

# 求解正则化最小二乘问题的一个非精确 交替方向乘子法<sup>\*1)</sup>

乐航睿

(南开大学数学科学学院, 天津 300071)

杨庆之

(喀什大学数学与统计学院, 新疆喀什 844006;

南开大学, 数学科学学院, 天津 300071)

## 摘 要

正则化最小二乘问题广泛出现在图像处理、统计学等领域中, 交替方向乘子法 (ADMM) 是求解这个问题的一种有效方法. ADMM 在每一步迭代过程中, 都需要求解两个子问题, 子问题能否有效地求解对整个算法的有效性有重要影响. 在有些情形, 精确求解子问题是不可能的, 或者是需要花费很大代价的. 由于这个原因, 非精确地求解子问题的一类算法得到了发展. 而已有的非精确类 ADMM 算法, 在迭代过程中需要不断提高子问题解的精度, 从而子问题迭代步数也越来越多, 这就影响了整个算法的效率. 这篇文章提出了一个不精确 ADMM 算法, 该算法的特点是在迭代过程中, 子问题的迭代步数是确定的, 这克服了之前算法的不足. 文章中的数值例子也说明了提出的算法是有效的.

**关键词:** 正则化最小二乘; 交替方向乘子法; 非精确算法

**MR (2000) 主题分类:** 65D17

## AN INEXACT ADMM FOR THE REGULARIZED LEAST SQUARES PROBLEM

Yue Hangrui

(School of Mathematical Sciences, Nankai University, Tianjin 300071, China)

Yang Qingzhi

(School of Mathematics and Statistics, Kashi University, Kashi 844006, Xinjiang, China;

School of Mathematical Sciences, Nankai University, Tianjin 300071, China)

## Abstract

The regularized least squares problems are widely used in image processing, statistics and other field. Alternating direction method of multipliers (ADMM) is an effective algorithm for the problem. The computation of ADMM is dominated by two subproblems at each step, and whether subproblems can be solved effectively have a major impact on the effectiveness of the algorithm. Since the subproblem could be too difficult to be solved exactly in many practical applications, the inexact versions of ADMM are developed. Most existing inexact ADMMs are based on solving subproblems approximately and gradually increasing

\* 2016 年 4 月 28 日收到.

<sup>1)</sup> 教育部博士点基金 (20120031110024) 和天津市自然科学基金 (12JCYBJC31200) 资助项目.

their precision, which means the precision can become very high after several iterations. In this paper, an easily implementable inexact ADMM with certain iteration number of subproblems is developed. Numerical examples illustrate the proposed algorithm is effective.

**Keywords:** regularized least squares problem; alternating direction method; inexact;  
**2000 Mathematics Subject Classification:** 65D17

## 1. 引言

本文研究如下正则化最小二乘问题:

$$\min \frac{1}{2} \|Ax - b\|_2^2 + r(x) \quad (1.1)$$

其中  $x \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{l \times n}$  为列满秩的矩阵,  $b \in \mathbb{R}^l$ , 正则项  $r: \mathbb{R}^n \rightarrow \mathbb{R}$  为凸函数.

实际应用中的正则项大部分都具有比较简单的形式. 例如图像处理中的 TV 模型使用  $r(x) = \|Px\|_1$  作为正则项, 其中  $P$  表示差分矩阵. 而矩阵优化中, 正则项取核范数  $\|X\|_*$  能得到低秩解<sup>[8,9]</sup>. 本文假设正则项  $r(x)$  有比较简单的形式, 具体而言, 我们假设  $r(x) = g(Px)$ ,  $g: \mathbb{R}^m \rightarrow \mathbb{R}$  的邻近点算子<sup>[7]</sup>有显示表达式,  $P$  是  $\mathbb{R}^{m \times n}$  中的一个矩阵.  $g(y)$  的邻近点算子定义为:

$$\text{prox}_{\alpha g}(y) = \underset{v}{\operatorname{argmin}} \left( g(v) + \frac{1}{2\alpha} \|y - v\|_2^2 \right), \quad \alpha > 0.$$

利用这个假设, 问题 (1.1) 可重新表述为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|Ax - b\|_2^2 + g(y) \\ \text{s.t.} \quad & y = Px. \end{aligned} \quad (1.2)$$

显然, 正则化最小二乘问题是一个凸问题, 因而可以利用凸规划中的内点法进行求解<sup>[5]</sup>. 但对于实际中存在的涉及大规模病态矩阵的问题, 内点法不是一个有效的算法, 反而是一些简单的梯度方法会更有效<sup>[6]</sup>. ADMM 作为一种简单而有效的方法经常被用于解决这类问题<sup>[3]</sup>.

