

期末简短实验报告

学号 10214810424

2024-12-27

主要发现 近端梯度下降法无论是否加速，迭代趋于稳定的点和函数自然对数值互相接近，且离最优点和最优函数自然对数值有一定距离。然而，交替方向乘子法迭代趋于稳定的点和函数自然对数值都和 CVXPY 软件包默认求解器得到的解高度接近。不加速的近端梯度下降法得到的点和函数自然对数值在大约 600 次迭代后趋于稳定，而加速的近端梯度下降得到的点和函数自然对数值能在大约 120 次迭代后趋于稳定。交替方向乘子法能让函数自然对数值在第一次迭代以比加速近端梯度下降更快的速度猛烈下降，之后下降速度放缓。

本文依靠的计算代码见：nbviewer 链接

A 参考资料

用到了作业信息附带的幻灯片 (Boyd, 2010) 和中文书 (刘浩洋等, 2022)，以及与 ADMM 方法相关的另外两组幻灯片。(Boyd, 2011; Tibshirani, 2019)

A.1 参考文献表

Boyd S, Parikh N, Chu E, et al., 2010. Distributed Optimization and Statistics via Alternating Direction Method of Multipliers[Z/OL]. https://web.stanford.edu/~boyd/papers/pdf/admm_talk.pdf.

刘浩洋, 户将, 李勇锋, 文再文, 2022. 最优化: 建模、算法与理论[M/OL]. 第二版. 北京: 高等教育出版社[2024-12-27]. <http://faculty.bicmr.pku.edu.cn/~wenzw/optbook.html>.

Boyd S, 2011. Alternating direction method of multipliers[Z/OL]. EE364b, Stanford University. https://web.stanford.edu/~boyd/papers/pdf/admm_slides.pdf.

Tibshirani R, 2019. Alternating direction method of multipliers[Z/OL]. Carnegie Mellon University. <https://www.stat.cmu.edu/~ryantibs/convexopt/lectures/admm.pdf>.

B 问题

经典的 Lasso 回归估计

$$\min_x \mu \|x\|_1 + \frac{1}{2} \|Ax - b\|_2^2$$

变量或常量	维度	备注
x	1024×1	竖向量
A	512×1024	按标准正态分布随机生成
u	1024×1	按标准正态分布随机生成，再按 10% 比例随机稀疏
b	512×1	按 $b = Au$ 得到
μ	1×1	正实数

在代码中将可微项 $\frac{1}{2} \|Ax - b\|_2^2$ 记作 $g(x)$ ，将不可微项 $\mu \|x\|_1$ 记作 $h(x)$ ，以求与本学期上课使用课件一致。

C 近端梯度下降法

C.1 不加速

Proximal Gradient Decent, PGD

目标函数值的自然对数值下降较平稳，是一条与直线相差不悬殊的下凸曲线。在第 800 次左右迭代趋于稳定，该值大于 0，小于 1。

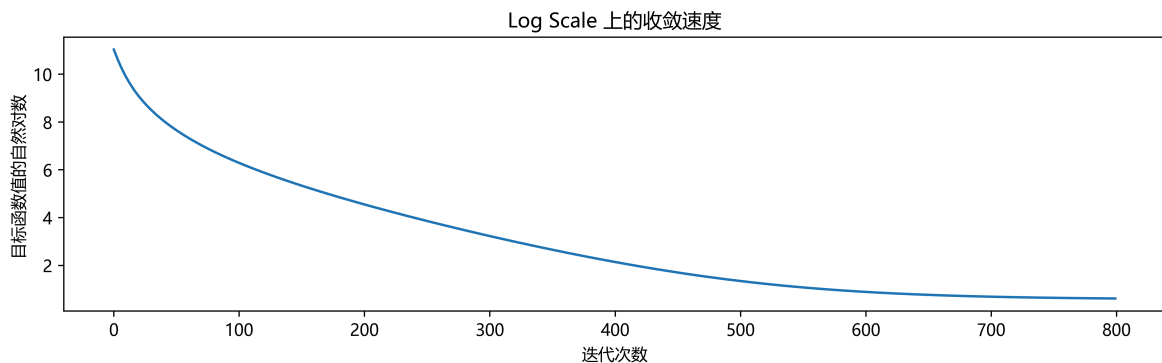


Figure 1: 不加速的近端梯度下降收敛示意

C.2 加速

Fast Iterative Shrinkage-Thresholding Algorithm, FISTA

目标函数值的自然对数值初始下降比不加速更陡峭，曲线的凹凸性多次来回变化（这可能与 FISTA 并不是每一步都下降有关）。在第 50 次左右迭代趋于稳定，该值仍大于 0，小于 1。

