

Master's Thesis, Zhengzhou University, No. 201012141481

**Conjugate gradient methods and Nesterov-type
methods for solving Lasso problem**

Candidate: Zhang Xue

Supervisor: Dong Yunda

Speciality: Operations Research and Cybernetics

School of Mathematics and Statistics , Zhengzhou University

Zhengzhou, 450001, P. R. China

April, 2013

原创性声明



本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

学位论文作者：张雪

日期：2013年5月30日

学位论文使用授权声明

本人在导师指导下完成的论文及相关的职务作品，知识产权归属郑州大学。根据郑州大学有关保留、使用学位论文的规定，同意学校保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权郑州大学可以将本学位论文的全部或部分编入有关数据库进行检索，可以采用影印、缩印或者其他复制手段保存论文和汇编本学位论文。本人离校后发表、使用学位论文或与该学位论文直接相关的学术论文或成果时，第一署名单位仍然为郑州大学。保密论文在解密后应遵守此规定。

学位论文作者：张雪

日期：2013年5月30日

摘 要

信号和图像恢复问题往往可以转化为求解一个大规模的待定的线性系统的稀疏解. 而这种待定的线性系统的一种逼近便是 Lasso 问题. Lasso 问题是凸极小化问题, 向前向后分裂算法是求解该问题的一种重要方法. 为了提高算法的收敛率, 向前向后分裂算法的加速变形被提出. 本文对 Salzo[20]等提出的一种加速变形进行了收敛性分析, 并通过数值试验检验其效果. 本文还考虑了 Lasso 问题的一个相关变形, 其目标函数是可微的, 此问题可以用非线性共轭梯度法来求解, 本文通过数值试验检验了 Dong[4] 提出的PR+算法和 Li[14]等提出的RMPRP算法的效果.

关键词 :光滑化; 共轭梯度; 向前向后分裂; 收敛率.

Abstract

Signal and image restoration problems are often related to finding the sparsest solution of large-scale and under-determined linear system. An approximation to under-determined linear systems is usually called Lasso problem. This problem is a convex minimization problem. The forward-backward splitting operator method is important for solving it. Accelerated version was proposed to improve the method's convergence rate. In this thesis, we prove the convergence rate of an inexact accelerated forward-backward splitting method proposed by Salzo[20] and show its effectiveness by preliminary experiment. We consider a variant of Lasso problem. Its objective function is differentiable. This problem can be solved by nonlinear conjugate gradient method. We show the effectiveness of PR+ method proposed by Dong[4] and RMPRP method proposed by Li[14] by preliminary experiment.

Key Words : smoothing; conjugate gradient; forward-backward splitting; convergence rate .

目 录

第一章 绪论	1
1.1 引言	1
1.2 符号和预备知识	5
1.3 主要研究内容	9
第二章 一类非精确加速向前向后分裂算法的收敛率	10
2.1 算法	10
2.2 收敛率	11
2.3 数值试验	17
第三章 Lasso 问题的一种变形	19
3.1 Lasso 问题的一种光滑变形	19
3.2 非线性共轭梯度法的一些近期结果	20
3.3 数值试验	21
参考文献	24
后 记	27
致 谢	28

求解Lasso问题的共轭梯度法和 Nesterov型算法

第一章 绪 论

1.1 引言

在图像形成、传送和存储的过程中, 由于摄像机与目标物的相对运动、设备造成的系统误差、畸变、噪声等因素的影响, 我们得到的图像往往不是真实的完善映像. 在成像过程中受到各种因素的影响, 导致了图像质量的降低, 称之为图像退化. 图像恢复就是通过计算机处理, 对质量下降的图像加以重建或恢复的处理过程. 图像恢复首先需建立造成图像质量下降的退化模型, 然后运用相反过程来恢复原来图像, 并运用一定准则来判定是否得到图像的最佳恢复.

图像恢复问题在医学、生物工程和其它科学与工程领域内都有很大的作用. 最常用的图像退化模型如下

$$b = Ax + \eta, \quad (1.1.1)$$

其中, $\eta \in R^m$ 是未知的扰动因素, A 是 $m \times n$ 的矩阵(通常 $m < n$), $b \in R^n$ 是观测到的图像 $x \in R^n$ 是未知的真实完善映像. 我们的目的就是找到系统 (1.1.1) 的稀疏解, 即非 0 元素个数最少的解. 对于系统 (1.1.1), 有两个重要因素: 首先, 如何根据已有的信息确定矩阵 A 中的元素; 其次, 在足够的信息下, 采用何种算法来有效快速的求解系统 (1.1.1). 本文只关注第二个因素.

由于 η 的未知性, 所以并不能直接求解系统(1.1.1), 而是来解下面的与之相关的线性约束问题

$$\min_{x \in R^n} \{\|x\|_0 : Ax = b\}, \quad (1.1.2)$$

其中 $\|x\|_0$ 表示向量 x 的非零分量的个数. 一般来说, 求解问题 (1.1.2) 需要用到组合优化, 这显然并不实用. 在某种程度上, 1-范数是 0-范数的凸松弛. 所以, 可以考虑与 (1.1.2)

密切相关的凸约束优化问题

$$\min_{x \in R^n} \{\|x\|_1 : Ax = b\}. \quad (1.1.3)$$

问题 (1.1.3) 对求解稀疏解有很重要的作用. 实际上, 在某些条件下 (见[5]) 问题 (1.1.2) 和 (1.1.3) 有相同的解.

实际上, 为了解决问题 (1.1.3), 许多人考虑它的各种变形来求解稀疏逼近解. 如

$$\min_{x \in R^n} \{\|x\|_1 : \|Ax - b\| \leq \delta\}$$

和

$$\min_{x \in R^n} \{\|Ax - b\| : \|x\|_1 \leq \epsilon\},$$

其中, δ 和 ϵ 是非负实参数. 前一个问题是二次约束线性规划, 后一个问题是二次规划. 而经常用的便是 Lasso 问题

$$\min_{x \in R^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \tau \|x\|_1 \right\},$$

其中, τ 是正则化参数. 显然, Lasso 问题的目标函数是凸函数, 但并不一定强凸且是不可微的.

对于 Lasso 问题(为了叙述方便, 记 $f(x) := \|Ax - b\|^2/2$, $g(x) := \tau \|x\|_1$). 在过去的一些年里, 很多算法被提出和研究. 虽然 Lasso 问题是结构简单的凸优化问题, 但在许多实际问题 (比如图像恢复) 中, Lasso 问题不仅规模大, 而且矩阵 A 中的元素往往并不稀疏. 所以一些算法, 比如内点法, 往往并不能有效快速的解决该问题.

在文献 [6] 中, Figueiredo, Nowak 和 Wright 提出的梯度投影算法是解决 Lasso 问题的最早的算法之一. 在这篇文章中, 作者将 Lasso 问题转化为箱型约束二次规划, 其过程如下: 令 $x = u - v$, $u, v \geq 0$, 其中 $u_i = \max(0, x_i)$, $v_i = -\min(0, x_i)$, 将上述变量分解应用于 Lasso 问题则有

$$\begin{aligned} \min \quad & \frac{1}{2} \|b - A(u - v)\|^2 + \tau 1_n^T u + \tau 1_n^T v \\ \text{s.t.} \quad & u \geq 0, v \geq 0. \end{aligned}$$

上述问题可以进一步转化为标准的箱型约束下的二次规划问题

$$\begin{aligned} \min_{u,v} \quad & \frac{1}{2} z^T B z + c^T z := f(z) \\ \text{s.t.} \quad & z \geq 0, \end{aligned}$$

其中

$$z = \begin{bmatrix} u \\ v \end{bmatrix}, c = \tau 1_{2n} + \begin{bmatrix} -A^T b \\ A^T b \end{bmatrix}, B = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}.$$

而求解上述箱型约束下的二次规划问题的投影梯度法为

1. 给定 $z_0 \in R^n$, 选择参数 $\beta \in (0, 1), \mu \in (0, 0.5)$, 置 $k := 0$.
2. 当 $k \geq 1$ 时, 若某停止准则满足, 停止. 否则, 计算

$$\rho_k = \frac{g_k^T g_k}{g_k^T B g_k},$$

选取 $\alpha_k = \max\{\rho_k t^j : j = 0, 1, 2, \dots\}$, 使得

$$f(P_+(z_k - \alpha_k g_k)) \leq f(z_k) - \mu g_k^T (z_k - P_+(z_k - \alpha_k g_k)).$$

3. 计算下一个迭代点 $z_{k+1} = P_+(z_k - \alpha_k g_k)$. $k := k + 1$, 转步1.

