Quiz-2 (CS-709)

19-Oct-2011

Note: Please provide *rigorous and short* answers. Max. Marks=10

1. Using KKT conditions[1], describe the optimal set[2] of the following program:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x, \\ \text{s.t.} \quad & \|x\|_p \leq 1, \end{aligned}$$

   where $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, p \in \mathbb{R}, p > 1$ and $c \in \mathbb{R}^n$. The final expression for the optimal set should only involve parameters for the above problem i.e., should involve c and/or p alone.

   [5 Marks]

2. A discrete random variable X takes on values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively[3]. For example, X could model the rolling of a die, in which case the $x_i = i$. In reality one does not know what numbers to assume for p_i given an application. In order to determine reasonable values for p_i, one usually performs simulations. Suppose in such simulations it was revealed that the j^{th} moment[4] of X lies in the interval $[l_j, u_j]$. Here, $j = 1, \ldots, m$.

[1] Do not employ ANY non-trivial result other than KKT conditions while attempting this problem.

[2] Optimal set of a program (P) is the set of all optimal solutions of (P).

[3] In other words, $P[X = x_i] = p_i \; \forall \; i = 1, \ldots, n$. Needless to say, we must have $p_i \geq 0 \; \forall \; i = 1, \ldots, n$ and $\sum_{i=1}^n p_i = 1$.

[4] The j^{th} moment of X is defined as $\sum_{i=1}^n x_i^j p_i$

The goal is to find the "simplest" distribution for X, i.e., the values for the probabilities p_i with maximum entropy[5], such that all the above simulation-determined constraints on the moments are satisfied. Express this problem as an ordinary convex program. Using any one of the duality schemes taught in lectures till now[6], write down a dual of this program[7].

[5 Marks]

[5] Entropy of the random variable X is defined as $-\sum_{i=1}^{n} x_i \log(x_i)$.

[6] Do not use Lagrange duality scheme, as I did not teach it yet.

[7] Also, after the quiz, attempt describing the optimal set for the primal and dual programs you wrote, in terms of x_i and/or l_i and/or u_i

Quiz-2 (CS-709) 19-Oct-11

Ans1 KKT conditions ~~can be applied~~ characterize optimality for regular differentiable OCPs.

So lets first pose the given problem as a regular differentiable OCP. It is obvious that the constraint ~~is only~~ can be written as $g(x)=\sqrt[p]{\sum_i |x_i|^p} - 1 \le 0$. However, this g is not differentiable. One way to do it is:

still not differentiable.

$$\left\{x \,\middle|\, \sqrt[p]{\sum_i |x_i|^p}-1\le 0\right\} = \left\{x \,\middle|\, \sum_i |x_i|^p - 1 \le 0\right\}$$

differentiable

$$= \left\{(x,t) \,\middle|\, \begin{array}{l} \sum_i t_i^p - 1 \le 0, \\ |x_i| - t_i \le 0 \end{array}\right\}$$

not differentiable

$$= \left\{(x,t) \,\middle|\, \sum_i t_i^p - 1 \le 0,\ \begin{array}{l} x_i - t_i \le 0, \\ -t_i - x_i \le 0 \end{array}\right\}$$

differentiable

Hence the given program is equal to:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n,\ t\in\mathbb{R}^n} & c^T x \\ \text{s.t.} & \sum_i t_i^p - 1 \le 0, \\ & x_i - t_i \le 0,\ -t_i - x_i \le 0,\ \forall\ i = 1 \text{ to } n \end{array} \qquad (I)$$

Note the following: the point $(0,t)$ where t is such that $\|t\|_p < 1$ is a point which ~~strictly~~ satisfies Slater's condition.

1. (I) is a regular, differentiable OCP
2. Till now we just know that the optimal values of (I) and the given program are equal. We will later realize the relationship between their optimal solutions.

The number of constraints $= 2n+1$. Hence the "dual" variable in KKT, say λ will be in $\mathbb{R}^{2n+1}$. So let the λ entry corresponding to $\sum_i t_i^p - 1 \le 0$ be 'ρ'

& the λ entries corresponding to $x_i - t_i \le 0 \;\; \forall i = 1 \text{ to } n$ be μ

& the λ entries corresponding to $-t_i - x_i \le 0 \;\; \forall i = 1 \text{ to } n$ be ν.

We have that:

$\left(\begin{bmatrix} x^* \\ t^* \end{bmatrix}, \begin{bmatrix} \rho^* \\ \mu^* \\ \nu^* \end{bmatrix}\right)$ is a KKT point iff.