用于求解问题 (1.2) 的经典的 ADMM 有如下格式:

$$\begin{cases} x^{k+1} := \underset{x}{\operatorname{argmin}} \frac{1}{2} \|Ax - b\|_2^2 - \lambda^{kT} (Px - y^k) + \frac{\beta}{2} \|Px - y^k\|_2^2 \\ y^{k+1} := \underset{y}{\operatorname{argmin}} g(y) - \lambda^{kT} (Px^{k+1} - y) + \frac{\beta}{2} \|Px^{k+1} - y\|_2^2 \\ \lambda^{k+1} := \lambda^k - \beta(Px^{k+1} - y^{k+1}). \end{cases}$$

其中  $\lambda^k \in \mathbb{R}^m$  是拉格朗日乘子,  $\beta > 0$  是惩罚参数. 利用最优性条件以及邻近点算子, 上述格式等价于:

$$\begin{cases} x^{k+1} = \left\{ x \in \mathbb{R}^n \mid (A^T A + \beta P^T P)x = A^T b + \beta P^T \left( y^k + \frac{1}{\beta} \lambda^k \right) \right\} \\ y^{k+1} = \text{prox}_{\frac{1}{\beta} g}(Px^{k+1} - \lambda^k) \\ \lambda^{k+1} = \lambda^k - \beta(Px^{k+1} - y^{k+1}). \end{cases} \quad (1.3)$$

因为我们假设正则项有比较简单的形式, 所以每一步 ADMM 的主要计算量花费在求解一个线性方程组. 相对于原问题, 解线性方程组要简单许多, 这也是 ADMM 适用于正则化最小二乘问题的原因.

包含大规模病态矩阵的线性方程组一般通过迭代法求解. 对于这一类问题, 得到线性方程组的精确解通常是困难的. 数值实验和理论分析表明, 在一定条件下, 不精度地求解子问题也可以保证一个算法收敛到问题的解<sup>[1]</sup>. 我们在本文中研究求解 (1.2) 的一种不精确 ADMM.

一般而言, 不精确求解子问题需要特定的误差准则来保证算法收敛. Eckstein 等人<sup>[2]</sup>证明了近似求解子问题并且逐步提高求解精度不会影响 ADMM 的收敛. NG 等人<sup>[1]</sup>提出一种能够实现的不精确 ADMM 的误差准则, 并且证明了收敛性. 但这些算法都要求在求解过程中, 子问题的求解精度必须越来越高.

本文提出了一种容易实现的不精确 ADMM 算法, 该算法用固定的迭代步数求解子问题. 本文将给出新算法的收敛性证明, 并且用数值实验说明新算法子问题迭代次数少, 总运行时间短.

## 2. 预备知识

本节概述下文中用到的符号和基本知识.

### 2.1. 求解线性方程组的迭代法

对于正则化最小二乘问题, 需要不精确求解的是如下方程组:

$$(A^T A + \beta P^T P)x = A^T b + \beta P^T \left( y^k + \frac{1}{\beta} \lambda^k \right). \quad (2.1)$$

我们定义  $H := (A^T A + \beta P^T P)$ . 由于  $H$  正定, 我们可把  $H$  分解为  $H = D - E - E^T$ , 其中  $D$  为对角矩阵而  $E$  为下三角矩阵.

利用迭代法求解 (2.1) 时, 设初始迭代点为  $x_{initial}^{k+1}$ , 迭代  $m$  次得到的点  $x_m^{k+1}$  记为精确解  $x_{exact}^{k+1}$  的近似. 记  $\epsilon_{initial}^{k+1} := x_{initial}^{k+1} - x_{exact}^{k+1}$ ,  $\epsilon_m^{k+1} := x_m^{k+1} - x_{exact}^{k+1}$ .

