文章编号:1005-9679(2019)05-0097-07

求解大规模 SCAD 回归问题的随机坐标下降算法研究

赵　磊[1]　陈　玎[2]　朱道立[1,2]

(1.上海交通大学 安泰经济与管理学院,上海　200030;2.上海交通大学 中美物流研究院,上海　200030)

摘　要:　回归方法是重要的数据分析工具。带平滑削边绝对偏离(smoothly clipped absolute deviation, SCAD)正则项的回归问题,以其在处理高维数据中的近似无偏性(见 Fan 和 Li,2001),在大数据分析中得到广泛应用。但在大数据背景下,待求解的 SCAD 回归问题的数据量往往很大,而且分布在不同地理位置,这使得在 SCAD 回归问题的求解算法设计中,需要重新考虑计算的内存使用量。常规用于求解 SCAD 回归问题的优化算法(LQA、LLA、ADMM 等)往往需要在每一次迭代中更新全部变量,从而造成计算的内存需求很大,难以适应大数据的求解要求。随机坐标下降方法(stochastic coordinate descent, SCD)以其子问题运算内存需求小(见 Nesterov,2012)的优势,在大规模分布式最优化问题中得到了广泛的应用。但目前理论上 SCD 算法仅能处理带凸惩罚项的回归问题,由于 SCAD 回归问题中惩罚项的非凸非光滑性,现有的随机坐标下降方法难以处理这一问题。首先对 SCAD 回归问题模型进行分析,得出 SCAD 回归模型的损失函数是导数 Lipschitz、惩罚函数是 semi-convex 的,此外根据已有结论,得到 SCAD 回归问题的稳定点即可保证良好的统计性质。基于这些性质的分析,介绍了一种新的随机坐标下降方法(variable bregman stochastic coordinate descent, VBSCD),这一方法能很好求解带 SCAD 惩罚项的回归问题,算法的收敛点是 SCAD 回归模型的稳定点。最后,通过计算实验进一步说明本算法在求解 SCAD 回归问题的有效性。对不同的变量分组数,算法迭代到稳定点所需的迭代回合数相对稳定。随着变量分块数的增加,单次迭代中计算的内存需求减少。该研究方法可广泛应用于大数据背景下 SCAD 回归问题的求解当中。

关键词:　平滑削边绝对偏离;回归问题;随机坐标下降方法

中图分类号:　C 935　　**文献标志码:**　A

A Randomized Coordinate Descent Algorithm for Large-scale SCAD Regression Problem

ZHAO Lei　CHEN Ding　ZHU Daoli

(1. Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China; 2. Sino-US Global Logistics Institute, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: The regression problem with the smoothed clipped absolute deviation (SCAD) penalty is widely used in big data analysis because of its approximate unbiasedness in processing high-dimensional data (see Fan and Li, 2001). However, in the context of big data, the amount of data is often large and distributed

收稿日期:2019-06-14

基金项目:国家自然科学基金资助项目(71471112;71871140)

作者简介:赵磊(1987—),男,上海交通大学管理科学与工程博士研究生,研究方向:最优化方法与智能决策,E-mail: l. zhao@sjtu. edu. cn;陈玎(1996—),男,上海交通大学物流工程硕士研究生,研究方向:机器学习与智能决策,E-mail: chen_d@sjtu. edu. cn;朱道立(1945—),男,上海交通大学教授,研究方向:大数据优化与智能决策,E-mail:dlzhu@sjtu. edu。

in different locations, which makes the conventional algorithms for SCAD (LQA, LLA, ADMM, etc.) difficult to adapt to the current needs of solving SCAD regression. Stochastic Coordinate Descent (SCD) has been widely used in large-scale distributed optimization problems because of its small memory requirements (see Nesterov, 2012). However, in theory, the SCD algorithm can only deal with the regression problem with convex penalty. Existing random coordinate descent method is difficult to deal with the non-convex non-smooth SCAD regression problem. This paper first analyzes the SCAD regression problem model, and concludes that the loss function of the SCAD regression model is the derivative Lipschitz, and the penalty function is semi-convex. In addition, according to the existing conclusions, the critical point of the SCAD regression problem can ensure good statistical properties. Based on the analysis above, this paper introduces a new method of Variable Bregman Stochastic Coordinate Descent (VBSCD), which can solve the regression problem with SCAD penalty. The accumulated point of the algorithm is the critical point of SCAD regression model. Finally, the effectiveness of the proposed algorithm in solving SCAD regression is further illustrated by computational experiments. The algorithm proposed in this paper can be widely applied to solve the SCAD regression in the big data era.

