Solution 7-9

Assignment 7

Q1

(a) Let $f(x) = \frac{1}{2}\|x\|^2$ and $h(x) = Ax - b$. Define a Lagrangian function by

$$L(x,\mu) = \frac{1}{2}\|x\|^2 + \mu^T(Ax - b)$$

where $x = (x_1, \ldots, x_n)$ and μ^T is a $1 \times m$ row vector. Then, we compute

$$\nabla L(x,\mu) = x + A^T\mu$$

Then, we have $\nabla L(x,\mu) = \mathbf{0} \iff A^T\mu = -x \iff x = -A^T\mu$. Since $\nabla_x^2 L(x,\mu) = 1 > 0$, so we have $\inf_{x\in\mathbb{R}^n : Ax=b} L(x,\mu) = L(-A^T\mu,\mu)$. Putting $x = -A^T\mu$ to $L(x,\mu)$, we have

$$\begin{aligned} L(-A^T\mu,\mu) &= \frac{1}{2}\|-A^T\mu\|^2 + \mu^T(A(-A^T\mu) - b) \\ &= \frac{1}{2}(A^T\mu)^T(A^T\mu) - \mu^T AA^T\mu - \mu^T b \\ &= -\frac{1}{2}\mu^T AA^T\mu - \mu^T b \end{aligned}$$

Thus, the dual problem (D) is

$$\max_{\mu\in\mathbb{R}^n}\left(-\frac{1}{2}\mu^T AA^T\mu - \mu^T b\right)$$

(b) Hence, we define $d(\mu) = -\frac{1}{2}\mu^T AA^T\mu - \mu^T b$. Then, we have

$$\nabla_\mu d(\mu) = -AA^T\mu - b$$

Then, we have $\nabla_\mu d(\mu) = \mathbf{0} \iff -AA^T\mu - b = \mathbf{0} \iff AA^T\mu = -b$. Since $\nabla_\mu^2 d(\mu) = -AA^T \prec 0$ which is negative definite, and it is clear that the dual problem has a unique solution if and only if AA^T is invertible so that $\mu = -(AA^T)^{-1}b$, or equivalently, A has rank m.

(c) In this case, from (b), this dual solution is given by $\mu = -(AA^T)^{-1}b$. Thus, the minimizer to the problem is

$$x^* = -A^T\left(-(AA^T)^{-1}b\right) = A^T(AA^T)^{-1}b$$

Q2

See the reference book *G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Spriner 2020.*

Q3

Define

$$\begin{aligned} f(x,y,z) &= \frac{1}{2}[(x-2)^2+y^2+z^2], \\ g_1(x,y,z) &= x^2+y^2-1, \\ g_2(x,y,z) &= y+z. \end{aligned}$$

Clearly, f, g_1, and g_2 are convex functions. Consider the feasible set K as

$$K := \{(x,y,z) : x^2+y^2 \le 1,\ y+z \le 0\}.$$

As $\lim_{x^2+y^2+z^2 \to +\infty} f(x,y,z) = +\infty$, so f is a coercive function, so there exists a minimizer $(x^*,y^*,z^*) \in K$ as a solution to the problem. Define the Lagrangian function as

$$L(x,y,z,\lambda_1,\lambda_2) = \frac{1}{2}[(x-2)^2+y^2+z^2] + \lambda_1(x^2+y^2-1) + \lambda_2(y+z)$$

for $\lambda_1, \lambda_2 \ge 0$. Now, we compute

$$\nabla_{(x,y,z)} L = \begin{pmatrix} x-2+2\lambda_1 x \\ y+2\lambda_1 y+\lambda_2 \\ z+\lambda_2 \end{pmatrix}.$$

By Euler's first-order condition, setting $\nabla_{(x,y,z)} L = \mathbf{0}$, we have

$$\begin{cases} x = \dfrac{2}{1+2\lambda_1}, \\ y = -\dfrac{\lambda_2}{1+2\lambda_1}, \\ z = -\lambda_2. \end{cases}$$

So, we have

$$\begin{aligned} &\min_{(x,y,z)} L(x,y,z,\lambda_1,\lambda_2) \\ &= L\left(\frac{2}{1+2\lambda_1}, -\frac{\lambda_2}{1+2\lambda_1}, -\lambda_2, \lambda_1, \lambda_2\right) \\ &= \frac{1}{2}\left[\left(\frac{-4\lambda_1}{1+2\lambda_1}\right)^2 + \left(-\frac{\lambda_2}{1+2\lambda_1}\right)^2 + \lambda_2^2\right] + \lambda_1\left[\left(\frac{2}{1+2\lambda_1}\right)^2 + \left(-\frac{\lambda_2}{1+2\lambda_1}\right)^2 - 1\right] + \lambda_2\left[-\frac{\lambda_2}{1+2\lambda_1} - \lambda_2\right] \\ &= \frac{-2\lambda_1^2 - 3\lambda_1 + \lambda_1\lambda_2^2 + \lambda_2^2}{1+2\lambda_1}. \end{aligned}$$