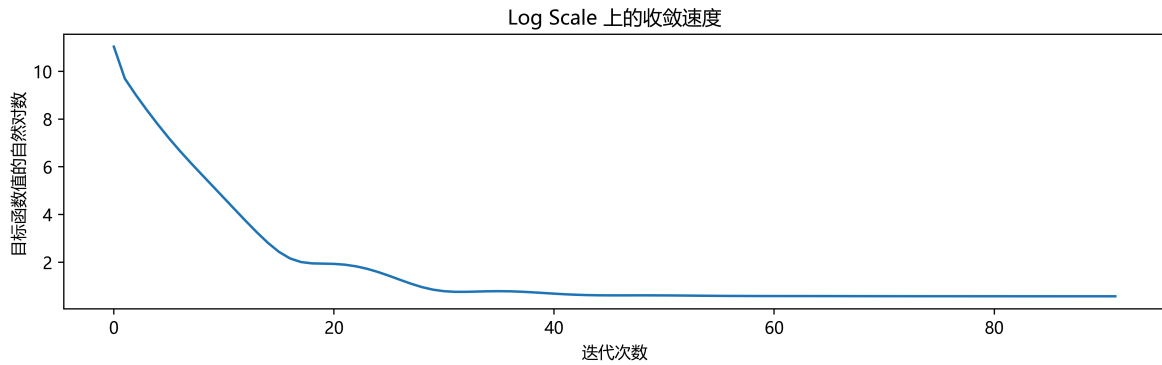


Figure 2: 加速的近端梯度下降收敛示意

D 交替方向乘子法

Alternating Direction Method of Multipliers, ADMM

D.1 算法自学

D.1.a 改写问题

这种方法是本学期上课没有讲过的，所以有一个自己推导的过程。

$$\text{记 } g(x) = \frac{1}{2} \|Ax - b\|_2^2 \quad h(x) = \mu \|x\|_1 \quad f(x) = g(x) + h(x)$$

则原问题写作 $\min_x f(x)$

现改写为 $\min_{x,z} g(x) + h(z) \quad \text{s. t.} \quad x - z = 0$

记 Lagrange 乘子为 y 罚因子为 ρ 固定步长为 τ

对改写形式构造增广 Lagrange 函数

$$L_\rho = g(x) + h(z) + y^T(x - z) + \frac{\rho}{2} \|x - z\|_2^2$$

接下来开始每次迭代内部的步骤

D.1.b 第一步

$$\begin{aligned}
x^{(k+1)} &= \operatorname{argmin}_x L_\rho(x, z^{(k)}, y^{(k)}) \\
&= \operatorname{argmin}_x \left\{ \left\| Ax - b \right\|_2^2 + \mu \left\| z \right\|_1 + (y^{(k)})^T (x - z^{(k)}) + \frac{\rho}{2} \left\| x - z^{(k)} \right\|_2^2 \right\} \\
&\Rightarrow \frac{\partial L}{\partial x} = 2A^T(Ax - b) + y^{(k)} + \rho(x - z^{(k)}) \\
\text{令 } \frac{\partial L}{\partial x} &= 0 \text{ 则 } 2A^T Ax - 2A^T b + y^{(k)} + \rho x - \rho z^{(k)} = 0 \\
&\Rightarrow (2A^T A + \rho I)x - 2A^T b + y^{(k)} - \rho z^{(k)} = 0 \\
&\Rightarrow (2A^T A + \rho I)x = 2A^T b + \rho z^{(k)} - y^{(k)} \\
&\Rightarrow x^{(k+1)} = (2A^T A + \rho I)^{-1} (2A^T b + \rho z^{(k)} - y^{(k)})
\end{aligned}$$

