S L. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities [J]. Journal of Optimization Theory and Applications, 2000, 106(2): 337–356.
- [44] 刘浩洋, 户将, 李勇锋, 等. 最优化: 建模、算法与理论[M]. 北京: 高等教育出版社, 2020.
- LIU H Y, HU J, LI Y F, WEN Z W, et al. Optimization: modeling, algorithm and theory [M]. Beijing: Higher Education Press, 2020. (in Chinese)
- [45] CHEN C H, HE B S, YE Y Y, et al. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent [J]. Mathematical Programming, 2016, 155(1/2): 57–79.
- [46] FASEL U, KUTZ J N, BRUNTON B W, et al. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control [J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2022, 478(2260): 20210904.
- [47] WENTZ J, DOOSTAN A. Derivative-based SINDy (DSINDy): Addressing the challenge of discovering governing equations from noisy data [J]. Computer Methods in Applied Mechanics and Engineering, 2023, 413: 116096.
- [48] MESSENGER D A, BORTZ D M. Weak SINDy: Galerkin-based data-driven model selection [J]. Multiscale Modeling Simulation, 2021, 19(3): 1474–1497.
- [49] MESSENGER D A, BORTZ D M. Weak SINDy for partial differential equations [J]. Journal of Computational Physics, 2021, 443: 110525.
- [50] RISH I, GRABARNIK G. Sparse modeling: theory, algorithms, and applications[M]. Boca Raton: CRC Press, 2014.
- [51] HASTIE T J, TIBSHIRANI R J, WAINWRIGHT M. Statistical learning with sparsity [M]. Boca Raton: CRC Press, Taylor & Francis Group, 2015.
- [52] PILLONETTO G, DINUZZO F, CHEN T S, et al. Kernel methods in system identification, machine learning and function estimation: a survey [J]. Automatica, 2014, 50(3): 657–682.
- [53] CHIUSO A, PILLONETTO G. System identification: a machine learning perspective [J]. Annual Review of Control, Robotics, and Autonomous Systems, 2019, 2(1): 281–304.
- [54] LJUNG L, CHEN T S, MU B Q. A shift in paradigm for system identification [J]. International Journal of Control, 2020, 93(2): 173–180.

附录

对于正则化系数为 μ , 正则项为 $R(\mathbf{x})$ 的临近算子的定义为

$$\text{Prox}_{1/\mu, R(\cdot)}(\mathbf{v}) \triangleq \mathbf{x}^* = \arg \min_x \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \mu R(\mathbf{x}) \right\}.$$

(1) 当 $R(\mathbf{x}) = \|\mathbf{x}\|_0$, 该临近算子著称为硬阈值算子, 解的分量形式定义为

$$x_j^* = [\text{Prox}_{1/\mu, R(\cdot)}(\mathbf{v})]_j = \begin{cases} v_j, & v_j < -\sqrt{2\mu}, \\ 0, & v_j \in [-\sqrt{2\mu}, \sqrt{2\mu}], \\ v_j, & v_j > \sqrt{2\mu}. \end{cases}$$

(2) 当 $R(\mathbf{x}) = \|\mathbf{x}\|_1$, 该临近算子著称为软阈值算子, 解的分量形式定义为

$$x_j^* = S_\mu(v_j) = \begin{cases} v_j - \mu, & v_j > \mu, \\ 0, & v_j \in [-\mu, \mu], \\ v_j + \mu, & v_j < -\mu. \end{cases}$$

(3) 对于弹性网正则项,即

$$R(\mathbf{x}) = \frac{1-\alpha}{2} \|\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_1,$$

该临近算子的求解可借助软阈值算子,将 $R(x)$ 的表达式代入,并通过配方法,得到

$$\begin{aligned} \mathbf{x}^* &= \arg \min_x \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \mu \left(\frac{1-\alpha}{2} \|\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_1 \right) \right\} \\ &\Rightarrow \arg \min_x \left\{ \left(\frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \frac{\mu(1-\alpha)}{2} \|\mathbf{x}\|_2^2 \right) + \mu\alpha \|\mathbf{x}\|_1 \right\} \\ &\Rightarrow \arg \min_x \left\{ \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{1+\mu(1-\alpha)} \right\|_2^2 + \frac{\mu\alpha}{1+\mu(1-\alpha)} \|\mathbf{x}\|_1 \right\}, \end{aligned}$$

进而可借助软阈值算子表示为

$$x_j^* = S_{\frac{\mu\alpha}{1+\mu(1-\alpha)}} \left(\frac{v_j}{1+\mu(1-\alpha)} \right).$$

(4) 对于加权 ℓ_1 正则项,即 $R(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_1$,该临近算子的求解仍可借助软阈值算子,将 $R(\mathbf{x})$ 表达式代入,得到

$$\mathbf{x}^* = \arg \min_x \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 + \mu \|\mathbf{W}\mathbf{x}\|_1 \right\},$$

该式的分量形式可表示为

$$\begin{aligned} x_j^* &= \arg \min_{x_j} \left\{ \frac{1}{2} (x_j - v_j)^2 + \mu W_{j,j} |x_j| \right\} \\ &= S_{\mu W_{j,j}}(v_j). \end{aligned}$$

(5) 对于 $\ell_{1/2}$ 正则项,即 $R(\mathbf{x}) = \|\mathbf{x}\|_{1/2}$,该临近算子著称为半阈值算子,其分量形式表示为

$$x_j^* = \begin{cases} \frac{2}{3} v_j \left\{ 1 + \cos \frac{2}{3} [\pi - \eta(v_j)] \right\}, & |v_j| > \frac{3\sqrt{2}}{4} (2\mu)^{2/3}, \\ 0, & \text{others,} \end{cases}$$

其中 $\eta(v_j) = \arccos \left[\frac{2\mu}{8} \left(\frac{|v_j|}{3} \right)^{-3/2} \right]$.