The Answer to Homework

1. Convex set

(1) Show that a polyhedron $\{x \in R^n : Ax \le b\}$ for some $A \in R^{m\times n}$, $b \in R^m$, is convex.

Proof: Suppose $x_1, x_2 \in \{x \in R^n : Ax \le b\}$ and $\theta \in [0,1]$.

Since $A(\theta x_1 + (1-\theta)x_2) = \theta Ax_1 + (1-\theta)Ax_2 \le \theta b + (1-\theta)b = b$,

$\theta x_1 + (1-\theta)x_2 \in \{x \in R^n : Ax \le b\}$

Thus a polyhedron $\{x \in R^n : Ax \le b\}$ for some $A \in R^{m\times n}$, $b \in R^m$, is convex.

(2) Consider a convex function $f : R^n \to R$. Prove that the set

$$\{(x,t) \mid f(x) \le t, x \in R^n, t \in R\}$$

is convex.

Proof: Suppose $(x_1,t_1),(x_2,t_2) \in \{(x,t) \mid f(x) \le t, x \in R^n, t \in R\}$ and $\theta \in [0,1]$.

Since $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2) \le \theta t_1 + (1-\theta)t_2$,

$[\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2] \in \{(x,t) \mid f(x) \le t, x \in R^n, t \in R\}$.

Thus $\{(x,t) \mid f(x) \le t, x \in R^n, t \in R\}$ is convex.

(3) Show that $\{x \in R_+^n \mid \prod_{i=1}^n x_i \ge 1\}$ is convex. (Hint: If $a,b \ge 0$ and $0 \le \theta \le 1$, than $a^\theta b^{(1-\theta)} \le \theta a + (1-\theta)b$)

Proof: Suppose $p,q \in \{x \in R_+^n \mid \prod_{i=1}^n x_i \ge 1\}$ and $\theta \in [0,1]$.

Since $\prod_{i=1}^n [\theta p_i + (1-\theta)q_i] \ge \prod_{i=1}^n p_i^\theta q_i^{1-\theta} = (\prod_{i=1}^n p_i)^\theta (\prod_{i=1}^n q_i)^{1-\theta} \ge 1$.

$\theta p + (1-\theta)q \in \{x \in R_+^n \mid \prod_{i=1}^n x_i \ge 1\}$.

Thus $\{x \in R_+^n \mid \prod_{i=1}^n x_i \ge 1\}$ is convex.

(4) Show that the set $\{x \mid \|x-a\|_2 \le \theta \|x-b\|_2\}$,where $a \ne b$ and $0 \le \theta \le 1$, is convex.

Proof: $\{x \mid \|x-a\|_2 \le \theta \|x-b\|_2\} = \{x \mid \|x-a\|_2^2 \le \theta \|x-b\|_2^2\}$

$= \{x \mid (x-a)^T (x-a) \le \theta (x-b)^T (x-b)\}$

$= \{x \mid (1-\theta) \mathrm{x}^T \mathrm{x} + (2\theta \mathrm{b}^T - 2\mathrm{a}^T) \mathrm{x} + \mathrm{a}^T \mathrm{a} - \theta \mathrm{b}^T \mathrm{b} \le 0\}$

Since $0 \le \theta \le 1$, the Hessian matrix of function

$$f(x) = (1-\theta)\ x^T x + (2\theta b^T - 2a^T) x + a^T a - \theta b^T b$$

is $(1-\theta)\ I$ is a positive semidefinite matrix. Thus, the function $f(x)$ is a convex function, the 0-sublevel set is convex.

2. Convex function

(1) Prove that that the entropy function, defined as

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i)$$

with $dom(f) = \{\mathrm{x} \in \mathrm{R}_{++}^n : \sum_{i=1}^n x_i = 1\}$, is strictly concave.

Proof:

$\nabla f = \left[-(\log x_1 + 1) \quad -(\log x_2 + 1) \quad \cdots \quad -(\log x_n + 1) \right]$

$$\nabla^2 f = \begin{bmatrix} -\frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & -\frac{1}{x_2} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & -\frac{1}{x_n} \end{bmatrix}$$

Which is a negative matrix since $\mathrm{x} \in \mathrm{R}_{++}^n$.

Moreover, the set $dom(f) = \{\mathrm{x} \in \mathrm{R}_{++}^n : \sum_{i=1}^n x_i = 1\}$ is a convex set. Thus, the entropy function, defined as

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i)$$

with $dom(f) = \{\mathrm{x} \in \mathrm{R}_{++}^n : \sum_{i=1}^n x_i = 1\}$, is strictly concave.

(2) Show that $f(x_1,x_2)=\frac{1}{x_1x_2}$ on R^2_{++} is convex.

Proof: $\nabla f=\begin{bmatrix}-\frac{1}{x_1^2x_2} & -\frac{1}{x_1x_2^2}\end{bmatrix}$

$\nabla^2 f=\begin{bmatrix}\frac{2}{x_1^3x_2} & \frac{1}{x_1^2x_2^2}\\ \frac{1}{x_1^2x_2^2} & \frac{2}{x_1x_2^3}\end{bmatrix}$, since $x\in R^2_{++}$, matrix $\nabla^2 f$ is a positive semidefinite matrix.