D.1.c 第二步

$$\begin{aligned}
z^{(k+1)} &= \operatorname{argmin}_z L_\rho(x^{(k+1)}, z, y^{(k)}) \\
&= \operatorname{argmin}_z \left\{ \left\| Ax^{(k+1)} - b \right\|_2^2 + \mu \left\| z \right\|_1 + (y^{(k)})^T (x^{(k+1)} - z) + \frac{\rho}{2} \left\| x^{(k+1)} - z \right\|_2^2 \right\} \\
&= \operatorname{argmin}_z \left\{ \mu \left\| z \right\|_1 + (y^{(k)})^T (x^{(k+1)} - z) + \frac{\rho}{2} \left\| x^{(k+1)} - z \right\|_2^2 \right\} \\
&\text{对要找最值点的目标乘以常数 } \frac{2}{\rho} \\
&= \operatorname{argmin}_z \left\{ \frac{2\mu}{\rho} \left\| z \right\|_1 + \frac{2}{\rho} (y^{(k)})^T (x^{(k+1)} - z) + \left\| x^{(k+1)} - z \right\|_2^2 \right\} \\
&= \operatorname{argmin}_z \left\{ \frac{2\mu}{\rho} \left\| z \right\|_1 + 2 \left(\frac{y^{(k)}}{\rho} \right)^T (x^{(k+1)} - z) + \left\| x^{(k+1)} - z \right\|_2^2 \right\} \\
&\text{加上与 } z \text{ 无关的项 } \left\| \frac{y^{(k)}}{\rho} \right\|_2^2 \\
&= \operatorname{argmin}_z \left\{ \frac{2\mu}{\rho} \left\| z \right\|_1 + \left\| \frac{y^{(k)}}{\rho} \right\|_2^2 + 2 \left(\frac{y^{(k)}}{\rho} \right)^T (x^{(k+1)} - z) + \left\| x^{(k+1)} - z \right\|_2^2 \right\} \\
&= \operatorname{argmin}_z \left\{ \frac{2\mu}{\rho} \left\| z \right\|_1 + \left\| \frac{y^{(k)}}{\rho} + x^{(k+1)} - z \right\|_2^2 \right\} \\
&= \operatorname{argmin}_z \left\{ \frac{1}{2} \left\| z - \left(x^{(k+1)} + \frac{y^{(k)}}{\rho} \right) \right\|_2^2 + \frac{\mu}{\rho} \left\| z \right\|_1 \right\} \\
&= S_{\left(\frac{\mu}{\rho}\right)} \left(x^{(k+1)} + \frac{y^{(k)}}{\rho} \right) \quad \text{其中 } S_{(\cdot)}(\cdot) \text{ 表示软阈值算子}
\end{aligned}$$

D.1.d 总结

$$x^{(k+1)} = (2A^T A + \rho I)^{-1} (2A^T b + \rho z^{(k)} - y^{(k)})$$

$$z^{(k+1)} = S_{\left(\frac{\mu}{\rho}\right)} \left(x^{(k+1)} + \frac{y^{(k)}}{\rho} \right)$$

$$y^{(k+1)} = y^{(k)} + \tau \rho (x^{(k+1)} - z^{(k+1)})$$

D.2 观察结果

目标函数值的自然对数值初始下降比加速的近端梯度下降法更陡峭，曲线下凸且即使平滑拟合线也很不平滑，存在拐点。在刚开始迭代时猛烈下降后，就缓慢下降，但可持续下降的次数比 FISTA 更多。在接近第两百次迭代时，趋于稳定，该值小于 0

