Advances in Applied Mathematics 应用数学进展, 2026, 15(11), 129-138
Published Online November 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.1511437

Hans汉斯

融合Lasso分位数回归的半光滑牛顿法

文广建，姚胜伟*，罗楚微

南宁师范大学数学与统计学院，广西 南宁

收稿日期：2026年10月9日；录用日期：2026年11月2日；发布日期：2026年11月17日

摘　要

目的：针对带融合Lasso惩罚的高维分位数回归中，原始空间交替方向乘子法(ADMM)仅具一阶收敛、迭代次数多、计算耗时长的问题，构造一种可利用解的稀疏与分段常数结构的二阶算法。方法：引入残差变量与系数差变量将问题分裂为可分离形式，借助Fenchel共轭推导对偶问题，把高维系数空间压缩到低维对偶变量空间；建立增广拉格朗日框架克服检验损失引起的对偶退化，并将内子问题归结为箱约束上的强半光滑残差方程，用投影半光滑牛顿法结合活跃集策略求解，再以Sherman-Morrison-Woodbury恒等式把牛顿系统由自由变量维数降到活跃变量维数。结果：在对偶广义雅可比非奇异的条件下，证明内层牛顿法局部超线性收敛、精确求解时二次收敛，外层增广拉格朗日迭代全局收敛且渐近线性。正态误差下的模拟实验表明，相较QFLASSO，所提QFL-SSNAL以更少的迭代次数和更短的运行时间达到相同的估计精度。结论：QFL-SSNAL为带空间或顺序结构的高维稀疏分位数回归提供了一种高效可靠的求解方法。

关键词

分位数回归，融合Lasso，半光滑牛顿法，增广拉格朗日方法，稀疏优化

Semismooth Newton Method for Fused-Lasso Quantile Regression

Guangjian Wen, Shengwei Yao*, Chuwei Luo

School of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

Received: October 9, 2026; accepted: November 2, 2026; published: November 17, 2026

Abstract

Objective: In high-dimensional quantile regression with the fused-Lasso penalty, primal-space alternating direction methods of multipliers (ADMM) enjoy only first-order convergence and require

*通讯作者。

文章引用: 文广建, 姚胜伟, 罗楚微. 融合 Lasso 分位数回归的半光滑牛顿法[J]. 应用数学进展, 2026, 15(11): 129-138. DOI: 10.12677/aam.2026.1511437

many iterations and long running time. This paper aims to develop a second-order algorithm that exploits both the sparsity and the piecewise-constant structure of the solution. Methods: Auxiliary variables for the residual and for the successive coefficient differences are introduced to split the problem into a separable form. The dual problem is derived through Fenchel conjugation, compressing the high-dimensional coefficient space into a low-dimensional dual variable space. An augmented Lagrangian framework is established to overcome the degeneracy of the dual problem caused by the check loss. The inner subproblems are reformulated as strongly semismooth residual equations over a box constraint and solved by a projected semismooth Newton method with an active-set strategy. The Newton system is further reduced from the dimension of free variables to that of active variables via the Sherman-Morrison-Woodbury identity. Results: Under the nonsingularity of a generalized dual Jacobian, the inner Newton method is shown to converge locally superlinearly, and quadratically when the Newton system is solved exactly; the outer augmented Lagrangian iterations converge globally with an asymptotic linear rate. Simulation experiments under normal errors show that, compared with QFLASSO, the proposed QFL-SSNAL attains the same estimation accuracy with markedly fewer iterations and shorter running time. Conclusion: QFL-SSNAL provides an efficient and reliable solver for high-dimensional sparse quantile regression with spatial or ordered structure.

Keywords

Quantile Regression, Fused-Lasso, Semismooth Newton Method, Augmented Lagrangian Method, Sparse Optimization

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

普通的最小二乘回归只刻画了条件均值，而且对异常值与尾部较重的误差也十分敏感。Koenker 和 Bassett [1]提出的分位数回归能在很大程度上分析在给定协变量下的响应变量的条件分布，刻画协变量在不同分位数水平上的效应之间的不同。给定样本 $\{(x_i, y_i)\}_{i=1}^n$，分位数回归通过极小化检验损失得到经验分位数。检验损失定义为 $\rho_\tau(u) := u\left(\tau - 1_{\{u<0\}}\right)$，其中当 $u \geq 0$ 时 $\rho_\tau(u) = \tau u$，以及当 $u < 0$ 时，$\rho_\tau(u) = (\tau - 1)u$。

它对重尾误差与异常值具有鲁棒性，且无需对误差分布做参数假设，在计量经济与生物医学等领域得到广泛应用。然而，检验损失在零点处存在折点，经典梯度法不能直接使用，所以当维数升高时，它的计算与存储代价会上升得比较多，使得在高维情形下给算法的设计带来了困难。

Tibshirani [2]提出的 Lasso 以 ℓ_1 惩罚项 $\lambda\|\beta\|_1$ 同时收缩系数并诱导稀疏，Belloni 和 Chernozhukov [3]将 ℓ_1 罚引入分位数回归，证明其在高维稀疏模型中具有近 oracle 的收敛速率。Zhang [4]提出的极大极小凹惩罚(MCP)则通过极小化最大凹性在无偏性与稀疏性之间取得平衡。上述惩罚仅刻画单个系数的大小，并不利用系数之间的顺序或空间结构。在许多实际问题中，相邻位置的系数天然地接近或相等，例如按基因组位置排列的拷贝数变异、按时间或空间排列的信号。针对这类问题，Tibshirani 等[5]提出融合 Lasso 惩罚，带融合 Lasso 惩罚的分位数回归模型为

