1 优化目标

考虑如下 group LASS 问题:

$$\min_{x\in\mathbb{R}^{n\times l}} \frac{1}{2}\|Ax-b\|_F^2+\mu\|x\|_{1,2}$$

其中 $A\in\mathbb{R}^{m\times n}, b\in\mathbb{R}^{m\times l}$，给定 $\mu>0$ ，$\|x\|_{1,2}$ 由如下给出:

$$\|x\|_{1,2}=\sum_{i=1}^{n}\|x(i,1:l)\|_2.$$

2 cvx 求解

利用 python 中的 cvxpy 包，如下为调用 mosek 和 gurobi 求解器得到的解，其中解为 $n\times l$ 矩阵，按列将解画在图上:

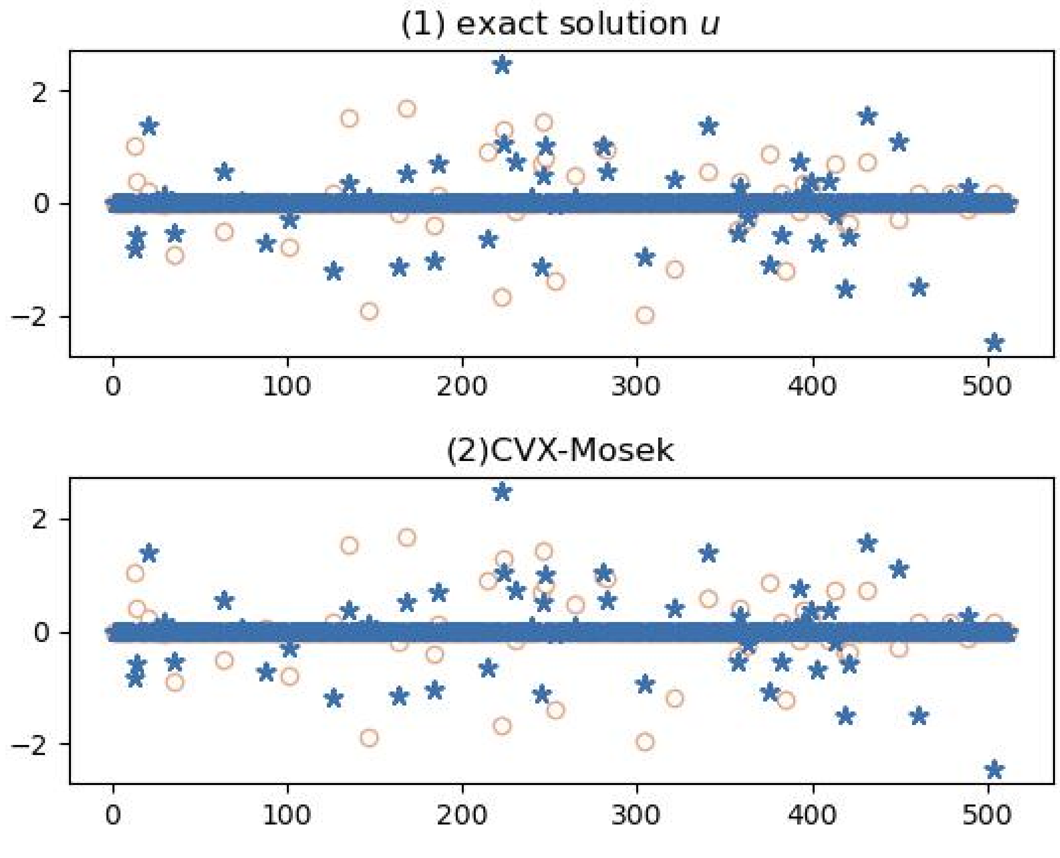

图 1: CVX-Mosek 解与精确解 u 对比

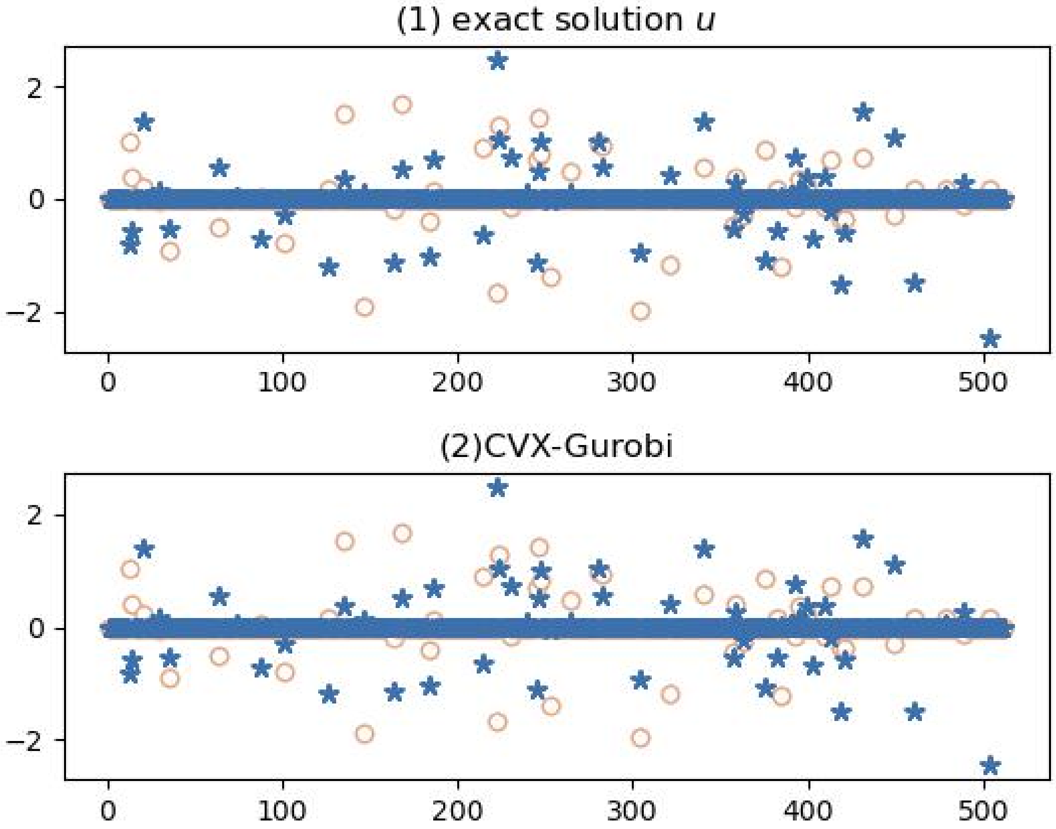

图 2: CVX-Gurobi 得到的解与精确解 u 对比

3 mosek 求解器和 gurobi 求解器

在调用 mosek 和 gurobi 前，现将原问题转化为如下等价的二次规划和二次锥约束问题：

$$\min_{x,t} \frac{1}{2}\|Ax-b\|_F^2 + \mu \mathbf{1}t$$
$$\text{s.t.}\|x(i,1:l)\| \le t_i, i = 1,2,\cdots,n$$