如果 Jacobi 矩阵  $J := D^{-1}(E + E^T)$  满足  $\rho(J) < 1$ , 我们可通过 Jacobi 迭代法来求解问题 (2.1)<sup>[4]</sup>. 那么第  $m$  次的误差  $\epsilon_m^{k+1}$  为:

$$\epsilon_m^{k+1} = J^m \epsilon_{initial}^{k+1}.$$

相似地, Gauss-Seidel 迭代法的误差  $\epsilon_m^{k+1}$  为:

$$\epsilon_m^{k+1} = G^m \epsilon_{initial}^{k+1}.$$

其中  $G := (D - E)^{-1} E^T$  满足  $\rho(G) < 1$ .

共轭梯度法 (PCG) 常被用来求解大规模线性方程组<sup>[10]</sup>. 其误差  $\epsilon_m^{k+1}$  满足:

$$\|\epsilon_m^{k+1}\|_H \leq 2 \left( \frac{\sqrt{K_2(P^{-1}HP^{-T})} - 1}{\sqrt{K_2(P^{-1}HP^{-T})} + 1} \right)^m \|\epsilon_{initial}^{k+1}\|_H,$$

其中  $P$  为非奇异的矩阵,  $K_2(P^{-1}HP^{-T})$  是  $P^{-1}HP^{-T}$  的条件数,  $\|\cdot\|_H$  定义为  $\|x\|_H = \sqrt{x^T H x}$ .

## 2.2. 算法以及最优性条件

通过简单的推导, 不难发现对于给定的  $M < 1$ , 上述的迭代方法均存在  $m_0$  满足:

$$\|\epsilon_{m_0}^{k+1}\|_H \leq M \|\epsilon_{initial}^{k+1}\|_H. \quad (2.2)$$

事实上, (2.2) 可看作是控制某种相对误差的收敛准则. 算法 1 给出了求解 (1.2) 的固定子问题迭代次数的不精确 ADMM.

### 算法 1.

取  $\begin{pmatrix} x^0 \\ y^0 \\ \lambda^0 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m, \beta > 0$

For  $k = 0, 1, 2, \dots$

1. 取初始点  $x_{initial}^{k+1} = x^k$ , 通过迭代法来求解问题 (2.1) 的近似解  $x^{k+1} = x_{m_0}^{k+1}$
2.  $y^{k+1} = \text{prox}_{\frac{1}{\beta}} g(Px^{k+1} - \lambda^k)$
3.  $\lambda^{k+1} = \lambda^k - \beta(Px^{k+1} - y^{k+1})$

Endfor

我们称该算法为不精确交替方向乘子法 (IADMM).

由于误差准则仅需子问题的迭代数等于一个容易计算的  $m_0$ , 所以 IADMM 相对于之前的不精确 ADMM 是更容易实现的. IADMM 和 ADMM 的区别在于 IADMM 子问题计算的是一个近似点而非精确点, 并且 IADMM 采用了上一个迭代得到的点作为求解线性方程组的初始点.

对于 IADMM 的每一步, 均有:

$$\begin{cases} Hx^{k+1} = A^T b + \beta P^T \left( y^k + \frac{1}{\beta} \lambda^k \right) + H \epsilon_{m_0}^{k+1} \\ \|\epsilon_{m_0}^{k+1}\|_H \leq M \|\epsilon_{initial}^{k+1}\|_H \\ 0 \in \partial g(y^{k+1}) + \lambda^k - \beta(Px^{k+1} - y^{k+1}) \\ \lambda^{k+1} = \lambda^k - \beta(Px^{k+1} - y^{k+1}) \end{cases} \quad (2.3)$$

其中  $\partial g(y^{k+1})$  表示  $g(y)$  在  $y^{k+1}$  的次梯度.

由于  $A$  列满秩, 问题 (1.1) 存在唯一解  $\begin{pmatrix} x^* \\ y^* \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m$ . 其最优性条件为:

$$\begin{cases} A^T(Ax^* - b) - P^T \lambda^* = 0 \\ 0 \in \partial g(y^*) + \lambda^* \\ y^* = Px^* \end{cases} \quad (2.4)$$

其中  $\lambda^*$  为线性约束的拉格朗日乘子,  $\partial g(y^*)$  为  $g(y)$  在  $y^*$  处的次梯度.

## 3. 收敛性分析

首先, 我们给出 IADMM 生成的点列所具有的收敛性质.

