arXiv:2601.02826v1 [stat.ME] 6 Jan 2026

可扩展超高维分位数回归与基因组学应用

Hanqing Wu[1], **Jonas Wallin**[2], **Iuliana Ionita-Laza**[3]

[1] *Department of Statistics*、 *Lund University*、 *Lund*、 *Sweden*、 *e-mail:* hanqing.wu@stat.lu.se

[2] *Department of Statistics*、 *Lund University*、 *Lund*、 *Sweden*、 *e-mail:*jonas.wallin@stat.lu.se

[3] *Department of Biostatistics*、 *Columbia University*、 *New York*、 *USA*、 *e-mail:* ii2135@cumc.columbia.edu

摘要： 现代社交媒体、基因组学和生物医学信息学产生的数据集通常具有异质性和（超）高维性，给传统建模技术带来了巨大挑战。分位数回归（QR）不仅提供了一种灵活的方法，用于捕捉结果变量条件分布中的异质性效应，还能自然地产生预测区间，从而帮助量化未来预测中的不确定性。然而，在大规模场景下，经典分位数回归方法可能面临严重的内存和计算约束。这些局限性促使人们采用并行计算以保持方法的可处理性。尽管已有大量研究考察了样本拆分策略，适用场景通常是观测数量 n 远大于特征数量 p，但同样重要的（超）高维情形（$p \gg n$）相对而言仍未得到充分研究。为弥补这一空白，我们针对高维情形下的惩罚分位数回归提出了一种特征拆分近端点算法 FS-QRPPA。借助变分分析领域的最新进展，我们建立了 FS-QRPPA 的 Q-线性收敛速率，并通过英国生物样本库中的大规模基因组学应用，证明了其相较于现有方法更优的可扩展性。此外，与当前方法相比，FS-QRPPA 能够得到更准确的系数估计，并为预测区间提供更好的覆盖率。我们在 R 软件包`fsQRPPA`中提供了并行实现，使惩罚分位数回归能够应用于大规模数据集。

关键词和短语： 高维分位数回归、特征拆分、近端点算法、基因组学应用。

1. 引言

自Koenker 和 Bassett（1978）以来，分位数回归（QR）已成为统计学中的主流技术，并应用于流行病学、经济学、社会科学和基因组学等多个领域（Yu、Lu 和 Stander （2003年）、Wang 等人 （2024年）。分位数回归将经典线性回归（LR）从对表型均值的建模扩展到对完整表型分布的建模，从而能够估计整个性状分布上的效应。分位数回归是一种稳健的建模技术；相较于线性回归，其主要优势之一是能够检测并刻画效应在结果分布不同分位数间的异质性。

这种分位数特定建模不仅揭示了经典线性回归无法捕捉的模式，还通过直接估计结果分布的边界，为构建预测区间提供了自然基础。本文关注数据处于（超）高维时，分位数回归框架下的推断。例如，我们将重点研究一项相关应用：在大规模基因组数据集（如生物样本库）中的应用。这些数据集包含数万（甚至数十万）个体和数百万个特征（例如遗传变异）。鉴于全球生物样本库的迅速扩张及其对生物医学研究的变革性影响，在这些场景中，可扩展的方法至关重要。

在这些高维设定中，出于识别和特征选择的考虑，我们采用惩罚分位数回归（PQR）。令 $y_i \in \mathbb{R}$ 为标量响应变量，$\boldsymbol{x}_i = (1, x_{i1}, \ldots\ldots, x_{ip})^\top \in \mathbb{R}^{p+1}$ 为包含截距的 $p+1$ 维特征向量。随后，我们求解以下优化问题：

$$\min_{\boldsymbol{\beta}} \quad \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}) + \Omega(\boldsymbol{\beta}). \tag{1}$$

这里，$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots\ldots, \beta_p)^\top \in \mathbb{R}^{p+1}$ 是系数向量，$\rho_\tau(z) = (\tau - 1\{z < 0\})z$ 是分位数回归的标准针式损失，而 $\Omega(\boldsymbol{\beta})$ 是惩罚函数。例如，$\Omega(\boldsymbol{\beta})$ 可以是加权 ℓ_1 惩罚项，即存在某个非负权重向量 $\boldsymbol{\lambda} = (0, \lambda_1, \ldots\ldots, \lambda_p)^\top \in \mathbb{R}^{p+1}$，使得 $\Omega(\boldsymbol{\beta}) = \sum_{i=1}^{p} \lambda_i |\beta_i|$。

求解（1）的一大难点在于针式损失的不可微性。对于凸惩罚分位数回归，有多种方法可用，包括内点法、坐标下降、基于平滑化的梯度下降，以及交替方向乘子法（ADMM）。采用 ADMM 时，不是直接最小化（1），而是将针式损失与惩罚项解耦，把原本困难的最小化问题分解为更易求解的子问题。在高维分位数回归的背景下，与其他方法（例如 Boyd（2010年）、Yu 和 Lin（2017）、Gu 等人（2018））相比，ADMM 天生适合并行化，因此在速度、估计准确性和特征选择方面尤其具有竞争力。

如今，研究人员经常面对一个（超）高维设计矩阵 $\boldsymbol{X} \in \mathbb{R}^{n\times(p+1)}$，其中 n 或 p，或二者均过大，即达到数十万甚至数百万的数量级。这会导致内存和计算瓶颈。通过交替方向乘子法，可以进一步将矩阵$\boldsymbol{X}$ 拆分为多个易处理的数据分区，将这些分区及相应的变量子集分配到多个核心或本地机器上，并同时更新变量（关于交替方向乘子法在高维回归问题中的若干应用，见Boyd（2010年）第8章）。通过交替方向乘子法实现并行化的思想也已扩展到分位数回归（Yu 和 Lin （2017））。事实上，关于如何部署分布式版本的交替方向乘子法来求解分位数回归问题的研究文献正日益丰富，其中 n 过大

通过拆分 $\boldsymbol{X}$ 行，或者换句话说，样本拆分（例如参见 Yu 和 Lin (2017)、Fan、Lin 和 Yin(2021)、Liu 和 Zeng (2024年)、Wu、Jiang 和 Zhang (2024年)、Wu 等(2025a)、Mirzaeifard、Ghaderyan 和 Werner (2025)）。

针对特征拆分的研究则少得多，即在特征数量 p 极其庞大的情形下，沿 $\boldsymbol{X}$ 的列进行分区，尽管这种情况在许多科学领域都很常见，尤其是在基因组学领域，高通量检测通常一次便可测量数百万个特征。例如，英国生物样本库（Bycroft 等（2018））是一个大规模群体资源，包含约五十万个个体的基因型和表型数据，以及数百万个测量得到的遗传变量。一个潜在原因是，在样本拆分中，每个并行工作节点所求解的正是集中式情形中的同一个子问题，只需找到适当的方法来聚合梯度、局部估计量或残差等量。相比之下，特征拆分“更难”，因为我们需要处理数据块之间的强耦合，并维护全局残差信息。Yu 和 Lin（2017）简要讨论了采用特征拆分对标准两块 ADMM 方案进行直接扩展。然而，由于缺乏理论依据和收敛保证，他们的方法本质上仍属于启发式方法。此外，他们的数值研究仅考虑了 $p = 1000$的情形，这远低于现实应用中遇到的（超）高维规模。鉴于此，同时为避免直接扩展标准两块 ADMM 不一定收敛这一陷阱（Chen 等（2016）），Wen 等人（2025）针对特征拆分的 PQR 提出了三块 ADMM 算法（FS-QRADMM），并给出了收敛保证和数值效率。然而，正如 Wu、Jiang 和 Zhang（2024）所指出的，FS-QRADMM 引入了总数为 $2Gn$ 的辅助变量，并在每次迭代中更新 $p+1+(3G-1)\,n$ 变量，其中 G 为分区数量。此外，FS-QRADMM 的迭代循环需要处理器或本地机器之间进行多次全局同步和通信，在大规模场景下可能代价高昂。当 G 或 n 较大时，这些因素会使 FS-QRADMM 的计算效能表现得不那么突出。此外，Wen 等人（2025）中的线性收敛速率是用一种相对复杂的度量来表示的，该度量结合了迭代误差与连续差分，使得算法的收敛行为不易直接解释。

最后但同样重要的是，我们注意到理论方法学进展与现有计算实现之间存在一个关键差距。据我们所知，目前尚无针对特征拆分 PQR 的并行化实现。此外，即使对于广受研究的样本拆分算法，高效的并行实现也十分稀缺；现有软件包通常依赖串行循环来模拟并行更新，而不是利用真正的多核或分布式计算。这种高效实现的缺乏，阻碍了这些拆分算法在（超）高维数据中的实际应用。

激发性应用：生物样本库规模基因组学数据。我们的主要动机来自大规模基因组研究中的应用。在过去二十年中，全基因组关联研究（GWAS）一直是识别基因型-表型关联以及进行基因组性状预测的主导方法。GWAS 是一种大规模、无假设的研究，用于检验在整个基因组范围内测量的数百万个遗传变异与目标表型之间的关联。传统上，GWAS 通过基于联盟的荟萃分析开展，将多个常规队列结合起来，以获得较大的样本量。如今，大型人群规模生物样本库，例如英国生物样本库、美国的“全民研究计划”以及世界各地其他全球性生物样本库，提供了统一且具有深度表型信息的数据集，其中包含数万乃至数十万名参与者，使得在单一资源内开展前所未有规模的 GWAS 成为可能。由于样本量庞大且维度很高，对生物样本库规模数据集的分析带来了重大的计算和内存挑战。

几乎所有全基因组关联研究（GWAS）都基于线性回归模型（Uffelmann 等人(2021））。这些模型采用静态视角看待遗传效应，默认遗传影响在不同环境中保持不变。这一假设忽略了基因型与表型关系中有充分记录的复杂性，包括基因与基因（G×G）以及基因与环境（G×E）之间的相互作用；这里的“环境”广义上涵盖各种生物学和外部因素（Pazokitoroudi 等人 (2024年）、Mackay (2014）、Mackay 和 Anholt(2024年）。这些潜在的相互作用会导致遗传关联呈现异质性，而 GWAS 中的线性模型并不适合解析这类异质关联（Wang 等人 (2024年）。尽管已有研究为 GWAS 提出了非线性模型，包括深度学习方法，但其大规模应用仍受若干关键挑战阻碍，例如计算需求高、变量选择不稳定、缺乏稳健的推断工具以及可解释性有限。相比之下，分位数回归（QR）等统计方法提供了一种有原则的途径，使分析能够超越线性假设，从而检测异质且依赖具体情境的遗传关联。分位数回归保留了线性回归的许多核心优势，而这些优势对 GWAS 至关重要：它是一种基于回归的框架，支持协变量调整并考虑样本亲缘关系，能够给出特定分位数水平下可解释的系数估计，并提供有力的统计保证；这些特性往往是深度学习方法所缺乏的。

除了遗传发现之外，基因组性状预测也是精准医学以及动植物育种的重点，旨在开发精准的表型预测模型，利用个体的全基因组遗传特征，这些特征通常概括为多基因风险评分。分位数回归为量化个体预测的不确定性提供了一种自然的方法：它通过给出预测区间，界定响应变量以高概率落入的下限和上限。

尽管具有诸多吸引人的特性，分位数回归方法在遗传学中的应用仍然有限，大多局限于小规模研究，因此其巨大潜力尚未得到充分发挥。一项主要障碍是缺乏可扩展的分位数回归方法，无法应对大规模全基因组关联研究数据的核心挑战，包括超高维问题——样本数达数十万、变异数达数千万。这些局限促使本文提出了相应的方法学进展。

我们的贡献。 受近期 Wu 等(2025b) 针对 Dantzig 选择器提出的特征分裂近端点算法（PPA）研究启发，本文提出 FS-QRPPA，这是一种专为带加权 ℓ_1 正则化的 PQR 设计的新型算法，同时还能自然地适用于平滑截断绝对偏差（SCAD）（Fan (1997)）和极小极大凹惩罚（MCP）（Zhang (2010年）等非凸惩罚项。FS-QRPPA 在结构上类似于经典的两块ADMM，仅使用 $2n$辅助变量，因此在内存使用和计算成本方面均具有优势。此外，FS-QRPPA 的每次迭代都由比 FS-QRADMM 更为简单的顺序更新组成，从而大幅降低了同步和通信开销。值得注意的是，Wu 等 (2025b) 仅报告了 PPA 用于 Dantzig 选择器时的最坏情况下的次线性收敛率。相比之下，基于近期变度量PPA的理论发展，我们以一种简洁的形式建立了该算法的 Q-线性收敛速率，从而清晰地揭示了算法的行为。从实践角度来看，本文的另一项贡献是开发了 FS-QRPPA 的多核并行 R 实现，并将其作为开源 R 软件包 `fsQRPPA`提供；该软件包使用 `RcppParallel` 和 `RcppArmadillo`开发。尽管已有多种分布式 QR 算法被提出，高效的并行实现仍然十分稀缺，限制了它们在大规模现代应用中的使用。因此，现有大多数应用都集中于规模相对适中的数据集。为解决这一局限性，我们将其应用于英国生物样本库的全基因组数据，展示了 FS-QRPPA 的可扩展性。

