HW2

wangxinshi47

February 2023

1 Problem

$f(x, x_0, \alpha) = \frac{1}{2\alpha}||x - x_0||_2^2 + ||x_1||_1$

$\partial f(x_i, (x_0)_i, \alpha) = \frac{1}{\alpha}(x_i - (x_0)_i) + \begin{cases} -1, & \text{if } x_i < 0 \\ [-1, 1], & \text{if } x_i = 0 \\ 1, & \text{if } x_i > 0 \end{cases}$

$\partial^2 f(x_i, x_{i0}, \alpha) = \begin{bmatrix} \frac{1}{\alpha} & 0 & \dots & 0 \\ 0 & \frac{1}{\alpha} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\alpha} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix}$

Since the hessian matrix is a positive definite matrix (identity matrix), we know it is a strictly convex function. Therefore there must be a unique solution.

From participation 4 we know that:

$$s_\lambda(\alpha) = \begin{cases} \alpha - \lambda, & \text{if } \alpha < -\lambda \\ \alpha + \lambda, & \text{if } \alpha > \lambda \\ 0, & \text{if } \alpha \in \lambda[-1, 1] \end{cases}$$

If $(x_0)_i$ is an element of $\mathbf{x_0}$, we have:

$$s_\alpha((x_0)_i) = \begin{cases} (x_0)_i - \alpha, & \text{if } (x_0)_i > \alpha \\ 0, & \text{if } (x_0)_i \in [-\alpha, \alpha] \\ (x_0)_i + \alpha, & \text{if } (x_0)_i < -\alpha \end{cases}$$

If $s_\alpha(x_0)_i$ is an element of $\mathbf{s}_\alpha(\mathbf{x_0})$, we have:

$$s_\alpha((x_0)_i) = \begin{cases} (x_0)_i - \alpha, & \text{if } (x_0)_i > \alpha \\ 0, & \text{if } (x_0)_i \in [-\alpha, \alpha] \\ (x_0)_i + \alpha, & \text{if } (x_0)_i < -\alpha \end{cases}$$

Therefore, we have: $s_\alpha((x_0)_i) = s_\alpha(x_0)_i$

2 Problem 2

Since x^* is a fixed point of (Prox), we know $X_t = X_{t+1} = X$. Therefore we can find the optimal solution for Prox by letting $0 \in \partial f$ (Fermat's condition).

$$argmin_x f = \frac{1}{2\alpha_t}||X - (X_t - \alpha_t \nabla f(x_t))||_2^2 + g(x)$$

By setting $AX = IX$ and $b = (X_t - \alpha_t \nabla f(x_t))$, we can find the derivative by $||AX - b||_2^2 = A^T(AX - b)$, Therefore

$$\begin{aligned}\partial f &= \frac{1}{\alpha_t}(X - (X_t - \alpha_t \nabla f(x_t))) + \partial g(x) \\ &= \nabla f(X_t) + \partial g(X)\end{aligned}$$

The optimal condition for Prox is therefore $0 \in \nabla f(X_t) + \partial g(X)$.

The minimizer of Comp is also $\nabla f(X) + \partial g(X)$ as f is differentiable, so the optimality condition is also $0 \in \nabla f(X_t) + \partial g(X)$. Since the optimality conditions are the same, therefore X^* is a minimizer of (Comp) if it is a fixed point of (Prox).

3 Problem 3

Using (Prox), we can find X_{t+1} in the composite gradient descent method by finding the gradient of f, which in this case is $\frac{1}{2}||AX - b||_2^2$. Therefore $\nabla f(X_t) = A^T(AX_t - b)$. Thus we have $X_{t+1} = argmin_x \frac{1}{2\alpha_t}||X - (X_t - \alpha_t A^T(AX_t - b))||_2^2 + \lambda||X||_1$. Dividing the objective by λ gives us $X_{t+1} = argmin_x \frac{1}{2\alpha_t\lambda}||X - (X_t - \alpha_t A^T(AX_t - b))||_2^2 + ||X||_1$. Therefore $X_{t+1} = S_{\alpha_t\lambda}(X_t - \alpha_t(A^T(AX_t - b)))$

4 Problem 4

The sub-differential of the LASSO objective function is the gradient of $\frac{1}{2}||AX - b||_2^2$ with the sub-gradient of $\lambda||X||_1$. Therefore the sub-differential is:

$$\nabla \frac{1}{2}||AX - b||_2^2 + \partial\lambda||X||_1 = A^T(AX - b) + \begin{cases} \lambda, & \text{if } X_i > 0 \\ -\lambda, & \text{if } X_i < 0 \\ [-\lambda, \lambda], & \text{if } x_i = 0 \end{cases}$$

What conclusions do you draw about the relative merits of the two approaches to solving the LASSO problem? What are the advantages and disadvantages of solving the LASSO problem vs the least squares problem?

Response: The graph for those two methods suggest the performance of these two methods are relatively the same. However, the running time for these two methods varies significantly. The lasso subgrad method takes around 2 minutes and the ista lasso method takes more than 15 minutes. Thus the subgrad method is more suitable for this problem.

The advantage of solving the LASSO problem is it encourages sparsity. Therefore it is suitable for this problem as we want x to be sparse. The disadvantage is there is no closed form solution to this problem. If we use the least squares method, we can find x by $A^{-1}b$, which is faster than iterative methods.