Now, we define $d(\lambda_1, \lambda_2) = \frac{-2\lambda_1^2 - 3\lambda_1 + \lambda_1\lambda_2^2 + \lambda_2^2}{1+2\lambda_1}$ for $\lambda_1, \lambda_2 \geq 0$. Then, we compute

$$\nabla d(\lambda_1, \lambda_2) = \begin{pmatrix} -\frac{4\lambda_1^2 + 4\lambda_1 - \lambda_2^2 - 3}{(1+2\lambda_1)^2} \\ -\frac{2(\lambda_1+1)\lambda_2}{1+2\lambda_1} \end{pmatrix}.$$

By Euler's first-order condition, we set $\nabla d(\lambda_1, \lambda_2) = \mathbf{0}$. We have

$$\begin{cases} 2(1+\lambda_1)\lambda_2 = 0, \\ 4\lambda_1^2 + 4\lambda_1 - \lambda_2^2 - 3 = 0. \end{cases}$$

Then, we have $\lambda_1 = -1$ or $\lambda_2 = 0$.
- **Case 1: $\lambda_1 = -1$**
Then putting in the second equation, we have

$$4(-1)^2 + 4(-1) - \lambda_2^2 - 3 = 0 \implies \lambda_2^2 = -3.$$

This is rejected since $\lambda_2 \geq 0$.
- **Case 2: $\lambda_2 = 0$**
Putting into the second equation, we have

$$4\lambda_1^2 + 4\lambda_1 - 3 = 0 \implies (2\lambda_1 - 1)(2\lambda_1 + 3) = 0 \implies \lambda_1 = \frac{1}{2} \quad \text{or} \quad \lambda_1 = -\frac{3}{2} \quad \text{(rejected)}.$$

Therefore, we have $(\lambda_1^*, \lambda_2^*) = \left(\frac{1}{2}, 0\right)$ to be the optimal solution to the dual problem. Thus, the optimal solution to the minimization problem is $(x^*, y^*, z^*) = (1, 0, 0)$ and

$$\min_{x^2+y^2\leq 1,\, y+z\leq 0} \frac{1}{2}\left[(x-2)^2 + y^2 + z^2\right] = \frac{1}{2}.$$

Assignment 8

Q1

(a) The feasible set is

$$\begin{aligned} K &= \left\{(x_1, x_2) : (x_1-1)^2 + (x_2-1)^2 \leq 1,\ (x_1-1)^2 + (x_2+1)^2 \leq 1\right\} \\ &= \left\{(x_1, x_2) : (x_1-1)^2 + (x_2-1)^2 \leq 1\right\} \cap \left\{(x_1, x_2) : (x_1-1)^2 + (x_2+1)^2 \leq 1\right\} \\ &= \{(1, 0)\}. \end{aligned}$$

and the optimal solution $x^* \in K = \{(1,0)\}$, so $x^* = (1, 0)$.
(b) For $(x_1^*, x_2^*) \in K$, and $\lambda_1, \lambda_2 \geq 0$, the KKT condition is

$$\begin{cases} \begin{pmatrix} 2x_1^* \\ 2x_2^* \end{pmatrix} + \lambda_1 \begin{pmatrix} 2(x_1^*-1) \\ 2(x_2^*-1) \end{pmatrix} + \lambda_2 \begin{pmatrix} 2(x_1^*-1) \\ 2(x_2^*+1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \lambda_1((x_1^*-1)^2 + (x_2^*-1)^2 - 1) = 0, \\ \lambda_2((x_1^*-1)^2 + (x_2^*+1)^2 - 1) = 0. \end{cases}$$

Suppose there exists such λ_1^*, λ_2^*. From (a), since $x^* = (x_1^*, x_2^*) = (1, 0)$ is an optimal solution. Plugging into the first equation, it gives

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_1^* \begin{pmatrix} 2(1-1) \\ 2(0-1) \end{pmatrix} + \lambda_2^* \begin{pmatrix} 2(1-1) \\ 2(0+1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

which gives a contradiction! So, there does not exist such λ_1^*, λ_2^* such that x^*, $(\lambda_1^*, \lambda_2^*)$ satisfy the above KKT conditions.