The domain R^2_{++} is a convex set. Therefore, the function $f(x_1,x_2)=\frac{1}{x_1x_2}$ on R^2_{++} is convex.

(3) Show that $f(X)=tr(X^{-1})$ is convex on $\mathrm{dom}\, f=S^n_{++}$

Proof:

Define $g(t)=f(Z+tV)$ where $Z\in S^n_{++}, V\in S^n$

$$g(t)=tr[(Z+tV)^{-1}]=tr[Z^{-1}(I+tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})^{-1}]$$
$$=\mathrm{tr}[Z^{-1}Q(I+t\Lambda)^{-1}Q^T]$$

where $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}}=Q^T\Lambda Q$, $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots\lambda_n)$

Thus $g(t)=\mathrm{tr}[Z^{-1}Q(I+t\Lambda)^{-1}Q^T]=\mathrm{tr}[Q^TZ^{-1}Q(I+t\Lambda)^{-1}]=\sum_{i=1}^{n}(Q^TZ^{-1}Q)_{ii}(1+t\lambda_i)^{-1}$

Function $g(t)$ is a positive weighted sum of convex function $\frac{1}{1+t\lambda_i}$. Therefor $g(t)$ is convex. The set $\mathrm{dom}\, f=S^n_{++}$ is a convex set. Thus function $f(X)=tr(X^{-1})$ is convex on $\mathrm{dom}\, f=S^n_{++}$.

3. Dual problem

(1) Formulate the dual problems of the following problems with one inequality constraint

$$\begin{aligned}&\min \quad c^Tx\\ &s.t. \quad f(x)\le 0\end{aligned}$$

Answer: For $\lambda = 0$, $g(\lambda) = \inf_x c^T x = -\infty$

For $\lambda > 0$, $g(\lambda) = \inf_x [c^T x + \lambda f(x)] = \lambda \inf_x [(\frac{c}{\lambda})^T x + f(x)]$

$= -\lambda \sup_x [-(\frac{c}{\lambda})^T x - f(x)] = -\lambda f^*(\frac{c}{\lambda})$

Thus, the dual problem is

$$\begin{aligned} &\max \quad -\lambda f^*(\frac{c}{\lambda}) \\ &\text{s.t.} \quad \lambda > 0 \end{aligned}$$

(2) Find the dual problem of the following general Linear programming

$$\begin{aligned} &\min \quad c^T x \\ &s.t. \quad Gx \le h \\ &\qquad\quad Ax = b \end{aligned}$$

Answer:

The Lagrangian function is

$$\begin{aligned} l(x,\lambda,v) &= \mathrm{c}^T x + \lambda^T (Gx - h) + v^T (Ax - b) \\ &= (c + G^T \lambda + A^T v)^T x - \lambda^T h + v^T b \end{aligned}$$

The dual function is

$$g(\lambda, v) = \inf_x L(x, \lambda, v) = \begin{cases} -\lambda^T h - v^T b, & c + G^T \lambda + A^T v = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Thus the dual problem is

$$\begin{aligned} &\max_{\lambda,\ v} \quad -h^T \lambda - b^T v \\ &s.t. \quad \lambda \ge 0 \\ &\qquad\quad c + G^T \lambda + A^T v = 0 \end{aligned}$$

4. KKT condition

(1) Give the KKT conditions of the following optimization problem

$$\begin{aligned} &\min \quad x_1^2 + x_2^2 \\ &\text{s.t.} \quad (x_1 - 1)^2 + (x_2 - 1)^2 \le 1 \\ &\qquad\quad (x_1 - 1)^2 + (x_2 + 1)^2 \le 1 \end{aligned}$$

Answer:

The KKT conditions are

$$(x_1-1)^2+(x_2-1)^2\le 1$$
$$(x_1-1)^2+(x_2+1)^2\le 1$$
$$\lambda_1\ge 0,\quad \lambda_2\ge 0$$
$$\lambda_1[(x_1-1)^2+(x_2-1)^2-1]=0$$
$$\lambda_2[(x_1-1)^2+(x_2+1)^2-1]=0$$
$$x_1+\lambda_1(x_1-1)+\lambda_2(x_1-1)=0$$
$$x_2+\lambda_2(x_2-1)+\lambda_2(x_2+1)=0$$

(2) Consider the equality constrained least square problem

$$\begin{aligned}&\min \quad \|Ax-b\|_2^2\\ &s.t.\quad Gx=h\end{aligned}$$

Where $A\in R^{m\times n}$ with $\mathrm{rank}A=n$ and $G\in R^{p\times n}$ with $\mathrm{rank}G=p$

Give the KKT conditions and derive expressions for the primal solution x^* and the dual solution v^*.