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right) + \lambda\left(\alpha\sum_{j=1}^p |\beta_j| + (1-\alpha)\sum_{j=2}^p |\beta_j - \beta_{j-1}| \right) \right\}, \tag{1}$$

其中 $\alpha \in (0,1)$，$\lambda > 0$ 为调优参数。记设计矩阵 $X = \left[x_1^\top, \cdots, x_n^\top\right]^\top \in \mathbb{R}^{n\times p}$，响应向量 $y \in \mathbb{R}^n$，并定义一阶差

分矩阵 $D\in\mathbb{R}^{(p-1)\times p}$，其相邻两行形如 $(0,\cdots,-1,1,\cdots,0)$，即

$$D=\begin{bmatrix}-1 & 1 & 0 & \cdots & 0\\ 0 & -1 & 1 & \cdots & 0\\ \vdots & & \ddots & \ddots & \vdots\\ 0 & 0 & \cdots & -1 & 1\end{bmatrix}, \tag{2}$$

则相邻系数差的绝对值之和可写为 $\|D\beta\|_1$。令 $\lambda_1=\lambda\alpha$，$\lambda_2=\lambda(1-\alpha)$，模型(1)可等价改写为

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} q_\tau(\mathbf{y}-X\boldsymbol{\beta})+p(\boldsymbol{\beta}),\quad p(\boldsymbol{\beta}):=\lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|D\boldsymbol{\beta}\|_1 \tag{3}$$

其中多元检验损失 $q_\tau(y-X\beta)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i-x_i^\top\beta)$。

由于检验损失非光滑、两个 ℓ_1 罚均为非光滑凸函数，问题(3)的目标函数不可微。许多学者采用交替方向乘子法(ADMM)求解，使各子问题具有闭式解。Gu 等[6]针对高维稀疏罚分位数回归提出快速 ADMM 并建立收敛性。Chen 等[7]系统讨论了线性约束凸问题 ADMM 的收敛条件。Wu 等[8]针对稀疏融合 Lasso 建模提出多块线性化 ADMM；Mirzaeifard 等[9]用光滑化 ADMM 处理非凸罚分位数回归。这类一阶方法结构简单而且单步代价低，但收敛速度本质上受限于对偶变量的一阶更新，为达到高精度往往需要大量迭代。融合 Lasso 问题同时包含两个 ℓ_1 罚，其对偶可行域是箱集与 ℓ_∞ 球的交，一阶方法在这种多面体可行域的面与面之间往往呈锯齿形推进，越接近解收敛越慢。与之不同，Li、Sun 和 Toh [10]在 Lasso 与融合 Lasso 子问题上证明，充分利用解的稀疏结构的半光滑牛顿法能够获得超线性甚至二次收敛，这一思路为突破一阶方法的速度瓶颈提供了途径。

本文提出一种基于半光滑牛顿增广拉格朗日的高效算法 QFL-SSNAL 求解问题(3)。主要工作如下：其一，通过引入残差变量与系数差变量将问题分裂为可分离形式，借助 Fenchel 共轭推导对偶问题，把高维系数空间压缩到低维对偶空间；其二，建立增广拉格朗日框架克服检验损失引起的对偶退化，将内子问题归结为箱约束上的强半光滑残差方程，用投影半光滑牛顿法结合样本活跃集与特征活跃集求解；其三，利用 Sherman-Morrison-Woodbury 恒等式将牛顿系统由自由变量维数降到活跃变量维数，并在对偶广义雅可比非奇异的条件下证明内层超线性或二次收敛、外层全局且渐近线性收敛；其四，通过模拟数据验证算法在迭代次数、运行时间与估计精度上的优势。

2. 预备知识

本节回顾凸分析中的共轭函数、邻近算子与 Moreau 分解，它们构成所提算法的理论基础。

2.1. 共轭函数

设 $\mathbb{R}^n$ 配备标准内积 $\langle\cdot,\cdot\rangle$ 及其诱导的 ℓ_2 范数。对于函数 $f:\mathbb{R}^n\to(-\infty,+\infty]$，其 Fenchel 共轭函数 $f^*:\mathbb{R}^n\to(-\infty,+\infty]$ 定义为

$$f^*(z)=\sup_{x\in\operatorname{dom} f}\{\langle z,x\rangle-f(x)\}=-\inf_{x\in\operatorname{dom} f}\{f(x)-\langle z,x\rangle\} \tag{4}$$

其中有效定义域 $\operatorname{dom} f=\{x\mid f(x)<+\infty\}$。无论 f 是否凸，f^* 总是凸的。Rockafellar [11]系统建立了凸分析与共轭对偶的理论框架。