求解 Lasso 问题等价于求解单调包含问题

$$0 \in \nabla f(x) + \partial g(x).$$

对于单调包含问题, 向前向后分裂算法是被研究的最多的算法之一, 其迭代格式为

$$x_{k+1} = (I + \alpha_k \partial g)^{-1} (I - \alpha_k \nabla f)(y_k),$$

其中 α_k 是某些合适的步长.

1999 年, Bakin[?] 提出了下面的群 Lasso 问题, 即

$$\min_{x \in R^n} \frac{1}{2} \|Ax - b\| + f(x), \quad f(x) = \sum_{J \in \mathcal{J}} \omega_J \|x_J\|_{K_J},$$

其中, $\|x_J\|_{K_J} = (x_J^T K_J x_J)^{\frac{1}{2}} = \|B_J x_J\|$ ($K_J = B_J^T B_J$), \mathcal{J} 中的元素是 $\{1, \dots, n\}$ 的一个划分, $x_J \in R^{|J|}$, B_J 是 $|J|$ 阶方阵. 当 $K_J = I$ 时, 上面的问题可以简化为

$$\min_{x \in R^n} \frac{1}{2} \|Ax - b\| + \sum_{J \in \mathcal{J}} \omega_J \|x_J\|. \quad (1.1.4)$$

当 $\forall J \in \mathcal{J}$ 有 $|J| = 1$ 且 $\omega_J = \tau$ 时, 问题 (1.1.4) 便简化为标准 Lasso 问题.

在 2010 年, Tseng[24] 证明了: 在不要求目标函数的强凸性的情况下, 求解问题 (1.1.4) 的向前向后分裂算法是线性收敛的. 在文献[26]中, 张海斌将 Tseng 的结果推广到了群 Lasso 问题上, 当每一个 $|J|$ 都比较小时, 文中建议用 Newton 法来求解子问题

$$(x_k - \alpha_k A^T(Ax_k - b))_J = x_J + \alpha_k B_J^T B_J x_J / \|B_J x_J\|.$$

为了提升算法求解问题的速度, 许多研究者给出了向前向后分裂算法的一些加速变形. 一类为

$$x_k = (I + \lambda_k \partial g)^{-1}(I - \lambda_k \nabla f)(y_k),$$

$$y_{k+1} = x_k + \alpha_k(x_k - x_{k-1}),$$

其中参数 $\lambda_k > 0$, $\alpha_k > 0$.

Nesterov 最早用上述加速变形来设计算法(见文献[17]). 在 2008 年, Tseng ([23]) 提出了另一个版本, 取

$$\alpha_k = \frac{\theta_k(1 - \theta_k)}{\theta_{k-1}},$$

其中 θ_k 满足条件

$$\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2},$$

并证明了其收敛率为 $\mathcal{O}(\sqrt{L_f/\epsilon})$.

在 2009 年, Beck 和 Teboulle [1] 提出了快速压缩阈算法(Fast Iterative Shrinkage-Thresholding Algorithm), 其迭代格式为

$$x_k = (I + \lambda_k \partial g)^{-1}(I - \lambda_k \nabla f)(y_k),$$

$$t_1 = 1, \quad t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2,$$

$$y_{k+1} = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}).$$

当 ∇f 是 Lipschitz 连续(记 Lipschitz 常数为 L_f)时, $\lambda_k \equiv 1/L_f$; 当 ∇f 不是 Lipschitz 连续时, 作者给出了一种自适应选取方法, 且在这两种情况下, 文中证明了目标函数值的复杂度均为

$$f(x_k) + g(x_k) - (f(x^*) + g(x^*)) \leq \mathcal{O}(1/k^2).$$

在2012年, Salzo 和 Villa[20] 提出了一个更一般的非精确加速形式

$$\begin{aligned}x_k &\cong_{\varepsilon_k} (I + \lambda_k \partial g)^{-1}(I - \lambda_k \nabla f)(y_k), \\t_1 &= 1, \quad t_{k+1} = (1 + \sqrt{1 + 4\lambda_k t_k^2 / \lambda_{k+1}}) / 2, \\y_{k+1} &= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}).\end{aligned}$$

并在两种非精确类型下证明了算法的收敛率.

在 2012 年, 李全燕[15]在其硕士毕业论文中构造了下面的加速变形

$$\begin{aligned}x_k &= (I + \lambda_k \partial g)^{-1}(I - \lambda_k \nabla f)(y_k), \\y_{k+1} &= x_k + \frac{k-1}{k+1}(x_k - x_{k-1}).\end{aligned}$$

且说明了上述算法并不是快速压缩阀算法的特例. 同样的, 当 ∇f 是 Lipschitz 连续 (记 Lipschitz 常数为 L_f) 时, $\lambda_k \equiv 1/L_f$; 当 ∇f 不是 Lipschitz 连续时, 类似文献[1]中, 作者给出了一种自适应选取方法, 且在这两种情况下, 文中证明了目标函数值的复杂度均为

$$f(x_k) + g(x_k) - (f(x^*) + g(x^*)) \leq \mathcal{O}(1/k).$$

虽然理论分析上的收敛率没有快速压缩阀算法快, 但她经过数值试验说明了: 在一些实际的例子中, 上述算法解决问题的速度并不比快速压缩阀算法慢.

1.2 符号和预备知识

空间 R^n 中任意一点 x , 本文设定其是列向量, 记其第 i 个分量为 x_i .

定义 1.2.1 空间 R^n 中任意两点 x, y , 定义它们的内积为

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

由上述内积诱导的范数

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\left(\sum_{i=1}^n x_i^2\right)}$$

是2-范数. 定义 x 的1-范数为

$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$

定理 1.2.2 对于空间 R^n 中任意两点 x, y , 有下面不等式成立

$$\langle x, y \rangle \leq \|x\| \|y\|,$$

这就是Cauchy-Schwarz不等式.

定义 1.2.3 空间 R^n 中的任意子集 C 是凸集当且仅当对任意的 $x, y \in C$ 及 $t \in [0, 1]$ 有下述关系成立

$$tx + (1 - t)y \in C.$$

若 C 还是闭集, 则称 C 是闭凸集.

可证明若 C 是凸集, 则 C 的内部和闭包也是凸集.

定义 1.2.4 设 C 是 R^n 中的非空的闭凸集, 如果函数 $f: C \rightarrow R \cup \{+\infty\}$ 对 C 中的任意一点 x 都有 $f(x) > -\infty$, 且存在一点 $x \in C$ 使得 $f(x) < +\infty$, 那么称函数 f 是正则函数.

定义 1.2.5 设 C 是 R^n 中的非空的闭凸集, 如果对 C 中的任意两点 x, y 和任意的 $t \in [0, 1]$, 正则函数 $f: C \rightarrow R \cup \{+\infty\}$ 满足

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y),$$

就称函数 f 是 C 上的凸函数.