Key words: smoothed clipped absolute deviation; regression; stochastic coordinate descent

回归作为一种常见的数据分析工具，已被广泛应用于量化金融分析、宏观经济研究、商务智能、医学与生物学研究等领域当中。在回归中，假设给定 n 组样本数据 $\{(x_i, y_i)\}_{i=1}^n$，其中每一个 x_i 是一个 p 维特征（或者因素）向量，y_i 为相对应的反应变量。回归的目标是找到特征向量 x_i 的线性关系用以近似反映变量 y_i，即

$$\eta(x_i)=\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j,\quad i=1,\cdots,n \tag{1}$$

即求出各特征（因素）的回归系数构成的向量 $\beta=(\beta_1,\cdots,\beta_p)^T\in R^p$，和截距项 β_0。最经典的回归系数向量 (β_0,β) 的估计模型为最小二乘估计模型（least squares estimator，LSE）：

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2\right\} \tag{2}$$

随着大数据技术的发展，数据产生的源头逐渐增加，数据收集的难度逐渐降低，从而造成数据规模和复杂性逐渐增加。此外，数据的维度（属性）通常很大，可以达到成百上千维甚至更高（例如文档数据、基因表达数据、图片数据、多媒体数据、贸易交易数据等）。在对应的回归问题中，样本数据的维数 p 往往很大，样本量 n 往往远远小于 p，即 $p\gg n$。以基因挖掘为例，通常是从成千上万维的核苷酸数据中挖掘出临床表现，但样本量往往很小。例如 Scheetz et. al. (2006) 的例子中，需要在 120 只 12 周大的小白鼠的 18976 条基因信息中进行挖掘，这一案例中 $p=18976$，$n=120$。

在这种情况下，特征向量 x_i 的很多属性往往线性相关，传统的 LSE 方法通常很难处理这样的数据。因此，一些统计学家在经典 LSE 方法的基础上加入变量筛选功能，以适应高维数据的处理，即将 LSE 模型推广为如下带稀疏惩罚项的最小二乘形式：

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2+\sum_{j=1}^{p}\varphi_\lambda(\beta_j)\right\} \tag{3}$$

其中，稀疏惩罚函数 $\varphi_\lambda(\theta)$ 可以有多种形式，常见的有岭回归（ridge regression）、套索回归（least absolute shrinkage and selection operator，LASSO）、平滑削边绝对偏离（smoothly clipped absolute deviation，SCAD）正则回归等。随着大数据时代的到来，这一类统计方法由于其在处理高维数据的优越，越来越受到数据科学家的重视，近几年在机器学习、统计学习、数据挖掘等领域被进一步研究。

岭回归中，$\varphi_\lambda(\theta)$ 通常选取如下平方形式：

$$\varphi_\lambda(\theta)=\lambda|\theta|^2 \tag{4}$$

LASSO 回归中，$\varphi_\lambda(\theta)$ 通常选取如下绝对值形式：

$$\varphi_\lambda(\theta)=\lambda|\theta| \tag{5}$$

SCAD 回归中，$\varphi_\lambda(\theta)$ 通常选取如下分段函数形式：

$$\varphi_\lambda(\theta)=\begin{cases}\lambda|\theta|, & |\theta|\leqslant\lambda,\\ \dfrac{-\theta^2+2a\lambda|\theta|-\lambda^2}{2(a-1)}, & \lambda<|\theta|\leqslant a\lambda,\\ \dfrac{(a+1)\lambda^2}{2}, & |\theta|>a\lambda,\end{cases} \tag{6}$$