2.2. 邻近算子与软阈值

命题 2.1 (邻近算子) 设 f 为正常闭凸函数。对任意 $x\in\mathbb{R}^n$ 与 $t>0$，参数为 t 的邻近算子定义为

$$\operatorname{prox}_{tf}(x)=\arg\min_{u\in\mathbb{R}^n}\left\{f(u)+\frac{1}{2t}\|u-x\|_2^2\right\} \tag{5}$$

一元检验损失的共轭是区间上的指示函数，即

$$\rho_\tau^*(s)=\delta_{[\tau-1,\tau]}(s)=\begin{cases}0, & s\in[\tau-1,\tau],\\ +\infty, & \text{否则}.\end{cases} \tag{6}$$

推广到经验损失，其共轭为 $q_\tau^*(u)=\delta_{[(\tau-1)/n,\tau/n]^n}(u)$，它把对偶变量逐分量限制在箱内。相应的邻近算子具有闭式解[8]：对任意 $x\in\mathbb{R}^n$，

$$\left[\operatorname{prox}_{tq_\tau}(x)\right]_i=\max\left\{x_i-\frac{t\tau}{n},\min\left\{0,x_i+\frac{t(1-\tau)}{n}\right\}\right\},\quad i=1,\cdots,n. \tag{7}$$

这是一个非对称分段阈值算子：超过上阈值的分量向下收缩，低于下阈值的分量向上收缩，中间区间的分量被压缩为零。

对于 $f(x)=\lambda\|x\|_1$，其共轭是对偶范数球的指示函数 $f^*(z)=\delta_{\{\|z\|_\infty\le\lambda\}}(z)$，其邻近映射为对称软阈值算子。根据 Combettes 和 Wajs [12]，

$$\operatorname{prox}_{\lambda\|\cdot\|_1}(x)=\operatorname{soft}(x,\lambda)=\operatorname{sign}(x)\odot\max\{|x|-\lambda,0\}. \tag{8}$$

软阈值算子是分片仿射的，其广义雅可比 B-次微分为对角 $\{0,1\}$ 矩阵 $\Theta=\operatorname{Diag}(\theta)$，当 $|x_i|>\lambda$ 时 $\theta_i=1$，否则 $\theta_i=0$。

2.3. 最优性条件与 Moreau 分解

命题 2.2 (最优性条件) 点 u 是邻近算子问题的唯一解当且仅当

$$u=\operatorname{prox}_{tf}(x)\Leftrightarrow x-u\in t\partial f(u), \tag{9}$$

其中 $\partial f(u)$ 为 f 在 u 处的凸次微分。

命题 2.3 (Moreau 分解) 对任意 $x\in\mathbb{R}^n$ 与 $t>0$，有

$$x=\operatorname{prox}_{tf}(x)+t\operatorname{prox}_{f^*/t}(x/t). \tag{10}$$

该恒等式建立了函数与其共轭的邻近算子之间的基本关系，是后文由软阈值构造 ℓ_∞ 球投影的关键。

3. 对偶问题与增广拉格朗日方法

3.1. 对偶问题的构造

为高效求解问题(3)，引入残差变量 $r\in\mathbb{R}^n$ 与系数差变量 $z\in\mathbb{R}^{p-1}$，将其分裂为可分离形式：

$$\min_{\beta,r,z} q_\tau(r)+\lambda_1\|\beta\|_1+\lambda_2\|z\|_1\quad\text{s.t.}\quad X\beta+r=y,\quad D\beta-z=0. \tag{11}$$

构造拉格朗日函数，其中 $u\in\mathbb{R}^n$，$w\in\mathbb{R}^{p-1}$ 为拉格朗日乘子：

$$L(\beta,r,z;u,w)=q_\tau(r)+\lambda_1\|\beta\|_1+\lambda_2\|z\|_1-\langle u,X\beta+r-y\rangle-\langle w,D\beta-z\rangle. \tag{12}$$

分别对 r、z、β 求下确界，利用共轭定义可得

$$\inf_r\{q_\tau(r)-\langle u,r\rangle\}=-q_\tau^*(u), \tag{13}$$

$$\inf_z\{\lambda_2\|z\|_1+\langle w,z\rangle\}=-\delta_{\{\|w\|_\infty\le\lambda_2\}}(w), \tag{14}$$

$$\inf_\beta\{\lambda_1\|\beta\|_1-\langle X^\top u+D^\top w,\beta\rangle\}=-\delta_{\{\|X^\top u+D^\top w\|_\infty\le\lambda_1\}}(X^\top u+D^\top w). \tag{15}$$

将对偶变量堆叠为 $\xi=[u;w]\in\mathbb{R}^m$，其中 $m=n+p-1$，并定义堆叠矩阵与线性项向量

$$B=\begin{bmatrix}X\\D\end{bmatrix}\in\mathbb{R}^{m\times p},\quad c=\begin{bmatrix}y\\0\end{bmatrix}\in\mathbb{R}^m, \tag{16}$$

则 $B^{\top}\xi=X^{\top}u+D^{\top}w$，且对偶约束统一为逐分量的箱约束。记箱集

$$\Xi=\left\{[u;w]\,|\,u\in\left[\frac{\tau-1}{n},\frac{\tau}{n}\right]^n,w\in[-\lambda_2,\lambda_2]^{p-1}\right\}, \tag{17}$$

