MAP6207 Final Exam

1. For the function $f(x) : \mathbb{R}^n \to \mathbb{R}$:

$$f(x) = \|Ax - b\|_2^2 + \lambda \|x\|_1 \ \text{ (here } A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ \lambda > 0),$$

(a) prove that $f(x)$ is convex and explain why $f(x)$ is not differentiable.

Solution: First of all, both $h(x) = \|x\|_1$ and $g(y) = \|y\|_2^2$ are convex functions. By the property of convex functions that the composite of a convex function and a linear function is still convex, $g(Ax)$ is convex with respect to x.

Since the sum of convex functions is still convex, $f(x) = h(x) + g(Ax)$ is convex.

However, it is not differentiable at $x = 0$ since $g(x)$ is not differential at $x = 0$.

(b) Write down the update formula for minimizing $f(x)$ by using
(1) the subgradient descent method

(2) the proximal gradient method

(3) the alternating direction method of multipliers

(4) the coordinate minimization method.

Solution:
(1) The subgradient is given by

$$x^{(k+1)} = x^{(k)} - \alpha \partial f(x^{(k)})$$

with sub-differential given by

$$\partial f(x) = 2A^T(Ax - b) + \lambda z,$$

where $z \in \partial \|x\|_1$, and the i-th coordinate of $\partial \|x\|_1$ is

$$\begin{cases} 1, \text{if } x_i > 0 \\ [-1, 1], \text{if } x_i = 0 \\ -1, \text{if } x < 0. \end{cases}$$

(2) The proximal gradient method is given by

$$x^{(k+1)} = \text{prox}_\alpha(x - \alpha \cdot 2A^T(Ax^{(k)} - b)),$$

with

$$\text{prox}_\alpha(z) = \frac{1}{2\alpha}(z - x)^2 + \lambda\|x\|_1.$$

The explicit formula of the proximal operator can be written as

$$[\text{prox}_\alpha(z)]_i = \max(0, |z_i| - \alpha\lambda)\text{sign}(z_i).$$

(3) Write the problem as

$$\arg\min_{x,y} \|Ax - b\|^2 + \lambda\|y\|_1, \ \text{subject to } x = y,$$

then the Lagrangian is given by

$$L(x, y, v) = \|Ax - b\|^2 + \lambda\|y\|_1 + \nu^T(x - y) + \frac{\rho}{2}\|x - y\|^2.$$

Then the update formula is given by

$$\begin{aligned}
x^{(k+1)} &= \arg\min_x L(x, y^{(k)}, \nu^{(k)}) \\
y^{(k+1)} &= \arg\min_y L(x^{(k+1)}, y, \nu^{(k)}) \\
\nu^{(k+1)} &= \nu^{(k)} + \rho(x^{(k+1)} - y^{(k+1)}).
\end{aligned}$$

(4) (follows from lecture note 22) Suppose we would like to update the i-th coordinate. Then write the j-th column of A as a_j, and the update formula is

$$x_i^+ = \frac{a_i(y_i - \sum_{j\neq i} a_j x_j)}{a_j^T a_j}.$$

2. Given a function $f(x) : \mathbb{R}^n \to \mathbb{R}$ such that $\nabla f(x)$ is Lipschitz continuous with a Lipschitz factor of $L > 0$ (i.e., $\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$ for all $x, y \in \mathbb{R}^n$), prove

(a)

$$f(y) \le f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2.$$

(b) if $x^* = \arg\min f(x)$, then

$$\frac{1}{2L}\|\nabla f(x)\|^2 \le f(x) - f(x^*)$$

(c)

$$f(y) \ge f(x) + \nabla f(x)^T (y - x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$$

Solution: (a)(b) Follows from HW 2, and (c) follows form HW 7 (see equation (1) in the solution.)

3. Consider the problem of projecting a point $a \in \mathbb{R}^n$ on the unit ball in ℓ_1 norm:

$$\min \frac{1}{2}\|x-a\|^2, \quad \text{subject to } \|x\|_1 \leq 1.$$

Derive the dual problem and describe an efficient method for solving it. Explain how you can obtain the optimal x from the solution of the dual problem.

