Homework 4

1. (An Inequality from KKT)

   Consider the inequality constrained convex optimization problem

$$\min_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad f_i(x) \le 0, \ \forall i \in [m].$$

and suppose that $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ satisfy the KKT conditions

$$\begin{aligned} f_i(x^*) &\le 0 \qquad \forall i \in [m] \\ \lambda_i^* &\ge 0 \qquad \forall i \in [m] \\ \lambda_i^* f_i(x^*) &= 0 \qquad \forall i \in [m] \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) &= 0. \end{aligned}$$

Show that

$$\nabla f_0(x^*)^T (x - x^*) \ge 0$$

for all feasible x.

Hint: Use the fact that $f_i(x^*) + \nabla f_i(x^*)^T(x - x^*) \le f_i(x) \le 0$ for all $i \in [m]$ and for all feasible x by the convexity of $\{f_i\}_{i=1}^m$. Then multiply each of these inequalities by λ_i^* and sum them over $i \in [m]$.

Since $\{f_i\}_{i=1}^m$ is convex and differentiable, we have the following inequality:

$$f_i(y) \ge f_i(x) + \nabla f_i(x)^T (y - x)$$

Let $y = x$ and $x = x^*$

$\Rightarrow$ we get $f_i(x) \ge f_i(x^*) + \nabla f_i(x^*)(x - x^*)$

By the inequality constraint $f_i(x) \le 0$,

we establish $f_i(x^*) + \nabla f_i(x^*)\{x - x^*\} \le f_i(x) \le 0$

consider $f_i(x^*) + \nabla f_i(x^*)\{x - x^*\} \le 0$

Multiply both sides by λ_i^*

$$\underbrace{\lambda_i^* f_i(x^*)}_{\substack{\shortparallel \\ 0}} + \lambda_i^* \nabla f_i(x^*)\{x - x^*\} \le 0$$

(by complementary slackness)

$$\lambda_i^* \nabla f_i(x^*)\{x - x^*\} \leq 0$$

sum over $i \in [m]$

$$\sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*)^T \{x - x^*\} \leq 0$$

stationarity of KKT states that

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0$$

$$\Rightarrow \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = -\nabla f_0(x^*)$$

Hence we get

$$-\nabla f_0(x^*)^T \{x - x^*\} \leq 0$$

$$\Rightarrow \nabla f_0(x^*)^T \{x - x^*\} \geq 0$$

(shown)

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. Consider the following problem:

$$(\text{P}) \qquad \min_x \ f(x) \quad \text{s.t.} \quad x \succeq 0. \qquad -x \leq 0$$

Let $v(x), w(x) : \mathbb{R}^n \to \mathbb{R}^n$ be two vector-valued functions of x. Show using KKT conditions that $\bar{x} \in \mathbb{R}^n$ is an optimal solution to problem (P) *if and only if* $\bar{x}$ is a solution to the following system:

$$\nabla f(x) \succeq 0 \quad ✓$$
$$x \succeq 0 \quad ✓$$
$$w(x)^T v(x) = 0$$

In the process, find the functions $v(x)$ and $w(x)$.

Advice: Be sure to justify the "if and only if".

With KKT conditions,

$\bar{x} \in \mathbb{R}$ is an optimal solution

$\Leftrightarrow$ KKT conditions are satisfied.

① Stationarity:

$L(x, \lambda) = f(x) + \lambda(-x)$

$\nabla f(x) - \lambda = 0$

② Complementary slackness: $\lambda(-x) = 0$

③ Primal feasibility: $-x \leq 0 \Leftrightarrow x \geq 0$

④ Dual feasibility: $\lambda \geq 0$

From ①, $\nabla f(x) = \lambda$

$\lambda \geq 0 \Leftrightarrow \nabla f(x) \geq 0$

From ②, $\lambda(-x) = 0 \Leftrightarrow \nabla f(x)^T(-x) = 0$

where $w(x)^T = \nabla f(x)$

and $v(x) = -x$

hence the 3 inequalities are satisfied.

$\bar{x}$ optimum to (P) $\Leftrightarrow$ $\bar{x}$ satisfies 3 inequalities.

since it will then satisfy KKT conditions, which is necessary and sufficient for $\bar{x}$ to be optimum to (P)

$\Rightarrow$ proving the iff.

3. (a) (15 points) Solve for the optimal x using the KKT conditions:

$$\max_{x_1,x_2} \; 14x_1 - x_1^2 + 6x_2 - x_2^2 + 7 \quad \text{s.t.} \quad x_1 + x_2 \le 2 \quad x_1 + 2x_2 \le 3$$

(b) (10 points) Consider the following problem

$$\max_{x} \; x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \quad \text{s.t.} \quad \sum_{i=1}^{n} x_i = 1 \quad x_i \ge 0, \quad i = 1, \dots, n$$

where a_i are given positive scalars. Find a global maximum and show that it is unique.