文章结构。 本文其余部分的组织结构如下。第 2 节我们介绍一般的 PQR 问题，重点讨论加权 ℓ_1惩罚项情形，并推导其相应的拉格朗日函数。第3节我们回顾现有的交替方向乘子法解法，并讨论其计算方面的不足。第 4 节我们详细介绍所提出的 FS-QRPPA 框架，并在第 5节正式证明其收敛性质。第 6 节介绍如何将 FS-QRPPA 无缝扩展到处理带非凸惩罚项的 PQR。我们通过广泛的数值研究以及第7节中的生物样本库规模应用，验证所提方法的可扩展性和实用性。最后，在第8 节我们总结全文，并讨论未来的研究方向。所有技术性证明和其他补充细节均见附录。

符号与定义。本文总结全文中使用的符号和定义。我们用 $q_{Y|X}(\tau \mid \boldsymbol{x}) = \inf\{y \in \mathbb{R} : F_{Y|X}(y \mid \boldsymbol{x}) \geq \tau\}$ 表示给定 $X = \boldsymbol{x}$ 时 Y 的第 τ 个条件分位数；在不会引起混淆的情况下，简写为 $q_\tau(\boldsymbol{x})$。令 $\mathbb{R}^m$ 表示 m 维欧几里得空间，令 $1\{\cdot\}$ 表示指示函数：当括号内的条件成立时取 1，否则取 0。符号 0 表示适当维度的零向量或矩阵。类似地，令 1 表示适当维度的全 1 向量。对于标量 $v \in \mathbb{R}$，记作 $v_+ = \max\{v, 0\}$；对于向量 $\boldsymbol{v} = (v_1, \ldots, v_m)^\top \in \mathbb{R}^m$，定义 $\boldsymbol{v}_+ = \big((v_1)_+, \ldots, (v_m)_+\big)^\top$。给定 $\boldsymbol{v}_1$、$\boldsymbol{v}_2 \in \mathbb{R}^m$，其逐元素（Hadamard）乘积记为 $\boldsymbol{v}_1 \circ \boldsymbol{v}_2$。对于实对称矩阵 $\boldsymbol{M} \in \mathbb{R}^{m\times m}$，分别用 $\Lambda_{\max}(\boldsymbol{M})$ 和 $\Lambda_{\min}(\boldsymbol{M})$ 表示其最大和最小特征值；若 $\boldsymbol{M}$ 为正定矩阵，则记为 $\boldsymbol{M} \succ 0$。记号 $\mathrm{diag}(d_1, \ldots, d_m)$ 表示对角（或分块对角）矩阵，其对角元素（或矩阵块）为 $d_1, \ldots, d_m$；$\boldsymbol{I}_m$ 表示 $m \times m$ 单位矩阵。我们用 $\|\cdot\|_1$ 和 $\|\cdot\|_2$ 分别表示 $\mathbb{R}^m$ 上的 ℓ_1 范数和 ℓ_2 范数。令 $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ 为对称正定矩阵。对于任意 $\boldsymbol{v} \in \mathbb{R}^m$ 和 $\mathcal{C} \subset \mathbb{R}^m$，定义由 $\boldsymbol{B}$ 诱导的范数 $\|\boldsymbol{v}\|_{\boldsymbol{B}} = \sqrt{\boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{v}}$ 以及相应的距离 $\mathrm{dist}_{\boldsymbol{B}}(\boldsymbol{v}, \mathcal{C}) = \inf_{\boldsymbol{s}\in\mathcal{C}} \|\boldsymbol{v} - \boldsymbol{s}\|_{\boldsymbol{B}}$。特别地，当 $\boldsymbol{B} = \boldsymbol{I}_m$ 为单位矩阵时，$\mathrm{dist}_{\boldsymbol{B}}(\boldsymbol{v}, \mathcal{C}) = \mathrm{dist}(\boldsymbol{v}, \mathcal{C}) = \inf_{\boldsymbol{s}\in\mathcal{C}} \|\boldsymbol{v} - \boldsymbol{s}\|_2$ 即为欧几里得距离。最后，令 $f : \mathbb{R}^m \to \mathbb{R}$。对于 $\boldsymbol{v} \in \mathbb{R}^m$，只要定义良好，我们将 $\partial f(\boldsymbol{v})$ 记为 f 在 $\boldsymbol{v}$ 处的次微分。

2. 惩罚分位数回归及其拉格朗日函数

一般而言，PQR问题（1）可等价地改写为以下约束优化问题：

$$\begin{aligned} \min_{\boldsymbol{\beta},\boldsymbol{z}} \quad & \hat{Q}_\tau(\boldsymbol{z}) + \Omega(\boldsymbol{\beta}), \\ \text{s.t.} \quad & \boldsymbol{z} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}, \end{aligned} \tag{2}$$

其中 $\hat{} \; Q_\tau(\boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^n \rho_\tau(z_i)$。

（2）的拉格朗日函数为：

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{z}; \boldsymbol{\theta}) = \Omega(\boldsymbol{\beta}) + \hat{Q}_\tau(\boldsymbol{z}) + \boldsymbol{\theta}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{z}), \tag{3}$$

其中 $\boldsymbol{\theta} \in \mathbb{R}^n$ 是对偶向量。当 $\Omega(\boldsymbol{\beta})$ 是凸惩罚项时，原始分位数回归问题（1）可等价地视为求解拉格朗日函数（3）的鞍点。这是一个标准结果（例如参见 Rockafellar andWets (1998)），为完整起见，我们在以下命题中陈述该结果。更多细节见附录C。

命题 1。假设 $\Omega(\boldsymbol{\beta})$ 是凸函数。那么，$\bar{\boldsymbol{u}} = (\bar{\boldsymbol{\beta}}^\top, \bar{\boldsymbol{z}}^\top)^\top \in \mathbb{R}^{p+n}$ 是¯最小化问题 (2) 的解，当且仅当存在某个 $\boldsymbol{\theta} \in \mathbb{R}^n$ 使得$(\bar{\boldsymbol{u}}, \bar{\boldsymbol{\theta}})$ 是拉格朗日函数 $\mathcal{L}$ 在 (3)中的鞍点，即对于所有的 $\boldsymbol{u} \in \mathbb{R}^{p+n}$，$\boldsymbol{\theta} \in \mathbb{R}^n$都有

$$\mathcal{L}(\bar{\boldsymbol{u}}; \boldsymbol{\theta}) \leq \mathcal{L}(\bar{\boldsymbol{u}}; \bar{\boldsymbol{\theta}}) \leq \mathcal{L}(\boldsymbol{u}; \bar{\boldsymbol{\theta}}).$$

在接下来的两节中，我们将重点讨论 $\Omega(\boldsymbol{\beta})$ 为加权 ℓ_1 惩罚项的具体情形，其中 LASSO 惩罚是其特例。具体而言，

$$\Omega(\boldsymbol{\beta}) = \|\boldsymbol{\lambda} \circ \boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} \lambda_j |\beta_j|, \tag{4}$$

其中 $\lambda_0 = 0$ 且 $\lambda_j \geq 0$ 对于 $j = 1, \ldots\ldots, p$。关于带惩罚项的 PQR（4）已有大量研究。这不仅是因为它与广受欢迎的 LASSO 惩罚直接相关，还因为一旦加权 ℓ_1 PQR 可以高效求解，就能通过局部线性近似（LLA）（Zou andLi（2008年）获得带一般折叠凹惩罚的 PQR 的高质量解；我们将在本文后文讨论此类非凸惩罚。

3. PQR 的交替方向乘子法

3.1. PQR 的典范交替方向乘子法及其多块扩展

直接求解（3）的鞍点可能是一个难题。标准的两块ADMM通过考虑原始拉格朗日函数（3）的增广版本，尝试求得其鞍点。

$$\mathcal{L}_\rho(\boldsymbol{\beta}, \boldsymbol{z}; \boldsymbol{\theta}) = \Omega(\boldsymbol{\beta}) + \hat{Q}_\tau(\boldsymbol{z}) + \boldsymbol{\theta}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{z}\|_2^2, \tag{5}$$

其中 $\rho > 0$ 为增广拉格朗日参数。（5）的鞍点也是（3）的鞍点，可以通过迭代求得。具体而言，设 $\boldsymbol{\beta}^k$、$\boldsymbol{z}^k$、$\boldsymbol{\theta}^k$ 为第 k 次迭代得到的更新值。那么，标准交替方向乘子法算法如下：

$$\begin{cases} \boldsymbol{\beta}^{k+1} & = \operatorname{argmin}_{\boldsymbol{\beta}} \mathcal{L}_\rho(\boldsymbol{\beta}, \boldsymbol{z}^k; \boldsymbol{\theta}^k) = \operatorname{argmin}_{\boldsymbol{\beta}} \Omega(\boldsymbol{\beta}) - \boldsymbol{\theta}^{k\top} \boldsymbol{X}\boldsymbol{\beta} + \frac{\rho}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{z}^k\|_2^2 \\ \boldsymbol{z}^{k+1} & = \operatorname{argmin}_{\boldsymbol{z}} \mathcal{L}_\rho(\boldsymbol{\beta}^{k+1}, \boldsymbol{z}; \boldsymbol{\theta}^k) = \operatorname{argmin}_{\boldsymbol{z}} \hat{Q}_\tau(\boldsymbol{z}) - \boldsymbol{\theta}^{k\top} \boldsymbol{z} + \frac{\rho}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{k+1} - \boldsymbol{z}\|_2^2 \\ \boldsymbol{\theta}^{k+1} & = \boldsymbol{\theta}^k + \rho(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{k+1} - \boldsymbol{z}^{k+1}) \end{cases} \tag{6}$$

对于一般的设计矩阵 $\boldsymbol{X}$，$\boldsymbol{\beta}^{k+1}$ 的更新没有闭式解。可以通过坐标下降法（Yu 和 Lin(2017）、Gu 等人(2018）求解，也可以通过引入更多辅助变量（Yu、Lin 和 Wang(2017）求解，或添加线性化项（Gu 等人 (2018）求解。当 $\boldsymbol{X}$ 大到难以处理时，内存和计算瓶颈会变得尤为严重；在极端情况下，单台机器甚至无法将完整矩阵 $\boldsymbol{X}$ 载入内存。在这种情况下，迭代优化所需的反复直接计算 $\boldsymbol{X}\boldsymbol{\beta}$($O(np)$ 操作）在计算上变得不切实际。并行计算使得将难以处理的庞大问题划分为可行的子任务成为可能。当特征数量 p 较大时，对 $\boldsymbol{\beta}^{k+1}$ 的更新采用特征拆分是理想的。在我们的加权ℓ_1 惩罚项设定中，可以通过引入更多原始辅助变量来进行拆分，因为 $\Omega(\boldsymbol{\beta})$ 可视为施加于各个坐标上的惩罚项之和，因而是“可分离的”。然而，这种扩展并非易事，因为由此产生的多块 ADMM 未必能够收敛（

因此“可分离”。然而，这种扩展并非易事，因为由此产生的多块 ADMM 未必能够收敛（Chen 等(2016)）。为此，Wen 等人(2025) 提出了用于 PQR 的特征分裂交替方向乘子法（FS-QRADMM），其中 $\Omega(\boldsymbol{\beta})$ 是加权的 ℓ_1-惩罚项，并采用具有收敛保证的三块交替方向乘子法[1]。注意，设计矩阵 $\boldsymbol{X}$ 可以按列划分为 G 个部分，使得

$$\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_G), \text{ where } \boldsymbol{X}_g \in \mathbb{R}^{n \times p_g}, \quad g = 1, \ldots, G \text{ and } \sum_{g=1}^{G} p_g = p + 1. \tag{7}$$

系数向量 $\boldsymbol{\beta}$ 可以相应地划分，即

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_G^\top\right)^\top, \text{ where } \boldsymbol{\beta}_g \in \mathbb{R}^{p_g}.$$

我们有 $\boldsymbol{X}\boldsymbol{\beta} = \sum_{g=1}^{G} \boldsymbol{X}_g \boldsymbol{\beta}_g$ 和

$$\Omega(\boldsymbol{\beta}) = \sum_{g=1}^{G} \Omega_g(\boldsymbol{\beta}_g), \text{ where } \Omega_g(\boldsymbol{\beta}_g) = \|\boldsymbol{\lambda}_g \circ \boldsymbol{\beta}_g\|_1 = \sum_{j=1}^{p_g} \lambda_{j,g} |\beta_{j,g}|. \tag{8}$$

在标准的两块交替方向乘子法表述中，包含$\boldsymbol{X}\boldsymbol{\beta}$ 的二次项会耦合不同的 $\boldsymbol{\beta}_g$，从而阻碍并行化。Wen 等人(2025) 引入了额外的松弛变量 $\boldsymbol{w}_g \in \mathbb{R}^n$，$g = 2$，……，$G$ 以实现解耦。他们将原始 PQR 问题（1）重新表述为

$$\begin{aligned} \min_{\boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\omega}_g} \quad & \hat{Q}_\tau(\boldsymbol{z}) + \Omega(\boldsymbol{\beta}), \\ \text{s.t.} \quad & \boldsymbol{z} + \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \sum_{g=2}^{G} \boldsymbol{\omega}_g = \boldsymbol{y}, \\ & \boldsymbol{X}_g \boldsymbol{\beta}_g = \boldsymbol{\omega}_g, \quad g = 2, \ldots, G. \end{aligned} \tag{9}$$