其中向量 $\mathbf{1} \in \mathcal{R}^n$，下面展示了具体结果：

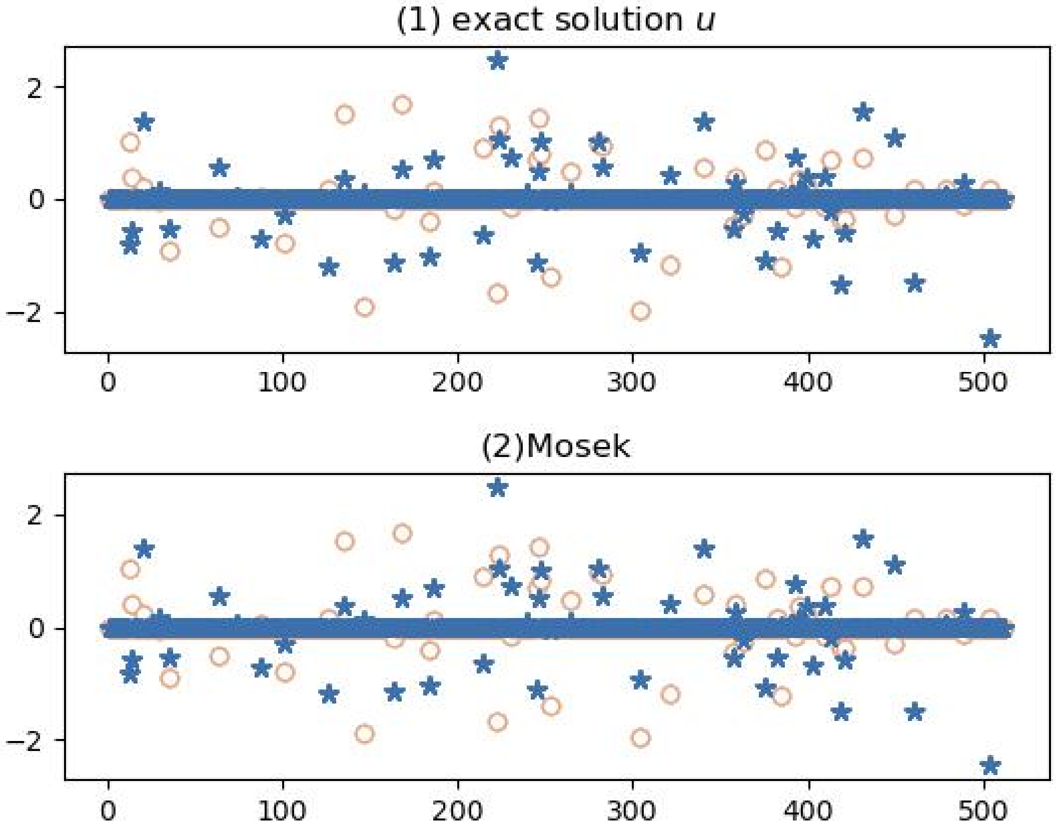

图 3: Mosek 得到的解与精确解 u 对比

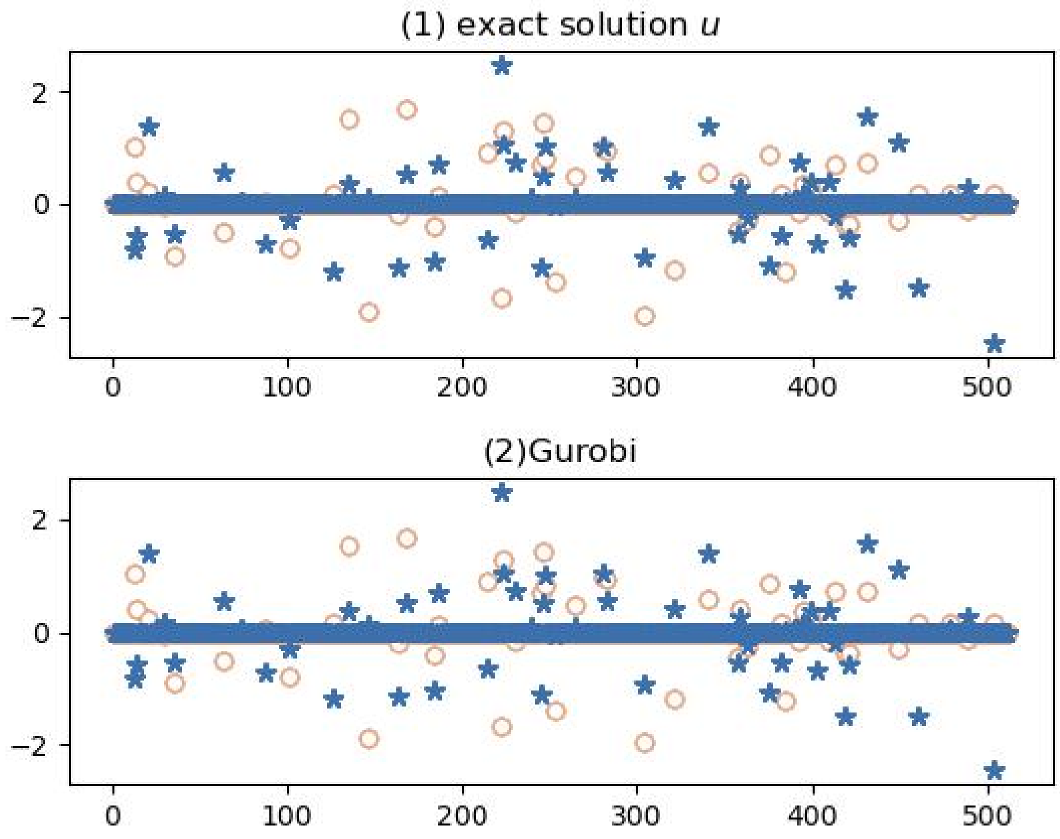

图 4: Gurobi 得到的解与精确解 u 对比

4 次梯度法

对于函数 $f(x) = \frac{1}{2}\|Ax - b\|_F^2 + \mu\|x\|_{1,2}$, 它的次梯度可以表示如下

$$\partial f(x) = A^T Ax - A^T b + \mu \partial \|x\|_{1,2}$$

其中

$$\partial \|x\|_{1,2} = \begin{bmatrix} \partial x(1, 1:l) \\ \partial x(2, 1:l) \\ \vdots \\ \partial x(n, 1:l) \end{bmatrix}$$

$$\partial x(i, 1:l) = \begin{cases} \frac{x(i,1:l)}{\|x(i,1:l)\|_2}, & x(i,1:l) \neq \mathbf{0} \\ \{w \mid \|w\|_2 \leq 1\}, & x(i,1:l) = \mathbf{0} \end{cases}$$

利用辅助函数确定当前步步长, 然后进行一步次梯度法迭代 $x^{k+1} = x^k - \alpha_k g\left(x^k\right)$ 。其中 $g\left(x^k\right) \in \partial f\left(x^k\right)$ 。题目中 μ 的值过小, 直接求解收敛速度过慢, 所以先从较大的 μ 开始, 采用连续化策略，直到 mu 衰减到题目中要求的值。

