Homework 3

Chen Yihang 1700010780

March 27, 2020

Contents

1 Problem 1

1.1 (a)

Due to the chain rule, as well as $f(x)$ is a nonnegative combination

$$\partial f(x) = A^\top \partial \|Ax - b\|_2 + \partial \|x\|_2$$

Obviously, when $x \neq 0$, $\partial \|x\| = \frac{x}{\|x\|_2}$. When $x = 0$, by definition, we have

$$\|y\|_2 \geq g^\top y, \forall y \in \mathbb{R}^n$$

which leads to $\|g\|_2 \leq 1$. Hence, $\partial \|x\|_2 = \{g | \|g\|_2 \leq 1\}$.

Hence,

$$\partial f(x) = \begin{cases} A^\top (Ax - b)/\|Ax - b\|_2 + x/\|x\|_2, & \text{if } x \neq 0, Ax - b \neq 0 \\ \{-A^\top b/\|b\|_2 + g | \|g\|_2 \leq 1\}, & \text{if } x = 0, b \neq 0 \\ \{A^\top g + x/\|x\|_2 | \|g\|_2 \leq 1\}, & \text{if } x \neq 0, Ax - b = 0 \\ \{A^\top g_1 + g_2 | \|g_1\|_2, \|g_2\|_2 \leq 1\}, & \text{if } x = 0, b = 0 \end{cases}$$

1.2 (b)

We first compute $\partial \|z\|_\infty$. Since $\|z\|_\infty = \sup_{\|w\|_1} z \top w$, by the **Danskin-Bertseka's Theorem for subdifferentials**, since $\{w : \|w\|_1 \leq 1\}$ is compact, and $w^\top z$ is convex in z for every w, we have

$$\begin{aligned} \partial \|z\|_\infty &= \operatorname{conv}\{w : \|w\|_1 \leq 1, z^\top w = \|z\|_\infty\} \\ &= \{w : \|w\|_1 \leq 1, z^\top w = \|z\|_\infty\} \end{aligned}$$

For given x, we minimize $\|Ay - x\|_\infty$ w.r.t. y, which is a LP problem. Then

$$\partial_{x,y} \|Ay - x\|_\infty = \partial_{x,y} \| \begin{pmatrix} -I & A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \|_\infty = \begin{pmatrix} -I \\ A^\top \end{pmatrix} = \{ \begin{pmatrix} -I \\ A^\top \end{pmatrix} w : \|w\|_1 \leq 1, w^\top (Ay - x) = \|Ay - x\|_\infty\}$$

Since we set $A^\top w = 0$, hence the subgradient is

$$\{w : A^\top w = 0, \|w\|_1 \leq 1, w^\top \hat{x} = \|A\hat{y} - \hat{x}\|_\infty\}$$

at point $\hat{x}$.

Now we need to prove that such w exist. We need first find $\hat{y}$ to minimize $\|Ay - x\|_\infty$. If $\|\hat{x} - A\hat{y}\|_\infty = 0$, we set $w = 0$. Otherwise, the problem can be reformulated as

$$\begin{aligned} \min_{t,y} \quad & t \\ \text{s.t.} \quad & -t\mathbb{1} \preceq Ay - \hat{x} \preceq t\mathbb{1} \end{aligned} \tag{1.1}$$

whose Lagrangian can be written as

$$L(t, y, a, b) = t + a^\top (Ay - \hat{x} - t\mathbb{1}) + b^\top (-t\mathbb{1} - Ay + \hat{x}) \tag{1.2}$$

where $a, b \succeq 0$.

The dual problem is

$$\begin{aligned} \min_{a,b} \quad & \hat{x}^\top (a-b) && \text{(1.3a)} \\ \text{s.t.} \quad & A^\top (a-b) = 0, \quad \mathbb{1}^\top (a+b) = 1 && \text{(1.3b)} \\ & a, b \succeq 0 && \text{(1.3c)} \end{aligned}$$

Assuming the optimal solution is $t^\star, y^\star, a^\star, b^\star$, then the KKT condition amounts to

$$A^\top (a^\star - b^\star) = 0 \tag{1.4a}$$

$$\mathbb{1}^\top (a^\star + b^\star) = 1 \tag{1.4b}$$

$$(a^\star)^\top (Ay^\star - \hat{x} - t^\star \mathbb{1}) = 0 \tag{1.4c}$$

$$(b^\star)^\top (-t^\star \mathbb{1} - Ay^\star + \hat{x}) = 0 \tag{1.4d}$$

We claim that $w = a^\star - b^\star$ satisfy 1.2. According to 1.4a, $A^\top w = 0$; and from 1.4b, 1.4c and 1.4d, we have

$$w^\top \hat{x} = (a^\star - b^\star)^\top (Ay^\star - \hat{x}) = (a^\star + b^\star)^\top \mathbb{1} t^\star = t^\star$$