相应的增广拉格朗日函数为

$$\begin{aligned} \mathcal{L}_\phi(\boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\omega}; \boldsymbol{\gamma}) = & \sum_{g=1}^{G} \Omega_g(\boldsymbol{\beta}_g) + \hat{Q}_\tau(\boldsymbol{z}) + \boldsymbol{\gamma}_1^\top \left(\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{z} + \boldsymbol{\omega}_2 + \cdots + \boldsymbol{\omega}_G - \boldsymbol{y}\right) \\ & + \frac{\phi}{2} \left\|\boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{z} + \boldsymbol{\omega}_2 + \cdots + \boldsymbol{\omega}_G - \boldsymbol{y}\right\|_2^2 + \sum_{g=2}^{G} \boldsymbol{\gamma}_g^\top \left(\boldsymbol{X}_g \boldsymbol{\beta}_g - \boldsymbol{\omega}_g\right) \\ & + \frac{\phi}{2} \sum_{g=2}^{G} \left\|\boldsymbol{X}_g \boldsymbol{\beta}_g - \boldsymbol{\omega}_g\right\|_2^2, \end{aligned} \tag{10}$$

[1]我们仅介绍Wen 等人(2025) 中的 FS-QRADMM-prox，该文还介绍了一个不含近端项的变体，即 FS-QRADMM-CD。不过，FS-QRADMM-CD 对 $\boldsymbol{\beta}$ 没有显式形式更新，而是依赖坐标下降；与 FS-QRADMM-prox 相比，其性能可能较差。更重要的是，FS-QRADMM-CD 缺乏收敛保证。

其中 $\boldsymbol{\gamma}_g \in \mathbb{R}^n$、$g = 1$、.....、$G$ 为对偶向量。对于（10），多块 ADMM 包含以下更新周期：$\boldsymbol{\beta}^k \to \boldsymbol{\omega}^{k+\frac{1}{2}} \to \boldsymbol{z}^k \to \boldsymbol{\omega}^{k+1} \to \boldsymbol{\gamma}^k$ 给出

$$
\begin{cases}
\boldsymbol{\beta}^{k+1} = \operatorname{argmin} \mathcal{L}_\phi \left(\boldsymbol{\beta}, \boldsymbol{z}^k, \boldsymbol{\omega}^k; \boldsymbol{\gamma}^k\right) + \frac{\phi}{2} \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^k\right\|_{\mathcal{T}_{\boldsymbol{\beta}}}^2 \\
\boldsymbol{\omega}^{k+\frac{1}{2}} = \operatorname{argmin} \mathcal{L}_\phi \left(\boldsymbol{\beta}^{k+1}, \boldsymbol{z}^k, \boldsymbol{\omega}; \boldsymbol{\gamma}^k\right) \\
\boldsymbol{z}^{k+1} = \operatorname{argmin} \mathcal{L}_\phi \left(\boldsymbol{\beta}^{k+1}, \boldsymbol{z}, \boldsymbol{\omega}^{k+\frac{1}{2}}; \boldsymbol{\gamma}^k\right) + \frac{\phi}{2} \left\|\boldsymbol{z} - \boldsymbol{z}^k\right\|_{\mathcal{T}_{\boldsymbol{z}}}^2 \\
\boldsymbol{\omega}^{k+1} = \operatorname{argmin} \mathcal{L}_\phi \left(\boldsymbol{\beta}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{\omega}; \boldsymbol{\gamma}^k\right) \\
\boldsymbol{\gamma}_1^{k+1} = \boldsymbol{\gamma}_1^k + \xi\phi \left(\boldsymbol{X}_1 \boldsymbol{\beta}_1^{k+1} + \boldsymbol{z}^{k+1} + \sum_{g=2}^G \boldsymbol{\omega}_g^{k+1} - \boldsymbol{y}\right) \\
\boldsymbol{\gamma}_g^{k+1} = \boldsymbol{\gamma}_g^k + \xi\phi \left(\boldsymbol{X}_g \boldsymbol{\beta}_g^{k+1} - \boldsymbol{\omega}_g^{k+1}\right), \quad g = 2, \ldots, G
\end{cases}
\tag{11}
$$

其中 $\xi \in (0, (\sqrt{5}+1)/2)$，$\mathcal{T}_{\boldsymbol{\beta}}$ 和 $\mathcal{T}_{\boldsymbol{z}}$ 是对收敛至关重要的半正定矩阵。请注意，为了改善收敛性，$\boldsymbol{\omega}_g$ 会更新两次。

3.2. 关于 ***FS-QRADMM*** 的一些说明

从式（10）和式（11）可以看出，通过引入松弛变量 $\boldsymbol{\omega}_g$ 作为 $\boldsymbol{X}_g\boldsymbol{\beta}_g$ 的局部副本，并引入块对角 $\mathcal{T}_{\boldsymbol{\beta}}$，增广拉格朗日函数在块变量 $\boldsymbol{\beta}_g$上变得可分离。因此，FS-QRADMM 通过 $\boldsymbol{\beta}_g$更新 G，各个独立子问题可在不同数据块间并行计算。给定 $\{\boldsymbol{\beta}_g\}_{g=1}^G$ 和 $\boldsymbol{z}$后，$\{\boldsymbol{\omega}_g\}_{g=2}^G$ 及其对偶变量 $\{\boldsymbol{\gamma}_g\}_{g=1}^G$的更新同样可按数据块分离，适合并行执行。利用这种并行性，FS-QRADMM 有效缓解了超大 p所带来的计算瓶颈，同时实现了较高的计算效率和准确的系数估计（Wen 等人（2025）。

然而，相较于两块 ADMM 方案，FS-QRADMM 的更新周期形成了更长且包含多个阶段的依赖链，因此需要额外更新松弛变量 $\boldsymbol{\omega}_g$和对偶变量 $\boldsymbol{\gamma}_g$，涉及 $g = 2$、.....、G（其中 $\boldsymbol{\gamma}_1$对应于两块方案中的单个对偶向量）。这还会增加内存负担，需要维护 $2(G-1)$ 个维度为 n的向量。此外，在并行环境中，每个阶段的实际计算速度取决于速度最慢的数据块 G，同时还要承担汇总和分发用于同步的全局量（例如 $\sum_{g=1}^G \boldsymbol{X}_g\boldsymbol{\beta}_g$ 和 $\boldsymbol{z}$）的成本。因此，尽管更新每个 $\boldsymbol{\omega}_g$ 或 $\boldsymbol{\gamma}_g$ 在代数上都很简单，但随着 G 增大，更新所有这些量的阶段往往会延长。结果是，当 G增加或 n 非常大时，内存、计算和协调成本可能会抵消并行执行 $\boldsymbol{\beta}$更新所带来的收益。一个自然的问题是：能否避免引入额外的 $\boldsymbol{\omega}_g$和 $\boldsymbol{\gamma}_g$，以进一步降低计算和内存成本？接下来我们将重点回答这个问题。

4. 一种用于 PQR 的特征分裂近端点算法

最近，Wu 等（2025b）针对一种类似于第 3.2节所讨论的挑战，进行了研究；具体而言，他们研究了通过多块 ADMM 求解的特征拆分 Dantzig 选择器（Wen、Yang和Zhao （2024年））。为减轻计算负担，他们开发了特征分裂近端点算法（PPA）。与多块 ADMM 相比，他们的 PPA 形式大幅减少了辅助变量的数量，并简化了更新逻辑。PPA（Martinet（1970）、Rockafellar （1976年））是一大类迭代方法，用于求解最优性条件表示为广义方程的优化问题。这一框架涵盖非光滑凸优化，因此可直接应用于我们关注的带加权 惩罚项的分位数回归。PPA 是许多统计优化方法的基础。全面的讨论请参见 Polson、Scott和Willard （2015年）。值得注意的是，ADMM 可视为 PPA 的特例（Gabay （1983年）、Eckstein和Bertsekas（1992年，参见 Boyd （2010年）第 3.5 节）。

受Wu 等 （2025b）的研究启发，我们提出了一种用于 PQR 的特征分裂 PPA（FS-QRPPA）。与经典的两块ADMM类似，我们的方法仅使用 $2n$ 个辅助变量，这有望提升内存效率并降低计算成本。FS-QRPPA 仅需使用经典两块 ADMM公式中的变量 $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$、 $\boldsymbol{z} \in \mathbb{R}^n$和 $\boldsymbol{\theta} \in \mathbb{R}^n$ ，并采用精简的更新方案，其每次迭代的步骤与式（6）中的 ADMM 更新相对应。因此，它避免了 FS-QRADMM 中额外的内存和计算开销，详见第 3.2 节。关于 PPA 的更多细节，以及 FS-QRPPA 如何归入称为变度量PPA（VMPPA）的子类别，见附录 D。

现在，我们介绍 FS-QRPPA 的更新框架，可概括如下：

$$
\begin{cases}
\boldsymbol{\beta}_g^{k+1} = \underset{\boldsymbol{\beta}_g \in \mathbb{R}^{p_g}}{\operatorname{argmin}} \Omega_g(\boldsymbol{\beta}_g) - \boldsymbol{\theta}^{k\top} \boldsymbol{X}_g \boldsymbol{\beta}_g + \dfrac{1}{2} \|\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^k\|_{\boldsymbol{M}_g}^2, \quad g = 1, \ldots, G \\
\boldsymbol{z}^{k+1} = \underset{\boldsymbol{z} \in \mathbb{R}^n}{\operatorname{argmin}} \hat{Q}_\tau(\boldsymbol{z}) - \boldsymbol{\theta}^{k\top} \boldsymbol{z} + \dfrac{\mu}{2} \|\boldsymbol{z} - \boldsymbol{z}^k\|_2^2 \\
\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \dfrac{\mu}{G+1} \left[2 \left(\sum_{g=1}^G \boldsymbol{X}_g \boldsymbol{\beta}_g^{k+1} + \boldsymbol{z}^{k+1} - \boldsymbol{y} \right) - \left(\sum_{g=1}^G \boldsymbol{X}_g \boldsymbol{\beta}_g^k + \boldsymbol{z}^k - \boldsymbol{y} \right) \right].
\end{cases}
\tag{12}
$$

这里 $\mu > 0$ 是预先指定的增广参数， $\boldsymbol{M}_g$ 是一个正定矩阵，使得 $\boldsymbol{M}_g - \boldsymbol{X}_g^\top \boldsymbol{X}_g \succ 0$。在本文中，我们取$\boldsymbol{M}_g = \eta_g \boldsymbol{I}_{p_g}$其中 $\eta_g > \mu \Lambda_{\max}(\boldsymbol{X}_g^\top \boldsymbol{X}_g)$ $g = 1$，……， G。人们可能会担心，对$\boldsymbol{X}_g^\top \boldsymbol{X}_g$进行特征分解的代价很高。不过，可以采用幂法（第 8.2 章，Golub和Van Loan(2013年）或 Lanczos迭代（第 10.1 章，Golub和Van Loan(2013年）等高效算法

来求得最大特征值，从而缓解对完整特征值分解成本过高的顾虑。在选择 $\boldsymbol{M}_g = \eta_g \boldsymbol{I}_{p_g}$ 的情况下， $\boldsymbol{\beta}^{k+1}$ 的更新可以进一步简化为

$$\boldsymbol{\beta}_g^{k+1} = \underset{\boldsymbol{\beta}_g}{\operatorname{argmin}}\, \Omega_g(\boldsymbol{\beta}_g) - \boldsymbol{\theta}^{k\top} \boldsymbol{X}_g \boldsymbol{\beta}_g + \frac{\eta_g}{2} \|\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^k\|_2^2, \quad g = 1, \ldots, G. \tag{13}$$

对 $\boldsymbol{\beta}^{k+1}$ 在（13）中的更新以及对$\boldsymbol{z}^{k+1}$ 在（12）中的更新有一个有利特性，即二者都具有显式形式。事实上，在加权 ℓ_1 惩罚项的情形下，对于 $g = 1, \ldots, G$,

$$\boldsymbol{\beta}_g^{k+1} = \underset{\boldsymbol{\beta}_g}{\operatorname{argmin}} \|\boldsymbol{\lambda}_g \circ \boldsymbol{\beta}_g\| - \boldsymbol{\theta}^{k\top} \boldsymbol{X}_g \boldsymbol{\beta}_g + \frac{\eta_g}{2} \|\boldsymbol{\beta}_g - \boldsymbol{\beta}_g^k\|_2^2$$