命题 1. 若  $\begin{pmatrix} x^* \\ y^* \\ \lambda^* \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  问题 (1.1) 的唯一解. 对于给定的  $\begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix} \in$

$\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ , 由 IADMM 生成的  $\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ . 满足:

$$(x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) + V_{k+1} - V_k + \frac{1}{2\beta} \|z_{k+1}\|_2^2 \leq (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1} \quad (3.1)$$

这里  $V_k = \frac{\beta}{2} \|y^k - y^*\|_2^2 + \frac{1}{2\beta} \|\lambda^k - \lambda^*\|_2^2$ ,  $z_{k+1} = \lambda^{k+1} - \lambda^k - \beta(y^{k+1} - y^k)$ .

证明. 由 (2.3) 与 (2.4), 可得:

$$\begin{aligned} & (x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) \\ &= (x^{k+1} - x^*)^T P^T (\lambda^{k+1} - \lambda^*) \\ & - \beta (x^{k+1} - x^*)^T P^T (y^{k+1} - y^k) + (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1}. \end{aligned} \quad (3.2)$$

因为  $-\lambda^{k+1} \in \partial g(y^{k+1})$  以及  $-\lambda^* \in \partial g(y^*)$ , 不难推导出:

$$-(y^{k+1} - y^*)^T (\lambda^{k+1} - \lambda^*) \geq 0. \quad (3.3)$$

利用 (3.2) 和 (3.3), 得:

$$\begin{aligned} & (x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) \\ & \leq (x^{k+1} - x^*)^T P^T (\lambda^{k+1} - \lambda^*) - (y^{k+1} - y^*)^T (\lambda^{k+1} - \lambda^*) \\ & - \beta (x^{k+1} - x^*)^T P^T (y^{k+1} - y^k) + (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1} \\ & \leq -\frac{1}{\beta} (\lambda^{k+1} - \lambda^k)^T (\lambda^{k+1} - \lambda^*) - \beta (y^{k+1} - y^k)^T (y^{k+1} - y^*) \\ & + (\lambda^{k+1} - \lambda^k)^T (y^{k+1} - y^k) + (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1}. \end{aligned}$$

由于  $(a - b)^T (a - c) = \frac{1}{2} (\|a - c\|_2^2 - \|b - c\|_2^2 + \|a - b\|_2^2)$ , 上式可化为:

$$\begin{aligned} & (x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) + \frac{\beta}{2} (\|y^{k+1} - y^*\|_2^2 - \|y^k - y^*\|_2^2 + \|y^{k+1} - y^k\|_2^2) \\ & + \frac{1}{2\beta} (\|\lambda^{k+1} - \lambda^*\|_2^2 - \|\lambda^k - \lambda^*\|_2^2 + \|\lambda^{k+1} - \lambda^k\|_2^2) - (\lambda^{k+1} - \lambda^k)^T (y^{k+1} - y^k) \\ & = (x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) + \frac{\beta}{2} (\|y^{k+1} - y^*\|_2^2 - \|y^k - y^*\|_2^2) \\ & + \frac{1}{2\beta} (\|\lambda^{k+1} - \lambda^*\|_2^2 - \|\lambda^k - \lambda^*\|_2^2) + \frac{1}{2\beta} \|\lambda^{k+1} - \lambda^k - \beta(y^{k+1} - y^k)\|_2^2 \\ & \leq (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1}. \end{aligned}$$

整理上式, 有:

$$(x^{k+1} - x^*)^T A^T A(x^{k+1} - x^*) + V_{k+1} - V_k + \frac{1}{2\beta} \|z_{k+1}\|_2^2 \leq (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1}.$$

由  $-\lambda^{k-1} \in \partial g(y^{k-1})$  以及  $-\lambda^k \in \partial g(y^k)$  可得:

$$-(y^{k-1} - y^k)^T (\lambda^{k-1} - \lambda^k) \geq 0.$$

记  $r_k = H^{-1}(\beta P^T(y^{k-1} - y^k + \frac{1}{\beta}(\lambda^{k-1} - \lambda^k))), k \geq 1$ , 则

$$\begin{aligned} \|r_k\|_2^2 &\leq \rho^2(H^{-1}P^T) \|\lambda^{k-1} - \lambda^k + \beta(y^{k-1} - y^k)\|_2^2 \\ &\leq \rho^2(H^{-1}P^T) \|\lambda^{k-1} - \lambda^k - \beta(y^{k-1} - y^k)\|_2^2 = \rho^2(H^{-1}P^T) \|z_k\|_2^2. \end{aligned} \quad (3.4)$$