(a) $\max_{x_1, x_2} 14x_1 - x_1^2 + 6x_2 - x_2^2 + 7$ s.t $x_1 + x_2 \le 2$

and $x_1 + 2x_2 \le 3$

$\Rightarrow \min_{x_1, x_2} x_1^2 + x_2^2 - 14x_1 - 6x_2 - 7$ s.t $x_1 + x_2 \le 2$

and $x_1 + 2x_2 \le 3$

$L(x_1, x_2, \lambda_1, \lambda_2)$

$= x_1^2 + x_2^2 - 14x_1 - 6x_2 - 7 + \lambda_1(x_1 + x_2 - 2) + \lambda_2(x_1 + 2x_2 - 3)$

stationarity:

$\frac{\partial}{\partial x_1} L(x_1, x_2, \lambda_1, \lambda_2) = 2x_1 - 14 + \lambda_1 + \lambda_2$

$= 0$

$\frac{\partial}{\partial x_2} L(x_1, x_2, \lambda_1, \lambda_2) = 2x_2 - 6 + \lambda_1 + 2\lambda_2$

$= 0$

complementary slackness:

$\lambda_1(x_1 + x_2 - 2) = 0$

$\lambda_2(x_1 + 2x_2 - 3) = 0$

Primal feasibility: $x_1 + x_2 \le 2$ and $x_1 + 2x_2 \le 3$

Dual feasibility: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$

① $\lambda_1 = \lambda_2 = 0$

$$\left.\begin{aligned} 2x_1 - 14 = 0 &\Rightarrow x_1 = 7 \\ 2x_2 - 6 = 0 &\Rightarrow x_2 = 3 \end{aligned}\right\} \text{violates PF}$$

② $\lambda_1 = 0$, $x_1 + 2x_2 - 3 = 0$

$x_1 + 2x_2 = 3$

$x_1 = 3 - 2x_2$

$2(3 - 2x_2) - 14 + \lambda_2 = 0$

$-4x_2 + \lambda_2 = 8$ — ①

$2x_2 - 6 + 2\lambda_2 = 0$

$2x_2 + 2\lambda_2 = 6$ — ②

$\underbrace{x_2 = -1,\ x_1 = 5}_{\text{violates PF}}$, $\lambda_2 = 4$

③ $\lambda_2 = 0$, $x_1 + x_2 - 2 = 0$

$x_1 + x_2 = 2$

$x_2 = 2 - x_1$

$2x_1 - 14 + \lambda_1 = 0$

$2x_1 + \lambda_1 = 14$ — ①

$2(2-x_1)-6+\lambda_1=0$

$-2x_1+\lambda_1=2$ — ②

$x_1=3$, $x_2=-1$, $\lambda_1=8$

④ $x_1+2x_2-3=0$

$x_1+x_2-2=0$

$x_1+2x_2=3$ — ①

$x_1+x_2=2$ — ②

$x_1=1$, $x_2=1$

$2(2)-14+\lambda_1+\lambda_2=0$

$\lambda_1+\lambda_2=10$ — ①

$2(1)-6+\lambda_1+2\lambda_2=0$

$\lambda_1+2\lambda_2=4$ — ②

$\lambda_1=16$, $\underbrace{\lambda_2=-4}_{\text{violates DF.}}$

The optimal x is $(x_1^*, x_2^*)=(3,-1)$

optimal solution is $14(3)-3^2+6(-1)-(-1)^2+7$

$=33$

(b) (b) (10 points) Consider the following problem $\sum_{i=1}^{n} x_i - 1 = 0$

$$\max_x \; x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n} \quad \text{s.t.} \quad \sum_{i=1}^{n} x_i = 1 \quad x_i \geq 0, \quad i = 1, \ldots, n$$

where a_i are given positive scalars. Find a global maximum and show that it is unique.

This is equivalent to

$$\max_x \underbrace{\ln (x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n})}_{f(x)}$$

$x_i \geq 0$ need not be considered as the inequality cannot be active at maximum.

$$\Rightarrow f(x) = \sum_{i=1}^{n} a_i \ln(x_i)$$

$$L(x,v) = \sum_{i=1}^{n} a_i \ln(x_i) - v\left(\sum_{i=1}^{n} x_i - 1\right)$$

stationary point: $\frac{\partial}{\partial x_i} L(x,v) = \frac{a_i}{x_i} - v$

$= 0$

$$a_i = v x_i \Rightarrow \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} v x_i$$

Given $\sum_{i=1}^{n} x_i = 1$, $\sum_{i=1}^{n} a_i = v$

Global max occurs when $x_i = \dfrac{a_i}{\sum_{i=1}^{n} a_i}$ $\forall i$

$$f(x) = \sum_{i=1}^{n} a_i \ln\left(\frac{a_i}{\sum_{i=1}^{n} a_i}\right)$$

$$\max_x f(x) \text{ s.t } \sum_{i=1}^{n} x_i = 1, \; x_i \geq 0, \; i = 1, \ldots n$$

since the new problem is concave, the lagrangian condition is also suffficent. Obj function f(x) is strictly concave hence maximum is unique.