这意味着

$$\boldsymbol{\beta}_g^{k+1} = \frac{1}{\eta_g} \left[\left(\eta_g \boldsymbol{\beta}_g^k + \boldsymbol{X}_g^\top \boldsymbol{\theta}^k - \boldsymbol{\lambda}_g \right)_+ - \left(-\eta_g \boldsymbol{\beta}_g^k - \boldsymbol{X}_g^\top \boldsymbol{\theta}^k - \boldsymbol{\lambda}_g \right)_+ \right] \tag{14}$$

and

$$\boldsymbol{z}^{k+1} = \left(\boldsymbol{z}^k + \frac{1}{\mu} \boldsymbol{\theta}^k - \frac{\tau}{n\mu} \right)_+ - \left(-\boldsymbol{z}^k - \frac{1}{\mu} \boldsymbol{\theta}^k - \frac{1-\tau}{n\mu} \right)_+ . \tag{15}$$

如（14）所示，FS-QRPPA 的结构本质在于其块可分性。与 FS-QRADMM 类似，每个块 $\boldsymbol{\beta}_g^{k+1}$的更新仅取决于局部数据（$\boldsymbol{X}_g$， $\boldsymbol{\beta}_g^k$）。这种解耦使得可以在G 个独立处理器或本地机器上并行执行。同样，计算 $\boldsymbol{X}\boldsymbol{\beta}$（更新 $\boldsymbol{\theta}^{k+1}$所需）可拆分为 G 个同时进行的子任务，先计算 $\boldsymbol{X}_g\boldsymbol{\beta}_g$，再求和。由于聚合步骤只需对 G n维向量求和，因此开销很低。除了并行效率之外，特征拆分还通过改善 $\boldsymbol{\beta}_g$子问题的条件数来加速计算，我们将在注1中详述这一性质。我们在算法1中总结了加权 ℓ_1 惩罚项下 FS-QRPPA 的更新方案，并正式定义了例程 FS-QRPPA(·)。

注释 1。 分区数量 G 会影响算法 1 的收敛表现，体现了与 Wen 等人 (2025) 中观察到的类似效应。一方面， G值较大意味着分区更精细，从而显著降低局部最大特征值 $\Lambda_{\max}(\boldsymbol{X}_g^\top \boldsymbol{X}_g)$，使其远小于双块近端交替方向乘子法所需的全局对应值$\Lambda_{\max}(\boldsymbol{X}^\top \boldsymbol{X})$（Gu 等人(2018)）。因此， η_g 可能远小于不采用特征拆分时的对应值。这一降低直接加速了 $\boldsymbol{\beta}$-子问题的求解，因为它允许采用显著更大的步长。另一方面，对于给定的 μ， $\boldsymbol{\theta}$-子问题的步长与 G成反比，因此随着 G 增大，其更新可能会变慢。此外，并行执行中的同步开销通常也会随着 G增大。因此， G 取适中值（例如 5-50）较为可取

[2]停止准则的详细说明见附录 D。

算法 1: FS-QRPPA with 加权 ℓ_1 惩罚项 (FS-QRPPA($\boldsymbol{X}$, $\boldsymbol{y}$, $\boldsymbol{\beta}^0$, $\boldsymbol{z}^0$, $\boldsymbol{\theta}^0$, μ, $\{\boldsymbol{\lambda}_g\}_{g=1}^G$, $\{\eta_g\}_{g=1}^G$))

输入: $\boldsymbol{X}=(\boldsymbol{X}_1、\ldots\ldots、\boldsymbol{X}_G)$, $\boldsymbol{y}$, $\boldsymbol{\beta}^0=(\boldsymbol{\beta}_1^{0\top}、\ldots\ldots、\boldsymbol{\beta}_G^{0\top})^\top$, $\boldsymbol{z}^0$、$\boldsymbol{\theta}^0$, μ, $\{\boldsymbol{\lambda}_g\}_{g=1}^G$, $\{\eta_g\}_{g=1}^G$)

$k \leftarrow 0$;

当 *Stopping criteria not satisfied*[2] **执行**

  并行执行: for $g=1$ **to** G **do**

    更新 $\boldsymbol{\beta}_g^{k+1}$, 依据（14）;

  end

  更新 $\boldsymbol{z}^{k+1}$, 依据（15）;

  更新 $\boldsymbol{\theta}^{k+1}$, 依据（12）;

  $k \leftarrow k+1$;

结束

输出: $\boldsymbol{w}^k=(\boldsymbol{\beta}^k, \boldsymbol{z}^k, \boldsymbol{\theta}^k)$, 其中 $\boldsymbol{\beta}^k=(\boldsymbol{\beta}_1^{k\top}, \ldots\ldots, \boldsymbol{\beta}_G^{k\top})^\top$

以便在计算资源允许的情况下平衡两个子问题的效率。附录 E 中的备注 A4 进一步讨论了 G 对收敛速率的双重影响。

备注 2. 选择 $\boldsymbol{M}_g=\eta_g\boldsymbol{I}_{p_g}$ 是出于计算方面的考虑。具体而言，我们只需计算 $\Lambda_{\max}(\boldsymbol{X}_g^\top\boldsymbol{X}_g)$，而且可以并行计算。然而，这种选择未必是最优的。通常存在某个对角矩阵 $\tilde{\boldsymbol{M}}_g$，它能提供更紧的界，即 $\eta_g\boldsymbol{I}_{p_g}\succ\tilde{\boldsymbol{M}}_g$ 且 $\tilde{\boldsymbol{M}}_g\succ\boldsymbol{X}_g^\top\boldsymbol{X}_g$，从而加速收敛。然而，通常要找到这样的更紧的 $\tilde{\boldsymbol{M}}_g$ 并非易事；事实上，这是图像处理领域中的一个重要问题，称为寻找“对角主化矩阵”（Muckley、Noll 和 Fessler (2015年)、McGaffin 和 Fessler(2015年)）。我们将把$\boldsymbol{M}_g$ 的改进设计留待未来研究。

5. FS-QRP 的 Q-线性收敛 PA

借助 VMPPA 的最新进展，我们建立了算法 1 的收敛性质。具体而言，我们证明了该算法具有全局收敛性，并且特别地达到了 Q-线性收敛速率。为便于说明，我们简要回顾收敛速率的定义（参见，例如，Ortega 和 Rheinboldt (2006年)、Nocedal 和 Wright (2006年)）。在我们的语境下，序列 $\{\boldsymbol{w}^k\}$ 被称为 Q-线性收敛于解集 $\mathcal{S}$，如果连续误差之比渐近地有界于 1 以下。相比之下，如果误差由一个 Q-线性收敛的序列控制，则称该序列 R-线性收敛；等价地说，如果误差被一个几何衰减序列所界定，也可作此称谓。

FS-QRPPA 的收敛性质总结如下定理，技术细节见附录 E。

定理 2。 记 $\left\{\boldsymbol{w}^k=\left(\boldsymbol{\beta}^k,\boldsymbol{z}^k,\boldsymbol{\theta}^k\right)\right\}$ 由算法*1*生成的序列。

1.（算法收敛）它收敛到某个 $\bar{\boldsymbol{w}}=\left(\bar{\boldsymbol{\beta}},\ \bar{\boldsymbol{z}},\ \bar{\boldsymbol{\theta}}\right)\in\mathcal{S}$，即拉格朗日函数的鞍点集合 （3）。*2.*（ $\boldsymbol{H}$诱导距离下的 Q-线性收敛速率）存在某个正定矩阵 $\boldsymbol{H}$，它仅取决于 $\boldsymbol{X}$，μ，G，$\{\eta_g\}_{g=1}^G$，以及一个常数$r\in$ （0，1），使得

$$\lim_{k\to\infty}\operatorname{dist}_{\boldsymbol{H}}(\boldsymbol{w}^k,\mathcal{S})=0\ \text{with}\ \limsup_{k\to\infty}\frac{\operatorname{dist}_{\boldsymbol{H}}(\boldsymbol{w}^{k+1},\mathcal{S})}{\operatorname{dist}_{\boldsymbol{H}}(\boldsymbol{w}^k,\mathcal{S})}\leq r.$$

推论 3 （欧几里得距离下的 R-线性收敛速率）。设 $\{\boldsymbol{w}^k\}$ 为定理 *2*中的序列。存在某个 $C>0$ 和 $\rho\in$ （0，1），使得对所有 $k=1$，2，.....，

$$\operatorname{dist}(\boldsymbol{w}^k,\mathcal{S})\leq C\rho^k.$$

注 3。加权 ℓ_1 惩罚项是加权弹性网惩罚项的一种特殊情形（Zou 和 Zhang （2009年），Ho、Sun 和 Xin （2015年）），当二次正则化参数为零时即可得到该情形。算法 1 可直接扩展，以适用于采用加权弹性网惩罚项的 PQR。在这一推广下，命题 1 和定理 2 仍然成立。详情见附录 B。

6. 扩展至非凸惩罚

算法 1 可无缝推广至采用一般折叠凹惩罚的特征拆分 QR $p_\lambda(|\beta|)$ $\beta\in\mathbb{R}$ (Fan, Xue and Zou (2014))。一般的折叠凹惩罚通常比标准 LASSO（ BrehenyandHuang (2011）、Fan, Xue andZou(2014) 能得到更准确的系数估计，因为它们避免了对较大系数的过度惩罚。这类惩罚项的另一项优势是，零并非“吸收态”（FanandLv (2008年）：在中间步骤被压至零的特征，可能会在后续迭代中重新进入活跃集。此类惩罚项中的典型例子包括 SCAD 和 MCP。导数 $p'_\lambda(|\beta|)$ 会根据惩罚项的选择而具有不同的形式。对于参数为 $a>2$ 的 SCAD，我们有

$$p'_\lambda(|\beta|)=\begin{cases}\lambda, & \text{if } |\beta|\leq\lambda,\\ \frac{a\lambda-|\beta|}{a-1}, & \text{if } \lambda<|\beta|<a\lambda,\\ 0, & \text{if } |\beta|\geq a\lambda,\end{cases}\tag{16}$$

对于参数为 $a>1$ 的 MCP 惩罚项，我们有

$$p'_\lambda(|\beta|)=\begin{cases}\lambda-\frac{|\beta|}{a}, & \text{if } |\beta|\leq a\lambda,\\ 0, & \text{if } |\beta|>a\lambda.\end{cases}\tag{17}$$

相应地，我们有 $\Omega(\boldsymbol{\beta}) = \sum_{i=1}^{p} p_\lambda(|\beta_i|)$，这使得目标函数 (1) 成为非凸函数。虽然这意味着可能存在多个局部极小点，但我们可以采用局部线性近似（Zou and Li（2008），Wang, Wu and Li（2012））来求解该优化问题。局部线性近似通过局部线性函数迭代地逼近凹惩罚，从而有效地将非凸问题转化为一系列带加权 ℓ_1 惩罚项的子问题，这些子问题可通过算法 1 直接求解。已有研究表明，在稀疏分位数回归情形下，满足温和条件时，以 LASSO 为初始值的 LLA 估计量仅需两次迭代，就能以极高概率收敛到 Oracle 解（见 Fan, Xue and Zou（2014）中的推论 8）。与这一理论结果一致，Wen 等人（2025）在数值实验中报告，即使仅进行一步局部线性近似，也能达到有竞争力的精度。我们在算法 2 中总结了用于 SCAD 或 MCP 惩罚项的 L 步 FS-QRPPA，其中 $\boldsymbol{\beta}$、$\boldsymbol{z}$ 和 $\boldsymbol{\theta}$ 的所有记号变体均遵循 (14)、(15) 和 (12) 中的相应更新。

算法 2： 采用 SCAD 或 MCP 惩罚的 FS-QRPPA

输入： $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_G)$, $\boldsymbol{y}$, $\boldsymbol{\beta}^0 = (\boldsymbol{\beta}_1^{0\top}, \ldots, \boldsymbol{\beta}_G^{0\top})^\top$, $\boldsymbol{z}^0, \boldsymbol{\theta}^0$, λ, μ, $\eta_g, g = 1, \ldots, G$

对于 $g = 1$ **到** G **执行**
  $\boldsymbol{\lambda}_g \leftarrow \lambda \circ \mathbf{1}$，其中 $\boldsymbol{\lambda}_g \in \mathbb{R}^{p_g}$;
结束
$\tilde{\boldsymbol{\beta}}^0$, $\tilde{\boldsymbol{z}}^0$, $\tilde{\boldsymbol{\theta}}^0 \leftarrow$ FS-QRPPA $\left(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\beta}^0, \boldsymbol{z}^0, \boldsymbol{\theta}^0, \mu, \{\boldsymbol{\lambda}_g\}_{g=1}^G, \{\eta_g\}_{g=1}^G\right)$;

对 $l = 1$ **到** L **执行**
  对 $g = 1$ **到** G **执行**
    $\tilde{\boldsymbol{\lambda}}_g^l \leftarrow \left(p'_\lambda\left(|\tilde{\boldsymbol{\beta}}_{g,1}^{l-1}|\right), \ldots, p'_\lambda\left(|\tilde{\boldsymbol{\beta}}_{g,p_g}^{l-1}|\right)\right)^\top$ 其中 $p'_\lambda(\cdot)$ 对应于 (16)，用于 SCAD 惩罚项或 (17)，用于 MCP 惩罚;
  end
  $\tilde{\boldsymbol{\beta}}^l, \tilde{\boldsymbol{z}}^l, \tilde{\boldsymbol{\theta}}^l \leftarrow$ FS-QRPPA$(\boldsymbol{X}, \boldsymbol{y}, \tilde{\boldsymbol{\beta}}^{l-1}, \tilde{\boldsymbol{z}}^{l-1}, \tilde{\boldsymbol{\theta}}^{l-1}, \mu, \{\tilde{\boldsymbol{\lambda}}_g^l\}_{g=1}^G, \{\eta_g\}_{g=\ \ 1}^G)$;
结束
输出： $\tilde{\boldsymbol{\beta}}^L$