根据定义可证明: 当 f, g 均是非空闭凸集 C 上凸函数时, 对任意的数 $t_1, t_2 \geq 0$, $t_1 f + t_2 g$ 也是 C 上的凸函数.

定义 1.2.6 设 C 是 R^n 中的非空的闭凸集, 如果正则函数 $f: C \rightarrow R \cup \{+\infty\}$ 可微, 且对 C 中的任意两点 x, y , 存在 $\mu > 0$ 使得

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2,$$

成立, 就称 f 是 C 上的 μ 强凸函数.

定义 1.2.7 函数 $f: R^n \rightarrow R \cup \{+\infty\}$, 如果其水平集

$$\{x \in R^n : f(x) \leq a\}$$

是闭集或其上图

$$\text{epi} f := \{(x, r) \in R^n \times R : f(x) \leq r, f(x) \neq \infty\}$$

是闭集, 或其下半连续, 即任对意的 $x \in R^n$ 有

$$\liminf_{y \rightarrow x} f(y) \geq f(x), \quad y \in R^n,$$

就称函数 f 是闭的.

定理 1.2.8 如果 $f : R^n \rightarrow R \cup \{+\infty\}$ 是可微函数, 则 f 是凸的当且仅当对任意的 $x, y \in R^n$ 有

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

定义 1.2.9 设 C 是 R^n 中的非空的闭凸集, $f : C \rightarrow R \cup \{+\infty\}$ 是凸函数, f 在点 $x \in R^n$ 处的次微分定义为

$$\partial f(x) := \{s \in R^n : f(y) \geq f(x) + \langle s, y - x \rangle, \quad \forall y \in R^n\}.$$

称 $s \in \partial f(x)$ 是 f 在点 x 处的次梯度. 当 f 可微时, s 就是 f 的梯度.

定义 1.2.10 设 C 是 R^n 中的非空的闭凸集, $f : R^n \rightarrow R \cup \{+\infty\}$ 是凸函数, f 在点 $x \in R^n$ 处的 ε -次微分定义为

$$\partial_\varepsilon f(x) := \{s \in R^n : f(y) \geq f(x) + \langle s, y - x \rangle - \varepsilon, \quad \forall y \in R^n\}.$$

定理 1.2.11 设 $f : R^n \rightarrow R$ 是凸函数, 若对 $x \in R^n$, $f(x) < +\infty$, 则 $\partial f(x)$ 是非空有界闭集.

定理 1.2.12 $C \subseteq R^n$, 称函数 $f : C \rightarrow R$ 在 C 上 lipschitz 连续, 如果存在常数 $L > 0$, 使得

$$\|f(x) - f(y)\| \leq L\|x - y\|, \quad \forall x, y \in C.$$

定义 1.2.13 算子 $T : R^n \rightarrow 2^{R^n}$ 是多值映射, 如果对 R^n 中的任意两点 x, y , 算子 T 满足

$$\langle x - y, \omega - \nu \rangle \geq 0, \quad \forall \omega \in T(x), \nu \in T(y),$$

则称算子 T 是单调算子. 如果对 R^n 中的任意两点 x, y , 算子 T 还满足

$$\langle x - y, \omega - \nu \rangle \geq 0, \quad \forall \omega \in T(x) \Rightarrow \nu \in T(y),$$

则称算子 T 是极大单调算子.

由定义可证: 若算子 T 是极大单调算子, 则 αT , $\alpha > 0$ 也是极大单调的.

定义 1.2.14 称函数 $f: C \rightarrow R, C \subseteq R^n$ 在 C 上是 μ -单调的, 如果存在常数 μ , 使得

$$\langle f(x) - f(y), x - y \rangle \geq \mu \|x - y\|^2, \quad \forall x, y \in C.$$

当 $\mu > 0$ 时, f 是 μ 强单调的; 当 $\mu = 0$ 时, f 是单调的; 当 $\mu < 0$ 时, f 是弱单调的. 当 f 是L-lipschitz 连续且 μ -单调时, 有

$$L\|x - y\|^2 \geq \langle f(x) - f(y), x - y \rangle \geq -\|f(x) - f(y)\|\|x - y\| \geq -L\|x - y\|^2,$$

得到 $|\mu| \leq L$.

定理 1.2.15 设 $f: R^n \rightarrow R$ 是正则闭凸函数, 则其次微分是极大单调的.

定义 1.2.16 若 T 是极大单调算子, 称 $J_T = (I + T)^{-1}$ 为 T 的预解式.

定理 1.2.17 若 $T: R^n \rightarrow 2^{R^n}$ 是极大单调算子, 对任意的 $\alpha > 0$, $J_{\alpha T} = (I + \alpha T)^{-1}$ 是非扩张的.

证明: 对任意的 $x, y \in R^n$, 记 $u = J_{\alpha T}(x)$, $v = J_{\alpha T}(y)$, 立即有

$$x - u \in \alpha T(u), \quad y - v \in \alpha T(v).$$

由算子 T 的单调性和 $\alpha > 0$ 可得

$$\langle u - v, x - u - y + v \rangle \geq 0,$$

即 $\|u - v\|^2 \leq \langle u - v, x - y \rangle$.

由Cauchy-Schwarz不等式进一步可得

$$\|J_{\alpha T}(x) - J_{\alpha T}(y)\| = \|u - v\| \leq \|x - y\|$$

则 $J_{\alpha T}$ 是非扩张的. □

对矩阵 $A \in R^m \times R^n$, A^T 表示矩阵的转置, 1_n 表示分量全为 1 的 n 维向量. $P_+(\cdot)$ 表示第一卦限上的投影. $\nabla f(x)$ 代表函数 $f(x)$ 的梯度, k 一般代表算法中的迭代数, f_k 代表 $f(x_k)$. $x \geq 0$ 表示向量 x 的每个分量都是非负的, $x > 0$ 表示向量 x 的每个分量都是正数.

1.3 主要研究内容

第二章证明了 Salzo[20]等提出的加速的向前向后分裂算法进行了收敛性分析, 通过数值试验检验其效果, 并与快速压缩阈算法进行比较.

第三章考虑了 Lasso 问题的一个相关变形, 其目标函数是可微的. 对于此问题, 本文通过数值试验检验了 Dong[4] 提出的PR+算法和 Li[14]等提出的RMPPRP算法的效果, 并与快速压缩阈算法进行比较。

第二章 一类非精确加速向前向后分裂算法的收敛率

2.1 算法

考虑下面问题

$$\min f(x) + g(x) := h(x)$$

其中 $g: R^n \rightarrow R$ 正则闭凸, $f: R^n \rightarrow R$ 可微凸且 ∇f Lipschitz 连续, 记其 Lipschitz 常数为 L . Lasso 问题 $\min \|Ax - b\|^2 + \tau \sum |x_i|$ 可以看作上述问题的一个特例.

上述问题的最优性条件为

$$0 \in \nabla f(x) + \partial g(x),$$

向前向后分裂算法是求解上述单调包含问题的有效算法. 它的一种加速变形为

$$x_k = (I + \lambda_k \partial g)^{-1}(I - \lambda_k \nabla f)(y_k),$$

$$y_{k+1} = x_k + \alpha_k(x_k - x_{k-1}),$$

其中 $\lambda_k > 0, \alpha_k > 0$. 在文献[1]中, Beck 和 Teboulle 提出了一个快速压缩熵算法(FISTA)在上述迭代中, 取

$$\lambda_k = 1/L_f, \quad \alpha_k = \frac{t_k - 1}{t_{k+1}},$$

其中数列 t_k 满足

$$t_1 = 1, \quad t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2, \quad k = 1, 2, \dots,$$

Beck 和 Teboulle 证明了算法的复杂度为

$$f(x_k) + g(x_k) - (f(x^*) + g(x^*)) \leq \mathcal{O}(1/k^2).$$

在文献[20]中, Salzo 和 Villa 提出了一个更一般的加速形式, 并在两种非精确情形下证明了算法的收敛率. 本章主要研究如下的非精确算法.