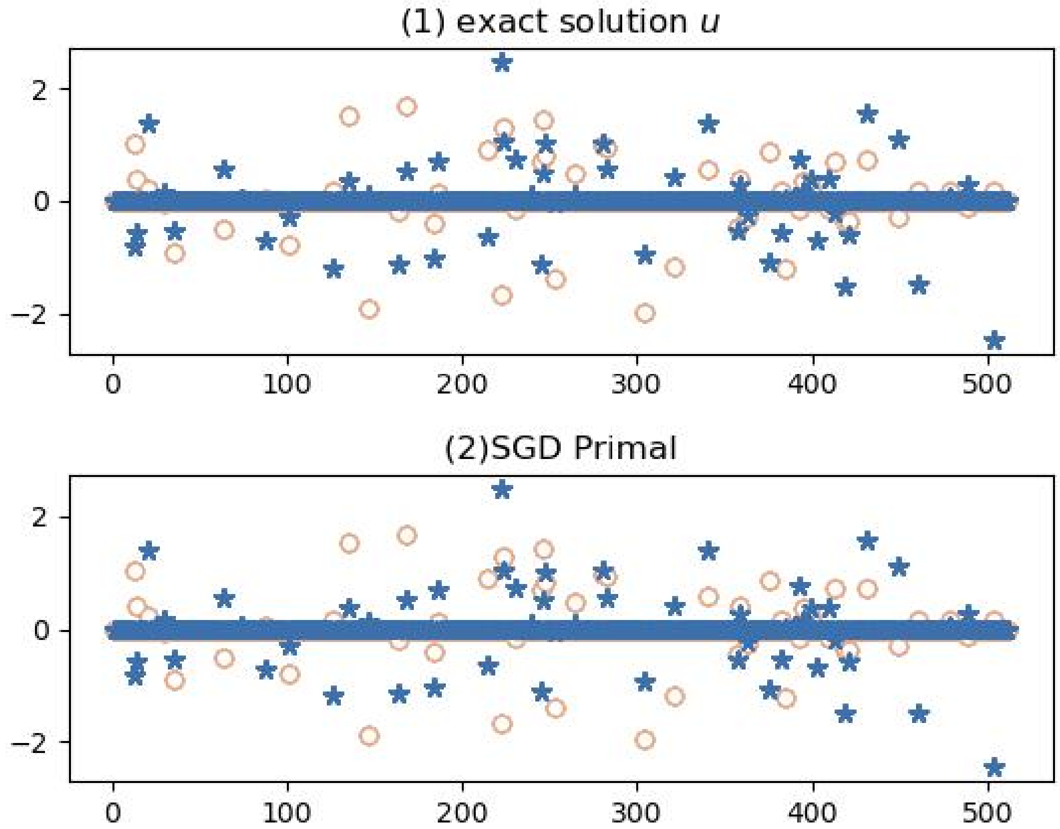

图 5: SGD 得到的解与精确解 u 对比

5 近似点梯度法

优化问题可以看作 $f(x) = \phi(x) + h(x)$，其中 $\phi(x) = \frac{1}{2}\|Ax - b\|_F^2$，$h(x) = \mu\|x\|_{1,2}$. 题目中 μ 的值过小, 直接求解收敛速度过慢, 采用连续化策略，连续化策略从较大的正则化参数 μ_t 逐渐减小到 μ (即 $\mu_1 \geq \cdots \geq \mu_{t-1} \geq \mu_t \geq \cdots \geq \mu$), 并求解相应的 LASSO 问题:

事实上，近似点梯度法的迭代格式根据定义可以写为

$$
\begin{aligned}
x^{k+1} &= \arg\min_u \left(\mu_k \|u\|_{1,2} + \frac{1}{2\alpha_k} \|u - x^k + \alpha_k \nabla\phi(x^k)\|_2^2 \right) \\
&= \arg\min_u \left(\mu_k \|u\|_{1,2} + \phi(x^k) + \nabla\phi(x^k)^\top (u - x^k) + \frac{1}{2\alpha_k} \|u - x^k\|_2^2 \right).
\end{aligned}
$$

检验是否满足非精确线搜索条件，针对 $f(x)$ 考虑线搜索准则，即为 $f(x^{k+1}(t)) \le C_k - \frac{1}{2}\rho t\|x^{k+1}(t) - x^k\|^2$，其中 $x^{k+1}(t) = \mathrm{prox}_{\mu_k\|\cdot\|_{1,2}}(x^k - \alpha_k \nabla\phi(x^k))$。

|nls| 记录线搜索循环的迭代次数，直到满足条件或进行 10 次步长衰减后退出线搜索循环，得到更新的 x^{k+1}。C_k 为 (Zhang & Hager) 线搜索准则中的量。如果不满足线搜索条件，对当前步长进行衰减，当前线搜索次数加一。线搜索结束，得到更新的 x, g。计算梯度范数和原始目标函数值。fvec 记录每一步对应的原 LASSO 问题的目标函数值。并进行内层循环的收敛判断：若当前梯度小于阈值或者目标函数变化小于阈值，内层迭代终止。

计算 BB 步长作为下一步迭代的初始步长。令 $s^k = x^{k+1} - x^k, y^k = g^{k+1} - g^k$, 这里在偶数与奇数步分别对应 $\frac{(s^k)' s^k}{(s^k)^\top y^k}$ 和 $\frac{(s^k)^\top y^k}{(y^k)^\top y^k}$ 两个 BB 步长。