其中，$a>2$，$\lambda>0$。如图 1 所示，二维空间中岭回归、LASSO 和 SCAD 的惩罚函数分别为图中虚

线、点划线、实线。

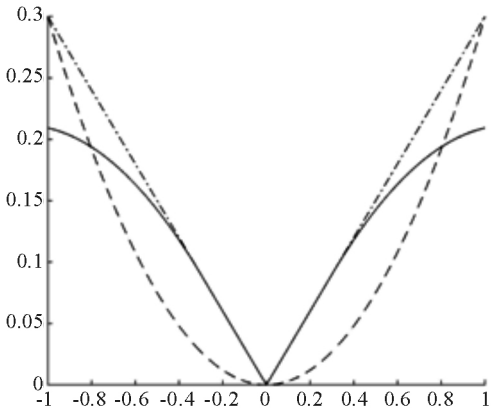

图 1　二维空间中岭回归、LASSO 和 SCAD 的惩罚函数图

前述三种带稀疏惩罚项的回归模型中，岭回归和 LASSO 回归模型在统计性质上是偏估计，SCAD 回归是近似无偏的，因此 SCAD 回归具有更好的统计性质，能够被用来处理更加复杂的数据分析问题，例如存在奇异值的数据。但从回归模型的求解上来讲，LASSO 回归、岭回归相应的最优化模型是凸的，可用一阶、二阶算法，随机坐标下降算法等进行求解（见 Boyd），而 SCAD 回归是非凸的，造成了求解上的困难。但 SCAD 惩罚方法作为非凸惩罚项的前瞻性工作，直到现在仍受到理论和应用领域的广泛重视，因此有效求解 SCAD 的方法近年仍在被讨论。在大数据背景下，待求解的回归问题的数据量往往很大，数据源分布在不同的地理位置。这种情况使得进行大数据分析时，往往难以获得/处理完整数据。此外，由于数据规模大，算法的每一次迭代中更新全部变量会造成计算机的内存不足。因此，在大数据环境下求解 SCAD 回归问题时，往往难以实现在算法的每一次迭代更新全部变量。而现有求解 SCAD 回归的算法（LQA、LLA、ADMM 等）在每一次迭代时均需要更新全部变量，这造成了现有求解 SCAD 回归的算法并不适用大数据背景下的应用。随机坐标下降方法（stochastic coordinate descent，SCD）以其每一次迭代中仅需要更新一部分变量的流程设计，减少了计算的内存需求，在大规模分布式最优化问题中得到了广泛的应用。但目前理论上 SCD 算法只能处理带凸惩罚项的回归问题，由于 SCAD 回归问题中的惩罚项是非凸和非光滑的，现有的随机坐标下降方法难以处理这一问题。

针对大数据时代高维数据、大规模数据、分布式存储的特点，本文研究大规模 SCAD 回归问题的随机坐标下降方法。为满足算法收敛的要求，我们分析得到 SCAD 回归模型的损失函数是导数 Lipschitz、惩罚函数是 semi-convex 的分析结论。根据已有结论，得到 SCAD 回归问题的稳定点可保证良好的统计性质。在分析的基础上，我们设计了一种随机坐标下降方法（variable bregman stochastic coordinate descent，VBSCD），这一方法能在内存可控的前提下，很好地求解 SCAD 回归问题，且该方法任意收敛点均是 SCAD 回归模型的稳定点。此外，在该算法每一次迭代的子问题中，更新变量数量可根据需求自由选择，这一点很好地控制了算法每一次迭代中的内存需求。本文通过计算实验说明这一方法在求解 SCAD 回归问题时的有效性，决策人可根据随机坐标集合大小来控制子问题运算所需要的内存需求。

本文结构如下：第 1 节对 SCAD 回归问题的模型性质进行分析；第 2 节介绍求解非凸问题的随机坐标下降方法及其主要的理论结论，以及求解 SCAD 回归的随机坐标下降算法流程；第 3 节进行数值实验，说明本文方法的有效性；最后给出总结。

1　SCAD 回归问题模型的性质分析

本节中考虑 SCAD 回归模型，即

$$\min_{\beta_0,\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2+\sum_{j=1}^{p}\varphi_\lambda(\beta_j)\right\} \tag{7}$$

其中，稀疏惩罚项 $\varphi_\lambda(\theta)$ 为如下形式：

$$\varphi_\lambda(\theta)=\begin{cases}\lambda|\theta|, & |\theta|\leqslant\lambda,\\ \dfrac{-\theta^2+2a\lambda|\theta|-\lambda^2}{2(a-1)}, & \lambda<|\theta|\leqslant a\lambda,\\ \dfrac{(a+1)\lambda^2}{2}, & |\theta|>a\lambda,\end{cases} \tag{8}$$

其中，$a>2$，$\lambda>0$。为了简化表达，我们做如下定义，令 $X=\begin{pmatrix}x_{11}\cdots x_{n1}\\ \vdots\ddots\vdots\\ x_{n1}\cdots x_{np}\end{pmatrix}$，$\hat{X}=\begin{pmatrix}1x_{11}\cdots x_{n1}\\ \vdots\vdots\ddots\vdots\\ 1x_{n1}\cdots x_{np}\end{pmatrix}$，$y=(y_1,y_n)^T$，$\hat{\beta}=(\beta_0,\beta_1,\beta_n)^T$，损失函数 $f(\hat{\beta})=f(\beta_0,\beta)=1/2n\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2=\frac{1}{2n}\|\hat{X}\hat{\beta}-y\|^2$，惩罚项 $g(\beta)=\sum_{j=1}^{p}\varphi_\lambda(\beta_j)$，则模型（3）可以表达为如下形式：

$$\min_{\beta}\{f(\hat{\beta})+g(\beta)\}=\min_{\beta}\left\{\frac{1}{2n}\|\hat{X}\hat{\beta}-y\|^2+\sum_{j=1}^{p}\varphi_\lambda(\beta_j)\right\} \tag{9}$$

