Existing problem

1.计算速度慢

2.参数选择方案

- 2.1.没有用公式统一

3.对比方法

- 3.1.起点不一致
- 3.2.方法数量不够

4.效果不够好

- 4.1.增大batchsize以超过uncertainty
- 4.2.设置参数更新最大值
- 4.3.对label set使用不同的权
- 4.4.更换计算样本难易度的方式

5.其他

- 5.1.核函数

Question review

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_l} v_i(y_i - \sum_{\boldsymbol{x}_k \in L} \theta_k k(\boldsymbol{x}_k, \boldsymbol{x}_i))^2 + \sum_{j=1}^{n_u} \left[v_j \cdot w_j \Big(\sum_{\boldsymbol{x}_k \in L} \theta_k k(\boldsymbol{x}_k, \boldsymbol{x}_j) \Big)^2 + 2v_j \cdot w_j \left| \sum_{\boldsymbol{x}_k \in L} \theta_k k(\boldsymbol{x}_k, \boldsymbol{x}_j) \right| \right] + \gamma \boldsymbol{\theta}^T K_{LL} \boldsymbol{\theta}.$$

$$\Rightarrow$$

$$\left\| \sqrt{diag(v_l)}(y_l - K_{LL}^T \theta) \right\|_2^2 + \left\| \sqrt{diag(\eta_u)}(K_{LU}^T \theta) \right\|_2^2 + \gamma \theta^T K_{LL} \theta + 2 \left\| diag(\eta_u) K_{LU}^T \theta \right\|_1$$

Where

$$\eta_i = v_i \cdot w_i$$

Let

$$\boldsymbol{J} = diag(\eta_u) K_{LU}^T \boldsymbol{\theta},$$
$$\boldsymbol{\theta} = (K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J}$$

We have

$$\min_{J} \left\| \sqrt{diag(v_l)}\left(y_l - K_{LL}^{T}\left(K_{LU}^{T}\right)^{+} diag(\eta_u)^{-1}\boldsymbol{J}\right) \right\|_2^2 + \left\| \sqrt{diag(\eta_u)}\left(diag(\eta_u)^{-1}\boldsymbol{J}\right) \right\|_2^2 + 2\|\boldsymbol{J}\|_1 + \gamma\left(\left(K_{LU}^{T}\right)^{+} diag(\eta_u)^{-1}\boldsymbol{J}\right)^{T} K_{LL}\left(K_{LU}^{T}\right)^{+} diag(\eta_u)^{-1}\boldsymbol{J}$$

L1-minimization

- **Proximal Gradient (PG) Methods**
- Gradient Projection (GP) Methods
- Iterative Shrinkage-Thresholding (IST) Methods
- Alternating direction method of multipliers (ADMM)

PGD

- Accelerated proximal gradient (APG) [IEEE Transactions on Image Processing, 2009]
- APG for nonconvex problem [NIPS, 2015]
- Fast stochastic PG [SIAM, 2014]
- Fast stochastic APG [NIPS, 2016]

$$\min_x f(x) + \lambda\|x\|_1$$

若 f(x) 连续可导，且f'(x)满足L-Lipschitz条件

Repeat until converge:

$$v_k = x_k - \frac{1}{L}\nabla f(x_k)$$

$$x_{k+1} = \arg\min_x \frac{\mathrm{L}}{2}\|x - v_k\|_2^2 + \lambda\|x\|_1$$

T =λ/L

软阈值(Soft Thresholding)函数

$$soft(x,T) = \begin{cases} x+T & x \le -T, \\ 0 & |x| \le T, \\ x-T & x \ge T. \end{cases}$$

$$f(\boldsymbol{J}) = \begin{array}{c} \left\| \sqrt{diag(v_l)}(y_l - K_{LL}^T (K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J}\ ) \right\|_2^2 + \\ \left\| \sqrt{diag(\eta_u)}(diag(\eta_u)^{-1} \boldsymbol{J}) \right\|_2^2 \end{array}$$

$$+\gamma((K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J})^T K_{LL} (K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J}$$

$$g(\boldsymbol{J}) = 2\|\boldsymbol{J}\|_1$$

$$\frac{1}{2} f'(\boldsymbol{J})$$
$$= (diag(\eta_u)^{-1} K_{LU}^+ K_{LL}^T diag(v_l) K_{LL}^T (K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J}$$
$$- (diag(\eta_u)^{-1} K_{LU}^+ K_{LL}^T diag(v_l) y_l + diag(\eta_u)^{-1} \boldsymbol{J}$$
$$+ \gamma(diag(\eta_u)^{-1} K_{LU}^+ K_{LL}^T (K_{LU}^T)^+ diag(\eta_u)^{-1} \boldsymbol{J}$$

$$\frac{1}{2} f''(\boldsymbol{J})$$
$$= (diag(\eta_u)^{-1} K_{LU}^+ K_{LL}^T diag(v_l) K_{LL}^T (K_{LU}^T)^+ diag(\eta_u)^{-1} + diag(\eta_u)^{-1}$$
$$+ \gamma(diag(\eta_u)^{-1} K_{LU}^+ K_{LL}^T (K_{LU}^T)^+ diag(\eta_u)^{-1}$$