Besides, if $a_i^\star \neq 0$, then from 1.4c we have $(Ay^\star - \hat{x})_i = t$, combined with 1.4d and $t > 0$, we have $b_i^\star = 0$. Similarly, if $b_i^\star \neq 0$, then $a_i^\star = 0$. Hence,

$$\|w\|_1 = \|a^\star - b^\star\|_1 = \|a^\star\|_1 + \|b^\star\|_1 = \mathbb{1}^\top (a^\star + b^\star) = 1$$

Therefore, the existence is proved. Form the proof, we can easily find that by solving the dual 1.3, $\partial f(\hat{x})$ is naturally obtained by setting $w = a^\star - b^\star$.

2 Problem 2

2.1 (a)

Define

$$\delta(x) = \begin{cases} 0, & \|x\|_\infty \leq 1 \\ +\infty, & \text{otherwise} \end{cases}$$

Hence, $f(x) = \|x\|_1 + \delta(x)$. If we optimize elementwisely, i.e.

$$\min_{u_i} \quad |u_i| + \delta(u_i) + \frac{1}{2}|u_i - x_i|^2$$

which means

$$\begin{aligned} \min_{u_i} \quad & |u_i| + \frac{1}{2}|u_i - x_i|^2 \\ \text{s.t.} \quad & -1 \leq u_i \leq 1 \end{aligned}$$

Taking the derivative w.r.t. u_i, we get

$$u_i - x_i + \partial |u_i| = 0$$

The solution is $u_i = \text{sign}(x_i) \max(|x_i| - 1, 0)$. Considering $|u_i| \leq 1$, since the problem is quadratic and have positive leading coefficient, the solution should be modified to $u_i = \text{sign}(x_i) \min(\max(|x_i| - 1, 0), 1)$.

Hence

$$\text{prox}_f(x)_i = \text{sign}(x_i) \min(\max(|x_i| - 1, 0), 1)$$

2.2 (b)

Take the subgradient, we have

$$\partial f(u) + u - x = 0$$

Assume e_i is the vector whose i-th item is 1 and the others are 0, according to the relation of taking maximum and subgradient, we have

$$\partial f(u) = \mathrm{conv}\{e_i | u_i = f(u)\}$$

where the equality holds since each x_k is linear (differentiable). Hence, we have

$$u_j = \begin{cases} x_j, & u_j \neq f(u) \\ x_j - c_j & u_j = f(u) \end{cases}$$

where c_j satisfies $\sum_{j:u_j=f(u)} c_j = 1$.

Assuming the solution of $\sum_k \max(x_k - t, 0) = 1$ w.r.t. $t \in \mathbb{R}$ is $\hat{t}$. Clearly, $\hat{t}$ exist and is unique. This equation can be solved by bisecting t with the initial interval $[\min_i x_i - \frac{1}{n}, \max_i x_i]$.

We thus have $\mathrm{prox}_f(x)_i = \min(x_i, \hat{t})$.

2.3 (c)

Since

$$\mathrm{prox}_f(x) = \min_u \left\{ \|Au - b\|_1 + \frac{1}{2}\|u - x\|_2^2 \right\}$$

which can be formulated as

$$\min_{u,z} \quad \|z\|_1 + \frac{1}{2}\|u - x\|_2^2 \tag{2.1a}$$

$$\text{s.t.} \quad z = Au - b \tag{2.1b}$$

Define the optimal solution of 2.2 by $(\hat{z}, \hat{u})$. The existence of such solution follows from the fact that $\|x\|_1$ is convex and positive. Clearly $\hat{u} = \mathrm{prox}_f(x)$, fixing $z = \hat{z}$, we have that $\hat{u}$ is the optimal solution of

$$\begin{aligned} \min_{u,z} \quad & \frac{1}{2}\|u - x\|_2^2 \\ \text{s.t.} \quad & Au = \hat{z} + b \end{aligned}$$