7. 数值研究

为使所提出的 FS-QRPPA 框架能够应用于超高维问题，我们开发了开源的 R 软件包 `fsQRPPA`。该软件包为 LASSO（算法 1）和非凸（算法 2）惩罚变异均实现了多核并行化。此实现充分利用了 FS-QRPPA 的并行结构，将矩阵—向量乘积等计算密集型操作 $\boldsymbol{X}\boldsymbol{\beta}$ 和 $\boldsymbol{X}^\top\boldsymbol{\theta}$，以及针对 $\boldsymbol{\beta}$的分块近端更新，分配到G 并行处理块中。本节通过实证研究验证 FS-QRPPA 的估计与预测性能，以及其可扩展性和计算效率。我们首先在稀疏和稠密设定下进行模拟研究。此外，我们将 `fsQRPPA` 应用于来自英国生物样本库的超高维基因组数据。该应用展示了

在生物样本库规模上进行分位数回归的可行性，并凸显了 fsQRPPA在捕捉异质性关联方面的作用。在这些数值研究中，我们将 fsQRPPA 与两种用于高维分位数回归的先进 R 软件包进行比较： conquer (Tan、Wang 和 Zhou (2022年),He 等人 (2023a), He 等人 (2023b), Man 等人 (2024年))以及 rqPen (Sherwood、Li 和 Maidman (2025))。请注意，由于没有公开的实现，本文不将第 3 节讨论的 FS-QRADMM 算法纳入比较。

conquer，使用 RcppArmadillo 实现，基于对钉球损失函数进行卷积型核平滑，从而能够利用平滑损失的梯度加速。它采用局部自适应的主化–最小化方案，与其他分位数回归算法相比，速度显著提升，例如迭代坐标下降法（ Peng 和 Wang (2015年)）以及两块ADMM（Gu 等人 (2018)）。默认情况下，conquer使用高斯核。

rqPen 将 hqreg (Yi 和 Huang (2017),Yi (2024年)) 用作加权 ℓ_1 惩罚分位数回归的后端求解器。 RcppArmadillo 也由其提供支持，hqreg通过 Huber近似对钉球损失进行平滑，并使用高效的半光滑牛顿坐标下降算法求解由此得到的优化问题。它还采用自适应强规则筛选加权ℓ_1 惩罚分位数回归中的非活跃特征，从而能够沿惩罚路径显著加快（超）高维设定下的计算 $\{\lambda_t\}_{t=1}^{T}$。 rqPen 还通过局部线性近似 hqreg扩展到非凸惩罚（SCAD 和 MCP）。

对于所有算法，我们在拟合模型时使用标准化设计矩阵，并将估计系数转换回原始尺度。由于 conquer 和 rqPen 均被硬编码为对带凹惩罚项的分位数回归使用单步局部线性近似（LLA），为确保比较公平，我们对 FS-QRPPA 也采用了相同策略。具体而言，我们对 SCAD 固定 $a = 3.7$，对 MCP 固定 $a = 3$。

我们采用数据驱动策略来调整超参数 λ，并根据问题的稀疏结构调整选择准则。在稀疏情形下，我们采用 Peng 和 Wang(2015年) 提出的适用于分位数回归的高维贝叶斯信息准则（HBIC），即：

$$\mathrm{HBIC}(\lambda) = \log\left(\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})\right) + |\mathcal{A}| \frac{\log\log n}{n} C_n, \tag{18}$$

其中 $|\mathcal{A}|$ 表示活跃集的基数，而 $C_n > 0$ 是一个随 n 增大而趋于无穷大的常数序列。按照Peng 和 Wang (2015年的建议，我们设 $C_n = \log p$。相反，在具有大量非零但信号较弱特征的稠密情形下，HBIC 往往会选出过于稀疏且次优的模型。为此，我们改为在独立验证集上评估预测针式损失，并选择使验证损失最小的 λ 值。候选 λ 值的序列通过以下方法生成：

由Belloni 和 Chernozhukov (2011年提出的枢轴量方法（详见附录F）。对于FS-QRPPA，我们采用了热启动策略（ Friedman 等人 (2007 年），Friedman、Hastie 和 Tibshirani(2010年），其中前一个 λ 的解作为当前 λ 的初始值（详见附录F）。

所有模拟均在配备 M4 Max CPU 和 64 GB 内存的 MacBook Pro 上进行。真实数据分析的计算则在配备 48 个 CPU核心和 256 GB 内存的高性能计算集群节点上完成。在数值研究中，我们为模拟设置了 $G = 5$ ，为真实数据分析设置了 $G = 48$ （这反映了真实数据应用中特征数量更多）。

7.1. 稀疏设定下的模拟

我们采用了与 Peng 和 Wang (2015年)、Fan、Lin 和 Yin (2021) 以及 Wen 等人 (2025) 相似的模拟设定，其中数据来自一个底层异方差回归模型：

$$Y = X_6 - X_{20} + X_{50} - X_{100} + X_{500} - X_{1000} + X_p + 0.7X_1\epsilon, \tag{19}$$

其中 $\epsilon \sim \mathcal{N}$（0，1）。设计矩阵 $\boldsymbol{X} = [\mathbf{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p]$ 按以下步骤生成：

~ ~ ~1. 从多元正态分布
$\mathcal{N}$ 0 $\boldsymbol{\Sigma}$ $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ 1 $\boldsymbol{\Sigma}_{ij} =$ 中生成
$0.5^{|i-j|}$
$\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p]$（，），其中 是一个 AR（ ）协方差矩阵，使得 。~ ~2. 令
$\boldsymbol{X}_1 = \Phi$（$\boldsymbol{X}_1$）且 $\boldsymbol{X}_j = \boldsymbol{X}_j$， $j = 2$，……，p，其中 Φ（·）是标准正态分布的累积概率函数。

我们考虑了三个不同的分位数： $\tau \in \{0.3, 0.5, 0.7\}$，并在三种截然不同的维度设定下进行分析：（$n$， p）=(1000，2000）、（2000，50000）和（30000，2000）。每种情形下，我们进行了100次独立重复试验。根据（19）， τ的第 Y 分位条件分位数为

$$q_\tau(\boldsymbol{x}) = X_6 - X_{20} + X_{50} - X_{100} + X_{500} - X_{1000} + X_p + 0.7X_1\Phi^{-1}(\tau). \tag{20}$$

由此可知，当 $\tau = 0.5$时， X_1 不会影响 Y 的条件分位数；但当 $\tau = 0.3$ 或0.7时，分别会产生负向或正向影响。

我们根据若干指标评估了上述算法的特征选择表现和系数估计精度：

1. 系数估计平均绝对误差（AE）：在 100 次重复试验中，估计系数向量与其真实对应向量之间 ℓ_1 距离的平均值和标准差： $\sum_{j=1}^{p} |\hat{}\beta_j - \beta_j^*|$，其中 β^* 为真实系数。注意， $\beta_1^* = 0.7\Phi^{-1}$（τ）。

2. P_1：在 100 次重复试验中，除 β_1之外的所有活跃系数均被选择的拟合模型所占的比例。对于 τ的三种选择，该比例预计都接近 1。3. P_2：在 100 次重复试验中，选择了 β_1 的拟合模型所占的比例。当 $\tau = 0.5$ 时，该比例预计接近 0，否则接近 1。4. 模型规模（规模）：在 100 次重复试验中，估计的活跃系数数量（不包括截距）的平均值。5. 时间：沿路径$\{\lambda_t\}_{t=1}^{50}$运行的每种算法的运行时间（以秒为单位），取 100 次重复试验的平均值。

模拟结果见表1 - 3，AE、规模和时间的标准差列于括号中。我们观察到，在系数估计精度和 P_2方面，带有MCP惩罚项的PQR在所有算法中始终优于LASSO和SCAD变异。此外，对于三种惩罚项，FS-QRPPA在所有竞争方法中均实现了最低的AE。在计算速度方面，FS-QRPPA的表现与`conquer` 和`rqPen` 相当；在 $n = 2000$ 和 $p = 50000$的设置下，特征拆分的优势则更加明显。值得注意的是，与不进行特征拆分的标准两块ADMM（Gu 等人(2018)）相比，FS-QRPPA的速度显著更快（见附录G）。

除了特征选择和系数估计精度之外，分位数回归的另一项重要应用是为响应变量构建预测区间。因此，我们评估了不同算法生成的预测区间的经验表现。对于三个模拟设置中的每一种，我们从相同的数据生成过程中额外生成了100个大小为 $0.5n \times p$的测试集，并将上一小节所述的100次重复试验用作训练集。随后，我们采用与特征选择和估计研究相同的流程：在分位数水平 $\tau = 0.1$ 和 $\tau = 0.9$ 下，沿着由上述枢轴方法构造的惩罚参数序列 $\{\lambda_t\}_{t=1}^{T}$，拟合带有LASSO、SCAD和MCP惩罚项的PQR，并通过HBIC选择最优惩罚参数。

对于每个测试观测 $\boldsymbol{x}_i$， Y_i的预测 τ 阶分位数为 $\hat{q}_\tau(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}(\tau)$，其中 $\hat{\boldsymbol{\beta}}(\tau)$ 是在训练集上拟合得到的，惩罚参数由 HBIC 选定。对于每种算法，我们构造预测区间$[\hat{q}_{0.1}(\boldsymbol{x}_i),\ \hat{q}_{0.9}(\boldsymbol{x}_i)]$。根据以下三个指标对这些区间进行评估：

1. 经验覆盖率：落在区间内的 y_i 所占的比例。特别是，为了确保不交叉性质，如果对于某个 i，我们有 $\hat{q}_{0.1}(\boldsymbol{x}_i) > \hat{q}_{0.9}(\boldsymbol{x}_i)$，则将其视为覆盖失败。理想情况下，该比例应接近 0.80。2. 下尾部覆盖不足：低于 $\hat{q}_{0.1}(\boldsymbol{x}_i)$ 的 y_i 所占的比例。理想情况下，该比例应接近 0.10。3. 上尾部覆盖不足：高于 $\hat{q}_{0.9}(\boldsymbol{x}_i)$ 的 y_i 所占的比例。理想情况下，该比例应接近 0.10。

TABLE 1 *Comparisons of algorithms for PQR when $n = 1000$, $p = 2000$. AE: the average and standard deviation of the L_1 distance from the estimated coefficient vector to its true counterpart; P_2: The proportion of fitted models where β_1 is selected over 100 replicates. It is expected to be close to 0; Size: the average number of estimated active coefficients excluding the intercept over 100 replicates; Time: The running time (in seconds) of each algorithm along the path $\{\lambda_t\}_{t=1}^{50}$ averaged over 100 replicates.*[*]

算法	τ	AE	P2	Size	Time
FS-QRPPA (LASSO)	0.3	0.254 (0.083)	0.98	8.42 (0.64)	0.41 (0.08)
	0.5	0.101 (0.034)	0	7.18 (0.46)	0.35 (0.01)
	0.7	0.238 (0.088)	0.98	8.45 (0.73)	0.41 (0.06)
conquer (LASSO)	0.3	0.523 (0.075)	0.98	8.61 (0.71)	0.20 (0.01)
	0.5	0.265 (0.036)	0	7.35 (0.58)	0.17 (0.01)
	0.7	0.498 (0.076)	0.99	8.62 (0.79)	0.20 (0.01)
rqPen (LASSO)	0.3	0.525 (0.090)	1	8.94 (1.08)	0.25 (0.06)
	0.5	0.259 (0.043)	0	7.61 (0.76)	0.23 (0.05)
	0.7	0.492 (0.098)	0.99	8.99 (0.96)	0.24 (0.04)
FS-QRPPA (SCAD)	0.3	0.110 (0.042)	1	8.00 (0.00)	0.46 (0.04)
	0.5	0.039 (0.012)	0	7.00 (0.00)	0.36 (0.02)
	0.7	0.110 (0.047)	1	8.00 (0.00)	0.46 (0.05)
conquer (SCAD)	0.3	0.231 (0.073)	0.97	8.01 (0.27)	0.23 (0.04)
	0.5	0.052 (0.015)	0	7.00 (0.00)	0.38 (0.05)
	0.7	0.232 (0.090)	0.94	7.97 (0.30)	0.23 (0.04)
rqPen (SCAD)	0.3	0.188 (0.053)	1	8.01 (0.10)	1.21 (0.12)
	0.5	0.037 (0.012)	0	7.00 (0.00)	1.15 (0.09)
	0.7	0.181 (0.061)	1	8.00 (0.00)	1.24 (0.12)
FS-QRPPA (MCP)	0.3	0.098 (0.054)	1	8.01 (0.10)	0.57 (0.06)
	0.5	0.039 (0.012)	0	7.00 (0.00)	0.42 (0.03)
	0.7	0.108 (0.063)	1	8.00 (0.00)	0.56 (0.04)
conquer (MCP)	0.3	0.176 (0.066)	0.99	8.01 (0.17)	0.23 (0.05)
	0.5	0.052 (0.015)	0	7.00 (0.00)	0.38 (0.05)
	0.7	0.174 (0.074)	0.99	7.99 (0.10)	0.23 (0.04)
rqPen (MCP)	0.3	0.157 (0.058)	1	8.01 (0.10)	1.27 (0.12)
	0.5	0.037 (0.012)	0	7.00 (0.00)	1.26 (0.09)
	0.7	0.149 (0.064)	1	8.00 (0.00)	1.23 (0.09)

