Optimization Theory and Algorithm	**Homework 1 - 10/11/2021**

Homework 1

Lecturer: Xiangyu Chang *Scribe:* 赵敬业

1 HW 1

1.1 Problem restatement

Consider the following convex optimization problem:

$$\min_x \ f(x),$$
$$s.t. \ \ x \preceq 0$$

Then,according to general optimality condition prove that the optimal point x^* satisfies

$$\nabla f(x^*) \preceq 0 \ \ and \ \ x_i^*(\nabla f(x^*))_i = 0, i = 1, ..., n$$

1.2 Proof

1. Proof of $\nabla f(x^*) \preceq 0$:

If $\exists \nabla f(x^*)_i > 0$, Let: $d = \begin{pmatrix} 0 \\ \vdots \\ \nabla f(x^*)_i \\ \vdots \\ 0 \end{pmatrix}$ which means only $d_i = \nabla f(x^*)_i$ and define the parameter $t \to 0_+$,
$y = x^* - td$ then,$y \preceq 0$

Then, apply the Taylor spansion: $f(y) = f(x^*) + \langle \nabla f(x^*), y - x^* \rangle + O((y - x^*)^2)$
so,

$$f(y) = f(x^*) - t(\nabla f(x^*)_i^2) + O((t^2 \nabla f(x^*)_i^2))$$

$t > 0$,so $t(\nabla f(x^*)^2) \geq 0$ Consider :

$$t\nabla f(x^*)^2 > O(t^2 \nabla f(x)^2) \Rightarrow f(y) < f(x^*)$$

Contradicts with $f(x^*)$ is the minimum，So$\nabla f(x^*) \preceq 0$

2. Proof $x_i^* \nabla f(x^*)_i = 0$:
(1) . $\nabla f(x^*)_i = 0$,then $x_i^* f(x^*)_i = 0$

(2) . $x_i^* = 0, then x_i^* f(x_i^*) = 0$

(3) . so ,only when $x_i^* < 0$ and $\nabla f(x^*)_i < 0$,$x_i^* f(x^*)_i \neq 0$

Consider the direction $d = \begin{pmatrix} 0 \\ \vdots \\ t\nabla f(x^*)_i \\ \vdots \\ 0 \end{pmatrix}$ which means only $d_i = t\nabla f(x^*)_i$ with the parameter $t \to 0_-$,

so the $x^* + d \prec 0$is in the feasible set

Use Taylor equation:

$$f(x^*+d) = f(x^*)+\langle \nabla f(x^*), x^* + d - x^* \rangle + O(t\nabla f(x^*))^2 = f(x^*)+t(\nabla f(x^*))^2+O(t\nabla f(x^*))^2 < f(x^*)$$

Contradicts with $f(x^*)$ is the minimum,so $x_i^* \nabla f(x^*)_i = 0$

2 HW 2

2.1 Problem restatement

Consider a linear programming

$$
\begin{aligned}
&min \;\; c^T x, \\
&s.t. \; Ax \leq b.
\end{aligned}
$$

Show its Lagrange dual problem

2.2 Answer

Step 1:

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = (c^T + \lambda^T A)x - \lambda^T b$$

Step 2:

$$
\begin{aligned}
g(\lambda) &= \inf_{x \in D} \left\{ (c^T + \lambda^T A)x - \lambda^T b \right\} \\
&= \begin{cases} -\lambda^T b & , c^T + \lambda^T A = 0 \\ -\infty & , otherwise \end{cases}
\end{aligned}
$$

Step 3:

$$
\begin{aligned}
\max \quad & -\lambda^T b \\
s.t. \quad & c^T + \lambda^T A = 0 \\
& \lambda \geq 0
\end{aligned}
$$