Thanks to the strong duality, there exsit w such that

$$\hat{u} \in \arg\min_u \left\{ \frac{1}{2}\|u - x\|_2^2 + \langle y, Au - \hat{z} - b\rangle \right\} \tag{2.3a}$$

$$A\hat{u} = \hat{z} + b \tag{2.3b}$$

By 2.3a, we have

$$\hat{u} = x - A^\top y \tag{2.4}$$

Substituting it into 2.3b, we have

$$A(x - A^\top y) = \hat{z} + b$$

Since $AA^\top = D$, and D is a diagonal matrix whose diagonal elements are positive,

$$y = D^{-1}(Ax - \hat{z} - b)$$

which combined with 2.4, we have

$$\hat{u} = x + A^\top D^{-1}(\hat{z} + b - Ax) \tag{2.5}$$

Substituting it into 2.2, we have

$$\hat{z} = \arg\min_z \left\{ \|z\|_1 + \frac{1}{2}\|A^\top D^{-1}(z + b - Ax)\|_2^2 \right\} \tag{2.6}$$

Taking the derivative w.r.t. z, assuming $d_i = D_{ii}$, we have

$$\partial\|z_i\|_1 + (z + b - Ax)_i/d_i = 0$$

which leads to

$$z_i = \operatorname{sign}(Ax - b)_i \max(|(Ax - b)|_i - d_i, 0)$$

Substituting it into 2.5

$$\operatorname{prox}_f(x) = \hat{u} = x + A^\top D^{-1}\left(b - Ax + \operatorname{sign}(Ax - b) \odot \max(|Ax - b| - \operatorname{diag}(D), 0)\right)$$

3 Problem 3

Taking the derivative w.r.t. y, we have

$$y - w + \frac{1}{(1-\alpha)\sigma n}\sum_{i=1}^n \partial(y_i - t)_+$$

where

$$\partial(x)_+ = \begin{cases} 1, & x > 0 \\ [0,1], & x = 0 \\ o, & x < 0 \end{cases}$$

WLOG, assuming $w_1 \le w_2 \le \cdots, w_n$, and $w_0 = -\infty, w_{n+1} = \infty$. For each t, we can readily find u, v, such that $w_u \le t < w_{u+1}, w_v \le t + \frac{1}{(1-\alpha)\sigma n} < w_{v+1}$.

Then, for $i \le u$, $y_i = w_i$; for $i \ge v+1$, $y_i = w_i - \frac{1}{(1-\alpha)\sigma n} > t$; for $u+1 \le i \le v$, y_i must be t. Substituting it into the original problem, we have

$$\min_t \quad t + \sum_{i=v+1}^n \frac{w_i - t}{(1-\alpha)n} + \sum_{i=u+1}^v (w_i - t)^2 + \left(\frac{1}{(1-\alpha)\sigma n}\right)^2 (n - v)$$

i.e. given u, v

$$\min_t \quad (v-u)t^2 + \left(1 - \frac{n-v}{(1-\alpha)n} - 2\sum_{i=u+1}^v w_i\right)t$$

Keeping u, v, it is a quadratic minimization. Denote $w'_i = w_i - \frac{1}{(1-\alpha)\sigma n}$. Then w_i, w'_i divide the real line into at most $2n+1$ interval. Each interval corresponds to a set of (u, v). For each set of (u, v), we can minimize t by performing a quadratic function minimization.

If we precompute and store $\sum_{i=1}^{m} w_i$ for $m = 1, 2, \cdots, n$, the algorithm's complexity can be $O(n)$.

4 Problem 4

Assuming $X = S\Lambda S^\top$, where Λ is diagonal, since

$$\log\det(X) = \log\det(S\Lambda S^\top) = \log\det\Lambda$$

$\log\det(X)$ is determined by the eigenvalues of X. A standard result shows that the optimal solution can be written as $U^\star = SDS^\top$, where D is diagonal. Denote $d_i = D_{ii}, \lambda_i = \Lambda_{ii}$. We have

$$\begin{aligned}
&f(U) + \frac{1}{2}\|U - X\|_F^2 \\
=&f(D) + \frac{1}{2}\|\Lambda - D\|_F^2 \\
=&\sum_{i=1}^{n}\left(-\log d_i + \frac{1}{2}|\lambda_i - d_i|^2\right)
\end{aligned}$$

For each i, we optimize seperately. Note that since $X, U \in S_{++}^n$, $\lambda_i, d_i > 0$, taking the derivative w.r.t. d_i and setting it to 0, we have

$$-\frac{1}{d_i} + d_i - \lambda_i = 0$$

which leads to $d_i = \frac{\lambda_i + \sqrt{\lambda_i^2 + 4}}{2}$. (omit the negative root)

If we denote $g(X) = Sg(\Lambda)S^\top$ and operations are taken element-wise on Λ. Define $g(x) = \frac{x + \sqrt{x^2+4}}{2}, x \in \mathbb{R}$, then

$$\mathrm{prox}_f(X) = g(X)$$