得到如下结果：

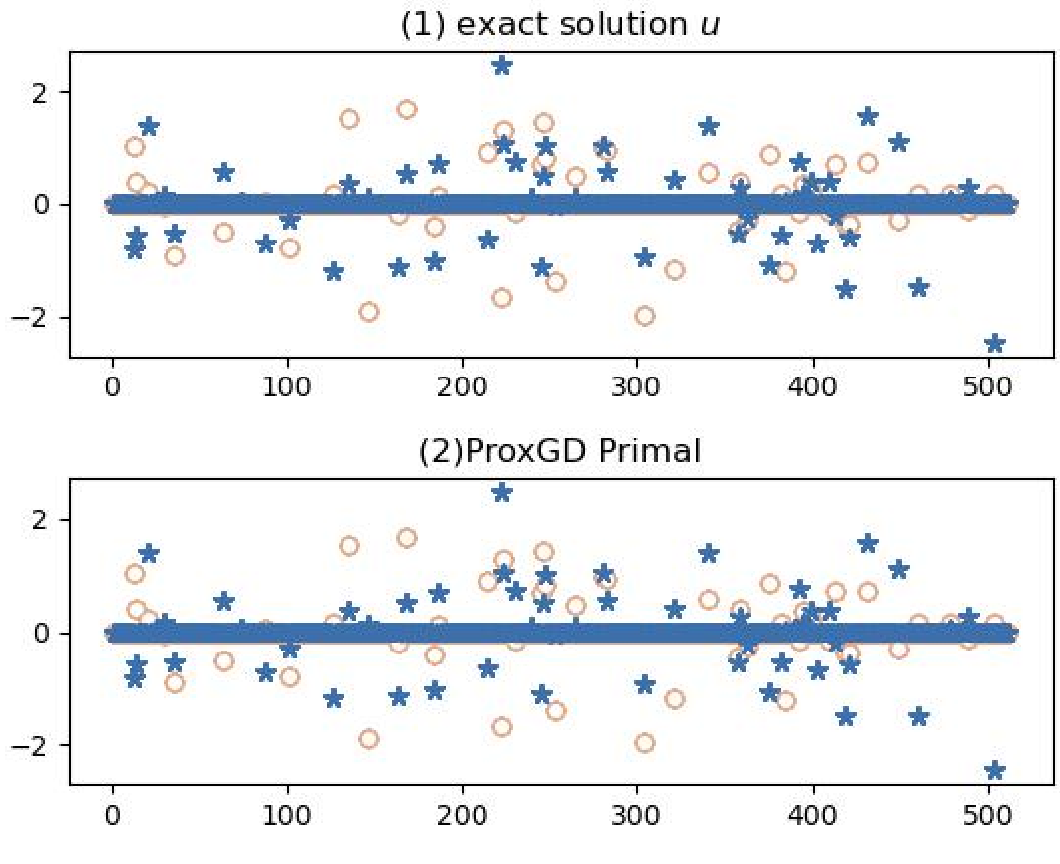

图 6: ProxGD 得到的解与精确解 u 对比

6 FProxGD 算法

利用 Nesterov 加速的近似点梯度法进行优化。该算法被外层连续化策略调用，在连续化策略下完成某一固定正则化系数的内层迭代优化。在每一步迭代时, 算法首先在之前两步迭代点的基础上进行一步更新 $y^k = x^{k-1} + \frac{k-2}{k+1}\left(x^{k-1} - x^{k-2}\right)$, 然后再在 y^k 处进行一步近似点梯度法, $x^k = \operatorname{prox}_{\mu_k\|\cdot\|_{1,2}}\left(y^k - t_k A^\top\left(Ay^k - b\right)\right.$ 。

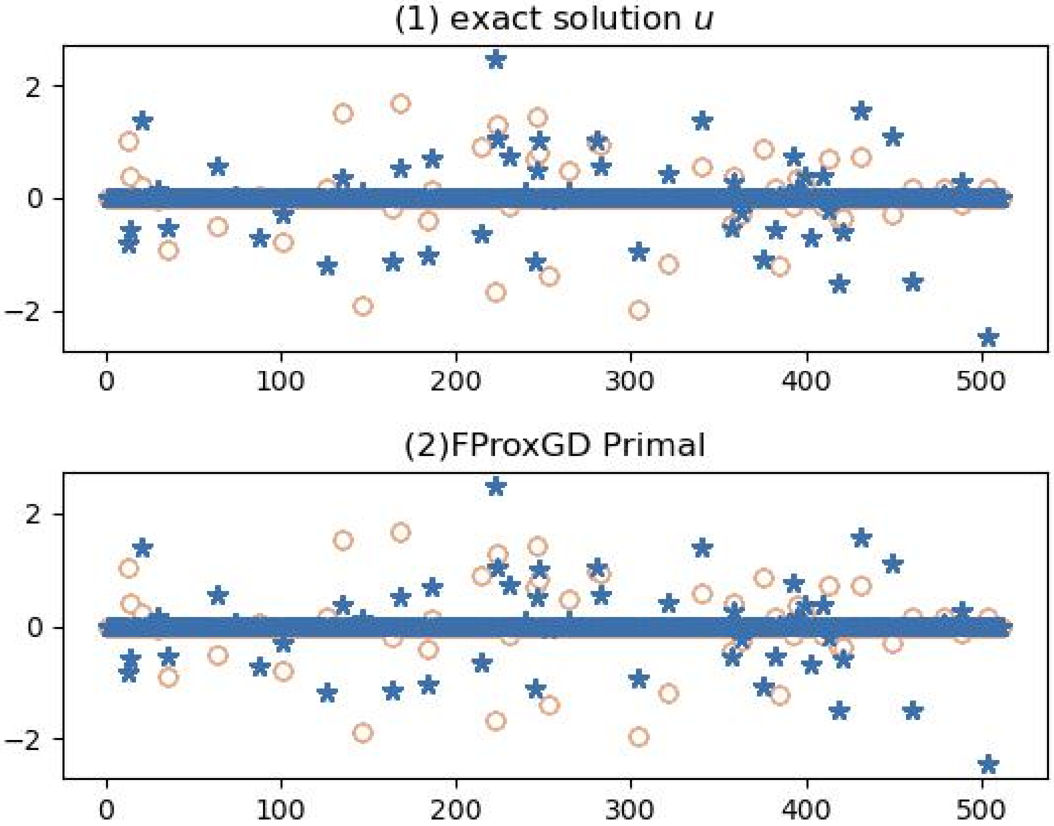

图 7: FProxGD 得到的解与精确解 u 对比

7 对偶问题的 ALM 算法

原问题等价于

$$\min_{x,z} \frac{1}{2}\|z - b\|_F^2 + \mu\|x\|_{1,2}$$
$$\text{s.t. } Ax - z = 0$$