3 HW 3

3.1 Problem restatement

Compute the conjugate function: (i)$f(x) = \delta_{B_{\|\cdot\|_\infty}}$.
(ii)$f(x) = \delta_{R^n_-}(x)$.
(iii)$f(x) = \log(1 + e^x)$.
(iv)$f(x) = g(x - a) + \langle x, a \rangle$.
(v)$f(x) = \inf\limits_z \left\{ \frac{1}{2}\|x - z\|^2 + g(z) \right\}$

3.2 Answer

(i)$f^*(y) = \sup\limits_x \left\{ y^T x - f(x) \right\} = \sup\limits_{\|x\|_\infty \le 1} \left\{ y^T x - f(x) \right\} = \sigma_{B_{\|\cdot\|_\infty}}(x)$
consider that:
$y^T x \le ||x||_\infty ||y||_1 \le ||y||_1$

if $x = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$ with $\quad x_i = \begin{cases} 1 & , y_i > 0 \\ -1 & , y_i \le 0 \end{cases}$

the " = "holds
In conclusion,$f^*(y) = ||y||_1$
(ii)

$$
\begin{aligned}
f^*(y) &= \sup_x \left\{ y^T x - f(x) \right\} \\
&= \sup_x \left\{ y^T x - \delta_{R^n_-}(x) \right\} \\
&= \sup_{x \in R^n_-} \left\{ y^T x \right\} \\
&= \begin{cases} 0 & , y \in R^n_+ \\ \infty & , otherwise \end{cases} \\
&= \delta_{R^n_+}(y)
\end{aligned}
$$

(iii)$f^*(y) = \sup\limits_x \left\{ y^T x - log(1 + e^x) \right\}$
Let $L(x, y) = y^T x - log(1 + e^x)$
$\nabla_x L(x, y) = y - \frac{e^x}{1+e^x}$, with the x goes up,the $\nabla_x L$ goes down.
Let $\nabla_x L = 0$, then $x = -\log(1 - y)$, $y \in (0, 1)$ so the $L(x, y)$ touch supreme at this point,and the value is:

$$
y \log\left(\frac{y}{1 - y}\right) + \log(1 - y)
$$

if $y > 1$,then $\nabla_x L \geq 0$ and L has no supreme.

if $y < 0$,then $\nabla_x L \leq 0$ and L has no supreme.

if $y = 0$ ， $\sup_x \left\{y^T x - log(1+e^x)\right\} = \sup_x \{-log(1+e^x)\} = 0$

if $y = 1$， $\sup_x \left\{y^T x - log(1+e^x)\right\} = \sup_x \{x - log(1+e^x)\} = 0$

In conclusion： $f^*(y) = \begin{cases} y\log(\frac{y}{1-y}) + \log(1-y) & , y \in (0,1) \\ 0 & , y=0, y=1 \\ \infty & , otherwise \end{cases}$

(iv)

$$\begin{aligned} f^*(y) &= \sup_x \left\{y^T x - f(x)\right\} \\ &= \sup_x \left\{y^T x - g(x-a) - b^T x\right\} \end{aligned}$$

Let $x - a = z$,then

$$\begin{aligned} f^*(y) &= \sup_z \left\{y^T(z+a) - g(z) - b^T(z+a)\right\} \\ &= \sup_z \left\{(y^T - b^T)(z+a) - g(z)\right\} \\ &= \sup_z \left\{(y^T - b^T)z - g(z)\right\} + (y^T - b^T)a \\ &= g^*(y^T - b^T) + (y^T - b^T)a \end{aligned}$$

(v)

$$\begin{aligned} f(x) &= \inf_z \left\{\frac{1}{2}\|x-z\|^2 + g(z)\right\} \\ &= -\sup_z \left\{-\frac{1}{2}\|x-z\|^2 - g(z)\right\} \end{aligned}$$

then,

$$\begin{aligned} f^*(y) &= \sup_x \left\{y^T x - f(x)\right\} \\ &= \sup_x \left\{\sup_z \left\{-\frac{1}{2}\|x-z\|^2 - g(z)\right\} + y^T x\right\} \\ &= \sup_x \sup_z \left\{-\frac{1}{2}\|x-z\|^2 - g(z) + y^T x\right\} \end{aligned}$$