Solution: (a) The Lagrangian is given by

$$\frac{1}{2}\|x-a\|^2 + \lambda(\|x\|_1 - 1),$$

so its dual function is given by

$$g(\lambda) = \min_x \frac{1}{2}\|x-a\|^2 + \lambda(\|x\|_1 - 1) = -\lambda + \sum_{i=1}^n \left[\min_{x_i} \frac{1}{2}(x_i - a_i)^2 + \lambda|x_i|\right],$$

and it can be shown that

$$\min_{x_i} \frac{1}{2}(x_i - a_i)^2 + \lambda|x_i| = \begin{cases} -\frac{1}{2}\lambda^2 + \lambda|a_i|, \text{ if } |a_i| > \lambda \\ \frac{1}{2}a_i^2, \text{ if } |a_i| \leq \lambda. \end{cases}$$

In summary, we have

$$g(\lambda) = -\lambda + \sum_{i=1}^n \max(\frac{1}{2}a_i^2, -\frac{1}{2}\lambda^2 + \lambda|a_i|).$$

This is a convex function that maps $\mathbb{R}$ to $\mathbb{R}$ (in fact, all dual problems are convex), so it can be solved by the subgradient method.

Alternatively, one may also set $g'(\lambda) = 0$, then the solution is given by

$$\sum_{|a_i|>\lambda} (|a_i| - \lambda) = 1.$$

One may use bisection method to solve it, note that the LHS is nonincreasing with respect to λ.

(b) If λ^* is known, then by KKT condition, we have

$$x_i^* = \arg\min_{x_i} \frac{1}{2}(x_i - a_i)^2 + \lambda^*|x_i| = \text{sign}(a_i)\max(|a_i| - \lambda^*, 0).$$

4. Consider a convex optimization problem

$$\min_x f_0(x), \quad \text{subject to } f_i(x) \leq 0,\ i = 1, 2, \cdots, m$$

and its dual

$$\max_\lambda g(\lambda), \quad \text{subject to } \lambda \geq 0. \tag{1}$$

The centering problem in the barrier method is

$$\min_x t\, f_0(x) - \sum_{i=1}^m \ln(-f_i(x)), \tag{2}$$

where t is a positive number.

(a) The centering problem can be written as

$$\min_{x,y} t\, f_0(x) - \sum_{i=1}^m \ln(y_i), \quad \text{subject to } f_i(x) + y_i \leq 0, i = 1, 2, \cdots, m$$

with variables x and y. Derive the Lagrange dual of this problem and express it in terms of the dual function $g(\lambda)$ in (1).

(b) Suppose the feasible set of the dual problem in (1) contains strictly positive λ. Show that the centering problem (2) is bounded below for any positive t.

Solution: (a) The dual function is given by

$$\begin{aligned}
\bar{g}(\lambda) =& \min_{x,y} t\, f_0(x) - \sum_{i=1}^m \ln(y_i) + \sum_{i=1}^m \lambda_i(f_i(x) + y_i) \\
=& t \min_x [\, f_0(x) + \sum_{i=1}^m \frac{\lambda_i}{t} f_i(x)] + \min_y [- \sum_{i=1}^m \ln(y_i) + \sum_{i=1}^m \lambda_i y_i] \\
=& t g(\frac{\lambda}{t}) + m + \sum_{i=1}^m \ln \lambda_i.
\end{aligned}$$

(b) By weak duality, the lower bound of the centering problem is larger than

$$\max_{\lambda \geq 0} \bar{g}(\lambda).$$

For strictly positive λ_0, $\bar{g}(\lambda_0)$ is well-defined. So if the feasible set of the dual problem in (1) contains strictly positive λ, the lower bound of the centering problem is larger than this well-defined $\bar{g}(\lambda_0)$. This means that the centering problem (2) is bounded below.

Space for showing more work

Space for showing more work

Space for showing more work