Answer:

The KKT conditions is

$$Gx=h$$
$$2A^TAx=2A^Tb-G^Tv$$

We can solve the linear equation to obtain the optimal solution

First, since $\mathrm{rank}A=n$, A^TA is inversible, we may obtain

$$x=\frac{1}{2}(A^TA)^{-1}(2A^Tb-G^Tv) \qquad (*)$$

Putting it into the first equation, we have

$$G[\frac{1}{2}(A^TA)^{-1}(2A^Tb-G^Tv)]=h$$
$$\frac{1}{2}G(A^TA)^{-1}G^Tv=G(A^TA)^{-1}A^Tb-h$$

Thus the optimal solution of dual problem $v^*=2[G(A^TA)^{-1}G^T]^{-1}[G(A^TA)^{-1}A^Tb-h]$

Putting it into (*), we obtain the optimal solution for the primal solution

$$x^*=(A^TA)^{-1}\{A^Tb-G^T[G(A^TA)^{-1}G^T]^{-1}[G(A^TA)^{-1}A^Tb-h]\}$$

5. Gradient and Newton Descent

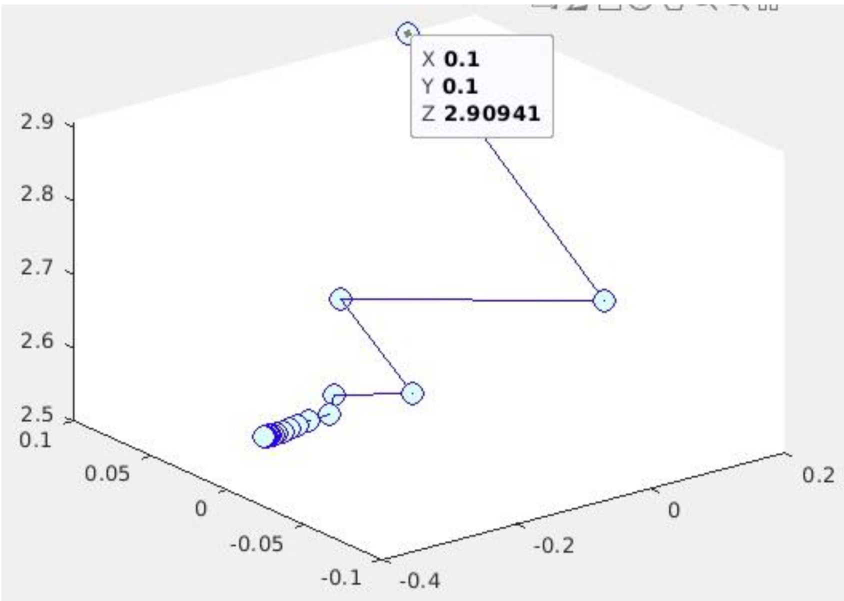
Consider the optimization

$$\min_{x_1,x_2} \quad e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$$

Write a code to solve this optimization using the gradient method and Newton method with the backtracking parameters $\alpha=0.1$ and $\beta=0.6$, draw $f(x_k)$ verses k for k =0, 1, 2……,50.

```
Alpha = 0.1;
Beta = 0.6;
nK = 50;
x0 = [0.1;0.1];

points = zeros(3,nK);
x = x0;
for iter = 1:nK
    val = exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)+exp(-x(1)-0.1);
    grad = [exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)-exp(-x(1)-0.1);
            3*exp(x(1)+3*x(2)-0.1)-3*exp(x(1)-3*x(2)-0.1)];
    points(1,iter)=x(1);
    points(2,iter)=x(2);
    points(3,iter)=val;
    v = -grad;
    x = x+Alpha*v;
end
plot3(points(1,:),points(2,:),points(3,:),'-o','Color','b','MarkerSize',10,...
    'MarkerFaceColor','#D9FFFF')
```



```
Alpha = 0.1;
Beta = 0.6;
nK = 50;
x0 = [0.1;0.1];

points = zeros(3,nK);
x = x0;
for iter = 1:nK
    val = exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)+exp(-x(1)-0.1);
    grad = [exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)-exp(-x(1)-0.1);
            3*exp(x(1)+3*x(2)-0.1)-3*exp(x(1)-3*x(2)-0.1)];
    hess = [exp(x(1)+3*x(2)-0.1)+exp(x(1)-3*x(2)-0.1)+exp(-x(1)-0.1),3*exp(x(1)+3*x(2)-0.1)-3*exp(x(1)-3*x(2)-0.1);
            3*exp(x(1)+3*x(2)-0.1)-3*exp(x(1)-3*x(2)-0.1),9*exp(x(1)+3*x(2)-0.1)+9*exp(x(1)-3*x(2)-0.1)];
    points(1,iter)=x(1);
    points(2,iter)=x(2);
    points(3,iter)=val;
    v = -hess\grad;
    x = x+Alpha*v;
end
plot3(points(1,:),points(2,:),points(3,:),'-o','Color','b','MarkerSize',10,...
    'MarkerFaceColor','#D9FFFF')
```