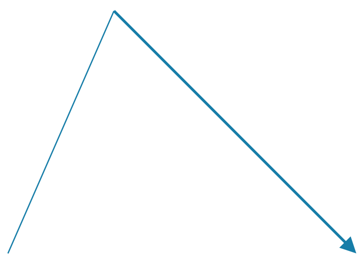
APG

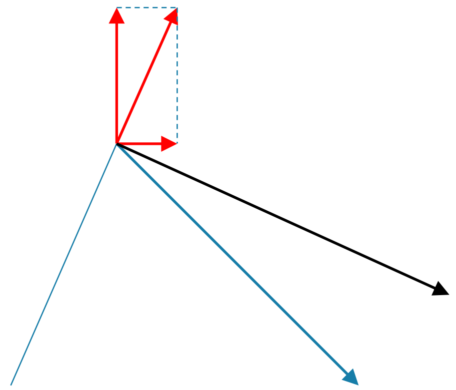
$$g_k = \mu \nabla f(x_k)$$

$$x_{k+1} = x_k - g_k$$

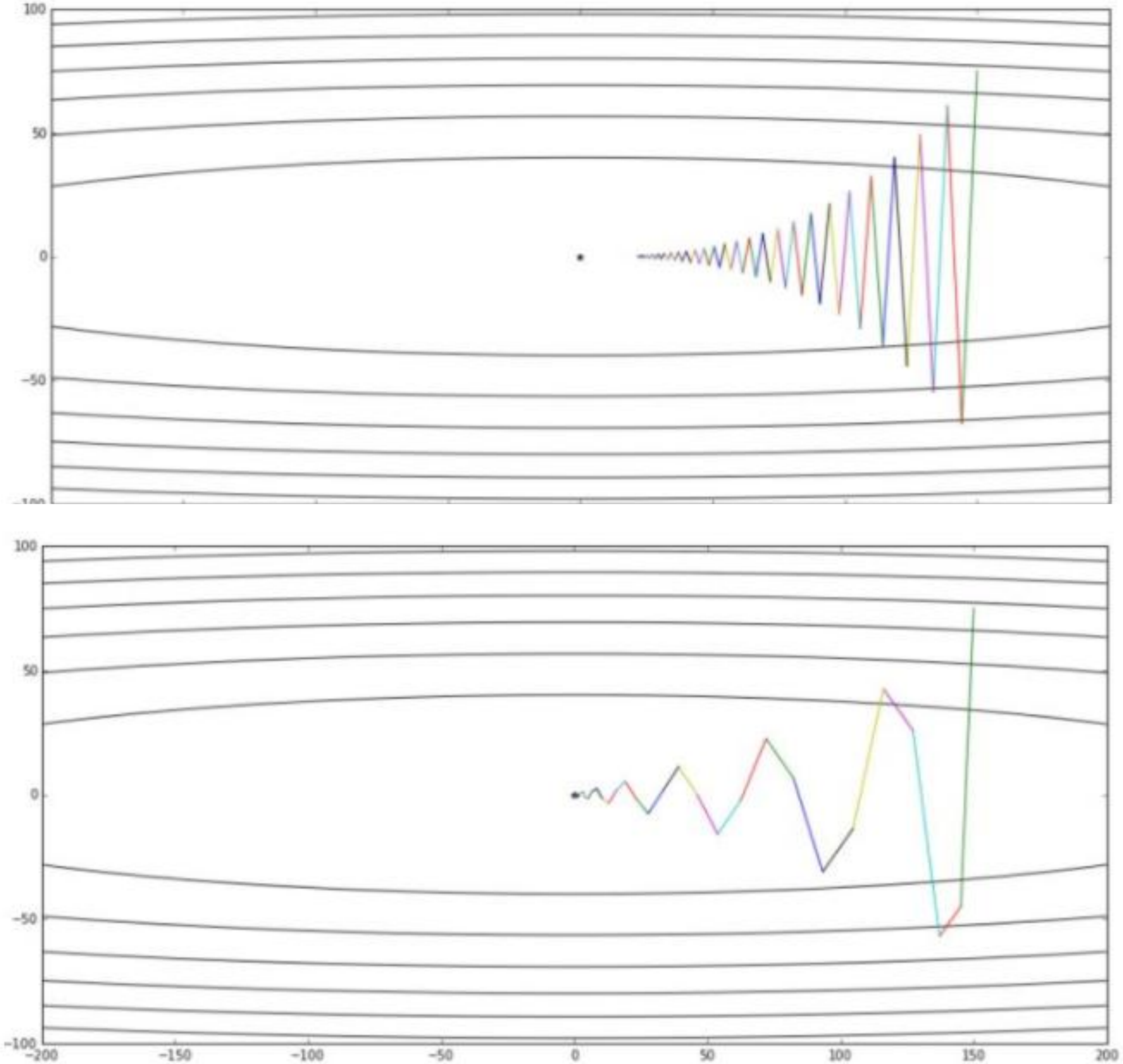
$$g_k = \mu \nabla f(x_k) + \sigma g_{k-1}$$

$$x_{k+1} = x_k - g_k$$





APG



MFISTA

Input: $L \geq L(f)$ - An upper bound on the Lipschitz constant of ∇f.

