**DATA130026.01 Optimization**
**Solution of Assignment 11**

1. Compute the projection of a given point $x$ to the second order cone $Q = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \le t\}$.
**Solution.** Consider the Lagrangian dual for the projection problem

$$\begin{array}{ll} \min_{u,v} & \frac{1}{2}\|(u,v)-(x,t)\|_2^2 \\ \text{s.t.} & -(u,v) \preceq_Q 0, \end{array}$$

where $(u,v) \succeq_Q 0$ denote $\|u\| \le v$. Suppose the dual variable is $(\mu,\nu) \succeq_Q 0$. Then we have

$$L = \frac{1}{2}\|(u,v)-(x,t)\|_2^2 - (\mu,\nu)^T(u,v).$$

And the optimality condition is

$$\begin{array}{rcl} u-\mu & = & x, \\ v-\nu & = & t, \\ (u,v) & \succeq_Q & 0, \\ (\mu,\nu) & \succeq_Q & 0, \\ u^T\mu + v\nu & = & 0. \end{array}$$

It follows from the Cauchy-Schwarz inequality that the last three conditions are satisfied if one of the following three cases holds.

- $u=0, v=0$, $\|\mu\| \le \nu$. The first two conditions give $\mu = -x$, $\nu = -t$. The forth implies $t < 0$ and $\|x\|_2 \le -t$. In this case $(x,t)$ is in the negative second-order cone, and its projection is the origin.
- $\mu = 0, \nu = 0$, $\|u\| \le v$. The first two conditions give $u = x$, $v = t$. The third implies $\|x\| \le t$.
  In this case $(x,t)$ is in the second-order cone, so it is its own projection.
- $\mu/\nu = -u/v$, $\|\mu\| = \nu$, $\|u\| = v$. We can express $\mu = -u\nu/v$. Then we have $u(1+\nu/v) = x$ and thus, due to $\|u\| = v$, $v + \nu = \|x\|$. Together with $v - \nu = t$, we have $v = \frac{t+\|x\|}{2}$ and $\nu = \frac{\|x\|-t}{2}$. And $v$ and $\nu$ are positive only if $|t| < \|x\|$. Now $u(1+\nu/v) = x$ implies
$$u = \frac{\|x\|+t}{2\|x\|}x.$$

From the above cases, we have

$$P_Q(x,t) = \begin{cases} 0, & \|x\| \le -t, \\ (x,t), & \|x\| \le t, \\ (\frac{\|x\|+t}{2\|x\|}x, \frac{t+\|x\|}{2}), & \|x\| \ge |t|. \end{cases}$$

2. Show that the projection onto the set $C = [l, u] = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ can be written as

$$P_C(x)_i = \begin{cases} l_i, & \text{if } x_i \leq l_i, \\ x_i, & \text{if } l_i \leq x_i \leq u_i, \\ i_i, & \text{if } x_i \geq u_i. \end{cases}$$

**Solution.** The projection problem is

$$\begin{array}{ll} \min_y & \frac{1}{2}\|y - x\|_2^2 \\ \text{s.t.} & y \in C. \end{array}$$

It can be rewritten as

$$\begin{array}{ll} \min_y & \frac{1}{2}\sum_{i=1}^{n}(y_i - x_i)^2 \\ \text{s.t.} & l_i \leq y_i \leq u_i, \quad i = 1, \ldots, n. \end{array}$$

Note that the problem is separable. For $i = 1, \ldots, n$, we have

$$\begin{array}{ll} \min_{y_i} & (y_i - x_i)^2 \\ \text{s.t.} & l_i \leq y_i \leq u_i. \end{array}$$

Thus, the projection onto $C$ can be written as

$$P_C(x)_i = \begin{cases} l_i, & \text{if } x_i \leq l_i, \\ x_i, & \text{if } l_i \leq x_i \leq u_i, \\ i_i, & \text{if } x_i \geq u_i. \end{cases}$$

It's easy to check by Lagrangian or by observation.