[*]我们不报告 P_1，因为所有算法都达到 $P_1 = 1$。

三种（n, p）设定下的结果汇总于图1，相应的测试针状损失见附录G中的表A1至A3。在各种设定下，采用三种惩罚项（LASSO、SCAD、MCP）的FS-QRPPA，其经验覆盖率中位数接近 0.80 的名义水平，下尾部和上尾部误覆盖率的中位数也接近 0.10。相比之下，`conquer` 和`rqPen` 呈现出过度覆盖的倾向，尤其是在使用非凸惩罚项时：经验覆盖率超过 0.80，而两端的误覆盖率均明显低于 0.10。一种可能的解释是，对于这两种方法，平滑诱导偏差在极端分位数水平（0.1 和 0.9）下更为显著。随着 n 增加，所有算法的覆盖率都更加稳定（重复试验间的离散程度更小）。在每种设定下，三种方法的离散程度

TABLE 2
Comparisons of algorithms for PQR when $n = 2000$, $p = 50000$. AE: the average and standard deviation of the L_1 distance from the estimated coefficient vector to its true counterpart; P_2: The proportion of fitted models where β_1 is selected over 100 replicates. It is expected to be close to 0; Size: the average number of estimated active coefficients excluding the intercept over 100 replicates; Time: The running time (in seconds) of each algorithm along the path $\{\lambda_t\}_{t=1}^{50}$ averaged over 100 replicates. *

算法	τ	AE	P2	Size	Time
FS-QRPPA (LASSO)	0.3	0.219 (0.060)	1	8.34 (0.50)	14.61 (0.50)
	0.5	0.105 (0.032)	0	7.13 (0.37)	12.25 (0.31)
	0.7	0.218 (0.055)	1	8.29 (0.50)	14.92 (0.56)
conquer (LASSO)	0.3	0.433 (0.048)	1	8.28 (0.49)	29.00 (0.69)
	0.5	0.218 (0.023)	0	7.14 (0.38)	25.24 (0.58)
	0.7	0.422 (0.049)	1	8.38 (0.58)	28.79 (0.70)
rqPen (LASSO)	0.3	0.411 (0.058)	1	8.50 (0.72)	13.29 (1.09)
	0.5	0.212 (0.031)	0	7.16 (0.39)	15.53 (1.37)
	0.7	0.408 (0.054)	1	8.38 (0.57)	13.06 (1.07)
FS-QRPPA (SCAD)	0.3	0.075 (0.033)	1	8.00 (0.00)	15.05 (0.47)
	0.5	0.029 (0.008)	0	7.00 (0.00)	12.98 (0.35)
	0.7	0.074 (0.029)	1	8.00 (0.00)	15.23 (0.52)
conquer (SCAD)	0.3	0.164 (0.045)	1	8.02 (0.14)	29.66 (0.86)
	0.5	0.035 (0.010)	0	7.00 (0.00)	41.35 (2.54)
	0.7	0.157 (0.049)	1	8.02 (0.14)	29.66 (0.99)
rqPen (SCAD)	0.3	0.139 (0.037)	1	8.01 (0.10)	68.90 (2.98)
	0.5	0.024 (0.008)	0	7.00 (0.00)	71.70 (2.16)
	0.7	0.128 (0.039)	1	8.00 (0.00)	68.59 (2.74)
FS-QRPPA (MCP)	0.3	0.055 (0.024)	1	8.00 (0.00)	16.83 (0.52)
	0.5	0.029 (0.008)	0	7.00 (0.00)	14.45 (0.36)
	0.7	0.053 (0.022)	1	8.00 (0.00)	17.02 (0.53)
conquer (MCP)	0.3	0.121 (0.040)	1	8.01 (0.10)	29.74 (0.89)
	0.5	0.035 (0.010)	0	7.00 (0.00)	40.44 (2.78)
	0.7	0.113 (0.042)	1	8.00 (0.00)	29.71 (0.90)
rqPen (MCP)	0.3	0.105 (0.035)	1	8.01 (0.10)	68.77 (2.86)
	0.5	0.024 (0.008)	0	7.00 (0.00)	73.19 (2.41)
	0.7	0.095 (0.036)	1	8.00 (0.00)	68.31 (2.61)

*我们不报告 P_1，因为所有算法均达到 $P_1 = 1$。

不同方法的指标总体上具有可比性。

7.2. 稠密设定下的模拟

我们还考虑了一个稠密设定下的模拟情景，这种设定在遗传学研究中很常见：许多复杂性状（例如身高）具有高度多基因性，即大量变异与目标表型的关联较弱。由于在此类情形下，可靠的变量选择和系数估计本质上是难以实现的，我们的主要目标是评估特定条件分位数处的预测准确性。

TABLE 3 *Comparisons of algorithms for PQR when $n = 30000$, $p = 20000$. AE: the average and standard deviation of the L_1 distance from the estimated coefficient vector to its true counterpart; P_2: The proportion of fitted models where β_1 is selected over 100 replicates. It is expected to be close to 0; Size: the average number of estimated active coefficients excluding the intercept over 100 replicates; Time: The running time (in seconds) of each algorithm along the path $\{\lambda_t\}_{t=1}^{50}$ averaged over 100 replicates.* [*]

算法	τ	AE	P2	Size	Time
FS-QRPPA (LASSO)	0.3	0.046 (0.011)	1	8.20 (0.45)	20.08 (0.65)
	0.5	0.021 (0.004)	0	7.08 (0.27)	17.56 (0.69)
	0.7	0.047 (0.011)	1	8.22 (0.48)	20.00 (0.81)
conquer (LASSO)	0.3	0.084 (0.011)	1	8.29 (0.59)	7.50 (0.26)
	0.5	0.036 (0.004)	0.01	7.16 (0.39)	7.63 (0.18)
	0.7	0.084 (0.010)	1	8.25 (0.48)	7.38 (0.18)
rqPen (LASSO)	0.3	0.058 (0.011)	1	8.23 (0.47)	4.20 (0.08)
	0.5	0.029 (0.004)	0	7.10 (0.30)	4.15 (0.07)
	0.7	0.058 (0.011)	1	8.22 (0.44)	4.26 (0.10)
FS-QRPPA (SCAD)	0.3	0.011 (0.005)	1	8.00 (0.00)	20.06 (0.59)
	0.5	0.004 (0.001)	0	7.00 (0.00)	18.32 (0.66)
	0.7	0.010 (0.005)	1	8.00 (0.00)	20.12 (0.76)
conquer (SCAD)	0.3	0.033 (0.009)	1	8.00 (0.00)	15.56 (1.52)
	0.5	0.008 (0.002)	0	7.00 (0.00)	19.29 (1.90)
	0.7	0.033 (0.009)	1	8.00 (0.00)	15.83 (1.43)
rqPen (SCAD)	0.3	0.013 (0.005)	1	8.00 (0.00)	26.92 (1.20)
	0.5	0.009 (0.003)	0	7.00 (0.00)	26.69 (1.23)
	0.7	0.012 (0.005)	1	8.00 (0.00)	27.43 (1.39)
FS-QRPPA (MCP)	0.3	0.011 (0.005)	1	8.00 (0.00)	21.03 (0.85)
	0.5	0.004 (0.001)	0	7.00 (0.00)	18.86 (0.67)
	0.7	0.010 (0.005)	1	8.00 (0.00)	21.21 (1.06)
conquer (MCP)	0.3	0.033 (0.009)	1	8.00 (0.00)	15.37 (1.37)
	0.5	0.008 (0.002)	0	7.00 (0.00)	18.81 (1.85)
	0.7	0.033 (0.009)	1	8.00 (0.00)	15.55 (1.39)
rqPen (MCP)	0.3	0.013 (0.005)	1	8.00 (0.00)	26.43 (0.82)
	0.5	0.009 (0.003)	0	7.00 (0.00)	26.32 (1.13)
	0.7	0.012 (0.005)	1	8.00 (0.00)	26.87 (1.07)

[*]由于所有算法均达到 $P_1 = 1$，因此我们不报告 P_1。

为此，我们将数据生成机制调整为类似于 Wang 等人 (2025a). 具体而言，我们考虑以下设定： $n = 10000$和 $p = 20000$。响应变量由具有随机系数的异方差回归模型生成：

$$Y = \sum_{j=1}^{m_1} a_j X_j + \left(\sum_{j=m_1+1}^{m_1+m_2} |b_j| X_j \right) \epsilon, \tag{21}$$

其中 $a_j \sim \mathcal{N}(0, \sigma_1^2/m_1)$ 对于 $j = 1, \ldots, m_1$, $b_j \sim \mathcal{N}(0, \sigma_2^2/m_2)$ 对于 $j = m_1 + 1, \ldots, m_1 + m_2$, 且 $\epsilon \sim \mathcal{N}(0, 1)$。我们按照 $\sigma_1^2 = 0.3$和 $\sigma_2^2 = 0.1$, 参照 Wang 等人 (2025a), 并固定 $m_1 = 500$ 和 $m_2 = 50$。设计矩阵的生成方式

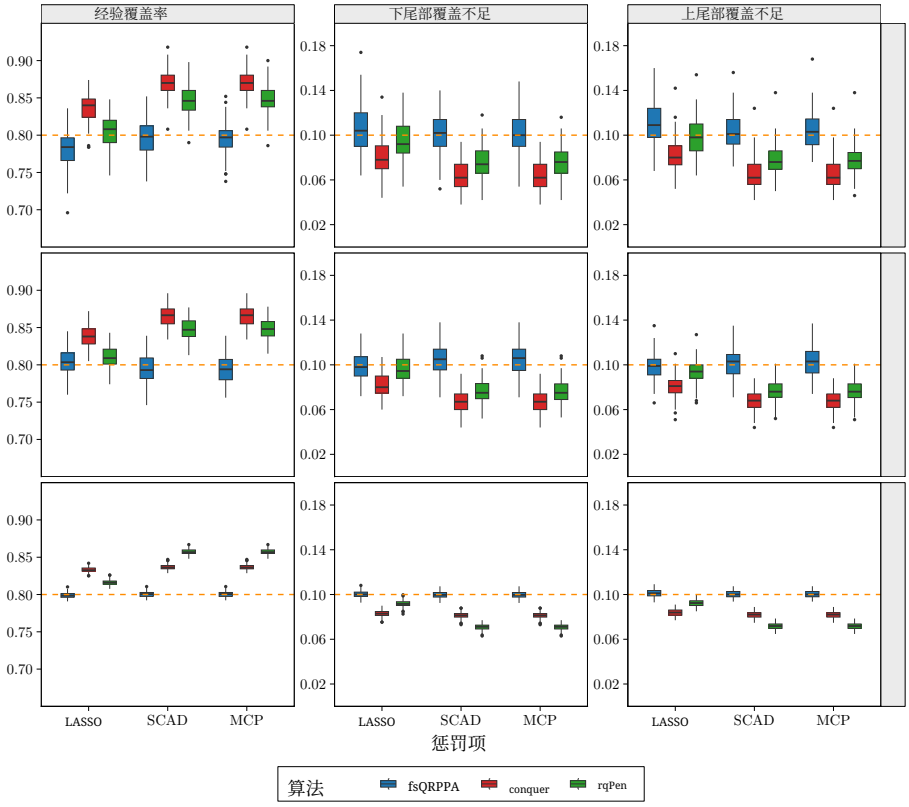

图 1：比较稀疏设定下，采用三种不同（n， p）以及不同惩罚项（LASSO、SCAD、MCP）时，预测分位数区间（$\tau = 0.1$–0.9）的经验测试集覆盖率和尾部比例。经验覆盖率：落在区间内的 y_i 所占比例；下尾部覆盖不足：低于$\hat{q}_{0.1}(\boldsymbol{x}_i)$的 y_i 所占比例；上尾部覆盖不足：高于$\hat{q}_{0.9}(\boldsymbol{x}_i)$的 y_i所占比例。

按照稀疏设定中的相同步骤进行，但第二步中例外的是，我们令 $\boldsymbol{X}_j = \Phi(\boldsymbol{X}_j)$ 对于 $j = m_1 + 1, \ldots, m_1 + m_2$。

在此，我们重点评估所考虑算法生成的预测区间的经验表现。100 次独立重复试验的结果汇总于图 2中，相应的测试针式损失见附录G中的表 A4。对于所有惩罚项，这三种算法得到的预测区间均略有覆盖不足。这一现象可能是由正则化偏差，以及高维分位数回归模型准确捕捉众多弱效应并将其与真正的零效应区分开的能力有限所致。这一局限与多基因风险评分文献中的发现相呼应