**命题 2.** 设  $\begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  为 IADMM 生成的点列, 则有:

$$\begin{cases} \|\epsilon^1\|_H \leq M \|\epsilon_{initial}^1\|_H, \\ \|\epsilon_{m_0}^{k+1}\|_H \leq \sum_{i=1}^k M^{k-i+1} \|H^{-1}P^T q_i\|_H + M^{k+1} \|\epsilon_{initial}^1\|_H, k \geq 1 \end{cases} \quad (3.5)$$

**证明.**  $\|\epsilon^1\|_H \leq M \|\epsilon_{initial}^1\|_H$  显然成立.

注意到  $x_{initial}^{k+2} = x^{k+1}, k \geq 0$ , 则利用 (2.3) 中的第一式可得:

$$Hx_{initial}^{k+2} = A^T b + \beta P^T(y^k + \frac{1}{\beta}\lambda^k) + H\epsilon_{m_0}^{k+1}, k \geq 0.$$

由定义,  $x_{exact}^{k+2}$  满足:

$$Hx_{exact}^{k+2} = A^T b + \beta P^T(y^{k+1} + \frac{1}{\beta}\lambda^{k+1}), k \geq 0.$$

则有:

$$\epsilon_{initial}^{k+2} = H^{-1}(\beta P^T(y^k - y^{k+1} + \frac{1}{\beta}(\lambda^k - \lambda^{k+1}))) + \epsilon_{m_0}^{k+1}, k \geq 0.$$

因此:

$$\begin{aligned} \|\epsilon_{initial}^{k+1}\|_H &= \|H^{-1}P^T q_k + \epsilon_{n_k}^k\|_H \leq \|H^{-1}P^T q_k\|_H + M \|\epsilon_{initial}^k\|_H \\ &\leq \sum_{i=1}^k M^{k-i} \|H^{-1}P^T q_i\|_H + M^k \|\epsilon_{initial}^1\|_H, k \geq 1. \end{aligned}$$

故:

$$\|\epsilon_{m_0}^{k+1}\|_H \leq M \|\epsilon_{initial}^{k+1}\|_H \leq \sum_{i=1}^k M^{k-i+1} \|H^{-1}P^T q_i\|_H + M^{k+1} \|\epsilon_{initial}^1\|_H.$$

命题得证.

下面的定理给出了本文的主要结论.

**命题 3.** 设  $\begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$  为 IADMM 生成的点列. 若  $M < \frac{1}{1 + \sqrt{\frac{\beta}{2\sigma}} \rho_1 \rho_2}$ ,  $K$

趋向于无穷, 则  $x^K - x^*$  与  $y^K - y^*$  趋向于 0. 其中  $\sigma$  是  $A^T A$  的最小奇异值,  $\rho_1 = \rho(H)$ ,  $\rho_2 = \rho(H^{-1}P)$ .

证明. 对于任意的  $v > 0$ , 由命题 2 得:

$$\begin{aligned}
 \Sigma_{k=0}^{K-1} (x^{k+1} - x^*)^T H \epsilon_{m_0}^{k+1} &= \Sigma_{k=0}^{K-1} \langle H x^{k+1} - x^*, \epsilon_{m_0}^{k+1} \rangle \\
 &\leq \Sigma_{k=1}^{K-1} \|x^{k+1} - x^*\|_H \|\epsilon_{m_0}^{k+1}\|_H + \|x^1 - x^*\|_H \|\epsilon^1\|_H \\
 &\leq \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|x^{k+1} - x^*\|_H \|H^{-1} P^T q_i\|_H \\
 &\quad + \Sigma_{k=1}^{K-1} M^{k+1} \|x^{k+1} - x^*\|_H \|\epsilon_{initial}^1\|_H + M \|x^1 - x^*\|_H \|\epsilon_{initial}^1\|_H \\
 &\leq \frac{v}{2} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|x^{k+1} - x^*\|_H^2 + \frac{1}{2v} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|H^{-1} P^T q_i\|_H^2 \\
 &\quad + \frac{v}{2} \Sigma_{k=0}^{K-1} M^{k+1} \|x^{k+1} - x^*\|_H^2 + \frac{1}{2v} \Sigma_{k=0}^{K-1} M^{k+1} \|\epsilon_{initial}^1\|_H^2.
 \end{aligned}$$