3. Compute the proximal mapping $\text{prox}_{tf}(t > 0)$ for the following function $f$.

- quadratic function ($A \succeq 0$): $f(x) = \frac{1}{2}x^T Ax + b^T x + c$.
- Euclidean norm: $f(x) = \|x\|_2$.
- logarithmic barrier: $f(x) = -\sum_{i=1}^{n} \log x_i$.

**Solution.**

- (quadratic funtion)

$$\begin{aligned} \text{prox}_{tf}(x) &= \text{argmin}_u \left( tf(u) + \frac{1}{2}\|u - x\|_2^2 \right) \\ &= \text{argmin}_u \left( \frac{t}{2}u^T Au + tb^T u + tc + \frac{1}{2}\|u - x\|_2^2 \right) \\ &= \text{argmin}_u \left( \frac{1}{2}u^T (tA + I)u + (tb - x)^T u + tc + \frac{1}{2}x^T x \right). \end{aligned}$$

Denote $g(u) := \frac{1}{2}u^T(tA+I)u + (tb-x)^T u + tc + \frac{1}{2}x^T x$. Since $A \succeq 0,\ t > 0 \Rightarrow tA + I \succ 0$, we know that the quadratic function $g(u)$ is convex. Therefore,

$$\begin{aligned} & \nabla g(u) = (tA+I)u + tb - x = 0 \\ \Rightarrow \quad & \operatorname{prox}_{tf}(x) = (tA+I)^{-1}(x - tb). \end{aligned}$$

- (Euclidean norm)

$$\begin{aligned} \operatorname{prox}_{tf}(x) &= \operatorname{argmin}_u \left( tf(u) + \frac{1}{2}\|u-x\|_2^2 \right) \\ &= \operatorname{argmin}_u \left( \|u\|_2 + \frac{1}{2t}\|u-x\|_2^2 \right). \end{aligned}$$

Denote $g(u) := \|u\|_2 + \frac{1}{2t}\|u-x\|_2^2$, we have

$$\partial g(u) = \begin{cases} \dfrac{u}{\|u\|_2} + \dfrac{1}{t}(u-x), & \text{if } u \neq 0, \\ \left\{ y - \dfrac{x}{t} \;\middle|\; \|y\|_2 \leq 1 \right\}, & \text{if } u = 0. \end{cases}$$

(i) When $u \neq 0$,

$$\begin{aligned} & \frac{u}{\|u\|_2} + \frac{1}{t}(u-x) = 0 \\ \Rightarrow \quad & (1 + \frac{t}{\|u\|_2})u = x. \end{aligned}$$

This implies $u$ and $x$ have the same direction (because $1 + \frac{t}{\|u\|_2} > 0$), i.e.,

$$x \neq 0 \quad \text{and} \quad \frac{u}{\|u\|_2} = \frac{x}{\|x\|_2}.$$

Therefore, we have

$$u = (1 - \frac{t}{\|x\|_2})x.$$

Note that $u$ and $x$ have the same direction, thus

$$1 - \frac{t}{\|x\|_2} > 0 \quad \Rightarrow \quad \|x\|_2 > t.$$

(ii) When $u = 0$,

$$\begin{aligned} & 0 \in \left\{ y - \frac{x}{t} \;\middle|\; \|y\|_2 \leq 1 \right\} \\ \Rightarrow \quad & \|\frac{x}{t}\|_2 \leq 1 \\ \Rightarrow \quad & \|x\|_2 \leq t. \end{aligned}$$

In conclusion,

$$\operatorname{prox}_{tf}(x) = \begin{cases} (1 - \frac{t}{\|x\|_2})x, & \text{if } \|x\|_2 > t, \\ 0, & \text{if } \|x\|_2 \le t. \end{cases}$$

- (logarithmic barrier)

$$\begin{aligned} \operatorname{prox}_{tf}(x) &= \operatorname{argmin}_u \left( tf(u) + \frac{1}{2}\|u - x\|_2^2 \right) \\ &= \operatorname{argmin}_u \left( -\sum_{i=1}^{n} \log u_i + \frac{1}{2t}\sum_{i=1}^{n} \|u_i - x_i\|_2^2 \right). \end{aligned}$$