通过理论分析，我们可以得到损失函数 $f(\hat{\beta})$ 和惩罚项 $g(\beta)$ 具有如下性质。

命题 1：

（1）损失函数 $f(\hat{\beta})$ 是导数 Lipschitz 的，即

$\|\nabla f(x)-\nabla f(y)\|\leqslant L\|x-y\|$，$\forall\ x,y\in\mathrm{dom}f$，

其导数 Lipschitz 常数可取$\frac{e_{max}(\hat{X}^T\hat{X})}{n}$，其中，$e_{max}(\hat{X}^T\hat{X})$为矩阵 $\hat{X}^T\hat{X}$ 的最大特征值。

(2)惩罚项 $g(\beta)$是 semi-convex 函数，即存在常数 $\rho\geqslant 0$，

$$g(y)-g(x)\geqslant(\xi,y-x)-\frac{\rho}{2}\|y-x\|^2,\ \forall\xi\in\partial g(x),\ x,y\in \mathrm{dom}g,$$

对 $g(\beta)$可取$\frac{1}{a-1}$作为 semi-convex 的常数。

证明：

(1)根据定义有

$$\|\nabla f(\hat{\beta})-\nabla f(\hat{\alpha})\|=\frac{1}{n}\|\hat{X}^T(\hat{X}\hat{\beta}-y)-\hat{X}^T(\hat{X}\hat{\alpha}-y)\|\leqslant\frac{e_{max}(\hat{X}^T\hat{X})}{n}\|\hat{\beta}-\hat{\alpha}\|。$$

根据导数 Lipschitz 的定义得到 $f(\hat{\beta})$的导数 Lipschitz 常数为$(\frac{e_{max}(\hat{X}^T\hat{X})}{n}$。

(2)这里可直接对 $\varphi_\lambda(\theta)$进行分析，当 $\theta>0$ 时，$\varphi_\lambda(\theta)$的次微分为

$$\partial\varphi_\lambda(\theta)=\begin{cases}\lambda, & \theta\leqslant\lambda\\ \frac{a\lambda-\theta}{a-1}, & \lambda<\theta\leqslant a\lambda\\ 0, & \theta>a\lambda\end{cases}$$

因此有

$$\partial\varphi_\lambda(\theta)+\frac{\theta}{a-1}=\begin{cases}\frac{\theta}{a-1}+\lambda, & \theta\leqslant\lambda\\ \frac{a\lambda}{a-1}, & \lambda<\theta\leqslant a\lambda\\ \frac{\theta}{a-1}, & \theta>a\lambda\end{cases}$$

不难看出，次微分为 $\partial\varphi_\lambda(\theta)+\frac{\theta}{a-1}$的函数是单调增函数。而当 $\theta>0$ 时，有

$$\partial\left[\varphi_\lambda(\theta)+\frac{\theta^2}{2(a-1)}\right]=\partial\varphi_\lambda(\theta)+\frac{\theta}{a-1}$$

根据[24]有 $\varphi_\lambda(\theta)+\frac{\theta^2}{2(a-1)}$在$(0,+\infty)$上是凸的。同理，有 $\varphi_\lambda(\theta)+\frac{\theta^2}{2(a-1)}$在$(0,-\infty)$上是凸的。因为函数 $\varphi_\lambda(\theta)+\frac{\theta^2}{2(a-1)}$在 $\theta=0$ 点连续，$\varphi_\lambda(\theta)$在 $\theta=0$ 点是绝对值函数，因此在 $\theta=0$ 点的邻域内 $\varphi_\lambda(\theta)+\frac{\theta^2}{2(a-1)}$是凸的。由此证得函数 $\varphi_\lambda(\theta)$是 semi-convex 的，其常数为$\frac{1}{a-1}$。由于 $g(\beta)=\sum_{j=1}^{p}\varphi_\lambda(\beta_j)$，因此 $g(\beta)$是带常数$\frac{1}{a-1}$的 semi-convex 函数。

此外，根据[23]的结论，我们可知 SCAD 回归模型的任意稳定点与全局最优点具有基本一致的统计性质。通过上述分析，不难看出，算法可求得 SCAD 回归问题的稳定点，即可得到原始参数的较好估计。下一节中将应用本节中分析到的 SCAD 回归模型的损失函数是导数 Lipschitz、惩罚函数是 semi-convex 的分析结论。

2　求解 SCAD 回归问题的随机坐标下降方法

2.1　求解非凸问题的随机坐标下降方法

最近 Zhao 和 Zhu(2019)给出了求解如下非凸问题的随机坐标下降方法，针对的问题如下：

$$\min_{x\in R^n}F(x)=f(x)+g(x)=f(x)+\sum_{i=1}^{N}g_i(x_i)\quad(8)$$

其中，$f:R^n\to(-\infty,+\infty]$是光滑函数(可以是非凸的)，$g:R^n\to(-\infty,+\infty]$是连续 semi-convex 函数，$g_i:R^{n_i}\to(-\infty,+\infty]$，$x_i\in R^{n_i}$，$\sum_{i=1}^{N}n_i=n$。此外，$f$ 和 g 满足假设 1。

