优化方法作业7

1. (a) 直接算 $\mathrm{Prox}_c f(x) = \arg\min_w \{\|w\|_2 + \frac{1}{2c}\|w-x\|^2\}$. $g(w)=\|w\|_2+\frac{1}{2}\|w-x\|^2$

对w分解 $w = t\frac{x}{\|x\|}+z$ 其中 $t\in\mathbb{R}$ $x^Tz=0$.

$\Rightarrow g(w)=\sqrt{t^2+\|z\|_2^2}+\frac{1}{2c}\|z\|_2^2+\frac{1}{2c}(\frac{t}{\|x\|_2}-1)^2\|x\|_2^2$

$\Rightarrow g(w)$ 最小时 $z=0$ $\Rightarrow g(w)=|t|+\frac{1}{2c}(t-\|x\|_2)^2 = \cancel{f(t)}\ h(t)$. 为t的函数

$\Rightarrow \arg\min_t \cancel{f(t)}\ h(t) = \max(\|x\|-c, 0)$

$\Rightarrow \mathrm{Prox}_c f(x)=\max(\|x\|-c,0)\cdot\frac{x}{\|x\|}$

Moreau $f^*(y)=0$ $\mathrm{dom} f^*=\{y \mid \|y\|_2\le 1\}$

$\mathrm{Prox}_{c^{-1}} f^*(c^{-1}x)=P_B(\frac{x}{c})$ $B=\{x \mid \|x\|_2\le 1\}$

$\Rightarrow \mathrm{Prox}_c f(x) = x - c\,\mathrm{Prox}_{c^{-1}}f^*(c^{-1}x)$

$= x - cP_B(\frac{x}{c}) = x - P_{cB}(x) = \begin{cases}0 & \|x\|\le c\\ x - c\frac{x}{\|x\|_2} & \|x\|>c\end{cases} = \max(\|x\|-c,0)\frac{x}{\|x\|}$

(b) $\mathrm{Prox}_c f(x)=\arg\min_w\{-\sum \ln w_i + \frac{1}{2c}\|w-x\|_2^2\}$

对每个i $g(w_i)=-\ln w_i+\frac{1}{2c}(w_i-x_i)^2$ $\arg\min_{w_i} g(w_i)=\frac{1}{2}(x_i+\sqrt{x_i^2+4c})$. ($\because w_i>0$)

$\therefore \mathrm{Prox}_c f(x)=\hat{w}$ $\hat{w}_i=\frac{1}{2}(x_i+\sqrt{x_i^2+4c})$ $1\le i\le n$

(c) $\mathrm{Prox}_c g(x)=\arg\min_w\{g(w)+\frac{1}{2c}\|w-x\|_2^2\}$

$=\arg\min_{w\ge 0}\{\lambda w^3+\frac{1}{2c}\|w-x\|_2^2\}$ 对w求导 $3\lambda w^2+\frac{1}{c}(w-x)=0$

① $\lambda=0 \Rightarrow \mathrm{Prox}_c g(x)=\begin{cases}0 & x<0\\ x & x\ge 0\end{cases}$

② $\lambda>0$. $\mathrm{Prox}_c g(x)=\begin{cases}\frac{-1+\sqrt{1+12cx\lambda}}{6c\lambda} & x\ge 0\\ 0 & x<0\end{cases}$

③ $\lambda<0$ $\mathrm{Prox}_c g(x)=$ $\cancel{\frac{-1}{6c\lambda}}$ $w\to+\infty$时 最小. $\Rightarrow \mathrm{Prox}_c g(x)=+\infty$.

2. $d(x)=\inf_{y\in C}\|x-y\|_2$

设 $u=\mathrm{Prox}_c d(x)$ ① $u\in C \Rightarrow u=P_C(x)$

② $u\notin C \Rightarrow \frac{1}{c}(x-u)\in\partial d(x) \Rightarrow u = x-\frac{c}{d(u)}(u-P_C(u))$

且 $P_C(u)=P_C(x)$

$\Rightarrow$ u是 $P_C(x)$ 和x的线性组合 $\Rightarrow u=\frac{t}{d(x)}P_C(x)+(1-\frac{t}{d(x)})x$

$\Rightarrow \mathrm{Prox}_c d(x)= \theta P_C(x)+(1-\theta)x$ $\theta=\begin{cases}t/d(x) & d(x)\ge t\\ 1 & \text{else}\end{cases}$

$f(x) = \frac{1}{2} d(x)^2$

$\mathrm{Prox}_c f(x) = \underset{u}{\arg\min} \inf_{y \in C} \frac{1}{2}\|u-y\|^2 + \frac{1}{2c}\|u-x\|^2.$