Note that the problem is separable, for $i = 1, \ldots, n$, denote

$$g_i(u) := -\log u_i + \frac{1}{2t}\|u_i - x_i\|_2^2.$$

It's a convex function. Let $\nabla g_i(u) = 0$, we have

$$\begin{aligned} -\frac{1}{u_i} + \frac{u_i - x_i}{t} = 0 \quad &\Rightarrow \quad u_i^2 - x_i u_i - t = 0 \\ \overset{(u_i > 0)}{\Rightarrow} \quad &u_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}. \end{aligned}$$

Therefore,

$$\operatorname{prox}_{tf}(x)_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}, \qquad i = 1, \ldots, n.$$

4. Consider the Lasso problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \tau \|x\|_1$$

where $\tau > 0$ is a weighting parameter, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are given data. Use the following Matlab code to generate the data:

```
m = 100; n = 500; s = 50;
A = randn(m,n);
xs = zeros(n,1); picks = randperm(n); xs(picks(1:s)) = randn(s,1);
b = A*xs;
```

Choose $x = 0$ as the starting point. Implement the proximal gradient method to solve the problem. Terminate your code after 1000 iterations or the norm of gradient mapping is less than `1e-6`. Implement your algorithm with fixed step size $1/L$, where $L$ is the Lipschitz constant for the smooth part. (You need to show how to compute $L$.)
Did you get a solution exactly equal to the given $xs$? How about try a different $\tau$, or different $m$? Give a summary how the parameters $\tau$ and $m$ affect the solution. Plot the results.

**Solution.** Please run different part of `codes/problem_4.m`.

5. (a) Reformulate the following problem as a semidefinite programming (SDP) problem

$$\max_{\|B\|\le 1} A \bullet B,$$

where $A \bullet B$ denotes the inner product of two $m \times n$ $(m \le n)$ matrixes $A$ and $B$, $\|\cdot\|$ is the spectral norm (largest singular value) of a matrix. Derive the dual problem of the SDP problem.

**Solution.** First note that $\|B\| \le 1$ is equivalent to $\begin{pmatrix} I & B \\ B^T & I \end{pmatrix} \succeq 0$ due to Schur complement. Hence the above problem is equivalent to

$$\max_{B} A \bullet B, \quad \text{s.t.} \begin{pmatrix} I & B \\ B^T & I \end{pmatrix} \succeq 0$$

Hence the Lagrangian dual problem is

$$\min_{S\succeq 0} \max_{B} A \bullet B + S \bullet \begin{pmatrix} I & B \\ B^T & I \end{pmatrix}$$

Let $S = \begin{pmatrix} S_1 & S_2 \\ S_2^T & S_3 \end{pmatrix}$. Then the inner problem becomes

$$\max (A + 2S_2) \bullet B + S_1 \bullet I + S_2 \bullet I = \begin{cases} S_1 \bullet I + S_2 \bullet I & \text{if } A + 2S_2 = 0 \\ \infty & \text{otherwise.} \end{cases}$$

This implies the dual problem is

$$\begin{array}{ll} \min_S & S_1 \bullet I + S_2 \bullet I \\ \text{s.t.} & A + 2S_2 = 0, S \succeq 0. \end{array}$$

(b) Let $X$ be rank $r$ $m \times n$ matrix with $1 \le r \le m \le n$. Suppose that

$$X = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^T$$

is a singular value decomposition of $X$, where $\Sigma = \mathrm{Diag}(\sigma_1(X), \dots, \sigma_r(X)) \in \mathbb{R}^{r\times r}$ is a diagonal matrix containing the nonzero singular values of $X$. Let $\|X\|_* = \sum_{i=1}^r \sigma_i$ denote the nuclear norm of $X$. Show that

$$\partial \|X\|_* = \left\{ U \begin{pmatrix} I_r & 0 \\ 0 & W \end{pmatrix} V^T : \|W\| \le 1 \right\},$$

where $I_r$ is the $r \times r$ identity matrix and $\|W\|$ is the spectral norm (largest singular value) of $W$.