算法2.1

Step 0. 取 $x_0 \in R^n$, $y_1 = x_0$, $t_1 = 1$, $\lambda_i \in (0, 1/L]$, $e_i \in R^n$, $i = 1, 2, \dots$

Step 1. ($k \geq 1$) 计算

$$\begin{aligned} x_k &= (I + \lambda_k \partial g)^{-1}(y_k - \lambda_k(\nabla f(y_k) + e_k)), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4\lambda_k t_k^2 / \lambda_{k+1}}}{2}, \\ y_{k+1} &= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}). \end{aligned}$$

Step 2. 如果某一停止准则成立, 算法停止.

Step 3. 否则, $k=k+1$.

2.2 收敛率

在本节, 我们证明算法2.1的收敛率. 先看下面三个引理.

引理 2.2.1 记

$$P_\lambda(y) = (I + \lambda \partial g)^{-1}(y - \lambda(\nabla f(y) + e)).$$

如果 $g: R^n \rightarrow R$ 正则闭凸, $f: R^n \rightarrow R$ 可微凸且 ∇f L -Lipschitz连续, 则对于任意的 $x, y \in R^n$, $\lambda \in (0, 1/L]$ 总有

$$f(x) - f(y) \leq \langle \nabla f(y), x - y \rangle + \frac{1}{2\lambda} \|x - y\|^2, \quad (2.2.1)$$

$$h(x) - h(P_\lambda(y)) \geq \frac{1}{2\lambda} \|P_\lambda(y) - y\|^2 + \frac{1}{\lambda} \langle y - x, P_\lambda(y) - y \rangle + \langle e, P_\lambda(y) - x \rangle \quad (2.2.2)$$

证明: 根据Taylor定理有

$$f(x) - f(y) = \int_0^1 df(y + t(x - y)) = \int_0^1 \langle \nabla f(y + t(x - y)), x - y \rangle dt.$$

考察最后一项的被积部分, 易得

$$\begin{aligned} & \langle \nabla f(y + t(x - y)), x - y \rangle \\ &= \langle \nabla f(y), x - y \rangle + \langle \nabla f(y + t(x - y)) - \nabla f(y), x - y \rangle \\ &\leq \langle \nabla f(y), x - y \rangle + tL\|x - y\|^2. \end{aligned}$$

对上式两边关于 t 积分就可以得到

$$\begin{aligned} f(x) - f(y) &\leq \langle \nabla f(y), x - y \rangle + \int_0^1 tL\|x - y\|^2 dt \\ &= \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2. \end{aligned}$$

又由已知得 $1/\lambda \geq L$, 则第一个不等式得证.

下面证第二个不等式成立. 对任意的 $y \in R^n$, $P_\lambda(y)$ 的定义等价于

$$\frac{1}{\lambda}(y - P_\lambda(y)) - \nabla f(y) - e \in \partial g(P_\lambda(y)).$$

由次微分的定义得到, 对任意的 $x \in R^n$ 有

$$g(x) \geq g(P_\lambda(y)) + \langle \frac{1}{\lambda}(y - P_\lambda(y)) - \nabla f(y) - e, x - P_\lambda(y) \rangle. \quad (2.2.3)$$

另一方面, 由 f 的凸性得到, 对任意的 $x, y \in R^n$ 有

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle.$$

并令(2.2.1)中的 $x = P_\lambda(y)$, 则有

$$f(y) \geq f(P_\lambda(y)) - \langle \nabla f(y), P_\lambda(y) - y \rangle - \frac{1}{2\lambda}\|P_\lambda(y) - y\|^2.$$

将上面两个不等式相加得到

$$f(x) \geq f(P_\lambda(y)) + \langle \nabla f(y), x - P_\lambda(y) \rangle - \frac{1}{2\lambda}\|P_\lambda(y) - y\|^2.$$

将上式与不等式 (2.2.3) 相加, 并且由 $h = f + g$, 我们得到

$$\begin{aligned} h(x) &\geq h(P_\lambda(y)) + \frac{1}{\lambda}\langle y - P_\lambda(y), x - P_\lambda(y) \rangle - \langle e, x - P_\lambda(y) \rangle - \frac{1}{2\lambda}\|P_\lambda(y) - y\|^2 \\ &= h(P_\lambda(y)) + \frac{1}{\lambda}\langle x - y, y - P_\lambda(y) \rangle + (\frac{1}{\lambda} - \frac{1}{2\lambda})\|P_\lambda(y) - y\|^2 - \langle e, x - P_\lambda(y) \rangle \\ &\geq h(P_\lambda(y)) + \frac{1}{2\lambda}\|P_\lambda(y) - y\|^2 + \frac{1}{\lambda}\langle y - x, P_\lambda(y) - y \rangle + \langle e, P_\lambda(y) - x \rangle. \end{aligned}$$

从而引理得证.

□

引理 2.2.2 $\lambda > 0$, $\{\lambda_k\} \subseteq [\lambda, 1/L]$, $t_1 = 1$,

$$t_{k+1} = \frac{1 + \sqrt{1 + 4\lambda_k t_k^2 / \lambda_{k+1}}}{2},$$

则对任意的 $k \geq 1$, $t_k \geq k\sqrt{\lambda L}/2$.

证明: 当 $k = 1$ 时, $t_1 = 1 \geq \sqrt{\lambda L}/2$.

$$\begin{aligned} t_{k+1} &\geq \frac{1}{2} + \sqrt{\frac{\lambda_k}{\lambda_{k+1}}} t_k \\ &\geq \frac{1}{2} + \sqrt{\frac{\lambda_k}{\lambda_{k+1}}} \left(\frac{1}{2} + \sqrt{\frac{\lambda_{k-1}}{\lambda_k}} t_{k-1} \right) \\ &= \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\lambda_k}{\lambda_{k+1}}} + \sqrt{\frac{\lambda_{k-1}}{\lambda_{k+1}}} t_{k-1} \\ &\geq \frac{1}{2} + \sum_{i=1}^k \frac{1}{2} \sqrt{\frac{\lambda_i}{\lambda_{k+1}}} \\ &\geq \frac{1}{2} + \frac{k\sqrt{\lambda L}}{2} \geq \frac{(k+1)\sqrt{\lambda L}}{2}. \end{aligned}$$

□

下面的引理来自于[21], 为了论文的完整性, 我们仍然给出证明.

引理 2.2.3 已知非负数列 $\{a_k\}_0^\infty$ 满足

$$a_k^2 \leq c_k + \sum_{i=1}^k a_i b_i, \quad k \geq 1,$$

其中 $\{c_k\}_0^\infty$ 是单调非减数列且 $c_0 \geq a_0^2$, $\{b_k\}_1^\infty$ 是非负数列. 则对任意的 $k \geq 0$, 下面的不等式成立

$$a_k \leq \frac{1}{2} \sum_{i=1}^k b_i + \left(c_k + \left(\frac{1}{2} \sum_{i=1}^k b_i \right)^2 \right)^{\frac{1}{2}} \quad (2.2.4)$$

证明: 用归纳法证明.

当 $k = 0$ 时, 由已知立即得到不等式 (2.2.4) 成立.