加入线性项 $\langle c,\xi\rangle$ 后，对偶问题写为最小化形式：

$$\min_{\xi\in\Xi}\delta_{\left\{\left\|B^{\top}\xi\right\|_{\infty}\le\lambda_1\right\}}\left(B^{\top}\xi\right)-\langle c,\xi\rangle. \tag{18}$$

3.2. 增广拉格朗日与内子问题

为处理对偶问题(18)，引入辅助变量 $v=B^{\top}\xi$ 使约束显式化：

$$\min_{\xi\in\Xi,v}\delta_{\{\|v\|_{\infty}\le\lambda_1\}}(v)-\langle c,\xi\rangle\quad\text{s.t.}\quad B^{\top}\xi-v=0. \tag{19}$$

给定惩罚参数 $\sigma>0$ 与乘子 $\beta\in\mathbb{R}^p$，构造增广拉格朗日函数

$$L_{\sigma}(\xi,v;\beta)=\delta_{\Xi}(\xi)+\delta_{\{\|v\|_{\infty}\le\lambda_1\}}(v)-\langle c,\xi\rangle+\langle\beta,v-B^{\top}\xi\rangle+\frac{\sigma}{2}\left\|v-B^{\top}\xi\right\|^2, \tag{20}$$

其中乘子 β 在最优处与原问题(3)的解一致。当 ξ、β、σ 固定时，v 子问题为

$$\min_{v}\delta_{\{\|v\|_{\infty}\le\lambda_1\}}(v)+\frac{\sigma}{2}\left\|v-\left(B^{\top}\xi-\beta/\sigma\right)\right\|^2. \tag{21}$$

令 $a=B^{\top}\xi-\beta/\sigma$。由 ℓ_1 范数与 ℓ_{∞} 球指示函数互为 Fenchel 共轭，对 Moreau 分解(10)取 $t=1$ 得

$$\operatorname{prox}_{\delta_{\{\|\cdot\|_{\infty}\le\lambda_1\}}}(a)=a-\operatorname{soft}(a,\lambda_1). \tag{22}$$

故 v 子问题的最优解为

$$v^{*}=a-\operatorname{soft}(a,\lambda_1)=\left(B^{\top}\xi-\beta/\sigma\right)-\operatorname{soft}\left(B^{\top}\xi-\beta/\sigma,\lambda_1\right). \tag{23}$$

将 v^{*} 代回增广拉格朗日函数并展开，线性项与二次项中的交叉项恰好抵消，得到关于 ξ 的内子问题

$$\Psi_{\sigma}(\xi)=\delta_{\Xi}(\xi)-\langle c,\xi\rangle+\frac{\sigma}{2}\left\|\operatorname{soft}\left(B^{\top}\xi-\beta/\sigma,\lambda_1\right)\right\|^2-\frac{1}{2\sigma}\|\beta\|^2. \tag{24}$$

在箱集内部 δ_{Ξ} 的次微分为零，由链式法则得梯度

$$\nabla\Psi_{\sigma}(\xi)=-c+\sigma B\operatorname{soft}\left(B^{\top}\xi-\beta/\sigma,\lambda_1\right). \tag{25}$$

3.3. 外层收敛性

增广拉格朗日方法可视为对极大单调包含 $0\in T(\beta)$ 的近端点法的不精确实现，其中 T 为问题(3)最优性条件所定义的极大单调算子。Rockafellar [13]证明了近端点法的全局收敛性与超线性潜力。结合 Li、Sun 和 Toh [14]对融合 Lasso 水平法子问题的分析，给出外层收敛结论。

定理 3.1 (外层收敛性) 设问题(3)的最优解集非空，内层子问题以满足相对误差准则的精度求解，惩罚序列 $\{\sigma_k\}$ 单调不减且有上界。则由 QFL-SSNAL 产生的序列 $\{\beta_k\}$ 全局收敛到问题(3)的一个最优解，乘子序列 $\{\xi_k\}$ 收敛到相应对偶最优解，并且在解附近，当内层精度与 σ_k 适当配合时，迭代渐近线性收敛，

即存在常数 $0\le\rho<1$ 使相邻误差满足 $\|e_{k+1}\|\le\rho\|e_k\|$。

证明 由 Ψ_σ 关于 ξ 为凸函数，一阶最优性条件给出的对偶残量是增广拉格朗日的次梯度，内层相对误差准则保证该残量受控，标准增广拉格朗日收敛理论[13]即给出全局收敛。局部地，箱集 Ξ 为多面体凸集，其法锥对应的投影映射为分片仿射。由 Robinson [15]关于多面体上多值映射的 Lipschitz 性质，增广拉格朗日迭代在该面上退化为固定线性映射的逐次逼近，其谱半径 $\rho<1$，从而渐近线性。证毕。

4. 内子问题的半光滑牛顿法

4.1. 残差方程与广义雅可比

内子问题(24)的最优性条件为 $-\nabla\Psi_\sigma(\xi)\in N_\Xi(\xi)$，其中 $N_\Xi(\xi)$ 为箱的集法锥。该条件等价于全局定义的半光滑残差方程

$$F(\xi):=\xi-\Pi_\Xi\left(\xi-\nabla\Psi_\sigma(\xi)\right)=0, \tag{26}$$

其中 Π_Ξ 为到箱集的欧氏投影。F 的非光滑性来自软阈值算子与投影算子，二者均为分片仿射，其 B-次微分为对角 $\{0,1\}$ 矩阵，分别产生两个活跃集。软阈值算子的 B-次微分为

$$\Theta=\mathrm{Diag}(\theta),\quad \theta_i=\begin{cases}1, & |a_i|>\lambda_1,\\ 0, & |a_i|\le\lambda_1,\end{cases}\quad a=B^\top\xi-\beta/\sigma. \tag{27}$$