假设 1：

(i)f 是一个光滑函数，$\mathrm{dom}f$ 是凸的。f 的梯度∇f 是 L-Lipschitz 的；

(ii)g 是一个 semi-convex 函数，其参数为 ρ；

(iii)F 是水平有界的，即对任意 $r\in R$，集合$\{x\in R^n\mid F(x)\leqslant r\}$是有界的。

针对非凸问题(8)的随机坐标下降方法(variable bregman stochastic coordinate descent, VB-SCD)流程如下：

选择初始解 $x^0\in R^n$

for $k=0,1,\cdots,$ do

从$\{1,\cdots,N\}$中等概率选择 $i(k)$

$$x^{k+1}=arg\min_{x\in R^n}\langle\nabla_{i(k)}f(x^k),x_{i(k)}\rangle+g_{i(k)}(x_{i(k)})+\frac{1}{\epsilon}D(x,x^k)\quad(9)$$

end for

在算法流程中 D 为 Bregman 函数，即 $D(x,y)=K(y)-[K(x)+\langle\nabla K(x),y-x\rangle]$，在 Bregman 函数中，$K$ 为核函数，ϵ 为步长参数。核函数 K 和步长参数 ϵ 服从假设 2。

假设 2：

(i)K 是 m-强凸函数，且其梯度∇K 是 M-Lipschitz 的。

(ii)参数 ϵ 服从如下关系：$0<\epsilon<\min\left\{\frac{m}{L},\frac{m}{\rho}\right\}$。

在非凸问题(8)满足假设 1,算法 VBSCD 满足假设 2 的情况下,得到了如下结论:算法 VBSCD 产生的任意收敛点是函数的一个稳定点。

结合本节中算法性质和第二节命题 1 中 SCAD 回归模型性质,利用 Zhao 和 Zhu 最近提出的 VBSCD 算法可以求解 SCAD 回归问题。

2.2 求解 SCAD 回归问题的随机坐标下降算法

本节我们给出求解 SCAD 回归问题的随机坐标下降方法。我们选择 Bregman 函数 $D(x,x^k)=\frac{1}{2}\|x-x^k\|^2$,即有 $m=M=1$。

根据命题 1,有 $L=\frac{e_{\max}(\hat{X}^T\hat{X})}{n}$,$\rho=\frac{1}{a-1}$,因此可以选择 $\epsilon=\min\left\{\frac{1}{L},\frac{1}{\rho}\right\}$。令 I_i 是一个包含 p_i 个元素的索引集合,$\sum_{i=1}^{N}p_i=p$,同时 $\bigcup^{N}{}_{i=1}I_i=\{1,\cdots,p\}$。则有如下算法流程:

求解 SCAD 回归问题的随机坐标下降方法如下:

选择初始解 $x^0\in R^n$

for $k=0,1,\cdots,$ do

$$\beta_0^{k+1}=\beta_0^k+\epsilon\frac{1}{n}\sum_{i=1}^{n}(y_i-\beta_0^k-\sum_{j=1}^{p}x_{ij}\beta_j^k) \tag{10}$$

从 $\{I_1,\cdots,I_N\}$ 中等概率选择 $I(k)$,对任意 $j\in I(k)$,计算

$$\beta_j^{k+1}=\arg\min_{\beta_j\in R}-\frac{1}{n}\sum_{i=1}^{n}x_{ij}(y_i-\beta_0^k-\sum_{j=1}^{p}x_{ij}\beta_j^k)\beta_j+\varphi_\lambda(\beta_j)+\frac{1}{2}\|\beta_j-\beta_j^k\|^2 \tag{11}$$

end for

其中,子问题(11)有如下闭合表达式(见 Fan 和 Li, 2001):

$$\beta_j^{k+1}=\begin{cases}(sign(r^k)\max\{0,|r^k|-\epsilon\lambda\}, & |r^k|\leqslant(1+\epsilon)\lambda,\\ \frac{(a-1)r^k-\epsilon a\lambda sign(r^k)}{a-1-\epsilon}, & (1+\epsilon)\lambda<|r^k|\leqslant a\lambda,\\ r^k, & |r^k|>a\lambda,\end{cases} \tag{12}$$

其中,$r^k=\beta_j^k+\epsilon\frac{1}{n}\sum_{i=1}^{n}x_{ij}(y_i-\beta_0^k-\sum_{j=1}^{p}x_{ij}\beta_j^k)$。