当 $k = n - 1$ 时, 假设不等式 (2.2.4) 成立. 记 $M_{n-1} = \max\{a_0, a_1, \dots, a_{n-1}\}$, 则当 $k = n$ 时, 由已知得

$$\left(a_n - \frac{b_n}{2} \right)^2 \leq c_n + \frac{b_n^2}{4} + M_{n-1} \sum_{i=1}^{n-1} b_i.$$

则有

$$a_n \leq \frac{b_n}{2} + \left(c_n + \frac{b_n^2}{4} + M_{n-1} \sum_{i=1}^{n-1} b_i \right)^{\frac{1}{2}}.$$

进一步可得

$$a_n \leq \max\{M_{n-1}, \frac{b_n}{2} + \left(c_n + \frac{b_n^2}{4} + M_{n-1} \sum_{i=1}^{n-1} b_i \right)^{\frac{1}{2}}\}.$$

如果上述不等式右边的两项相等, 可得 $M_{n-1}^2 = c_n + M_{n-1} \sum_{i=1}^{n-1} b_i$, 又 $M_{n-1} \geq 0$, 则方程的根为

$$M_{n-1}^* = \frac{1}{2} \sum_{i=1}^n b_i + \left(c_n + \left(\frac{1}{2} \sum_{i=1}^n b_i \right)^2 \right)^{\frac{1}{2}}.$$

若 $M_{n-1} \leq M_{n-1}^*$, 由于 M_{n-1} , $\frac{b_n}{2} + \left(c_n + \frac{b_n^2}{4} + M_{n-1} \sum_{i=1}^{n-1} b_i \right)^{\frac{1}{2}}$ 都是关于 M_{n-1} 的递增函数, 所以 $M_n \leq M_{n-1}^*$.

若 $M_{n-1} > M_{n-1}^*$, 则有

$$M_{n-1} \geq \frac{b_n}{2} + \left(c_n + \frac{b_n^2}{4} + M_{n-1} \sum_{i=1}^{n-1} b_i \right)^{\frac{1}{2}}.$$

因此 $M_n \leq M_{n-1}$, 进一步可得

$$a_n \leq M_n \leq M_{n-1} \leq \frac{1}{2} \sum_{i=1}^k b_i + \left(c_k + \left(\frac{1}{2} \sum_{i=1}^k b_i \right)^2 \right)^{\frac{1}{2}}$$

综上, 引理得证. □

定理2.2.4 假设问题的解集 X^* 非空. 设 x_k 为该方法生成的序列, $\lambda_k \in [\lambda, 1/L] \subseteq (0, 1/L]$, 则对于 $k \geq 1$, 必有

$$h(x_k) - h(x^*) \leq \frac{2}{k^2 L \lambda^2} \left(2A_k + (\|x_0 - x^*\|^2 + 2\lambda_1 \|e_1\| \|x_1 - x^*\|)^{\frac{1}{2}} \right)^2.$$

其中 $A_k = \sum_{i=1}^k \lambda_i t_i \|e_i\|$.

证明: 为了叙述方便, 记

$$\sigma_k = h(x_k) - h(x^*), \quad u_k = t_k x_k - (t_k - 1)x_{k-1} - x^*.$$

首先, 令引理 2.2.1 中的 (x, y) 分别取为 (x_k, y_{k+1}) , (x^*, y_{k+1}) , $x^* \in X^*$, 并根据算法2.1中的迭代 $x^{k+1} = P_{\lambda_{k+1}}(y^{k+1})$ 有

$$\begin{aligned} h(x_k) - h(x_{k+1}) &\geq \frac{1}{2\lambda_{k+1}} \|x_{k+1} - y_{k+1}\|^2 + \frac{1}{\lambda_{k+1}} \langle y_{k+1} - x_k, x_{k+1} - y_{k+1} \rangle + \langle e_{k+1}, x_{k+1} - x_k \rangle, \\ h(x^*) - h(x_{k+1}) &\geq \frac{1}{2\lambda_{k+1}} \|x_{k+1} - y_{k+1}\|^2 + \frac{1}{\lambda_{k+1}} \langle y_{k+1} - x^*, x_{k+1} - y_{k+1} \rangle + \langle e_{k+1}, x_{k+1} - x^* \rangle. \end{aligned}$$

令第一个不等式乘以 $t_{k+1} - 1$ 后, 与第二个不等式相加, 则有

$$\begin{aligned} &(t_{k+1} - 1)\sigma_k - t_{k+1}\sigma_{k+1} \\ &\geq \frac{t_{k+1}}{2\lambda_{k+1}} \|x_{k+1} - y_{k+1}\|^2 + \frac{1}{\lambda_{k+1}} \langle x_{k+1} - y_{k+1}, t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^* \rangle \\ &\quad + \langle e_{k+1}, t_{k+1}x_{k+1} - (t_{k+1} - 1)x_k - x^* \rangle. \end{aligned}$$

两边乘以 $\lambda_{k+1}t_{k+1}$ 后, 利用 $\lambda_k t_k^2 / \lambda_{k+1} = t_{k+1}^2 - t_{k+1}$, 我们就得到了

$$\begin{aligned} &\lambda_k t_k^2 \sigma_k - \lambda_{k+1} t_{k+1}^2 \sigma_{k+1} \\ &\geq \frac{1}{2} \|t_{k+1}x_{k+1} - t_{k+1}y_{k+1}\|^2 + \langle t_{k+1}x_{k+1} - t_{k+1}y_{k+1}, t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^* \rangle \\ &\quad + \lambda_{k+1} t_{k+1} \langle e_{k+1}, t_{k+1}x_{k+1} - (t_{k+1} - 1)x_k - x^* \rangle \\ &= \frac{1}{2} \|u_{k+1} - u_k\|^2 + \langle u_{k+1} - u_k, u_k \rangle + \lambda_{k+1} t_{k+1} \langle e_{k+1}, u_{k+1} \rangle \\ &= \frac{1}{2} (\|u_{k+1}\|^2 - \|u_k\|^2) + \lambda_{k+1} t_{k+1} \langle e_{k+1}, u_{k+1} \rangle \end{aligned}$$

其中第一个等式利用了

$$u_k = t_k x_k - (t_k - 1)x_{k-1} - x^*, t_{k+1}y_{k+1} = t_{k+1}x_k + (t_k - 1)(x_k - x_{k-1}).$$

第二个等式利用了

$$\|a\|^2 + 2\langle a, b \rangle = \|a + b\|^2 - \|b\|^2.$$

进一步, 由归纳法得到: 对于 $k \geq 1$, 总有

$$\lambda_k t_k^2 \sigma_k + \frac{1}{2} \|u_k\|^2 \leq \lambda_1 t_1^2 \sigma_1 + \frac{1}{2} \|u_1\|^2 - \sum_{i=2}^k \lambda_i t_i \langle e_i, u_i \rangle.$$

利用Cauchy-Schwarz不等式得

$$\lambda_k t_k^2 \sigma_k + \frac{1}{2} \|u_k\|^2 \leq \lambda_1 \sigma_1 + \frac{1}{2} \|u_1\|^2 + \sum_{i=2}^k \lambda_i t_i \|e_i\| \|u_i\|. \quad (2.2.5)$$

由于 λ_k, t_k, σ_k 的非负性得

$$\|u_k\|^2 \leq 2\lambda_1\sigma_1 + \|u_1\|^2 + 2 \sum_{i=2}^k \lambda_i t_i \|e_i\| \|u_i\|$$