Step 0. Take $\mathbf{y}_1 = \mathbf{x}_0 \in \mathbb{E}$, $t_1 = 1$.

Step k. $(k \geq 1)$ Compute

$$\mathbf{z}_k = p_L(\mathbf{y}_k),$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \tag{5.2}$$

$$\mathbf{x}_k = \operatorname{argmin}\{F(\mathbf{x}) : \mathbf{x} = \mathbf{z}_k, \mathbf{x}_{k-1}\} \tag{5.3}$$

$$\mathbf{y}_{k+1} = \mathbf{x}_k + \left(\frac{t_k}{t_{k+1}}\right)(\mathbf{z}_k - \mathbf{x}_k) + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}_k - \mathbf{x}_{k-1}). \tag{5.4}$$

Test on LASSO

Method	Iterations	Time (s)	$p^\star$	Error (abs)	Error (rel)
CVX	15	26.53	16.5822	—	—
Proximal gradient	127	0.72	16.5835	0.09	0.01
Accelerated	23	0.15	16.6006	0.26	0.04
ADMM	20	0.07	16.6011	0.18	0.03

$$\boldsymbol{\theta}^{k+1} = A^{-1}\boldsymbol{r}^T ,$$
$$\boldsymbol{z}^{k+1} = diag(\boldsymbol{\epsilon})^{-1}\boldsymbol{\zeta},$$
$$\boldsymbol{\delta}^{k+1} = \boldsymbol{\delta}^k + \rho \cdot diag(\boldsymbol{\eta})(\boldsymbol{z}^{k+1} - K_{LU}^T(\boldsymbol{\theta}^{k+1})) ,$$

$$A = K_{LL}\sqrt{diag(\boldsymbol{v_l})}K_{LL}^T + \frac{\rho}{2}K_{LU}diag(\boldsymbol{\eta})^2K_{LU}^T + \gamma K_{LL},$$
$$\boldsymbol{r} = \boldsymbol{y_l}^T\sqrt{diag(\boldsymbol{v_l})}K_{LL}^T + \frac{1}{2}\boldsymbol{\delta}^{k^T}diag(\boldsymbol{\eta})K_{LU}^T + \frac{\rho}{2}\boldsymbol{z}^{k^T}diag(\boldsymbol{\eta})^2K_{LU}^T,$$
$$\boldsymbol{\epsilon} = \sqrt{(\boldsymbol{\eta} + \frac{\rho}{2}\boldsymbol{\eta}\circ\boldsymbol{\eta})},$$
$$\boldsymbol{\zeta} = \arg\min \frac{1}{2}||\boldsymbol{\zeta} - \boldsymbol{v}||_2^2 + \sum_{j=1}^{n_u}\xi_j|v_j| = sign(\boldsymbol{v})(|\boldsymbol{v}| - \boldsymbol{\xi})_+,$$
$$\boldsymbol{v} = \frac{1}{2}diag(\boldsymbol{\epsilon})^{-1}(\rho \cdot diag(\boldsymbol{\eta})^2K_{LU}^T\boldsymbol{\theta}^{k+1} - diag(\boldsymbol{\eta})\boldsymbol{\delta}^k),$$
$$\boldsymbol{\xi} = diag(\boldsymbol{\epsilon})^{-1}\boldsymbol{\eta} .$$

filter out some less important constraints for efficiency by multiplying the sample weights *vj* · *wj*.

$$s.t. \quad v_j \cdot w_j(z_j - \sum_{\boldsymbol{x}_k \in L}\theta_k k(\boldsymbol{x}_k, \boldsymbol{x}_j)) = 0 \quad \forall j = 1 \cdots n_u .$$

Optimizing only with the samples whose weights are larger than 0.001.

- ADMM_origin : 放松约束条件
- ADMM_fast : 不放松约束，但仅用权大于0.001的未标记样本优化目标
- MFISTA :仅用权大于0.001的未标记样本优化目标

对不同初始化，重复120次优化，记录迭代至退出循环的CPU时间与退出循环时的目标函数数值

	CPU time	Object value
ADMM_origin	0.9109±1.5798	4.5355±8.2427
ADMM_fast	0.0410±0.0433	4.5116±8.2044
MFISTA	0.0132±0.0330	2.9878±5.2008