通过上述算法流程,不难看出本文给出的 VBSCD 算法,在算法每一次更新中仅随机选择一部分变量进行更新,这一点很好地适应了大数据时代求解 SCAD 回归问题时数据规模大、分布在不同区域的需求。更新变量个数与所选择随机坐标集合 $I(k)$ 的大小紧密相关。每一次迭代中,所需要的内存大小也和所选择随机坐标集合 $I(k)$ 的大小紧密相关。下一节,我们将通过仿真实验,说明每一次迭代中选择随机坐标集合 $I(k)$ 的大小与子问题求解内存需求、算法总迭代次数之间的关系。

3 数值实验

本节对两组仿真数据进行数值实验,通过数值实验说明:

(1)本文随机坐标下降算法可有效求解 SCAD 回归问题;

(2)本文随机坐标下降算法对不同规模算例求解时,所需迭代次数相对稳定;

(3)本文随机坐标下降算法在不同变量分组情况下,算法迭代到稳定点所需的迭代回合数(epoch)相对稳定;

(4)本文随机坐标下降算法随着变量分组的增加[随机坐标集合 $I(k)$ 变小],每一次迭代中子问题计算内存需求减少。

3.1 数据集与计算环境

(1)仿真数据:样本因素数据 $X\in R^{n\times p}$ 的选择服从 $N(0,1)$ 高斯分布,构造真实系数向量 $\beta^*\in R^p$,维数为 p,稀疏度为 s,β^* 中随机的 s 个元素为非 0 元素,其余元素为 0。非 0 元素的值服从 $N(0,1)$ 高斯分布。反应变量 $y=X\beta^*+\varepsilon$,其中各元素的噪声向量独立同分布,服从 $N(0,\sigma^2)$,用 σ 来控制噪声大小。

根据上述规则,生成两组仿真实验,分别为 $\sigma=0.01$、$n=100$、$p=1000$、$s=10$;以及 $\sigma=0.01$、$n=200$、$p=2000$、$s=10$。

(2)计算环境:Intel Core i5-6200U CPUs (2.40GHz),8.00 GB RAM

3.2 实验结论

针对 2 组仿真数据,我们选择 $\lambda=0.2\sigma\sqrt{n\log p}$,$a=3.7$(见 Fan 和 Li,2001),把变量分成 5、10、50、100 组(即 $N=5$、10、50、100)。即在 $p=1000$ 的实验中,每组分别有 200、100、20、10 个变量;在 $p=2000$ 的实验中,每组分别有 400、200、40、20 个变量。得到如下三组图。

图 2 和图 4 中实线表示将变量分为 100 组时目标函数的下降情况,点线表示将变量分为 50 组时目标函数的下降情况,点划线表示将变量分为 10 组时目标函数的下降情况,虚线表示将变量分为 5 组时的目标函数下降情况。

图 3 和图 5 中圆圈(o)表示原始系数,星号(*)表示通过 SCAD 回归模型计算出的系数。

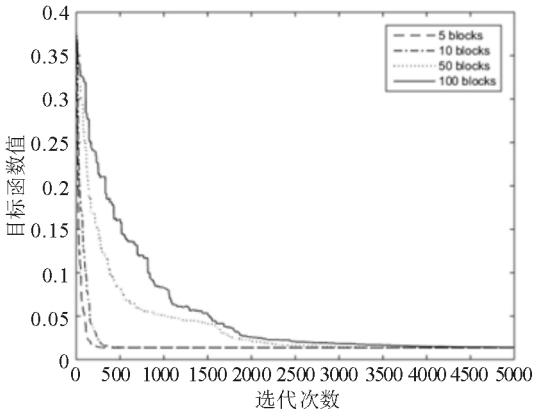

图 2　$n=100$、$p=1000$、$s=10$ **时变量划分块数与算法下降效果之间的关系**

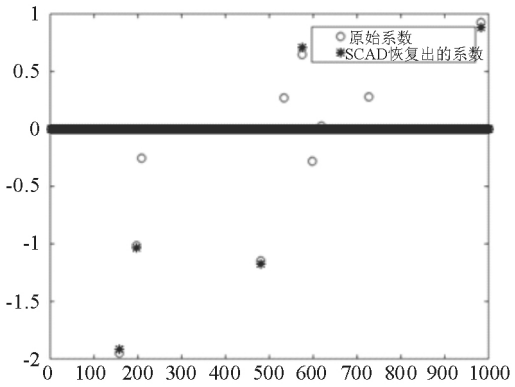

图 3　$n=100$、$p=1000$、$s=10$ **时信息恢复效果**

图 6 和图 7 表示变量划分块数、算法迭代回合(epoch)、目标函数下降效果之间的关系。其中，算法迭代回合(epoch)指的是 $\frac{\text{算法迭代次数}}{\text{分块数}}$ **。**

图 8 表示对 $n=100$、$p=1000$、$s=10$ **问题将变量分为 1、2，5、10、50、100 组时，求解子问题所需的内存大小。图 9 表示对** $n=200$、$p=2000$、$s=10$ **问题将变量分为 1、2、5、10、50、100 组时，求解子问题所需的内存大小。**