（Zhao and Zou (2022年)），该研究表明，随着因果信号数量相对于样本量的增加，将弱因果变异与零效应（噪声）变异区分开来会变得愈发困难。

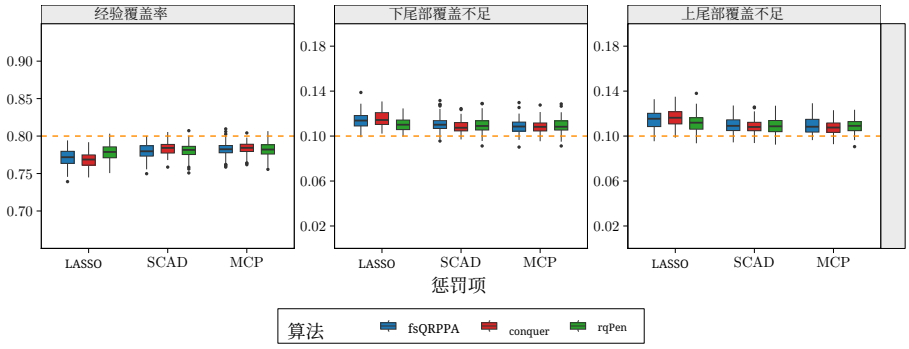

图2：在稠密设定下，使用不同惩罚项（LASSO、SCAD、MCP）时，预测分位数区间（$\tau = 0.1$–0.9）的经验测试集覆盖率和尾部比例比较。经验覆盖率：落在区间内的 y_i 所占的比例；下尾部覆盖不足：低于$\hat{}q_{0.1}(\boldsymbol{x}_i)$的 y_i 所占的比例；上尾部覆盖不足：高于$\hat{}q_{0.9}(\boldsymbol{x}_i)$的 y_i 所占的比例。

7.3. 英国生物样本库数据的应用

这里我们回到最初促使我们开展研究的应用：在英国生物样本库中进行全基因组关联研究。我们聚焦于具有欧洲血统的个体，以及两种数量性状：身高和脂蛋白(a)。我们分析了 39699 名具有欧洲血统的个体中 737401 个遗传变异（即单核苷酸多态性（SNPs））的数据。应用标准质量控制程序（Mbatchou 和 Marchini (2021)）筛选变异和样本后，再移除线性相关的变异（在遗传学中称为连锁不平衡或连锁不平衡剪枝，即 s.t. $r^2 < 0.6$），最终数据集包含 482502 个 SNPs 和 39699 名个体。最终数据集中缺失的基因型通过基于均值的插补进行填补。在这一（超）高维情境下，`conquer` 和`rqPen` 均未能在集群上完成计算，处理如此规模的矩阵时均以错误告终。因此，在这一情境下，我们仅报告 FS-QRPPA 的结果。

身高数据中没有缺失表型。我们将数据分别划分为包含30000、5000和4699个体的训练、验证和测试子集。对于脂蛋白(a)，数据集包含30560个体，我们将其分别划分为规模为25000、3000和2560的训练、验证和测试子集。在控制性别和年龄的情况下，我们在相应的训练集上，针对分位数水平 $\tau = 0.1$和 $\tau = 0.9$，分别采用LASSO、SCAD和MCP惩罚项拟合了FS-QRPPA。与模拟中一样，我们考察了经验覆盖率、

下尾部和上尾部覆盖不足率。此外，我们报告了不同模型所选择的单核苷酸多态性数量。这些结果总结于表4中，相应的测试针状损失报告于附录G中的表A5和A6。

TABLE 4 *Applications to height and Lipoprotein(a). Comparison of empirical test-set coverage and tail proportions for predicted quantile bands ($\tau = 0.1$–0.9) and number of selected SNPs for, FS-QRPPA prediction intervals with different penalties; Empirical coverage rate: the proportion of y_i that fall within the interval; Below-0.1 rate: the proportion of y_i below $\hat{q}_{0.1}(\boldsymbol{x}_i)$; Above-0.9 rate: the proportion of y_i above $\hat{q}_{0.9}(\boldsymbol{x}_i)$。*

性状	惩罚项	经验覆盖率	下尾部覆盖不足	上尾部覆盖不足	#单核苷酸多态性 ($\tau = 0.1$)	#单核苷酸多态性 ($\tau = 0.9$)
身高	LASSO	0.769	0.119	0.112	21689	18213
	SCAD	0.751	0.104	0.145	16086	16076
	MCP	0.724	0.142	0.134	15076	12598
脂蛋白(a)	LASSO	0.793	0.095	0.112	16926	16692
	SCAD	0.832	0.101	0.068	15077	8856
	MCP	0.771	0.097	0.132	10479	10720

根据表 4，所有模型的经验覆盖率都略微偏离名义水平 0。8，表明存在覆盖误差。对于这两种性状，所有模型都选出了约一万个非零单核苷酸多态性，这表明在这些真实数据应用中，预测依赖于大量微小遗传效应，而非极为稀疏的模型，这与这些性状已知的多基因特性一致。

FS-QRPPA 揭示了不同分位数间异质性的遗传效应。分位数回归的一项关键优势是能够揭示表型条件分布中异质性的特征效应。为可靠地检测这种异质性，我们重点关注在分布两端均保持活跃的遗传变异，具体而言是在分位数水平 $\tau = 0.1$ 和 $\tau = 0.9$处。在所有惩罚项（LASSO、SCAD 和 MCP）下，我们筛选出 FS-QRPPA 在两个分位数处的估计系数均非零的单核苷酸多态性。筛选得到身高的 226 个候选单核苷酸多态性和脂蛋白(a)的 86 个。

使用这些变异后，我们通过`conquer`在训练集上拟合无惩罚分位数回归，并对性别和年龄进行调整，分位数网格范围为 0.1 至 0.9（步长为 0.05）。对于身高，我们发现 13 个单核苷酸多态性在 0.1 和 0.9 分位数处的系数符号相反，且对应的 95% 渐近置信区间至少有一个不包含零。类似地，我们在脂蛋白(a)中发现 14 个具有此类符号差异的单核苷酸多态性。为展示这种异质性，图3展示了每种性状中三个代表性单核苷酸多态性的估计系数及其在不同分位数处对应的渐近置信区间。其余单核苷酸多态性的图见附录G。

FS-QRPPA 可提供个体化预测区间。除了识别异质性效应之外，QR 的另一项主要优势在于，相较于传统线性回归模型，它能够提供更为现实的预测区间。具体而言，我们利用在训练数据上基于 LASSO、SCAD 和 MCP 惩罚项估计得到的系数向量

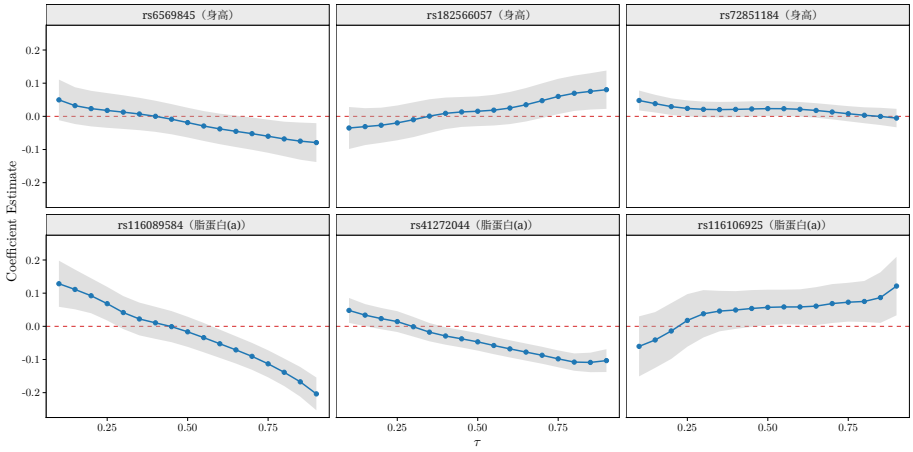

图3：FS-QRPPA 检测到的身高（上排）和脂蛋白(a)（下排）中选定单核苷酸多态性在不同分位数水平 τ 上的异质性效应。实线蓝线连接点估计值，红色虚线表示零，阴影区域表示 95% 渐近置信区间。

为留出测试数据构建了预测区间（$n = 4699$ 用于身高； $n = 2560$ 用于脂蛋白(a)），所使用的是 β 在训练数据上基于 LASSO、SCAD 和 MCP 惩罚项估计得到的系数向量。我们选取了三个具有代表性的子样本组，每组包含 100 个体：低组（对应性状值最小的 100 个个体）、中位数组（最接近中位数的 100 个个体）和高组（性状值最大的 100 个个体）。图4 展示了两种性状基于 LASSO 的预测区间。SCAD 和 MCP 的结果在定性上相似，详见附录G。标准线性模型通常仅能提供在个体间长度近乎恒定的对称预测区间，而 QR 预测区间则可能高度不对称，且长度因个体而异，从而反映潜在的生物异质性。这种个体化预测区间使我们能够区分两类人群：一类人的区间较窄，其性状主要由遗传因素解释；另一类人的区间较宽，非遗传因素或环境影响可能发挥更大作用。

8. 讨论

在本研究中，我们开发了一种特征分裂近端点算法（FS-QRPPA），用于求解（超）高维惩罚分位数回归，并提供了理论上的线性收敛保证。我们还在 R 软件包 fsQRPPA 中提供了并行实现，使惩罚分位数回归能够处理设计矩阵条目数达数十亿量级的数据集。通过利用按特征并行，

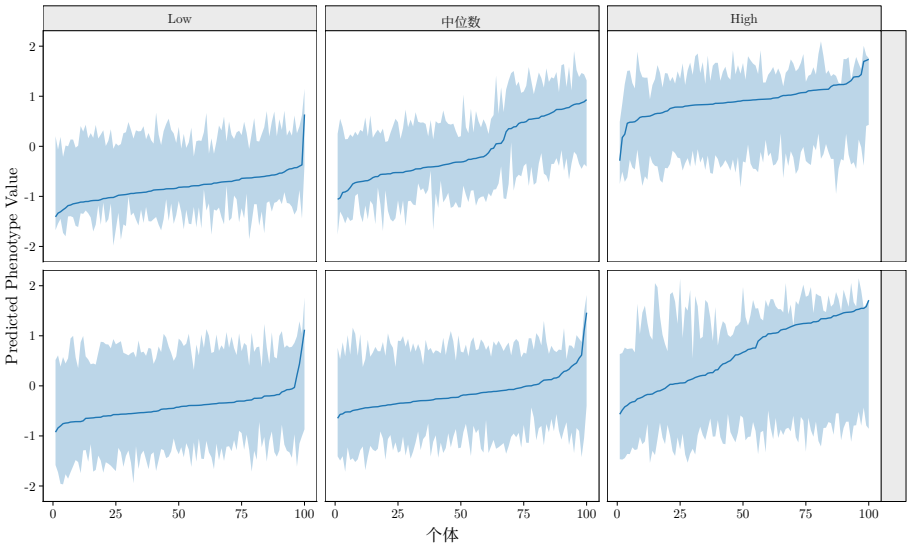

图4：使用 LASSO 惩罚估计的身高和脂蛋白(a)的90%预测区间及预测中位数，并按表型亚组（低、中位数和高）分层。蓝色实线表示预测中位数，浅蓝色阴影区域表示90%预测区间。每个面板中的个体均按其预测中位数排序。

FS-QRPPA 缓解了现代（超）高维设定下的内存和计算瓶颈。这显著拓宽了分位数回归在大规模问题中的适用性，例如全基因组关联研究中的基因组发现和基因组性状预测。

未来研究有几个值得探索的方向。具体而言，我们目前的并行实现针对共享内存环境下的单个计算节点进行了优化。这种设计已能支持涉及数万个个体的大规模全基因组关联研究分析。然而，对于生物样本库规模的完整分析（例如英国生物样本库，其中有$\sim$ 500K 个个体），单个节点仍可能不足。因此，开发 FS-QRPPA 的多节点实现具有重要的实际意义，它能够实现分布式内存扩展，并提升其在现代大规模分位数回归中的灵活性。除这一架构改进外，还可通过整合安全筛选技术进一步加速（Ghaoui、Viallon 和 Rabbani (2011年)，Ndiaye 等人(2017))，这些技术可确定性地剔除回归系数β 可保证为零的特征（即非活跃特征）。直观而言，此类筛选可显著降低计算成本，尤其是在稀疏情形下。将这些技术纳入 FS-QRPPA 方案，并为由此带来的计算效率提升提供严谨的理论依据，是未来研究的一个自然方向。除了减少

特征数量外，通过优化分块结构也能显著提升 FS-QRPPA 的收敛速度。例如，对设计矩阵的列进行策略性重排，以最小化每个块的最大特征值，可有效加快迭代更新。总而言之，这些策略共同构成了进一步改进 FS-QRPPA 的协同路线图。