合并有关  $x$  的项, 有:

$$\begin{aligned}
 &\frac{v}{2} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|x^{k+1} - x^*\|_H^2 + \frac{v}{2} \Sigma_{k=0}^{K-1} M^{k+1} \|x^{k+1} - x^*\|_H^2 \\
 &= \frac{v}{2} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|x^{k+1} - x^*\|_H^2 + \frac{v}{2} \Sigma_{k=0}^{K-1} M^{k+1} \|x^{k+1} - x^*\|_H^2 \\
 &= \frac{v}{2} (\Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} + \Sigma_{k=1}^{K-1} M^{k+1}) \|x^{k+1} - x^*\|_H^2 + M \|x^1 - x^*\|_H^2 \\
 &= \frac{v}{2} \Sigma_{k=0}^{K-1} \frac{M - M^{k+2}}{1 - M} \|x^{k+1} - x^*\|_H^2 \\
 &\leq \frac{v}{2} \Sigma_{k=0}^{K-1} \frac{M}{1 - M} \|x^{k+1} - x^*\|_H^2.
 \end{aligned}$$

注意到  $\|H^{-1} P^T q_i\|_H \leq \rho_2 \|z_i\|_H$ , 有:

$$\begin{aligned}
 &\frac{1}{2v} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|H^{-1} P^T q_i\|_H^2 \\
 &\leq \frac{1}{2v} \Sigma_{k=1}^{K-1} \Sigma_{i=1}^k M^{k-i+1} \|H^{-1} P^T q_i\|_H^2 \\
 &\leq \frac{1}{2v} \Sigma_{i=1}^{K-1} \Sigma_{k=i}^{K-1} M^{k-i+1} \|H^{-1} P^T q_i\|_H^2 \\
 &\leq \frac{1}{2v} \Sigma_{i=1}^{K-1} \frac{M - M^{k-i+1}}{1 - M} \|H^{-1} P^T q_i\|_H^2 \\
 &\leq \frac{\rho_2^2}{2v} \Sigma_{i=1}^{K-1} \frac{M}{1 - M} \|z_i\|_H^2.
 \end{aligned}$$

以及:

$$\begin{aligned}
 &\frac{1}{2v} \Sigma_{k=0}^{K-1} M^{k+1} \|\epsilon_{initial}^1\|_H^2 \\
 &\leq \frac{1}{2v} \Sigma_{k=0}^{K-1} M^{k-i+1} \|\epsilon_{initial}^1\|_H^2 \\
 &\leq \frac{1}{2v} \frac{M - M^{k+1}}{1 - M} \|\epsilon_{initial}^1\|_H^2 \\
 &\leq \frac{1}{2v} \frac{M}{1 - M} \|\epsilon_{initial}^1\|_H^2 \\
 &\leq \frac{1}{2v} \frac{M}{1 - M} \|\epsilon_{initial}^1\|_H^2.
 \end{aligned}$$

利用命题 1, 有:

$$\begin{aligned} & \Sigma_{k=0}^{K-1} (x^{k+1} - x^*)^T A^T A (x^{k+1} - x^*) + V_K - V_0 + \frac{1}{2\beta} \Sigma_{k=1}^K \|z_k\|_2^2 \\ & \leq \frac{v}{2} \Sigma_{k=0}^{K-1} \frac{M}{1-M} \|x^{k+1} - x^*\|_H^2 + \frac{\rho_2^2}{2v} \Sigma_{i=1}^{K-1} \frac{M}{1-M} \|z_i\|_H^2 + \frac{1}{2v} \frac{M}{1-M} \|\epsilon_{initial}^1\|_H^2 \end{aligned}$$

由  $M < \frac{1}{1 + \sqrt{\frac{1}{2\sigma} \rho_1 \rho_2}}$  可得:

$$1 > \frac{\beta M^2 \rho_1^2 \rho_2^2}{2\sigma(1-M)^2} = \left( \frac{1}{\sigma} \frac{v}{2} \frac{M}{1-M} \right) (2\beta \rho_2^2 \frac{1}{2v} \frac{M}{1-M}).$$

所以存在  $v_0$  满足:

$$\begin{aligned} \sigma & > \frac{v_0}{2} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p} \\ \frac{1}{2\beta} & > \rho_2^2 \frac{1}{2v_0} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p} \end{aligned}$$