In definition of $\sup_x f(x)$,it is obvious that

$\sup_x \sup_z f(x,y) = \sup_z \sup_x f(x,y)$, so

$$
\begin{aligned}
f^*(y) &= \sup_z \sup_x \left\{ -\frac{1}{2}\|x-z\|^2 - g(z) + y^T x \right\} \\
&= \sup_z \sup_x \left\{ -\frac{1}{2}x^T x + x^T z - \frac{1}{2}z^T z - g(z) + y^T x \right\} \\
&= \sup_z \left\{ \sup_x \left\{ -\frac{1}{2}x^T x + x^T z + y^T x \right\} - \frac{1}{2}z^T z - g(z) \right\} \\
&= \sup_z \left\{ \frac{(y+z)^T(y+z)}{2} - \frac{1}{2}z^T z - g(z) \right\} \\
&= \sup_z \left\{ y^T z - g(z) \right\} + \frac{y^T y}{2} \\
&= g^*(y) + \frac{y^T y}{2}
\end{aligned}
$$

4 HW 4

4.1 Problem restatement

Consider the following problem:

(i)Define the negative entropy function is $f(x) = x\log(x)$ and $x \geq 0, 0log0 = 0$,compute its conjugate function.

(ii)Consider the following entropy maximization problem:

$$\begin{aligned}
&\min_x \sum_i x_i \log(x_i), \\
&s.t.\ Ax \succeq b, \\
&\qquad \sum_i x_i = 1
\end{aligned}$$

Please compute its Lagrange dual problem.

(iii)Suppose the strong duality holds for the entropy maximization problem,and we have obtained the optimal dual variablesλ^* and v^*,then compute the optimal primal variable x^* by λ^* and v^*

4.2 Answer

(1)

$$\begin{aligned}
f^*(y) &= \sup_x \{yx - f(x)\} \\
&= \sup_x \{yx - x\log(x)\}
\end{aligned}$$

$$\begin{aligned}
L(x, y) &= yx - x\log(x) \\
\nabla_x L(x, y) &= y - \log(x) - 1
\end{aligned}$$

with the growth of x, the$\nabla_x L$ goes down, take:

$$\begin{aligned}
\nabla_x L(x, y) = 0 &\Rightarrow x = e^{y-1} \\
\Rightarrow f^*(y) = ye^{y-1} - e^{y-1}(y-1) &= e^{y-1}
\end{aligned}$$

(2)

The primal problem:

$$\begin{aligned}
&\min_x \sum_i x_i \log(x_i), \\
&s.t.\ b - Ax \preceq 0, \\
&\qquad \sum_i x_i = 1
\end{aligned}$$

Lagrange:

$$L(x, \lambda, v) = \sum_i x_i \log(x_i) + \lambda^T (b - Ax) + v(\mathbf{1^T} x - 1)$$

the variable v is a scaler.

Lagrange Dual Function:

$$\begin{aligned}
g(\lambda, v) &= \inf_x L(x, \lambda, v) \\
&= \inf_x \left\{ \sum_i x_i \log(x_i) + \lambda^T (b - Ax) + v(\mathbf{1^T} x - 1) \right\} \\
&= \inf_x \left\{ (v\mathbf{1^T} - \lambda^T A)x + \sum_i x_i \log(x_i) \right\} + \lambda^T b - v \\
&= -\sup_x \left\{ (\lambda^T A - v\mathbf{1^T})x - \sum_i x_i \log(x_i) \right\} + \lambda^T b - V \\
&= -f^*(A^T \lambda - \mathbf{1} v) + \lambda^T b - v
\end{aligned}$$

$$f^*(y) = \sup_x \sum (y_i x_i - x_i \log x_i) = \sum_i \sup_x (y_i x_i - x_i \log x_i) = \sum_i e^{y_i - 1}$$

$$\Rightarrow g(\lambda, v) = -\sum_i e^{a_i^T \lambda - v - 1} + \lambda^T b - v$$