令引理 2.2.3 中的

$$a_0 = 0, \quad a_k = \|u_k\|, \quad b_1 = 0, \quad b_{k+1} = 2\lambda_{k+1}t_{k+1}\|e_{k+1}\|, \quad k = 1, 2, \dots$$

$$c_k \equiv 2\lambda_1\sigma_1 + \|u_1\|^2, \quad k = 0, 1, \dots$$

记 $A_k = \sum_{i=1}^k \lambda_i t_i \|e_i\|$, $B_k = A_k - \lambda_1 \|e_1\|$ 得到

$$\begin{aligned} \|u_k\| &\leq B_k + (2\lambda_1\sigma_1 + \|u_1\|^2 + B_k^2)^{\frac{1}{2}} \\ &\leq 2B_k + (2\lambda_1\sigma_1 + \|u_1\|^2)^{\frac{1}{2}}. \end{aligned}$$

将上式代入不等式 (2.2.5) 中

$$\begin{aligned} \lambda_k t_k^2 \sigma_k &\leq \lambda_1 \sigma_1 + \frac{1}{2} \|u_1\|^2 + B_k \left(2B_k + (2\lambda_1\sigma_1 + \|u_1\|^2)^{\frac{1}{2}} \right) \\ &= \frac{1}{2} \left(2B_k + (2\lambda_1\sigma_1 + \|u_1\|^2)^{\frac{1}{2}} \right)^2. \end{aligned}$$

由引理 2.2.2 可得

$$\sigma_k \leq \frac{2}{k^2 L \lambda^2} \left(2B_k + (2\lambda_1\sigma_1 + \|x_1 - x^*\|^2)^{\frac{1}{2}} \right)^2.$$

在引理 2.2.1 中, 令 $x = x^*, y = y_1$ 得到

$$h(x^*) \geq h(P_{\lambda_1}(y_1)) + \frac{1}{\lambda_1} \langle y_1 - x^*, P_{\lambda_1}(y_1) - y_1 \rangle + \frac{1}{2\lambda_1} \|P_{\lambda_1}(y_1) - y_1\|^2 + \langle e_1, x_1 - x^* \rangle.$$

则有

$$\begin{aligned} \sigma_1 &= h(P_{\lambda_1}(y_1)) - h(x^*) \\ &\leq -\frac{1}{\lambda_1} \langle y_1 - x^*, P_{\lambda_1}(y_1) - y_1 \rangle - \frac{1}{2\lambda_1} \|P_{\lambda_1}(y_1) - y_1\|^2 - \langle e_1, x_1 - x^* \rangle \\ &\leq \frac{1}{2\lambda_1} (\|y_1 - x^*\|^2 - \|x_1 - x^*\|^2) + \|e_1\| \|x_1 - x^*\|. \end{aligned}$$

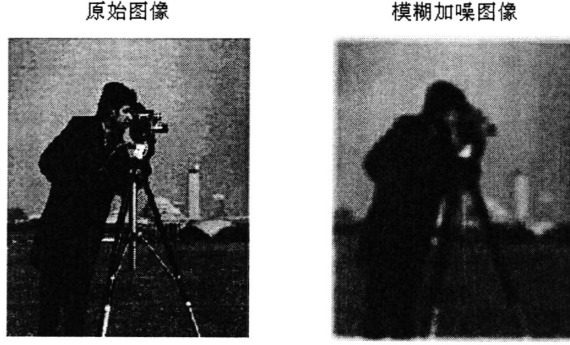


图 1 cameraman图像

最终我们得到

$$h(x_k) - h(x^*) \leq \frac{2}{k^2 L \lambda^2} \left(2A_k + (\|x_0 - x^*\|^2 + 2\lambda_1 \|e_1\| \|x_1 - x^*\|)^{\frac{1}{2}} \right)^2.$$

定理证毕.

□

由上述定理可知, 当 $\sum_{i=1}^{\infty} \lambda_i t_i \|e_i\|$ 是可和时, 有

$$h(x_k) - h(x^*) \leq \mathcal{O}\left(\frac{1}{k^2}\right).$$

2.3 数值试验

在本节中, 将通过实际的图像恢复问题来查看算法2.1对求解实际问题的效果. 并与FISTA算法进行比较.

试验图片 (“Cameraman.pgm”, “Lena.jpg”) 的像素成比例压缩到 $(0, 1)$ 之间. 所有图片对应的模糊加噪图像通过以下方式获得: 在Gaussian 模糊尺度为 9×9 , 标准差为4的图像上添加一个标准差为 10^{-3} 的Gaussian白噪声. 最终得到的模糊加噪图片便是问题 (2.1.1) 中的 b . 问题 (2.1.1) 中的 $A = RW$, 其中 R 代表模糊算子矩阵, W 表示小波变换的逆变换. 参数选取如下

$$\tau = 10^{-4}, \lambda_k \in (0.1/L, 1/L), e_k = 0.$$



图 2 FISTA: 左 $r_{100} = 0.09010$, 右 $r_{200} = 0.08341$



图 3 算法2.1: 左 $r_{100} = 0.11036$, 右 $r_{200} = 0.10562$

不失一般性, 我们用下面的比式来衡量算法恢复后的图片质量

$$r = \frac{\|x_k - x_{ori}\|^2}{\|x_k\|^2}.$$

值得注意的是, r 并不是实际图像恢复的一种可行的停止准则, 它仅仅是衡量图片质量的准则, 当然是在原有的清晰图片已知的情况下. 每运行100步, 我们比较算法恢复的图片质量.

例1 本例将对比较算法2.1和FISTA算法对灰度图像“Cameraman”的试验效果(见图2和3). 清晰的图像和模糊加噪的图像见图1.

从图2和图3可以看出, 虽然两种算法的复杂性一样, 但算法2.1对于这两个例子的实际效果不如FISTA算法, 但差别不大.

第三章 Lasso 问题的一种变形

3.1 Lasso 问题的一种光滑变形

Lasso 问题

$$\min_{x \in R^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \tau \|x\|_1 \right\}$$

不可微是因为1范数在0点是不可微的. 那么我们考虑下面的问题

$$\min_{x, y \in R^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \tau \sum_{i=1}^n \sqrt{x_i^2 + y_i^2} \right\} := h(x, y). \quad (2.1.1)$$

问题 (2.1.1) 仍然是无约束凸规划. $y > 0$ 时, $h(x, y)$ 是关于 x 的连续可微函数.

$\min_{x \in R^n} \{h(x, 0)\}$ 就是 Lasso 问题. 当 $y > 0$ 并且 $\|y\|_\infty$ 很小时, 我们期待问题 (2.1.1) 的解逼近 Lasso 问题的解. 那么求解 Lasso 问题时, 我们可以考虑求解与之相关的光滑问题 $\min_{x \in R^n} \{h(x, y), y > 0, \|y\|_\infty < \varepsilon\}$, 其中 ε 是一个很小的正数.

下面我们讨论两个问题的解的关系.

定理 3.1.1 x^* 是问题 (2.1.1) 的解当且仅当它是 Lasso 问题的解.

证明: \Rightarrow 设 (x^*, y^*) 是问题 (2.1.1) 的解, 则有 $f(x^*, y^*) \leq f(x^*, 0)$, 即

$$\tau \sum_{i=1}^n \sqrt{(x_i^*)^2 + y_i^{*2}} + \frac{1}{2} \|Ax^* - b\|^2 \leq \tau \|x^*\|_1 + \frac{1}{2} \|Ax^* - b\|^2.$$

从上式可得 $y^* = 0$. 那么对任意的 $x, y \in R^n$ 有 $f(x^*, y^*) = f(x^*, 0) \leq f(x, 0)$, 即

$$\tau \|x^*\|_1 + \frac{1}{2} \|Ax^* - b\|^2 \leq \tau \|x\|_1 + \frac{1}{2} \|Ax - b\|^2,$$

所以 x^* 是 Lasso 问题的解.