③ Feasibility:

$$\sum_i t_i^{*p} - 1 \le 0, \quad x_i^* - t_i^* \le 0 \;\; \forall i, \quad -t_i^* - x_i^* \le 0 \;\; \forall i; \quad \rho \in \mathbb{R} \ge 0, \; \mu \in \mathbb{R}^n \ge 0, \; \nu \in \mathbb{R}^n \ge 0$$

④ Complementary slackness:

$$\rho\left(\sum_i t_i^p - 1\right) = 0, \quad \mu_i(x_i - t_i) = 0 \;\; \forall i, \quad \nu_i(t_i + x_i) = 0 \;\; \forall i$$

Gradient condition:

$$\frac{\partial L\left(\begin{bmatrix} x^* \\ t^* \end{bmatrix}, \begin{bmatrix} \rho^* \\ \mu^* \\ \nu^* \end{bmatrix}\right)}{\partial x_i} = 0 \;\; \forall i \quad ①$$

$$\frac{\partial L\left(\begin{bmatrix} x^* \\ t^* \end{bmatrix}, \begin{bmatrix} \rho^* \\ \mu^* \\ \nu^* \end{bmatrix}\right)}{\partial t_i} = 0 \;\; \forall i \quad ②$$

where

$$L\left(\begin{bmatrix} x^* \\ t^* \end{bmatrix}, \begin{bmatrix} \rho \\ \mu \\ \nu \end{bmatrix}\right) = c^T x + \rho\left(\sum_i t_i^p - 1\right) + \sum_{i=1}^n \mu_i (x_i - t_i) - \sum_{i=1}^n \nu_i (t_i + x_i)$$

① $\Leftrightarrow \; c_i + \mu_i^* - \nu_i^* = 0 \;\; \forall i$

② $\Leftrightarrow \; \rho p t_i^{*p-1} - \mu_i^* - \nu_i^* = 0 \;\; \forall i$

Now suppose $C_j = 0$ for some j,

then $\mu_j^* = \nu_j^* = \frac{\xi^* p t_j^{*p-1}}{2}$ $(\because ① \& ②)$

Also from ④ we get: $t_j^{*p-1} x_j - t_j^* = 0$, $t_j^* + t_j^{*p-1} x_i^* = 0$ $\Rightarrow t_j^* = 0$ $\Downarrow (\because ③)$ $x_j^* = 0$ — (R1)

(this case actually is easy to see from the problem itself.)

Suppose $C_j \neq 0$ for some j, then,

$$① \& ② \Leftrightarrow \left.\begin{aligned} \frac{C_j + \xi^* p t_j^{*p-1}}{2} &= \nu_j^*, \\ \frac{\xi^* p t_j^{*p-1} - C_j}{2} &= \mu_j^*. \end{aligned}\right\} ⑤$$

Now because of ④, it is easy to see that both ν_j^* & μ_j^* cannot be non-zero: if they were non-zero, then $\left.\begin{aligned} x_j^* &= t_j^* \\ x_j^* + t_j^* &= 0 \end{aligned}\right\} \Leftrightarrow \begin{aligned} x_j^* &= 0, \\ t_j^* &= 0, \end{aligned}$

which is impossible by ⑤ as $C_j \neq 0$.

$\therefore$ At least one of them is zero $\Rightarrow$ $C_j = \xi^* p t_j^{*p-1}$ (here $\mu_j^* = 0$)

$C_j = -\xi^* p t_j^{*p-1}$ (here $\nu_j^* = 0$)

This shows both cannot be zero $\therefore$ either $\mu_j^* = 0$ & $\nu_j^* \neq 0$ or $\nu_j^* = 0$ & $\mu_j^* \neq 0$

Also $\xi^* \neq 0$

~~In case $\mu_j^* = 0 \Rightarrow \nu_j^* \neq 0 \Rightarrow x_j^* = t_j^*$~~

Also we know from (3), that $t_j^* \geq 0$

Hence, if $c_j^* > 0$, then $c_j = \rho p t_j^{*\,p-1}$ and hence $\mu_j^* = 0 \Rightarrow \nu_j^* \neq 0$

$$\Downarrow$$

$$\text{i.e. } x_j^* = -\left(\frac{c_j}{\rho^* p}\right)^{1/p-1} \qquad x_j^* = -t_j^*$$

if $c_j^* < 0$, then $c_j = -\rho p t_j^{*\,p-1}$ & hence $\nu_j^* = 0 \Rightarrow \mu_j^* = 0$

$$\Downarrow$$

$$\Rightarrow x_j^* = \left(\frac{-c_j}{\rho^* p}\right)^{1/p-1} \qquad x_j^* = t_j^*$$

~~Now $\rho^* \neq 0$, hence from (4), we have $\sum_i t_i^p = 1$~~

~~In other words,~~

From above we get, $|x_j^*| = t_j^* = \left(\frac{|c_j|}{\rho^* p}\right)^{1/p-1}$