称 $\theta_i=1$ 的指标集为 ℓ_1 活跃集 $\mathcal{A}=\{i:|a_i|>\lambda_1\}$。箱投影算子的 B-次微分为

$$\Gamma=\mathrm{Diag}(\gamma),\quad \gamma_i=\begin{cases}1, & \xi_i-\left[\nabla\Psi_\sigma(\xi)\right]_i\in\mathrm{int}(\Xi),\\ 0, & \xi_i-\left[\nabla\Psi_\sigma(\xi)\right]_i\text{落在}\Xi\text{边界上}\end{cases} \tag{28}$$

称 $\gamma_i=1$ 的指标集为自由集 $\mathcal{F}=\{i:\gamma_i=1\}$。由链式法则，$F$ 的 Clarke 广义雅可比满足

$$\partial F(\xi)\ni I-\Gamma\left(I-\sigma B\Theta B^\top\right)=(I-\Gamma)+\sigma\Gamma B\Theta B^\top\Gamma. \tag{29}$$

为严格保证正定性，加入小正则项 εI，取广义雅可比的一个特定元素

$$V=(I-\Gamma)+\Gamma\left(\varepsilon I+\sigma B\Theta B^\top\right)\Gamma. \tag{30}$$

令 $B_{\mathcal{A}}=B[:,\mathcal{A}]$，则 $B\Theta B^\top=B_{\mathcal{A}}B_{\mathcal{A}}^\top$，其秩等于活跃集大小 $|\mathcal{A}|$；在稀疏情形下 $|\mathcal{A}|\ll m$。

4.2. 约化牛顿系统与 SMW 加速

对 $\gamma_i=0$ 的约束的活跃分量，牛顿方向由残量结构直接确定；只需在自由集 $\mathcal{F}$ 上求解约化牛顿系统

$$\left(\varepsilon I_{|\mathcal{F}|}+\sigma B_{\mathcal{F},\mathcal{A}}B_{\mathcal{F},\mathcal{A}}^\top\right)d_{\mathcal{F}}=-F(\xi)_{\mathcal{F}}, \tag{31}$$

其中 $B_{\mathcal{F},\mathcal{A}}$ 为以 $\mathcal{F}$ 索引行、$\mathcal{A}$ 索引列的子矩阵。利用低秩结构，采用 Sherman-Morrison-Woodbury (SMW) 恒等式

$$\left(\varepsilon I+\sigma UV^\top\right)^{-1}=\frac{1}{\varepsilon}\left(I-U\left(\varepsilon/\sigma I+V^\top U\right)^{-1}V^\top\right). \tag{32}$$

令 $U=V=B_{\mathcal{F},\mathcal{A}}$，原 $|\mathcal{F}|$ 维系统降为求解 $|\mathcal{A}|$ 维系统

$$\left(\frac{\varepsilon}{\sigma}I_{|\mathcal{A}|}+B_{\mathcal{F},\mathcal{A}}^\top B_{\mathcal{F},\mathcal{A}}\right)z=B_{\mathcal{F},\mathcal{A}}^\top F(\xi)_{\mathcal{F}}, \tag{33}$$

回代得到自由集上的牛顿方向

$$d_{\mathcal{F}} = \frac{1}{\varepsilon}\left(-F(\xi)_{\mathcal{F}} + B_{\mathcal{F},\mathcal{A}} z\right). \tag{34}$$

由此出现双重二阶稀疏：样本侧由箱活跃集把 n 维压缩到自由集维数，特征侧由 SMW 公式压缩到 ℓ_1 活跃集维数，避免了装配和分解大规模矩阵。

4.3. 线搜索与算法

为保证全局收敛，采用基于残差范数的 Armijo 线搜索。令 $\xi^{+} = \xi + \alpha d$，$\alpha = \delta^{m}$，$\delta \in (0,1)$，取使价值函数 $\frac{1}{2}\|F(\xi)\|^2$ 充分下降的最大步长：

$$\|F(\xi + \alpha d)\|^2 \le \|F(\xi)\|^2 + \mu\alpha \left\langle \nabla\left(\frac{1}{2}\|F(\xi)\|^2\right), d \right\rangle, \quad \mu \in (0, 1/2). \tag{35}$$

完整算法 QFL-SSNAL 总结如下。

算法 1 QFL-SSNAL。输入 $\lambda_1, \lambda_2 > 0$，$\tau \in (0,1)$；初始化 $\xi_0, \beta_0, \sigma_0 > 0$，$\sigma_{\text{inc}} > 1$，$\sigma_{\max}$。当终止准则不满足时循环：步骤 1 (内更新)给定 β_k, σ_k，用上述半光滑牛顿法近似求解 $\min_{\xi} \Psi_{\sigma_k}(\xi)$，得 ξ_{k+1}；步骤 2 按式(23)计算 v_{k+1}；步骤 3 更新 $\sigma_{k+1} = \min\{\sigma_k \sigma_{\text{inc}}, \sigma_{\max}\}$；步骤 4 更新 $\beta_{k+1} = \beta_k - \sigma_k\left(B^{\top}\xi_{k+1} - v_{k+1}\right)$。结束循环，输出 β_{k+1}。