构造拉格朗日函数 $L(x,z,y) = \frac{1}{2}\|y - b\|_F^2 - \langle z, y\rangle + \mu\|x\|_{1,2} + \left\langle A^T y, x\right\rangle$ 通过最小化 x 和

z，我们有：

$$g(y)=\inf_{x,z}L(x,z,y)=\inf_{x,z}\frac{1}{2}\|y-b\|_F^2-\langle z,y\rangle+\mu\|x\|_{1,2}+\left\langle A^Ty,x\right\rangle$$
$$=-\frac{1}{2}\|y\|_F^2-\langle b,y\rangle+I\left[\left\|A^Ty\right\|_{\infty,2}^*\le\mu\right]$$

因此原问题的对偶问题为：

$$\begin{array}{ll}\min_y & \frac{1}{2}\|y\|_F^2+\langle b,y\rangle\\ \text{s.t} & \left\|A^Ty\right\|_{\infty,2}\le\mu.\end{array}$$

其中矩阵的 $(\infty,2)$ 范数定义为 $\|A\|_{\infty,2}=\max_i\|A(i,1:l)\|_2$，为了避免复杂的符号 $\left\|A^Tz\right\|_{\infty,2}$，引入等式约束，使得 ALM 算法更加的平滑。

$$\begin{array}{ll}\min_y & \frac{1}{2}\|y\|_F^2+\langle b,y\rangle\\ \text{s.t} & \|s\|_{\infty,2}\le\mu, s=A^Ty\end{array}$$

它的增广拉格朗日函数为：

$$L^d(y,s,\lambda,\mu)=\frac{1}{2}\|y\|_F^2+\langle b,y\rangle+\left\langle\lambda,A^Ty-s\right\rangle+\frac{\sigma}{2}\left\|A^Ty-s\right\|_F^2$$

以及约束 $\|s\|_{\infty,2}\le\mu$。因此对偶问题的增广拉格朗日函数法对于每个固定的 λ 解决如下子问题：

$$\begin{array}{ll}\min_{y,s} & \frac{1}{2}\|y\|_F^2+\langle b,y\rangle+\left\langle\lambda,A^Ty-s\right\rangle+\frac{\sigma}{2}\left\|A^Ty-s\right\|_F^2\\ \text{s.t} & \|s\|_{\infty,2}\le\mu\end{array}$$

7.1 迭代格式

上述对偶问题的增广拉格朗日函数法的迭代格式为

$$\begin{cases}(y^{k+1},s^{k+1})=\underset{y,\|s\|_{\infty,2}\le\mu}{\arg\min}L^d(y,s,\lambda^k,\mu)=\underset{y,\|s\|_{\infty,2}\le\mu}{\arg\min}\left\{\frac{1}{2}\|y\|_F^2+\langle b,y\rangle+\frac{\sigma_k}{2}\left\|A^Ty-s+\frac{\lambda}{\sigma_k}\right\|_F^2\right\}\\ \lambda^{k+1}=\lambda^k+\sigma_k(A^\top y^{k+1}-s^{k+1})\\ \sigma_{k+1}=\min\left\{\rho\sigma_k,\bar{\sigma}\right\}\end{cases}$$

其中 $\rho>1$ 和 $\bar{\sigma}<+\infty$ 为算法参数，由于 (y^{k+1},s^{k+1}) 的显式表达式是未知的，我们需要利用迭代算法来进行求解。

首先关于 y 的极小化问题为：

$$\min\frac{1}{2}\|y\|_F^2+\langle b,y\rangle+\left\langle\lambda,A^Ty-s\right\rangle+\frac{\sigma}{2}\left\|A^Ty-s\right\|_F^2$$

进一步化简可知：

$$y=\arg\min_y\operatorname{tr}\left(\frac{1}{2}\|y\|_F^2+\frac{1}{2}y^\top(\sigma_kAA^\top)y+(b^\top+\lambda^\top A^\top-\sigma_ks^\top A^\top)y\right)$$

设 $h(y)=\frac{1}{2}^{\top}(\sigma_k AA^{\top})y+(b^{\top}+\lambda^{\top}A^{\top}-\sigma_k s^{\top}A^{\top})y$,那么 $y=\mathrm{prox}_h(0)=\arg\min_u \mathrm{tr}\left\{h(u)+\frac{1}{2}\|y\|_F^2\right\}$,有二次函数的近似算法可知 $y=\left(I+\sigma_k AA^{\top}\right)^{-1}(\sigma_k As-A\lambda-b)$

关于 s 的极小化问题为:

$$\min_s \frac{\sigma_k}{2}\left\|A^T y-s+\frac{\lambda}{\sigma_k}\right\|_F^2$$
$$\text{s.t.}\|s\|_{\infty,2}\le\mu$$