\Leftarrow 设 x^* 是 Lasso 问题的解. 那么对任意的 $x, y \in R^n$ 有 $f(x^*, 0) \leq f(x, 0) \leq f(x, y)$

. 所以 $(x^*, 0)$ 是问题 (2.1.1) 的解.

□

3.2 非线性共轭梯度法的一些近期结果

对于光滑问题 $\min_{x \in R^n} \{h(x, y), y > 0, \|y\|_\infty < \varepsilon\}$, 其中 ε 是一个很小的正数, 我们可以用非线性共轭梯度法来求解. 在这里简述一下非线性共轭梯度法背景.

考虑极小化问题: 在 R^n 中, 找一点 x 使得

$$g(x) = 0,$$

其中, $g: R^n \rightarrow R^n$ 是非线性连续的. 当 n 是一个很大的数时, 非线性共轭梯度法是求解上述问题的简单可行且常用的方法.

非线性共轭梯度法最早由Fletcher 和Reeves[7] 在1964 年提出, 称为FR 方法. 对任意给定的初始点 $x_0 \in R^n$, 其基本迭代格式为

$$x_{k+1} = x_k + \alpha_k d_k,$$

$\alpha_k > 0$ 是步长, d_k 是搜索方向满足

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_{k+1} d_k, \quad k = 0, 1, 2, \dots,$$

其中 $\beta_{k+1}^{FR} = \|g^{k+1}\|^2 / \|g^k\|^2$.

在1969 年, Polak 和Ribière [18] 给出了一个 β_{k+1} 的更加实用的变形

$$\beta_{k+1}^{PR} = \frac{(g^{k+1} - g^k)^T g^{k+1}}{\|g^k\|^2}.$$

容易验证, 在精确线搜索下, 当 g 线性的且强单调时, FR方法和PR方法就是HS[13] 线性共轭梯度法. 因此, FR公式和PR公式可以看作是线性情形的一个推广. 在实际执行中, β_k^{PR} 可能会是负值, Powell 建议取 $\beta_k^{PR+} = \max\{0, \beta_k^{PR}\}$.

在实际中, 用精确线搜索来决定步长往往并不实用, 所以在保证收敛的情况下, 许多人提出了各种形式的步长准则. 下面概述一些近期的结果.

2006年, Zhang [27]等给出了PR方法中的搜索方向的一个变形

$$\tilde{d}_0 = -g_0, \quad \tilde{d}_k = -g_k + \beta_k^{PR} \tilde{d}_{k-1} - \left(g_k^T \tilde{d}_{k-1} / \|g_{k-1}\|^2 \right) (g_k - g_{k-1}), \quad k = 1, 2, \dots.$$

在 2011 年, Li 和 Tian [14] 对上述搜索方向的基础上进行了进一步的研究, 提出了一个新的步长准则: 当 $k \geq 0$ 时, 计算

$$\rho_k = \frac{h\|g_k\|^2}{|[g(x_k + h\tilde{d}_k) - g_k]^T \tilde{d}_k|},$$

选取 $\alpha_k = \max\{\rho_k t^j : j = 0, 1, 2, \dots\}$, 使得

$$f(x_k + \alpha_k \tilde{d}_k) \leq f(x_k) + \sigma \alpha_k \tilde{d}_k^T g_k, \quad \sigma \in (0, 1),$$

在停止准则 $\|g_k\| \leq 10^{-6}$ 下, 他们得到了非常好的数值结果.

最近, Dong[4]提出了只利用梯度的实用PR+方法, 并在只假设梯度是Lipschitz连续的情况下, 证明了算法的全局收敛性. PR+方法对只有梯度信息, 而原函数比较难估计的问题, 有很大的实用性. 其步长准则为: 当 $k \geq 0$ 时, 计算

$$\mu_k = \frac{(g(x_k \alpha_{k-1} d_k) - g_k)^T d_k}{\alpha_{k-1} \|d_k\|^2}$$

和

$$\bar{\rho}_k = \frac{1}{\max\{10^{-9}, |\mu_k|\}} \min \left\{ 10^9, \frac{\|g_k\|^2 - \beta_k^{PR+} g_k^T d_{k-1}}{\|g_k\|^2} \right\}, \quad k = 0, 1, \dots \quad (2.2.1)$$

令 $\rho_k = \bar{\rho}_k \|g_k\|^2 / \|d_k\|^2$. 取 $\alpha_k = \max\{\rho_k t^j : j = 0, 1, 2, \dots\}$ 使得

$$g(x_k + \alpha_k d_k)^T d_k - 1/2 \mu_k \alpha_k \|d_k\|^2 \leq c g_k^T d_k, \quad k \leq K; \quad (2.2.2)$$

$$g(x_k + \alpha_k d_k)^T d_k + 1/2 \max\{-\mu_k, 0\} \alpha_k \|d_k\|^2 \leq c g_k^T d_k, \quad k > K. \quad (2.2.3)$$

3.3 数值试验

在本章中我们给出了 Lasso 问题的一种变形, 使其目标函数连续可微, 从而可以用非线性共轭梯度法来求解. 在这一章中我们使用数值试验来检验这种方法是否可行有效.

本文中的数值试验均在Matlab7.1中编写, 在以下环境中运行: 操作系统Windows3, 处理器是Intel(R) Core(TM) Duo CPU T2450@2.00GHz.

在本节中, 对 Dong[4] 提出的PR+方法, Li 和Tian [14] 提出的RMPPR 方法, Beck 和Teboulle [1] 提出的快速压缩阈算法(FISTA)进行数值比较. 对于PR+方法, 取参数为

$t = 0.5, c = 10^{-4}, K = 10^6$. 对于RMPRP方法, 取参数为 $\sigma = 10^{-4}, t = 0.5, h = 10^{-8}$. 对于FISTA方法, 去参数为 $\lambda_k \equiv 1/L$, L 是 $A^T A$ 的最大特征值.

对于问题(2.1.1)中的 y , 在FISTA方法中直接取为0; 在PR+方法和RMPRP方法中, 取 $y_0 = 10^{-20} * \text{ones}(n, 1)$, 之后进行自适应选取:

$$y = (\text{abs}(x_{\text{old}}) \leq 10^{-20}) . * y_0.$$

例1 信号恢复问题

本例来自于[?], [25]. 问题构造方法为: 矩阵 A 是一个随机的 $m \times n$ 矩阵, 其中 $m = 2^8, n = 2^{10}$. A 中元素服从期望为0, 方差为 $1/(2n)$ 的高斯分布. 观察到的向量 $b = Ax_{\text{true}} + \eta$, 其中噪声 η 中的元素服从期望为0, 方差为 10^{-4} 的高斯分布. x_{true} 中的元素满足 $x_{2^5 * k} = 1, k = 1, \dots, 32$.

上述问题是一个典型的信号恢复问题. 令 $\tau = 0.1$, 初始点 $x_0 = \text{ones}(n, 1)$. 对于这三个算法, 并无公平有效的停止准则, 所以每运行20步时, 比较它们的结果. 数值结果如表1.

非线性共轭梯度法是求解无约束光滑问题的首选. 非线性共轭梯度法适应非凸问题, 比分离算法更适应一般情形. RMPRP方法和PR+方法是近年来提出的两种算法, 相应的在文献[14]和[4]中, 作者做了大量的数值试验且两种算法都有好的数值效果. 从表1可以看出, 对于该例, RMPRP方法每20步目标函数值下降量较少. PR+方法可以解决该问题, 目标函数值下降比较快. 在80步时, 目标函数值已经接近函数值的稳定值, 最多在120步时, 目标函数值已经稳定. FISTA方法高效快速的解决了该问题. 非线性共轭梯度法的数值结果依赖于预处理技术, 如何进行预处理从而使得算法更快速, 需要进一步的研究.