4.4. 内层收敛性

命题 4.1 (广义雅可比非奇异) 残差映射 F 在 R^m 上强半光滑，且其任一按式(30)构造的广义雅可比元素 V 均非奇异。

证明 软阈值与箱投影均为分片仿射，二者与光滑仿射映射的复合保持强半光滑性，故 F 强半光滑。由式(30)，V 在约束活跃子块上为单位阵，在自由子块上为 $\varepsilon I + \sigma B_{\mathcal{F},\mathcal{A}} B_{\mathcal{F},\mathcal{A}}^{\top}$。后者因 $\varepsilon > 0$ 而严格正定，故整体 V 非奇异。证毕。

定理 4.1 (内层收敛性) 设 ξ^{*} 为内子问题(24)的唯一解。则半光滑牛顿迭代 $\{\xi^{j}\}$ 局部超线性收敛到 ξ^{*}，即存在 j 使对 j 以后有

$$\|\xi^{j+1} - \xi^{*}\| \le O\left(\|\xi^{j} - \xi^{*}\|^{1+\varrho}\right), \quad \varrho \in (0,1]. \tag{36}$$

当牛顿系统(31)被精确求解 $\varrho = 1$ 时收敛为二次，配合 Armijo 线搜索，迭代从任意初始点全局收敛。

证明 命题 4.1 给出强半光滑性与广义雅可比一致非奇异，由 Qi 和 Sun [16]的半光滑牛顿理论直接得到局部超线性或二次收敛。价值函数 $\frac{1}{2}\|F\|^2$ 沿牛顿方向在解附近充分下降，Armijo 准则保证全局收敛。该结论与 Li、Sun 和 Toh [10]关于 Lasso 与融合 Lasso 子问题的半光滑牛顿收敛定理相对应，区别在于本文以小正则项 εI 保证自由子块正定，这正是检验损失经对偶线性化后结构的特殊之处。证毕。更一般的非光滑方程与互补问题理论可参见 Facchinei 和 Pang [17]。

5. 数值实验

5.1. 模拟设置

所有实验固定 $\lambda_1 = \lambda_2 = 0.05$，$\sigma_0 = 1.0$，$\sigma_{\text{inc}} = 1.3$，$\sigma_{\max} = 50.0$，$\varepsilon = 10^{-3}$，$\delta = 10^{-10}$，结果在 5 次独立试验上取平均，并与记为 QFLASSO 的融合 Lasso 一阶法[8]这个先进算法比较。观测模型按 Lima 等[18]的方案生成。对每个 i，令 $(\tilde{x}_{i,1}, \cdots, \tilde{x}_{i,P})^{\top} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$，其中 $\Sigma_{p,q} = 0.5^{|p-q|}$；定义 $x_{i,1} = \Phi(\tilde{x}_{i,1})$，对 $p \ge 2$ 取

$x_{i,p}=\tilde{x}_{i,p}$，Φ 为标准正态累积分布函数。观测模型为

$$y_i=\sum_{p=1}^{P}\zeta_p x_{i,p}+x_{i,6}+x_{i,12}+x_{i,15}+x_{i,20}+0.7\varepsilon_i x_{i,1}, \tag{37}$$

其中 $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,1)$，$\xi_p \overset{\text{i.i.d.}}{\sim} N\left(0,10^{-6}\right)$，样本量与维数取 $(n,P)=(100,500)$。第 τ 个条件分位数线性函数为

$$\sum_{p=1}^{P}\xi_p x_{i,p}+x_{i,6}+x_{i,12}+x_{i,15}+x_{i,20}+0.7\Phi^{-1}(\tau)x_{i,1}. \tag{38}$$

采用运行时间(Time，秒)、迭代次数(Iter)、平均绝对误差(MAE)、相对 ℓ_2 误差(RLNE)与均方误差(MSE)评价算法，其中

$$\mathrm{MAE}=\frac{1}{P}\sum_{j=1}^{P}\left|\hat{\beta}_j-\beta_j^*\right|,\ \mathrm{RLNE}=\frac{\left\|\hat{\beta}-\beta^*\right\|_2}{\left\|\beta^*\right\|_2},\ \mathrm{MSE}=\frac{1}{P}\sum_{j=1}^{P}\left(\hat{\beta}_j-\beta_j^*\right)^2 \tag{39}$$

以 KKT 残量 $\eta_{\mathrm{KKT}}=\max\{e_a,e_b,e_c\}$ 作为停止准则，当 $\eta_{\mathrm{KKT}}<10^{-3}$ 或迭代达到 20,000 次时终止，其中 $e_a=\dfrac{\left\|z-\mathrm{prox}_{q_\tau/n}(z+\hat{u})\right\|_2}{1+\|z\|_2+\|\hat{u}\|_2}$，$e_b=\dfrac{\left\|\beta-\mathrm{prox}_{\lambda_1\|\cdot\|_1+\lambda_2\|D\cdot\|_1}\left(\beta+X^{\mathrm{T}}\hat{u}\right)\right\|_2}{1+\|\beta\|_2+\left\|X^{\mathrm{T}}\hat{u}\right\|_2}$，$e_c=\dfrac{\|X\beta+z-y\|_2}{1+\|y\|_2}$。