**Hint:** You may first prove the dual norm of the nuclear norm is the spectral norm and then use subgradient calculus rule to prove the result. Hint for proving the dual norm: you may first prove that dual norm of spectral norm "$\ge$" nuclear norm and then construct a dual problem of $\max_{\|B\|\le 1} A \bullet B$, and show that dual norm of spectral norm "$\le$" nuclear norm.

**Solution.** Note that the dual norm of spectral norm is defined by

$$\|X\|_d = \max\{X \bullet Y : \|Y\| \le 1\}. \tag{1}$$

Suppose that

$$X = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^T$$

is a singular value decomposition of $X$. Then $Y_0 = U \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} V^T$ is a feasible solution for (1). Hence $\|X\|_d \le X \bullet Y_0 = \mathrm{trace}(\Sigma) = \|X\|_*$.
On the other hand, from problem (a), the dual problem of (1) is

$$\begin{array}{ll} \max_S & S_1 \bullet I + S_3 \bullet I \\ \text{s.t.} & X + 2S_2 = 0, S \succeq 0. \end{array}$$

A feasible solution of this problem is

$$S = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \frac{1}{2}\Sigma & -\frac{1}{2}\Sigma \\ -\frac{1}{2}\Sigma & \frac{1}{2}\Sigma \end{pmatrix} \begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix}.$$

Denoting the dual optimal value as $d^*$, we have

$$d^* \ge \mathrm{tr}(\Sigma) = \|X\|_*$$

Because the SDP problem satisfies the Slater condition (check it by yourself), we have $\|X\|_* \le d^* = p^* \le \|X\|_*$. This implies that the dual norm of the spectral norm is the nuclear norm. So from the dual norm expression and subdifferential calculus rule, we have

$$\partial\|X\|_* = \{Y : \|Y\| \le 1,\ Y \bullet X = \|X\|_*\}.$$

Now let $\Lambda = U^T Y V$. From $\|Y\| \le 1$, we have $\|\Lambda\| \le 1$. Hence

$$Y \bullet X = \mathrm{tr}(Y^T X) = \mathrm{tr}(\Lambda^T \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}) = \sum_{i=1}^{r} \Lambda_{ii}\sigma_i \le \sum_{i=1}^{r} \sigma_i = \|X\|_*,$$

where the inequality takes the equality if and only if $\Lambda_{ii} = 1$, $i = 1, \ldots, r$. Also $\|\Lambda\| \le 1$ implies that $I - \Lambda \succeq 0$. This yields

$$\Lambda = \begin{pmatrix} I_r & 0 \\ 0 & W \end{pmatrix}.$$

Substitute $Y = U\Lambda V^T$ to the expression of $\partial\|X\|_*$, we have

$$\partial\|X\|_* = \left\{ U \begin{pmatrix} I_r & 0 \\ 0 & W \end{pmatrix} V^T : \|W\| \le 1 \right\}.$$

(c) Use the subdifferential $\partial\|X\|_*$ to compute the proximal mapping of $t\|X\|_*$.

**Solution.** Now consider the proximal problem

$$\min_Z \frac{1}{2t}\|Z - X\|_F^2 + \|Z\|_*.$$

First order optimality condition is $0 \in Z - X + t\partial\|Z\|_*$. By noting that $Z = U\begin{pmatrix} (\Sigma - tI_r)_+ & 0 \\ 0 & 0 \end{pmatrix} V^T$ satisfying this condition and the uniqueness of the, we obtain

$$\mathrm{prox}_{t\|\cdot\|_*}(X) = U\begin{pmatrix} (\Sigma - tI_r)_+ & 0 \\ 0 & 0 \end{pmatrix} V^T,$$

where $(A)_+ = \begin{pmatrix} \cdots & \cdots & \cdots \\ \vdots & \max\{A_{ij}, 0\} & \vdots \\ \cdots & \cdots & \cdots \end{pmatrix}$.