则在

$$\begin{aligned} & \Sigma_{k=0}^{K-1} (\bar{x}^{k+1} - x^*)^T (A^T A - \frac{v_0}{2} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p} I) (\bar{x}^{k+1} - x^*) + \Sigma_{k=1}^{K-1} \left( \frac{1}{2\beta} - \rho_2^2 \frac{1}{2v_0} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p} \right) \|z_k\|_2^2 \\ & + \frac{1}{2\beta} \|z_K\|_2^2 + V_K \leq \frac{\rho_1 \mu_1^p}{2v_0(1 - \mu_1^p)} \|\epsilon_{initial}^1\|_2^2 + V_0 \end{aligned}$$

中, 有  $A^T A - \frac{v_0}{2} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p} I$  正定, 以及  $\frac{1}{2\beta} - \rho_2^2 \frac{1}{2v_0} \frac{\rho_1 \mu_1^p}{1 - \mu_1^p}$  大于零. 故, 当  $K$  趋向于无穷时,  $\bar{x}^K - x^*$  和  $\|z_K\|_2$  趋向于零.

由于  $-\lambda^{K-1} \in \partial g(y^{K-1})$  and  $-\lambda^K \in \partial g(y^K)$ , 则:

$$(y^{K-1} - y^K)(\lambda^{K-1} - \lambda^K) < 0,$$

也就是  $\|\lambda^{K-1} - \lambda^K\|_2^2 + \beta^2 \|y^{K-1} - y^K\|_2^2 < \|z_K\|_2^2$ . 所以当  $K$  趋向于无穷时,  $\lambda^K - \lambda^{K-1}$  趋向于零.

故当  $K$  趋向于无穷时,  $y^K - y^* = \frac{1}{\beta}(\lambda^K - \lambda^{K-1}) + P(x^K - x^*)$  趋向于零.

#### 4. 数值算例

$l_1$  正则化线性回归是一种重要的正则化最小二乘问题, 在统计中也被称为 *lasso*. 即:

$$\min \frac{1}{2} \|Ax - b\|_2^2 + r \|x\|_1,$$

其中  $r$  正则化参数. 在本文的例子中,  $r = 0.1 * \|A^T b\|_\infty$ . 所有的子问题均通过 Gauss-Seidel 迭代法来求解.

我们分别生成一个包含稠密矩阵的算例和一个包含稀疏矩阵的算例. 其中稠密矩阵取  $A_1 = U * V$ ,  $U$  是一个  $1000 \times 1000$  的对角矩阵,  $V$  是一个用随机矩阵正交化后得到的矩阵. 我们取对角矩阵的对角元是最小值为 1 最大值为 50 的等差数列, 则  $A_1^T A_1$  的条件数为 2500. 我们取  $w_1$  与  $v_1$  为随机向量,  $b_1 = A_1 w_1 + 0/1 v_1$ . 稀疏矩阵  $A_2$  取稀疏度为 0.001