通过简单的推导可知:

$$s=\mathcal{P}_{\|s\|_{\infty,2}\le\mu}\left(A^T y+\frac{\lambda}{\sigma_k}\right)$$

7.2 优化结果

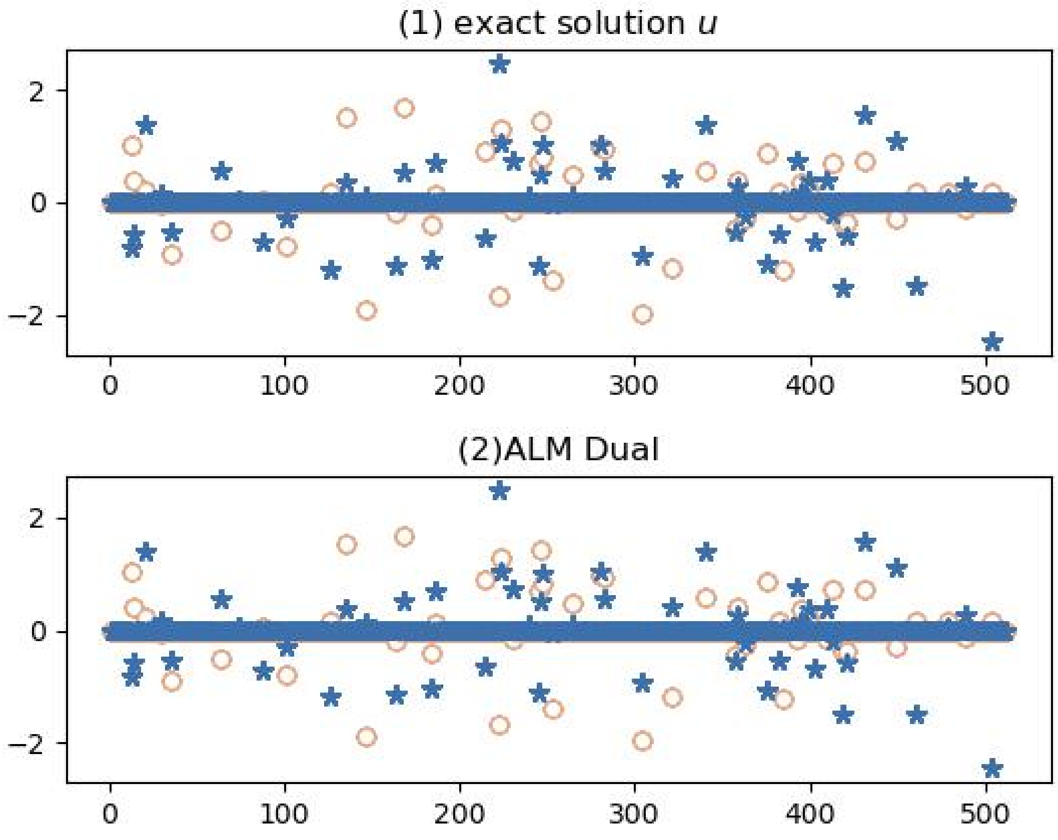

图 8: ALM Dual 得到的解与精确解 u 对比

8 对偶问题的 ADMM 算法

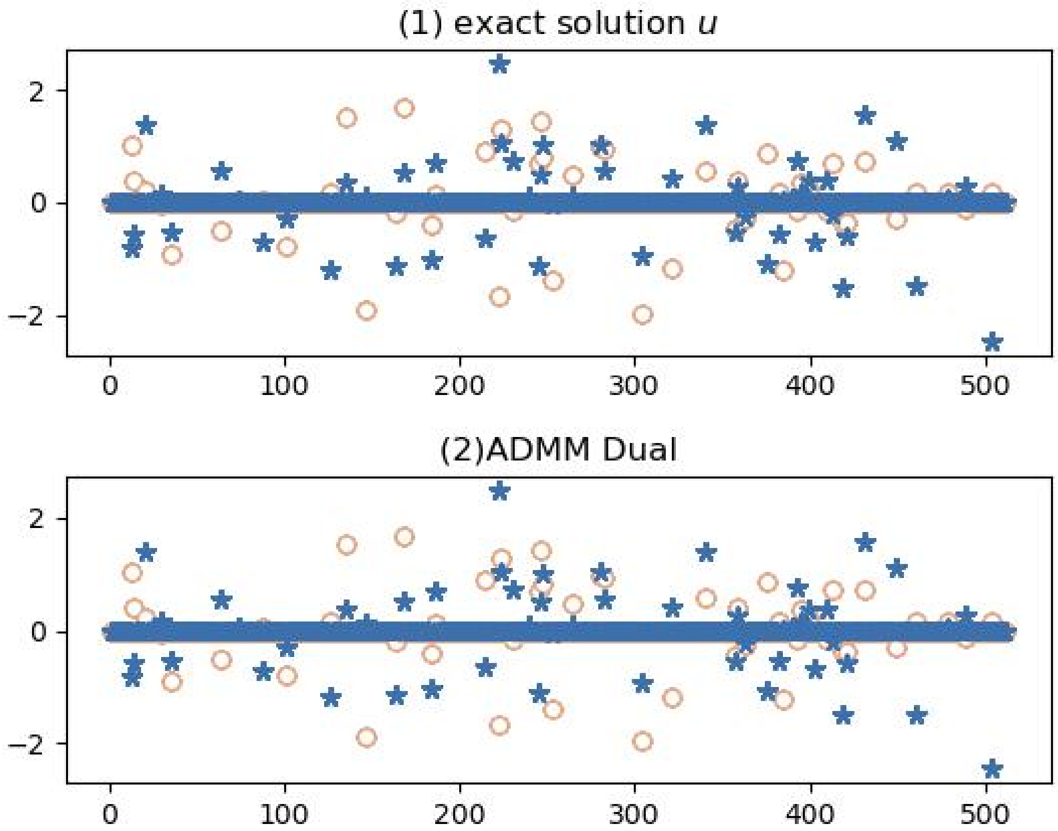

图 9: ADMM Dual 得到的解与精确解 u 对比

9 原问题的 ADMM 算法

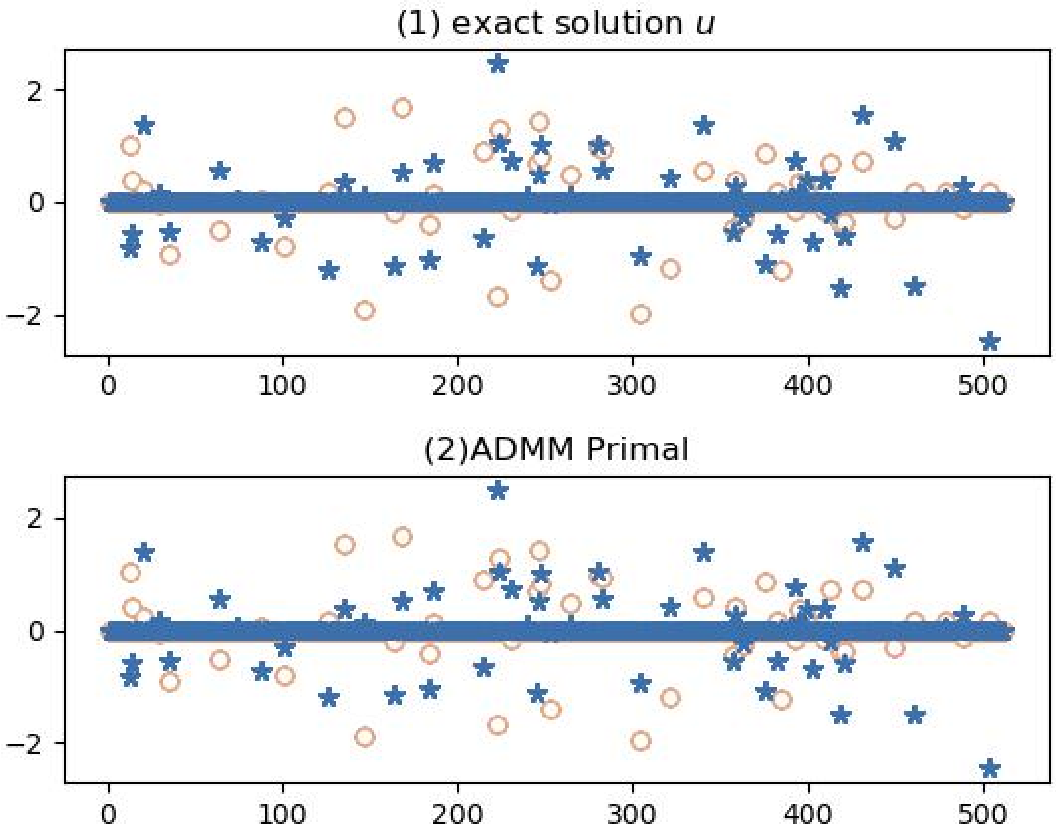

图 10: ADMM Primal 得到的解与精确解 u 对比

10 总结

表 1: 比较不同算法的效率

Solver	Time(s)	Iter	Fval	Sparisity	Errfun-Exact	Errfun-CVX-Mosek	Errfun-CVX-Gurobi
CVX-Mosek	0.46	9	5.42980E-01	0.294	4.21E-05	0.00E+00	1.47E-06
CVX-Gurobi	3.86	16	5.42980E-01	0.328	4.33E-05	1.47E-06	0.00E+00
Mosek	0.25	10	5.42980E-01	0.291	4.19E-05	2.94E-07	1.72E-06
Gurobi	4.20	14	5.42986E-01	0.750	4.72E-05	6.43E-06	5.14E-06
SGD Primal	19.12	100	5.42983E-01	0.578	5.79E-05	1.90E-05	1.79E-05
ProxGD Primal	22.34	13	5.42980E-01	0.291	4.19E-05	3.11E-07	1.73E-06
FProxGD Primal	4.99	6	5.42980E-01	0.291	4.19E-05	3.19E-07	1.74E-06
ALM Dual	1.54	283	5.42997E-01	0.248	7.59E-05	4.33E-05	4.26E-05
ADMM Dual	0.69	76	5.42998E-01	0.248	8.07E-05	4.67E-05	4.59E-05
ADMM Primal	9.52	2392	5.42980E-01	0.291	4.19E-05	3.21E-07	1.71E-06