5.2. 模拟结果

三种算法均从相同初值出发并在相同准则下终止，分位数水平取 $\tau=0.1,0.5,0.9$，结果列于表 1。

Table 1. Comparison results of QFL-SSNAL and QFLASSO
表 1. QFL-SSNAL 与 QFLASSO 的结果比较

方法	Iter	Time	RLNE	MAE	MSE
QFL-SSNAL (τ = 0.1)	221	0.365	0.7008	0.01714	0.0079
QFLASSO (τ = 0.1)	2200	9.111	0.7013	0.01719	0.0080
QFL-SSNAL (τ = 0.5)	268	0.378	0.6109	0.01324	0.00527
QFLASSO (τ = 0.5)	1824	7.518	0.6142	0.01332	0.00532
QFL-SSNAL (τ = 0.9)	235	0.407	0.6825	0.01732	0.00766
QFLASSO (τ = 0.9)	1885	7.680	0.6830	0.01736	0.00767

由表 1 可见，QFL-SSNAL 在低，中和高三个分位数水平上的迭代次数仅为 221~268 次，而对比算法 QFLASSO 则需要 1824~2200 次；相应地，QFL-SSNAL 的运行时间稳定在 0.365~0.407 秒，仅为 QFLASSO 的 4%~5%。以分位数水平为低的场景为例，QFL-SSNAL 仅用 221 次迭代、0.365 秒即达到预设停机准则，迭代次数约为 QFLASSO 的 10.0%，计算耗时仅为 QFLASSO 的 4.0%。

在估计精度方面，两种算法在各分位数水平下的 RLNE、MAE 与 MSE 数值几乎一致，差异仅出现在小数点后第三至第四位，说明 QFL-SSNAL 的计算效率提升并非以放松停机准则或牺牲解的精度为代价，其加速效果源于半光滑牛顿步的局部超线性收敛特性，以及双重活跃集策略带来的求解维度压缩。这一结果与第 4 节的收敛性理论分析完全吻合：当迭代进入活跃面稳定阶段后，半光滑牛顿法仅需极少步数即可将残差降至收敛容差以下；而作为一阶方法的 QFLASSO 只能沿对偶可行域逐次线性逼近最优解，因此需要更多迭代次数与更长计算时间。

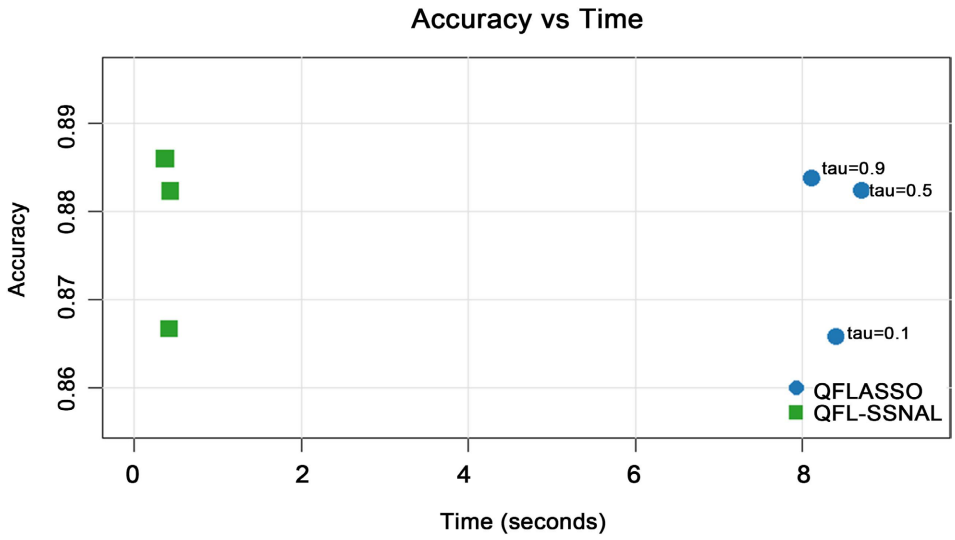


Figure 1. Final variable-selection accuracy versus stopping time at τ = 0.1, 0.5, 0.9
图 1. 在 τ = 0.1、0.5、0.9 时，变量选择准确率与停止时间的关系

第二个实验比较各算法识别活跃与非活跃系数的准确率，其定义为正确识别的系数个数与总系数个数之比。图 1 表明，QFL-SSNAL 在三个分位数水平下均能以更短的停止时间达到相同的最终识别准确率，其准确率曲线随时间更快地趋于平台，反映出二阶方法在确定稀疏支撑集上的优势。