Q2

Since $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$ satisfy the KKT conditions, then we have

$$\begin{aligned} \langle \nabla f(x^*), x - x^* \rangle &= \left\langle -\sum_{i=1}^m \lambda_i^* g_i(x^*), x - x^* \right\rangle \\ &= -\sum_{i=1}^m \lambda_i^* \langle g_i(x^*), x - x^* \rangle \\ &= -\sum_{i=1}^m \lambda_i^* \langle \nabla g_i(x^*), x - x^* \rangle. \end{aligned}$$

As g_i are convex, so for any $t \in (0,1)$, we have

$$g_i(tx + (1-t)x^*) \leq t g_i(x) + (1-t) g_i(x^*).$$

And by Taylor's expansion, we have

$$\begin{aligned} g_i(x) &= g_i(x^*) + t\langle \nabla g_i(x^*), x - x^* \rangle + \frac{t^2}{2}(x - x^*)^T \operatorname{Hess}(g_i)(x - x^*) \\ &\geq g_i(x^*) + t\langle \nabla g_i(x^*), x - x^* \rangle \qquad (\because \operatorname{Hess}(g_i) \geq 0 \text{ by convexity}) \end{aligned}$$

$$\langle \nabla g_i(x^*), x - x^* \rangle \leq \frac{1}{t}\big(g_i(x) - g_i(x^*)\big).$$

Thus, we have

$$\begin{aligned} \langle \nabla f(x^*), x - x^* \rangle &= -\sum_{i=1}^m \lambda_i^* \langle \nabla g_i(x^*), x - x^* \rangle \\ &\geq \frac{1}{t} \sum_{i=1}^m \lambda_i^* \big(g_i(x^*) - g_i(x)\big) \\ &= \frac{-1}{t} \sum_{i=1}^m \underbrace{\lambda_i^*}_{\geq 0} \underbrace{g_i(x)}_{\leq 0} \\ &\geq 0 \end{aligned}$$

for all feasible x.

Q3

(a)

$$f^*(d) = \begin{cases} -N - \sum_{i=1}^N \log(-d_i) & \text{if } d \in \mathbb{R}^N_-, \\ +\infty & \text{otherwise.} \end{cases}$$

(b) $f^*(d) = \frac{1}{4}(d-b)^T A^{-1}(d-b) - c$

(c) $f^*(d) = \begin{cases} 0 & \text{if } \|d\| \le 1, \\ +\infty & \text{otherwise.} \end{cases}$

Assignment 9

Q1

(i) Define $f(x) = \frac{1}{2}x^T Qx + c^T x$ and the feasible region

$$K := \left\{ x \in \mathbb{R}^N : l \le x_k \le u, \, \forall k = 1, \dots, N \right\}.$$

Now, we check the following conditions:

- K **is convex and closed:**

For any $x, y \in K$ and $t \in (0,1)$, then

$$\begin{aligned} tx + (1-t)y &= t \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} + (1-t) \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \\ &= \begin{pmatrix} tx_1 + (1-t)y_1 \\ \vdots \\ tx_N + (1-t)y_N \end{pmatrix}. \end{aligned}$$

Since for each $k = 1, \dots, N$, we have $\ell \le x_k, y_k \le u$, and so we have

$$\ell \le tx_k + (1-t)y_k \le u,$$

this implies that $tx + (1-t)y \in K$, so K is convex. The closeness of K is due to $K = \bar{K}$.

- $\text{Hess}(f) \succeq \alpha I_N$ **for some** $\alpha > 0$**:**

Note that $\nabla f(x) = Qx + c$ and $\text{Hess}(f) = Q$, which is symmetric positive-definite. So all its eigenvalues are strictly positive, ordering its eigenvalues as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N > 0$. Now, for any $v \ne 0$, we have

$$0 < \lambda_N \le \frac{v^T Qv}{v^T v} \le \lambda_1.$$

Choose $\alpha = \frac{1}{2}\lambda_N > 0$ such that for any $v \ne 0$, we have

$$\begin{aligned} v^T \left(\text{Hess}(f) - \alpha I_N \right) v &= v^T \left(Q - \alpha I_N \right) v \\ &= v^T Qv - \alpha v^T v \\ &\ge \frac{1}{2}\lambda_N v^T v > 0. \end{aligned}$$

So, $\text{Hess}(f) - \frac{1}{2}\lambda_N I_N$ is positive definite.

- ∇f is M-Lipschitz:

For any $x, y \in K$, note that

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\| &= \|Q(x-y)\| \\ &\leq \|Q\|\|x-y\|.\end{aligned}$$

By choosing $M = \|Q\|$, then the condition holds.

- From the above, we require that

$$\rho \in \left(0, \frac{2\alpha}{M^2}\right) = \left(0, \frac{\lambda_N}{\|Q\|^2}\right).$$

This problem satisfies the condition of the Projected Gradient Algorithm.

Q2

Define $f(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}x^T x$ and $K := \{x : Ax = b\}$. Now, it remains to check the following conditions:

- f is strongly convex and coercive:

Note that f is clearly coercive since

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

To show that f is strongly convex, we compute

$$\nabla f(x) = x, \quad \text{and} \quad \text{Hess}(f) = I_N,$$

which is positive definite, so f is strongly convex.

- Linear constraint:

It is clear that $Ax = b$ is a linear constraint.

Thus, the problem satisfies the condition of the Uzawa Algorithm.