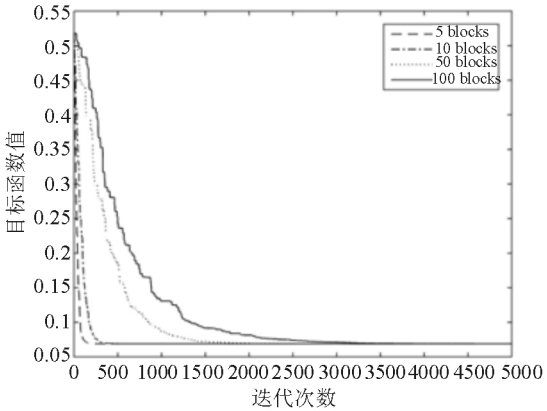

图 4　$n=200$、$p=2000$、$s=10$ **时变量划分块数与算法下降效果之间的关系**

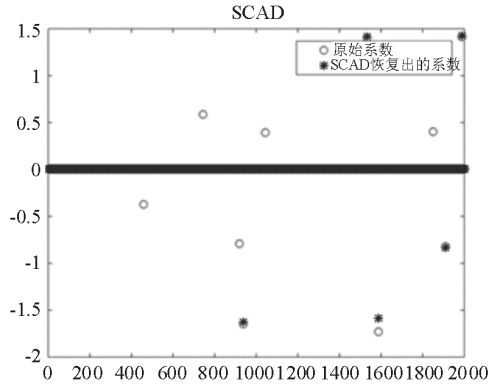

图 5　$n=200$、$p=2000$、$s=10$ **时信息恢复效果**

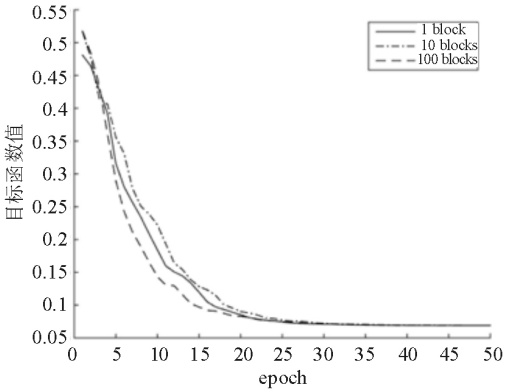

图 6　$n=200$、$p=2000$、$s=10$ **时变量划分块数、算法迭代回合(epoch)、目标函数下降效果之间的关系**

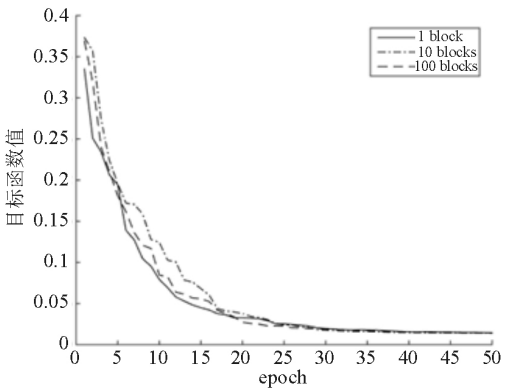

图 7　$n=200$、$p=2000$、$s=10$ **时变量划分块数、算法迭代回合(epoch)、目标函数下降效果之间的关系**

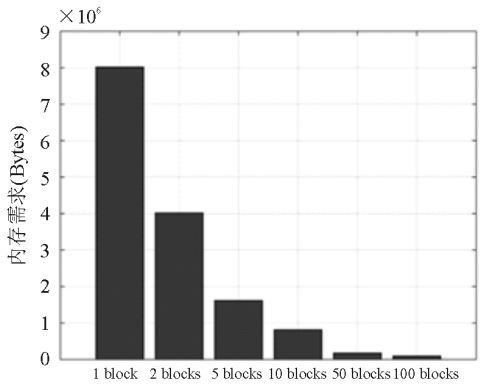

图 8　$n=100$、$p=1000$、$s=10$ **算例，当把变量分成 1、2、5、10、50、100 组时子问题内存需求**

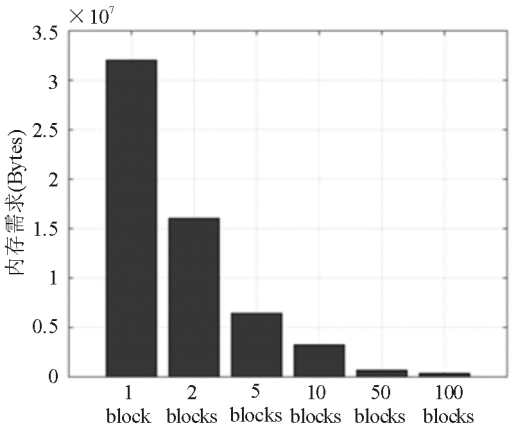

图 9　$n=200$、$p=2000$、$s=10$ 算例，当把变量分成 1、2、5、10、50、100 组时子问题内存需求

通过对图 2 至图 9 的分析有如下结论：

(1)本文随机坐标下降算法可有效求解 SCAD 回归问题：从图 2 和图 4 可以看出算法在迭代 3000 次左右时，即可收敛到较为稳定的解。从图 3 和图 5可以看出 SCAD 回归具有很好的信息回归效果。图中圆圈(o)表示原始系数，星号($*$)表示通过 SCAD 回归模型计算出的系数。