表1 CPU代表cputime, obj表示 $f(x_k, 0)$

	PR+		RMPRP		FISTA	
	CPU	obj	CPU	obj	CPU	obj
k=20	9.1719	202.6076	10.4219	102.3603	0.2188	9.1650e-4
k=40	16.2656	0.9981	21.1719	102.2904	0.6094	9.1650e-4
k=60	23.3594	0.0058	31.8125	102.2205	0.9844	9.1650e-4
k=80	30.5000	9.4138e-4	42.3906	102.1507	1.4219	9.1650e-4
k=100	37.6250	9.1663e-4	52.9375	102.0808	1.8594	9.1650e-4
k=120	44.7344	9.1650e-4	63.3281	102.0109	2.3750	9.1650e-4
k=140	51.9063	9.1650e-4	73.7969	101.9410	2.7969	9.1650e-4
k=160	58.9531	9.1650e-4	84.3750	101.8711	3.2344	9.1650e-4
k=180	66.0938	9.1650e-4	94.9219	101.8013	3.6563	9.1650e-4
k=200	75.0625	9.1650e-4	105.5938	101.7314	4.0781	9.1650e-4

参考文献

- [1] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.
- [2] Cheng L, Dong Y. A note on convergence of PRP method with new nonmonotone line search[J]. Applied Mathematics and Computation, 2010, 216(7): 2232-2234.
- [3] Dai Y H. Conjugate gradient methods with Armijo-type line searches[J]. Acta Mathematicae Applicatae Sinica, 2002, 18(1): 123-130.
- [4] Dong Y. A practical PR+ conjugate gradient method only using gradient[J]. Applied Mathematics and Computation, 2012.
- [5] Donoho D L. For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution[J]. Communications on pure and applied mathematics, 2006, 59(6): 797-829.
- [6] Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems[J]. Selected Topics in Signal Processing, IEEE Journal of, 2007, 1(4): 586-597.
- [7] Fletcher R, Reeves C M. Function minimization by conjugate gradients[J]. The computer journal, 1964, 7(2): 149-154.
- [8] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization[J]. SIAM Journal on Optimization, 1992, 2(1): 21-42.
- [9] Grippo L, Lampariello F, Lucidi S. A truncated Newton method with nonmonotone line search for unconstrained optimization[J]. Journal of Optimization Theory and Applications, 1989, 60(3): 401-419.
- [10] Grippo L, Lucidi S. A globally convergent version of the Polak-Ribière conjugate gradient method[J]. Mathematical Programming, 1997, 78(3): 375-391.
- [11] Hager W W, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search[J]. SIAM Journal on Optimization, 2005, 16(1): 170-192.

- [12] Hager W W, Zhang H. A survey of nonlinear conjugate gradient methods[J]. Pacific journal of Optimization, 2006, 2(1): 35-58.
- [13] Hestenes M R, Stiefel E. Methods of conjugate gradients for solving linear systems[J]. 1952.
- [14] Li D H, Tian B S. n-step quadratic convergence of the MPRP method with a restart strategy[J]. Journal of computational and applied mathematics, 2011, 235(17): 4978-4990.
- [15] 李全艳. 求解凸优化问题向前后分裂算法的两种变形[D]. 郑州大学, 2012.
- [16] Moré J J, Garbow B S, Hillstom K E. Testing unconstrained optimization software[J]. ACM Transactions on Mathematical Software (TOMS), 1981, 7(1): 17-41.
- [17] Nesterov Y. Smooth minimization of non-smooth functions[J]. Mathematical Programming, 2005, 103(1): 127-152.
- [18] Polak E, Ribière G. Note on the convergence of methods of conjugate directions[J]. Revue Francaise d' Informatique et de Recherche Operationnelle, 1969, 3(16): 35-43.
- [19] Powell M J D. Convergence properties of algorithms for nonlinear optimization[J]. SIAM review, 1986, 28(4): 487-500.
- [20] Salzo S, Villa S. Inexact and accelerated proximal point algorithms[J]. J. Convex Anal, 2012, 19(4).
- [21] Schmidt M, Roux N L, Bach F. Convergence rates of inexact proximal-gradient methods for convex optimization[J]. arXiv preprint arXiv:1109.2415, 2011.
- [22] Shi Z J, Shen J. Convergence of PRP method with new nonmonotone line search[J]. Applied mathematics and computation, 2006, 181(1): 423-431.
- [23] Tseng P. On accelerated proximal gradient methods for convex-concave optimization[J]. submitted to SIAM Journal on Optimization, 2008.
- [24] Tseng P. Approximation accuracy, gradient methods, and error bound for structured convex optimization[J]. Mathematical programming, 2010, 125(2): 263-295.
- [25] Wright S J, Nowak R D, Figueiredo M A T. Sparse reconstruction by separable approximation[J]. Signal Processing, IEEE Transactions on, 2009, 57(7): 2479-2493.

- [26] Zhang H, Wei J, Li M, et al. On proximal gradient method for the convex problems regularized with the group reproducing kernel norm[J]. Journal of Global Optimization, 2013: 1-20.
- [27] Zhang L, Zhou W, Li D H. A descent modified Polak - Ribière - Polyak conjugate gradient method and its global convergence[J]. IMA Journal of Numerical Analysis, 2006, 26(4): 629-640.

后 记

Lasso 问题是近年来比较受关注的问题, 国内外众多学者对它进行了研究, 提出了许多解决方法. 本文的主要工作及推广如下:

向前向后分裂算法是求解 Lasso 问题的可行方法, 为了提高算法的收敛速度, 许多加速变形被提出. Salzo[20]等提出了一种非精确的加速变形, 并在两类非精确形式下, 证明了其收敛率为 $\mathcal{O}(1/k^2)$, 本文证明了在另一种非精确形式下, 该算法的收敛率仍为 $\mathcal{O}(1/k^2)$.

本文考虑的 Lasso 问题的相关光滑变形, 该问题可以用非线性共轭梯度法求解. 变形后的问题的最优解与 Lasso 问题的最优解之间有什么关系? 是否可以构造一些等价的光滑变形? 从第三章的数值结果可以看出, PR+方法求解光滑变形问题得到的目标函数最优值与快速压缩阈算法求解 Lasso 问题得到的目标函数最优值相差不大. 那么PR+方法得到的迭代序列或其子列是否可以收敛到 Lasso 问题的解点? 这些问题都是需要进一步研究的.

致 谢

本文的完成除了我自己三年来的认真学习和研究,也获益于他人。在论文的撰写过程中遇到的一些困难,都在老师和同学的帮助下解决了。在论文即将付梓之际,我思绪万千。我急切地想要把我的感谢献给每一位帮助过我的人。尤其要感谢我的指导老师—董老师。我不是您最优秀的学生,但您却是最尊敬和感谢的老师。您治学严谨,思想深邃,学识渊博,视野雄阔。授人以鱼不如授人以渔,置身其间,耳濡目染,潜移默化,从论文题目的选定到论文写作过程中的指导,您总是悉心的点拨,再经认真思考后的领悟,常常让我有“山重水复疑无路,柳暗花明又一村”的感受。感谢柳朝阳老师在编程方面对我的无私指导。感谢我的父母,养育之恩,无以回报,感谢你们对我的永远支持。你们永远健康快乐是我最大的愿望。从开始进入课题到论文的顺利完成,有多少可敬的师长、同学、朋友给了我无言的帮助,在这里请接受我诚挚谢意!同时也感谢学校和数学系为我提供的良好的做毕业设计的环境。最后再一次感谢在毕业设计中曾经帮助过我的所有良师益友和同学,以及论文中引用或参考的论著的作者。