Now since, $\rho^* \neq 0$, we get from (4), that $\sum_i t_i^p = 1$

$$\Leftrightarrow \frac{\sum_{j=1}^{n} (|c_j|)^{p/p-1}}{\rho^{*\,p/p-1}\, p^{p/p-1}} = 1$$

$$\Leftrightarrow \rho^* = \frac{\left(\sum_{j=1}^{n} |c_j|^{p/p-1}\right)^{p-1/p}}{p}$$

i.e. $x_j^* = \dfrac{-c_j^{1/p-1}}{\left(\sum_{i=1}^{n} |c_i|^{p/p-1}\right)^{1/p}}$ when $c_j^* > 0$ & $x_j^* = \dfrac{(-c_j)^{1/p-1}}{\left(\sum_{i=1}^{n} |c_i|^{p/p-1}\right)^{1/p}}$ when $c_j < 0$

(R2) (R3)

~~Summary of KKT conditions are~~:
Combining (R1) _ (R2) _ (R3), we get:

$$\boxed{x_j^* = \begin{cases} \dfrac{-\left(|c_j|\right)^{1/p-1}}{\left(\sum_{j=1}^{n} |c_i|^{p/p-1}\right)^{1/p}} & \text{if } c_j \geq 0 \\[2ex] \dfrac{\left(|c_j|\right)^{1/p-1}}{\left(\sum_{i=1}^{n} |c_i|^{p/p-1}\right)^{1/p}} & \text{else} \end{cases}}$$

Since this x^* is such that $\|x^*\|_p \leq 1$ & archives ~~optimal value~~ minimum ~~of the given problem~~ value for $c^T x$, we have that this is the optimal solution of the given problem also! Here the optimal set is the singleton having the above x^*.

(From this x^*, it is also clear that the optimal value of the given problem is $-\|c\|_q$, where $\frac{1}{p}+\frac{1}{q}=1$.)

---

(2) $$\min_{p_i > 0} \sum_{i=1}^{n} p_i \log p_i$$ (I)

$$\underline{\text{s.t.}} \quad \sum_{i=1}^{n} p_i = 1, \; l_i \leq \bar{a}_i^T x \leq u_i \quad \forall \; i = 1 \text{ to } m$$

where $\bar{a}_i := \begin{bmatrix} x_1^i \\ \vdots \\ x_n^i \end{bmatrix}$

~~Combining~~ ~~(R1)~~ ~~Th~~ (I) is a PCCP. i.e.

$$\text{(I)} = \min_{p \in \mathcal{P}} f(p)$$

<u>s.t.</u> $a_i^T p \le b_i$ $i = 1$ to $2m$,
$a_0^T p = b_0$

where $f(p) = \sum_{i=1}^{m} p_i \log p_i$, $\mathcal{P} = \{p \in \mathbb{R}^n / p_i > 0 \; \forall i\}$

$a_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1} = \mathbb{1}$; $b_0 = 1$

we actually know, that $\mathbb{1}^T p = 1 \iff \begin{matrix} \mathbb{1}^T p \le 1 \\ -\mathbb{1}^T p \le -1 \end{matrix}$

~~So we can get two~~
$\longrightarrow$ let corresponding dual variable be $\rho \rightarrow$ sign will be unrestricted

$a_i = \bar{a}_i$, $b_i = u_i$ } for $i = 1$ to m $\xrightarrow{\text{dual variable}}$ $\lambda \in \mathbb{R}^m$

$a_{i+m} = -\bar{a}_i$, $b_{i+m} = -l_i$ } for $i = 1$ to m. $\xrightarrow{\text{dual variable}}$ $\mu \in \mathbb{R}^m$

By the PCCP duality derived in lecture, we get.

$$\text{(I)} = \max_{\rho \in \mathbb{R}, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^m} -f^*\left(-\sum_{i=1}^{m} \lambda_i \bar{a}_i + \sum_{i=1}^{m} \mu_i \bar{a}_i - \rho \mathbb{1}\right) - \sum_{i=1}^{m} u_i \lambda_i + \sum_{i=1}^{m} l_i \mu_i - \rho \qquad \text{(II)}$$

<u>s.t.</u> $\lambda \geq 0$, $\mu \geq 0$

Now $f^*(y) = \max_{p_i > 0} y^T p - \sum_i p_i \log p_i \xrightarrow{\text{at optimality}} y_i - 1 - \log p_i = 0 \;\forall i \Rightarrow p_i = e^{y_i - 1}$

$$\therefore f^*(y) = \sum_{i=1}^{n} y_i e^{y_i - 1} - \sum_{i=1}^{n} (y_i - 1) e^{y_i - 1} = \sum_{i=1}^{n} e^{y_i - 1}.$$

Substituting in (II) gives the dual.