此外，我们当前的特征分割框架仅适用于采用加权弹性网惩罚的 PQR。具体而言，我们关于线性收敛的理论保证依赖于该惩罚项的分段线性二次（PLQ）结构。一个值得探索的拓展方向是引入一般的组结构惩罚，例如组 LASSO（Yuan 和 Lin (2006年)）或稀疏组 LASSO（Simon 等人 (2013年)）。这类拓展对于基因组应用尤为重要，因为单核苷酸多态性在基因内天然呈现分组结构。事实上，近期统计学文献已证明组惩罚分位数回归在遗传学领域的实用价值（Mendez-Civieta、Aguilera-Morillo 和 Lillo (2021)、Ouhourane 等人 (2022年)）。然而，尽管这些惩罚项的组可分性保留了特征分割架构在计算上的可行性，它们却不具备 PLQ 性质。因此，要为这些情形建立收敛速率，就需要开展新的理论分析。

最后，在我们的模拟研究和真实数据分析中，我们观察到分位数回归预测区间可能存在覆盖误差，实际覆盖率偏离名义水平。这种现象并非偶然；统计学文献已对覆盖偏差的普遍性进行了严格研究（Romano、Patterson 和 Cand`es (2019年); Bai 等人; Gibbs、Cherian 和 Cand`es (2025)），近期研究也凸显了保形方法在遗传性状预测中的效用（Wang 等人 (2025b)）。重要的是，FS-QRPPA 的优化方案与这些校正策略兼容。未来一项颇具前景的工作方向是将保形预测整合到 FS-QRPPA 流程中，并将这一统一框架应用于大规模遗传数据。这种方法旨在提供准确的个性化预测和不确定性量化，从而为临床应用带来潜在益处。

致谢。 感谢 Ma-lgorzata Bogdan 就此主题提供的宝贵讨论。本研究使用英国生物样本库资源开展，申请编号为 41849。

资助。 本研究得到瑞典研究理事会的资助。

代码可用性。该R 软件包`fsQRPPA` 可从以下网址获取：https://anonymous.4open.science/r/fsQRPPA-6764，其中支持采用 LASSO、SCAD 和 MCP 惩罚项的 FS-QRPPA。

参考文献

Bai, Y., Mei, S., Wang, H. 和 Xiong, C. 理解不确定性估计中的覆盖偏差。载于第35届神经信息处理系统国际会议论文集。NIPS '21 18307–18319。Curran Associates Inc.Belloni, A.和 Chernozhukov, V. (2011)。高维稀疏模型中的 ℓ1惩罚分位数回归。《统计学年鉴》39.Boyd, S.(2010)。通过交替方向乘子法实现分布式优化与统计学习。《机器学习基础与趋势®》3 1–122。Breheny, P.和 Huang, J.(2011)。非凸惩罚回归的坐标下降算法及其在生物特征筛选中的应用。《应用统计年鉴》5.Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T.,Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J.,Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G.,Leslie, S., Allen, N., Donnelly, P. 和 Marchini, J.(2018)。英国生物样本库资源：深度表型分析与基因组数据。《自然》562 203–209。Chen, C., He, B., Ye, Y. 和 Yuan, X. (2016)。用于多块凸最小化问题的 ADMM 直接扩展未必收敛。《数学规划》155 57–79。Eckstein, J. 和 Bertsekas, D. P. (1992)。论最大单调算子的 Douglas—Rachford 分裂方法与近端点算法。《数学规划》55 293–318。Fan, J. (1997)。对 A. Antoniadis《统计中的小波：综述》的评论。《意大利统计学会期刊》6 131。Fan, Y., Lin, N.和 Yin, X. (2021)。采用松弛变量表示法对分布式大数据进行惩罚分位数回归。《计算与图形统计杂志》30 557–565。Fan, J.和 Lv, J. (2008)。超高维特征空间的确定性独立筛选。《皇家统计学会杂志B辑：统计方法学》70 849–911。Fan, J., Xue, L. 和 Zou, H. (2014)。折叠凹惩罚估计的强预言最优性。《统计学年鉴》42. Friedman, J., Hastie, T. 和 Tibshirani, R. (2010)。通过坐标下降求解广义线性模型的正则化路径。《统计软件杂志》33.Friedman, J., Hastie, T., Höfling, H. 和 Tibshirani, R. (2007)。逐路径坐标优化。《应用统计年鉴》1.Gabay, D. (1983)。乘子法在变分不等式中的应用。载于《数学及其应用研究》15 299–331。爱思唯尔。Ghaoui, L. E., Viallon, V.和 Rabbani, T. (2011)。LASSO 与稀疏监督学习问题的安全特征消除。

Gibbs, I., Cherian, J. J. 和 Candès, E. J. (2025). 分位数回归覆盖偏差的校正。Golub, G. H. 和 Van Loan, C. F. (2013).《矩阵计算》第四版。约翰斯·霍普金斯数学科学研究。约翰斯·霍普金斯大学出版社，巴尔的摩。Gu, Y., Fan, J., Kong, L., Ma, S. 和 Zou, H. (2018). 用于高维稀疏惩罚分位数回归的交替方向乘子法。《技术计量学》60 319–331。He, X., Pan, X., Tan, K. M.和 Zhou, W.-X. (2023a). 结合大规模推断的平滑分位数回归。《计量经济学杂志》232 367–388。He, X., Pan, X., Tan, K. M. 和 Zhou, W.-X. (2023b). conquer：卷积型平滑分位数回归 R 软件包，版本 1.3.3。Ho, M., Sun, Z. 和 Xin, J. (2015). 加权弹性网络惩罚的均值-方差投资组合设计与计算。《SIAM 金融数学杂志》6 1220–1244。Koenker, R. 和 Bassett, G. (1978). 回归分位数。《计量经济学》46 33。Liu, Y. 和 Zeng, P. (2024). 大数据中带线性约束的一般惩罚分位数回归扩展交替方向乘子法。《统计通讯：模拟与计算》53 4268–4289。Mackay, T. F. C. (2014). 上位性与数量性状：利用模式生物研究基因-基因相互作用。《自然综述：遗传学》15 22–33。Mackay, T. F. C.和 Anholt, R. R. H. (2024). 数量性状的多效性、上位性与遗传结构。《自然综述：遗传学》25 639–657。Man, R., Pan, X., Tan, K. M. 和 Zhou, W.-X. (2024). 惩罚卷积平滑分位数回归的统一算法。《计算与图形统计学杂志》33 625–637。Martinet, B. (1970). 简短通信。通过逐次逼近对变分方程进行正则化。《法国计算机科学与运筹学杂志》，红色系列 4 154–158。Mbatchou, J.和 Marchini, J. (2021). 英国生物样本库分析建议。https://rgcgithub.github.io/regenie/recommendations/.
McGaffin, M. G. 和 Fessler, J. A.(2015). 大规模逆问题中主化函数的算法设计。Mendez-Civieta, A., Aguilera-Morillo, M. C. 和 Lillo, R. E. (2021). 分位数回归中的自适应稀疏组 LASSO。《数据分析与分类进展》15 547–573。Mirzaeifard, R., Ghaderyan, D. 和 Werner, S. (2025). 用于分位数回归的非凸稀疏惩罚去中心化平滑交替方向乘子法。《IEEE 信号处理快报》32 1915–1919。Muckley, M. J., Noll, D. C.和 Fessler, J. A. (2015). 基于 B1、采用自适应重启的迭代软阈值算法（BARISTA）实现快速并行磁共振图像重建。《IEEE 医学影像汇刊》34 578–588。

Ndiaye, E., Fercoq, O., Gramfort, A. 和 Salmon, J. (2017). 稀疏性惩罚的间隙安全筛选规则。《机器学习研究杂志》18 1–33。Nocedal, J. 和 Wright, S. J.(2006).《数值优化》第二版。Springer运筹学与金融工程丛书。Springer，纽约，纽约州。Ortega, J. M. 和 Rheinboldt, W. C. (2000). 多变量非线性方程的迭代求解：最初出版于纽约：学术出版社，1970年。应用数学经典丛书 30。工业与应用数学学会（SIAM，费城市场街3600号6层，宾夕法尼亚州19104），费城，宾夕法尼亚州。Ouhourane, M., Yang, Y., Benedet, A. L.和 Oualkacha, K.(2022). 分组惩罚分位数回归。《统计方法与应用》31 495–529。Pazokitoroudi, A., Liu, Z., Dahl, A., Zaitlen, N., Rosset, S.和Sankararaman, S. (2024). 一种可扩展且稳健的方差组分方法揭示复杂性状背后基因-环境相互作用遗传结构的新见解。《美国人类遗传学杂志》111 1462–1480。Peng, B. 和 Wang, L. (2015). 一种用于高维非凸惩罚分位数回归的迭代坐标下降算法。《计算与图形统计学杂志》24 676–694。Polson, N. G., Scott, J. G.和 Willard, B. T.(2015). 统计学与机器学习中的近端算法。《统计科学》30。Rockafellar, R. T. (1976). 单调算子与近端点算法。《SIAM控制与优化杂志》14 877–898。Rockafellar, R. T. 和 Wets, R. J. B. (1998). 变分分析。数学科学基础丛书 317。Springer Berlin Heidelberg，柏林，海德堡。Romano, Y., Patterson, E. 和 Candès, E. J. (2019). 共形化分位数回归。载于：第33届神经信息处理系统国际会议论文集 318 3543–3553。Curran Associates Inc.，美国纽约州雷德胡克。Sherwood, B., Li, S. 和 Maidman, A.(2025). rqPen：惩罚分位数回归 R 软件包，版本 4.1.3。Simon, N., Friedman, J., Hastie, T. 和 Tibshirani, R. (2013). 稀疏组 LASSO。《计算与图形统计学杂志》22 231–245。Tan, K. M., Wang, L. 和 Zhou, W.-X. (2022). 高维分位数回归：卷积平滑与凹正则化。《皇家统计学会杂志B辑：统计方法学》84 205–233。Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T. 和 Posthuma, D. (2021). 全基因组关联研究。《自然综述：方法学基础》1 59。Wang, L., Wu, Y. 和 Li, R. (2012). 用于分析超高维异质性的分位数回归。《美国统计学会杂志》

协会 107 214–222。Wang, C., Wang, T., Kiryluk, K., Wei, Y., Aschard, H. 和 Ionita-Laza, I.（2024年）。利用生物样本库规模的分位数回归发现全基因组生物标志物。《自然·通讯》15 6460。Wang, F., Wang, C., Wang, T., Masala, M., Fiorillo, E., Devoto, M.,Cucca, F. 和 Ionita-Laza, I. （2025a）。Regenie.QRS：适用于生物样本库规模、计算高效的全基因组分位数回归。bioRxiv：生物学预印本服务器 2025.05.02.651730。Wang, C., Wang, F., Bogdan, M., Masala, M., Fiorillo, E., De-voto, M., Cucca, F., Belsky, D. 和 Ionita-Laza, I. （2025b）。利用保形化分位数回归对多基因风险评分进行个体化不确定性量化。Wen, J., Yang, S. 和 Zhao, D. （2024）。非凸 Dantzig 选择器及其并行计算算法。《统计与计算》34 180。Wen, J., Yang, S., Wang, C. D., Jiang, Y. 和 Li, R. （2025）。超高维分位数回归的特征拆分算法。《计量经济学杂志》249 105426。Wu, X., Jiang, J. 和 Zhang, Z. （2024）。适用于高维线性模型的分区无关并行交替方向乘子法算法。Wu, X., Guo, D., Liang, R. 和 Zhang, Z. （2025a）。结合高斯回代的并行交替方向乘子法算法，用于高维分位数回归与分类。《统计与计算》35 139。Wu, X., Chao, Y., Liang, R., Tang, S.和 Zhang, Z.（2025b）。用于 Dantzig 选择器的特征拆分并行算法。《统计与计算》35 116。Yi, C. （2024）。hqreg：用于 LASSO 或弹性网络惩罚 Huber 损失回归和分位数回归的正则化路径，R 软件包版本 1.4-1。Yi, C.和 Huang, J. （2017）。用于弹性网络惩罚 Huber 损失回归和分位数回归的半光滑牛顿坐标下降算法。《计算与图形统计学杂志》26 547–557。Yu, L.和 Lin, N.（2017）。大数据中的惩罚分位数回归交替方向乘子法。《国际统计评论》85 494–518。Yu, L., Lin, N. 和 Wang, L. （2017）。大规模非凸惩罚分位数回归的并行算法。《计算与图形统计学杂志》26 935–939。Yu, K., Lu, Z. 和 Stander, J. （2003）。分位数回归：应用与当前研究领域。《皇家统计学会杂志：D辑（统计学家）》52 331–350。Yuan, M. 和 Lin, Y. （2006）。含分组变量回归中的模型选择与估计。《皇家统计学会杂志B辑：统计方法学》68 49–67。Zhang, C.-H. （2010）。最小最大非凹惩罚下的近乎无偏变量选择。《统计学年鉴》38。Zhao, B.和 Zou, F.（2022）。复杂性状预测的多基因风险评分。《生物统计学》78 499–511。Zou, H. 和 Li, R. （2008）。非凹惩罚下的单步稀疏估计

似然模型。《统计学年鉴》**36**。Zou、 H. 和 Zhang、 H. H. （2009年）。关于参数数量发散时的自适应弹性网络。《统计学年鉴》**37**。