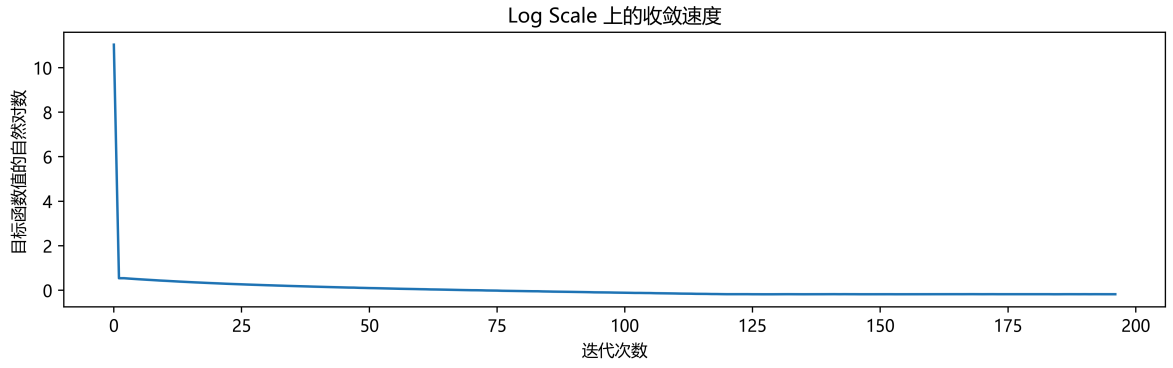


Figure 3: 交替方向乘子法

E 三种方法对比

E.1 收敛速度

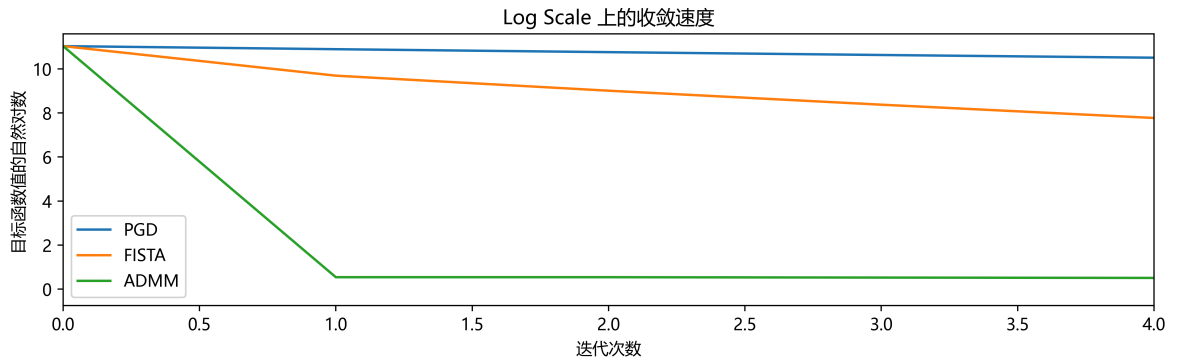


Figure 4: 前 4 次迭代

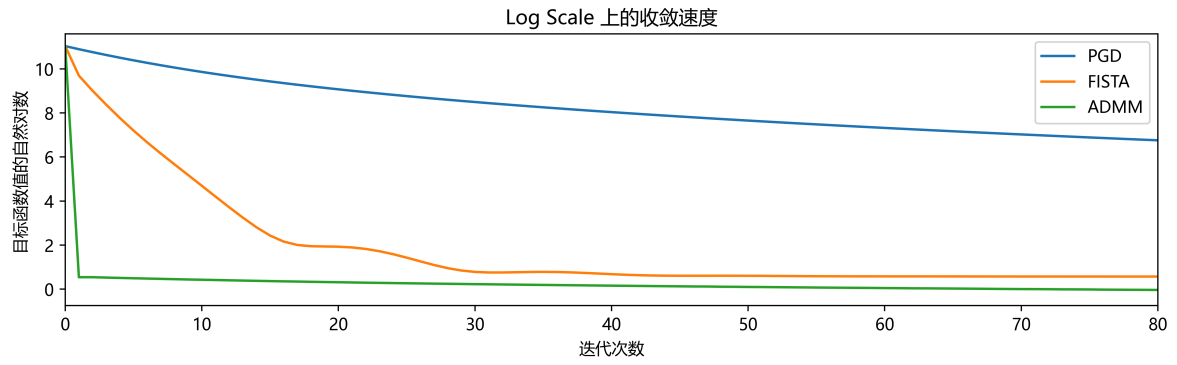


Figure 5: 前 80 次迭代

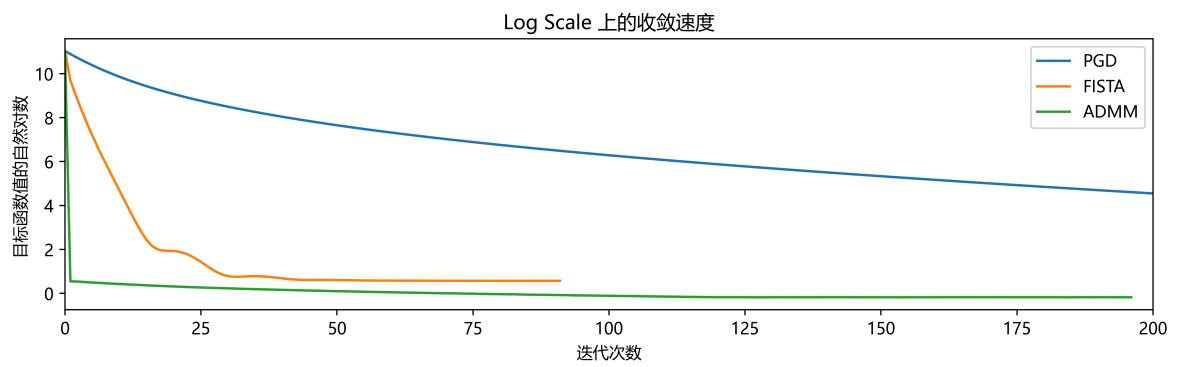


Figure 6: 前 200 次迭代

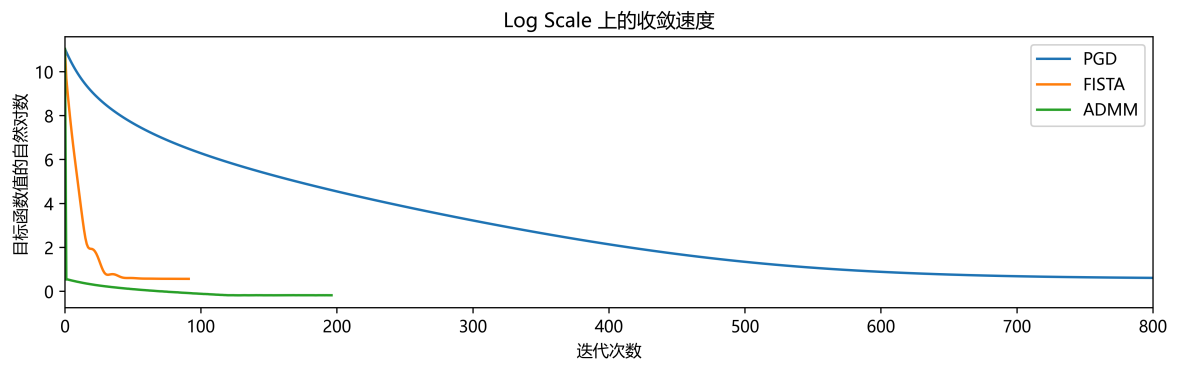


Figure 7: 前 800 次迭代

观察图象可看出，与 ADMM 方法相比，近端梯度下降法不仅下降较慢，而且容易在某个次优点停滞不动，触发点更新距离过近的停机准则。

E.2 准确性评估

虽然 CVXPY 默认求解器不一定得到最优解,但是由于 Lasso 问题足够简易,所以仍考虑将三种方法停机时的点与 CVXPY 默认求解器得到的解作比较,如下表所示:

(表中的数字已经尽可能地保留小数点后三位)

方法	点与 CVXPY 点欧氏距离	点与 CVXPY 点余弦相似度
PGD	8.264	0.696
FISTA	8.240	0.698
ADMM	0.005	0.99999992

可见 ADMM 比另两种方法更准确。另外,两种近端梯度下降方法得到的函数值自然对数都大于 0,但是 ADMM 和 CVXPY 默认求解器得到的都小于 0,可见 ADMM 停机时的的函数值也更准确。