Lagrange Dual Problem:

$$\begin{aligned}
\max_{\lambda, v} \quad & -\sum_i e^{a_i^T \lambda - v - 1} + \lambda^T b - v \\
s.t. \quad & \lambda \geq 0
\end{aligned}$$

(3)

$$L(x, \lambda^*, v^*) = \sum_i x_i \log(x_i) + \lambda^{*T} (b - Ax) + v^* (\sum_i x_i - 1)$$

$$\frac{\partial L(x, \lambda^*, v^*)}{\partial x_i} = 1 + \log(x_i) - (A^T \lambda^*)_i + v^* = 0 \Rightarrow x_i = \exp\left(-v^* + (A^T \lambda^*)_i - 1\right)$$

means the ith position of x^* satisfy $x_i = \exp\left(-v^* + (A^T \lambda^*)_i - 1\right)$

5 HW 5

5.1 Problem restatement

Show the Lagrange dual problems for

(i)

$$\min_x f(x) + g(Ax)$$

(ii)Ridge Regression:

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_2^2$$

5.2 Answer

(1)primal problem:

$$\min_x f(x) + g(Ax)$$

equal to:

$$\begin{aligned} &\min_x f(x) + g(z) \\ &s.t. \quad Ax = z \end{aligned}$$

Lagrange:

$$L(x, v) = f(x) + g(z) + v^T(Ax - z)$$

Lagrange Dual Function:

$$\begin{aligned} t(v) &= \inf_x L(x, v) \\ &= \inf_x \left\{ f(x) + g(z) + v^T(Ax - z) \right\} \\ &= \inf_x \left\{ f(x) + v^T Ax \right\} + \inf_x \left\{ g(z) - v^T z \right\} \\ &= -\sup_x \left\{ (-f(x) - v^T Ax) \right\} - \sup_x \left\{ (-g(z) + v^T z) \right\} \\ &= -f^*(-A^T v) - g^*(v) \end{aligned}$$

Lagrange Dual Problem:

$$\max_v -f^*(-A^T v) - g^*(v)$$

(2)

$$
\begin{aligned}
&\min_x \frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_2^2 \\
&= \min_x \sup_y \left((Ax-b)^T y - \frac{1}{2}\left|\left|y\right|\right|^2 \right) + \lambda \left|\left|x\right|\right|_2^2 \\
&\geq \sup_y \min_x \{ (Ax)^T y - b^T y - \frac{1}{2}\left|\left|y\right|\right|^2 + \lambda \left|\left|x\right|\right|_2^2 \} \\
&= \sup_y \{ \min_x \{ (A^T y)^T x + \lambda \left|\left|x\right|\right|_2^2 \} - b^T y - \frac{1}{2}\left|\left|y\right|\right|^2 \} \\
&= \sup_y \{ -\sup_x \{ (-A^T y)^T x - \lambda \left|\left|x\right|\right|_2^2 \} - b^T y - \frac{1}{2}\left|\left|y\right|\right|^2 \} \\
&= \sup_y \{ -2\lambda \sup_x \{ (-\frac{A^T y}{2\lambda})^T x - \frac{1}{2}\left|\left|x\right|\right|_2^2 \} - b^T y - \frac{1}{2}\left|\left|y\right|\right|^2 \} \\
&= \sup \{ -\frac{1}{4\lambda}\left|\left|A^T y\right|\right|^2 - b^T y - \frac{1}{2}\left|\left|y\right|\right|^2 \}
\end{aligned}
$$

So,the lagrange dual problem:

$$
\begin{aligned}
&\max_y \ \ -\frac{1}{2}\|y\|_2^2 - \frac{1}{4\lambda}\|A^T y\|_2^2 - y^T b \\
&s.t. \quad \lambda \geq 0
\end{aligned}
$$

参考文献