4. (a) (20 points) Consider the gradient descent method with bounded error,

$$x^{(k+1)} = x^{(k)} - s(\nabla f(x^{(k)}) + \epsilon^{(k)})$$

where s is a constant stepsize, $\epsilon^{(k)}$ are error terms satisfying

$$\|\epsilon^{(k)}\| \leq \delta$$

for all k, and f is the positive definite quadratic function

$$f(x) = \frac{1}{2}(x - x^*)^T Q(x - x^*)$$

Let

$$q := \max\{|1 - s\lambda_{\min}(Q)|, |1 - s\lambda_{\max}(Q)|\}$$

and assume that $q < 1$. Show that for all k, we have

$$\|x^{(k)} - x^*\| \leq \frac{s\delta}{1-q} + q^k \|x^{(0)} - x^*\|$$

Consider lemma A:

Given a value $0 \leq c \leq 1$ and that

$$\|x^{(k+1)} + x^*\| \leq s\delta + c\|x^{(k)} - x^*\|$$

we can find a bound for $\|x^{(k)} - x^*\|$ to be

$$\|x^{(k)} - x^*\| \leq \frac{s\delta}{1-c} + c^k\|x^{(0)} - x^*\|$$

Proof:

$$\|x^{(k)} - x^*\| \leq s\delta + c\|x^{(k-1)} - x^*\|$$

$$\leq (1+c)s\delta + c^2\|x^{(k-2)} - x^*\|$$

$$\vdots$$

$$\leq \left(\sum_{i=0}^{k} c^i\right) s\delta + c^k\|x^{(0)} - x^*\|$$

$$\leq \left(\sum_{i=0}^{\infty} c^i\right) s\delta + c^k\|x^{(0)} - x^*\|$$

$$= \frac{s\delta}{1-c} + c^k\|x^{(0)} - x^*\|$$

consider a 2nd lemma, Lemma B

For PSD matrix A and $s > 0$, the following holds

$$\|I - sA\| \leq \max\{|1 - s\lambda_{min}(A)|, |1 - s\lambda_{max}(A)|\}$$

Proof

$A = U^T \Lambda U \in \mathbb{R}^{n \times n}$ where U is unitary matrix of eigenvalues and Λ is diagonal matrix with +ve eigenvalues. Note $\|B\| = \sqrt{\lambda_{max}(B^T B)}$

Hence
$$
\begin{aligned}
\|I - sA\| &= \|U^T U - s U^T \Lambda U\| \\
&= \|U^T (I - s\Lambda) U\| \\
&\leq U^T \|I - s\Lambda\| \|U\| \\
&= \|I - s\Lambda\| \\
&= \sqrt{\max\{(1 - s\lambda_1)^2, \dots (1 - s\lambda_n)^2\}} \\
&= \max\{|1 - s\lambda_1|, |1 - s\lambda_2|, \dots |1 - s\lambda_n|\} \\
&= \max\{|1 - s\lambda_{min}(A)|, |1 - s\lambda_{max}(A)|\}
\end{aligned}
$$

Hence we can proof the question using the above lemmas

$$\|x^{(k+1)} - x^*\| = \|(x^{(k)} - x^*) - s(\nabla f(x^{(k)}) + \varepsilon_k)\|$$

$$= \|(x^{(k)} - x^*) - s(Q(x_k - x^*) + \varepsilon_k)\|$$

$$= \|(I - sQ)(x^{(k)} - x^*) - s\varepsilon_k\|$$

$$\leq \|(I - sQ)(x^{(k)} - x^*)\| + s\|\varepsilon_k\|$$

$$\leq \|I - sQ\| \, \|x^{(k)} - x^*\| + s\delta$$

$$\leq \max\{|1 - s\lambda_{max}(Q)|, |1 - s\lambda_{min}(Q)|\}.$$
$$\|x^{(k)} - x^*\| + s\delta$$

$$= q\|x^{(k)} - x^*\| + s\delta$$

since we assume s to be small enough s.t $q < 1$

using lemma A we show:

$$\|x^{(k+1)} - x^*\| \leq q\|x^{(k)} - x^*\| + s\delta$$
$$\leq \frac{s\delta}{1-q} + q^{k+1}\|x^{(0)} - x^*\| \quad \text{(shown)}$$

(b) (5 points) Consider the function f in part (a) of this problem. Suppose we now use Newton's method to find the minimum. Does Newton's method converge? If so how many iterations (Newton steps) are required for convergence to within $\varepsilon = 10^{-10}$ of x^*, i.e., find the minimum k such that $\|x^{(k)} - x^*\| \leq \varepsilon$?

f is the positive definite quadratic function

$$f(x) = \frac{1}{2}(x-x^*)^T Q (x-x^*)$$

Newton's method should converge because:

(a) $f(x)$ is PD

(b) Q is also PD

$\nabla f(x) = Q(x-x^*) \Rightarrow \nabla^2 f(x) = Q$ where Q is constant and PD

Hence the hessian is non-singular at all points x^k in sequence of iterates

First iteration $\Rightarrow x^{(1)} = x^{(0)} - Q^{-1} Q \{x^{(0)} - x^*\}$

$= x^*$ (which is minimum)

Hence for $\|x^{(k)} - x^*\| \leq q$, min $k = 1$