(2)本文随机坐标下降算法对不同算例规模所需迭代次数相对稳定：如图 2 和图 4 中，实线、点线、点划线、虚线的对比可以得到，随着拆分成的组(block)数量减少[随机坐标集合 $I(k)$规模增加]，算法收敛到所需要结果的迭代次数减少。此外，对比图 2 和图 4，可以得到不同规模算例，在分组数一致的情况下，迭代到所需要点的所需迭代次数基本相同。

(3)本文随机坐标下降算法在不同变量分组情况下，算法迭代到稳定点所需的迭代回合数(epoch)相对稳定：如图 6 和图 7 所示，在两组算例中，目标函数值随迭代回合数增加而下降，其下降效率受变量分组数(随机坐标集合 $I(k)$大小)影响不大。

(4)本文随机坐标下降算法随着变量分组的增加[随机坐标集合 $I(k)$变小]，每一次迭代中子问题计算内存需求减少：如图 8 和图 9 中可以得到，随着变量分组的增加[随机坐标集合 $I(k)$变小]，算法子问题所需计算内存减少。对比图 8 和图 9，可以看出每组中变量的个数是子问题求解内存需求的关键因素，在每组选择变量个数相近的情况下，子问题的计算内存需求相近。

4　总结

SCAD 回归问题，以其在处理高维数据中的近似无偏性，在大数据分析中得到广泛应用。但由于其惩罚项的非凸性，现有的随机坐标下降方法难以处理这一问题。本文首先对 SCAD 回归模型进行分析，得到 SCAD 回归模型的损失函数是导数 Lipschitz，惩罚函数是 semi-convex 的，并根据已有结论得出 SCAD 回归模型的任意稳定点有良好的统计性质。基于这些分析，本文介绍一种新的随机坐标下降方法，可应用于求解大规模 SCAD 回归问题，算法的任意收敛点均是 SCAD 回归模型的稳定点。这一方法能很好地求解 SCAD 回归问题。本文通过计算实验说明这一方法在求解 SCAD 回归问题时的有效性，同时说明算法在不同变量分组情况下，算法迭代到稳定点所需的迭代回合数(epoch)相对稳定。随着变量分块数的增加，单次迭代中计算的内存需求减少。这一方法可被广泛应用于大数据分析和决策实践当中。

在大数据分析工作里，求解非凸非光滑的随机坐标下降算法的计算复杂度分析是个重要问题，Zhao 和 Zhu 证明了在水平集次微分误差界条件下，随机坐标下降算法可以达到线性收敛，而 Li 和 Pong 证明了 SCAD 回归问题的目标函数满足该误差界条件。因文章篇幅问题，本文算法的计算复杂度分析不在此处做进一步讨论。

参考文献：

[1] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American statistical Association, 2001, 96(456): 1348-1360.

[2] NESTEROV Y. Efficiency of coordinate descent methods on huge-scale optimization problems[J]. SIAM Journal on Optimization, 2012, 22(2): 341-362.

[3] 叶五一，肖丽华，缪柏其. 基于变系数分位点回归的金砖四国金融稳定分析[J]. 管理科学学报，2018(5)：44-52.

[4] 陈雨露，马勇，阮卓阳. 金融周期和金融波动如何影响经济增长与金融稳定[J]. 金融研究，2016(2)：1-22.

[5] 李梅，柳士昌. 对外直接投资逆向技术溢出的地区差异和门槛效应——基于中国省际面板数据的门槛回归分析. 管理世界，2012(1)：21-32.

[6] NAIFAR N. Do global risk factors and macroeconomic conditions affect global Islamic index dynamics? A quantile regression approach[J]. The Quarterly Review of Economics and Finance，2016(61)：29-39.

[7] 左文明，王旭，樊偿. 社会化电子商务环境下基于社会资本的网络口碑与购买意愿关系[J]. 南开管理评论，2014，17(4)：140-150.

[8] 杨海明，孟宪杰，吴薇，等. CKD-MBD 患者血清骨代谢标记物与中医证候特征的回归分析[J]. 中国中药杂志，2017，42(20)：4027-4034.