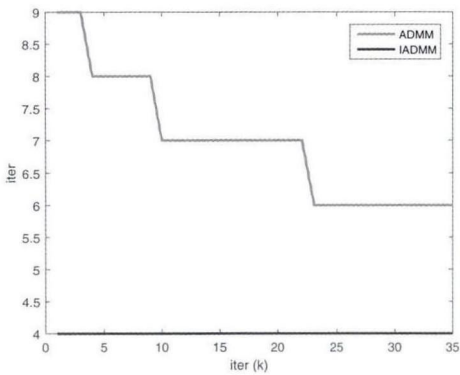


图 1 稠密矩阵情形下在每一步的子问题迭代次数

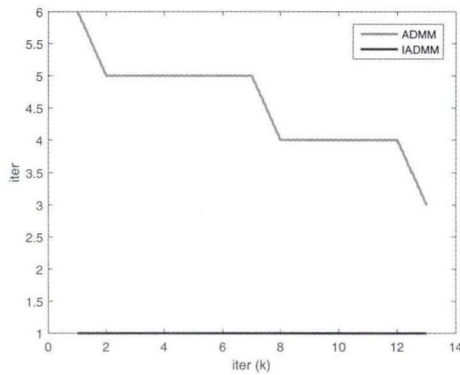


图 2 稀疏矩阵情形下在每一步的子问题迭代次数

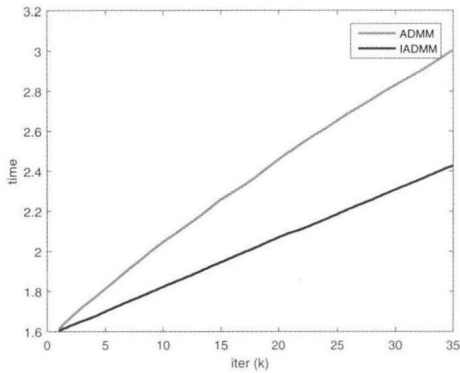


图 3 稠密矩阵情形下在每一步的运行时间

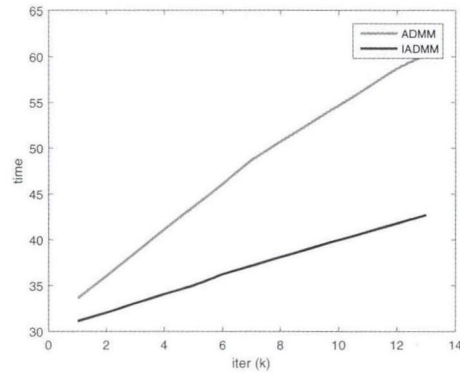


图 4 稀疏矩阵情形下在每一步的运行时间

的  $10^6 \times 10^4$  的矩阵.  $b_2$  的生成方式与稠密矩阵的情形类似, 除了  $w_2$  取稀疏度为 0.01 的随机向量.

我们取初始点  $x^0 = 0, y^0 = 0, \lambda^0 = 0$ , 参数  $\beta = 0.5 * \|A^T b\|_\infty, M = \frac{1}{1 + \sqrt{\frac{\beta}{2\sigma}} \rho_1 \rho_2}$ . 对于精确的 ADMM, 其收敛准则为  $\|x^{k+1} - x^k\|_\infty < 10^{-6}$ . 对于非精确 ADMM 可计算出子问题固定的迭代步数, 稠密矩阵的迭代步数为 4, 而稀疏矩阵的迭代步数为 1.

表 1 IADMM 与 ADMM 的收敛性质比较

$A^T A$	$10^3 \times 10^3$ 稠密矩阵		$10^6 \times 10^4$ 稀疏矩阵	
条件数	2500		52.7894	
算法	ADMM	IADMM	ADMM	IADMM
外循环迭代次数	35	35	13	13
子问题最大迭代次数	9	4	6	1
总时间	3.0671	2.4380	60.2561	42.7101
循环所用时间	1.4684	0.8814	29.8261	12.6674

表 1 中给出了 IADMM 与 ADMM 的收敛性质比较. 其中总时间包含准备用时间和循环用时间两部分, 而准备用时间大部分用于计算矩阵乘法  $A^T A$ . 可以看出, 我们的 IADMM 的子问题迭代步数较少, 解问题所用的时间也较少.

致谢. 感谢马耀楠和罗刚帮助检查论文中的证明, 还感谢审稿人提出的宝贵修改意见.

## 参 考 文 献

- [1] Ng M K, Wang F, Yuan X. Inexact alternating direction methods for image recovery[J]. SIAM Journal on Scientific Computing, 2011, 33(4): 1643-1668.
- [2] Eckstein J, Bertsekas D P. On the Douglas Rachford splitting method and the proximal point algorithm for maximal monotone operators[J]. Mathematical Programming, 1992, 55(1-3): 293-318.
- [3] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends® in Machine Learning, 2011, 3(1): 1-122.
- [4] Varga R S. Matrix iterative analysis[M]. Springer Science & Business Media, 2009.
- [5] Ben-Tal A, Nemirovski A. Lectures on modern convex optimization: analysis, algorithms, and engineering applications[M]. Siam, 2001.
- [6] Hansen P C, Nagy J G, O'leary D P. Deblurring images: matrices, spectra, and filtering[M]. Siam, 2006.
- [7] Parikh N, Boyd S P. Proximal algorithms[J]. Foundations and Trends in optimization, 2014, 1(3): 127-239.
- [8] Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM review, 2010, 52(3): 471-501.
- [9] Toh K C, Yun S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems[J]. SIAM review, 2010, 15(6): 615-640.
- [10] Trefethen L N, Bau III D. Numerical linear algebra[M]. Siam, 1997.