6. 结论

本文针对带融合 Lasso 惩罚的高维分位数回归问题，提出了一种基于半光滑牛顿增广拉格朗日的高效求解算法 QFL-SSNAL。通过引入残差变量与系数差变量实现问题分裂，借助 Fenchel 共轭对偶将高维原系数空间的优化问题压缩至低维对偶变量空间；建立增广拉格朗日框架克服检验损失导致的对偶退化问题，将内子问题转化为箱约束下的强半光滑残差方程，采用投影半光滑牛顿法结合样本侧与特征侧双重活跃集策略求解，并通过 Sherman-Morrison-Woodbury 恒等式将牛顿系统的求解维度进一步降至活跃变量维数。理论分析表明，内层半光滑牛顿迭代具有局部超线性收敛性，精确求解牛顿系统时可达到二次收敛；外层增广拉格朗日迭代全局收敛，且在解附近具有渐近线性收敛速率。

数值模拟实验结果与理论分析完全一致：在本文设定的高维稀疏场景下，与一阶 ADMM 基准算法 QFLASSO 相比，QFL-SSNAL 在保持 RLNE、MAE、MSE 等估计精度指标基本相当的前提下，迭代次数减少约 90%，运行时间仅为基准算法的 4%~5%，计算效率提升显著。该算法在特定模拟条件下显示出巨大的应用潜力，尤其适用于系数具有顺序或一维空间结构的高维稀疏分位数回归与变量选择问题。

需要指出的是，本文的算法构建与收敛性分析均建立在凸的融合 Lasso 惩罚基础之上，仍存在若干可拓展的研究方向。其一，将算法框架推广至非凸融合惩罚、组融合结构、二维及更高维空间结构等更复杂的正则化场景，其收敛性质与实际表现有待进一步验证；其二，针对重尾误差、异方差结构等更复杂的数据分布，开展显式建模与算法适配性研究；其三，探索超大规模数据下的分布式、在线式版本算法，拓展算法的应用边界。此外，如何在不损失二阶收敛速度的前提下进一步降低单步迭代的存储开销与矩阵装配代价，也是后续值得深入研究的方向。

基金项目

广西自然科学基金(2025GXNSFAA069505)。

参考文献

[1] Koenker, R. and Bassett, G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50. https://doi.org/10.2307/1913643

[2] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[3] Belloni, A. and Chernozhukov, V. (2011) ℓ1-Penalized Quantile Regression in High-Dimensional Sparse Models. *The Annals of Statistics*, **39**, 82-130. https://doi.org/10.1214/10-aos827

[4] 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

[5] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005) Sparsity and Smoothness via the Fused Lasso. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **67**, 91-108. https://doi.org/10.1111/j.1467-9868.2005.00490.x

[6] Gu, Y., Fan, J., Kong, L., Ma, S. and Zou, H. (2018) ADMM for High-Dimensional Sparse Penalized Quantile Regression. *Technometrics*, **60**, 319-331. https://doi.org/10.1080/00401706.2017.1345703

[7] Chen, L., Sun, D. and Toh, K. (2016) A Note on the Convergence of ADMM for Linearly Constrained Convex Optimization Problems. *Computational Optimization and Applications*, **66**, 327-343. https://doi.org/10.1007/s10589-016-9864-7

[8] Wu, X., Liang, R., Zhang, Z. and Cui, Z. (2025) Multi-Block Linearized Alternating Direction Method for Sparse Fused Lasso Modeling Problems. *Applied Mathematical Modelling*, **137**, Article 115694. https://doi.org/10.1016/j.apm.2024.115694

[9] Mirzaeifard, R., Venkategowda, N.K.D., Gogineni, V.C. and Werner, S. (2023) Smoothing ADMM for Sparse-Penalized Quantile Regression with Non-Convex Penalties. *IEEE Open Journal of Signal Processing*, **5**, 213-228. https://doi.org/10.1109/ojsp.2023.3344395

[10] Li, X., Sun, D. and Toh, K. (2018) A Highly Efficient Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems. *SIAM Journal on Optimization*, **28**, 433-458. https://doi.org/10.1137/16m1097572

[11] Rockafellar, R.T. (1997) Convex Analysis. Princeton University Press.

[12] Combettes, P.L. and Wajs, V.R. (2005) Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, **4**, 1168-1200. https://doi.org/10.1137/050626090

[13] Rockafellar, R.T. (1976) Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, **14**, 877-898. https://doi.org/10.1137/0314056

[14] Li, X., Sun, D. and Toh, K. (2018) On Efficiently Solving the Subproblems of a Level-Set Method for Fused Lasso Problems. *SIAM Journal on Optimization*, **28**, 1842-1866. https://doi.org/10.1137/17m1136390

[15] Robinson, S.M. (1981) Some Continuity Properties of Polyhedral Multifunctions. *Mathematical Programming Studies*, **14**, 206-214. https://doi.org/10.1007/bfb0120929

[16] Qi, L. and Sun, J. (1993) A Nonsmooth Version of Newton's Method. *Mathematical Programming*, **58**, 353-367. https://doi.org/10.1007/bf01581275

[17] Facchinei, F. and Pang, J.-S. (2003) Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer.

[18] Lima, M.V.S., Ferreira, T.N., Martins, W.A. and Diniz, P.S.R. (2014) Sparsity-Aware Data-Selective Adaptive Filters. *IEEE Transactions on Signal Processing*, **62**, 4557-4572. https://doi.org/10.1109/tsp.2014.2334560