对u求导 $\Rightarrow$ $\hat{u}$满足 $u = \frac{1}{t+1} y + \frac{1}{t+1} x$ 代入有 $\frac{1}{2}\|\frac{t}{t+1} y + \frac{1}{t+1} x - u\|_2^2 + \frac{1}{2t}\|\frac{t}{t+1} y + \frac{1}{t+1} x - x\|_2^2$

$= \frac{1}{2(t+1)}\|y-x\|_2^2$ $y = P_C(x)$ 时最小

$\Rightarrow \mathrm{Prox}_c f(x) = u = \frac{1}{c+1} P_C(x) + \frac{1}{c+1} x.$

3. (a) $\mathrm{Prox}_c g(x - ca) = \underset{w}{\arg\min}\ g(w) + \frac{1}{2c}\|w-(x-ca)\|_2^2.$

$= \underset{w}{\arg\min}\ f(w) - a^T w + \frac{1}{2c}\|w-x\|_2^2 + \frac{1}{2c}\|ca\|_2^2 + \frac{1}{c}(w-x)^T ca$

$= \underset{w}{\arg\min}\ f(w) + \frac{1}{2c}\|w-x\|_2^2$ (去除与w无关的常数).

$= \mathrm{Prox}_c f(x).$

(b) $\mathrm{Prox}_\lambda g(\lambda(c^{-1}x + \mu^{-1}a)) = \underset{w}{\arg\min}\ g(w) + \frac{1}{2\lambda}\|w - \lambda(c^{-1}x + \mu^{-1}a)\|_2^2$

$= \underset{w}{\arg\min}\ f(w) - \frac{1}{2\mu}\|w-a\|_2^2 + \frac{1}{2\lambda}\|w - x + (x - \frac{\lambda}{c}x - \frac{\lambda}{\mu}a)\|_2^2.$

~~最后一项展开~~ $\frac{1}{2\lambda}\|w-x\|_2^2 + \frac{1}{2\lambda}(w-x)^T(\frac{2\lambda}{\mu}x - \frac{\lambda}{\mu}a) + C_1$ ~~C_1, C_2是与w无关常数~~

$= \frac{1}{2c}\|w-x\|_2^2 + \quad w^T(x-a) + C_2$

$= \frac{1}{2c}\|w-a\|_2^2 + \frac{1}{2c}(w-a)^T(a-x) +$

w的二次项系数 $\frac{1}{2\mu} + \frac{1}{2\lambda} = \frac{1}{2c}$

一次项系数

$-\frac{1}{2\mu}\cdot 2a + \frac{1}{2\lambda}\cdot -2\lambda(c^{-1}x + \mu^{-1}a) = \frac{1}{2c}\cdot 2x$

$\therefore$ $\mathrm{Prox}_\lambda g(\cdots) = \underset{w}{\arg\min}\ f(w) + \frac{1}{2c}\|w-x\|_2^2 = \mathrm{Prox}_c f(x)$

去与w无关的常数 可忽略.

4. $x_k = \mathrm{Prox}_\eta g(\cdots) \Rightarrow \frac{1}{\eta}(y_k - \eta\nabla f(y_k) - x_k) \in \partial g(x_k)$

由次梯度定义 $g(y_k) \geq g(x_k) + \langle \frac{1}{\eta}(y_k - x_k - \eta\nabla f(x_k)), y_k - x_k\rangle$

$= g(x_k) + \frac{1}{\eta}\|y_k - x_k\|_2^2$

由定义知 $g(x_k) + \frac{1}{2\eta}\|x_k - y_k + \eta\nabla f(y_k)\|_2^2 \leq g(y_k) + \frac{1}{2\eta}\|y_k - y_k + \eta\nabla f(y_k)\|_2^2$

展开有 $g(x_k) \leq g(y_k) - \frac{1}{2\eta}\|x_k - y_k\|^2 - \langle\nabla f(y_k), x_k - y_k\rangle$ 证毕!

5. 由 $f(x) = \chi_L(x)$ $\quad f^*(y) = \sup_{x} y^T x - f(x) = \sup_{x} y^T x - \chi_L(x) = \sup_{x \in L} y^T x = \begin{cases} 0 & y \in L^{\perp} \\ \infty & \text{else} \end{cases}$

$= \chi_{L^\perp}(y)$

又 $\text{Prox}_c f(x) = P_L(x)$

$\therefore x = \text{Prox}_c f(x) + \text{Prox}_c f^*(x)$

$= P_L(x) + P_{L^\perp}(x)$

6. μ-strongly $\Rightarrow$ $f(x_1) \geq f(x_2) + \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\mu}{2} \|x_1 - x_2\|_2^2$

$f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \frac{\mu}{2} \|x_1 - x_2\|_2^2$

相加有 $\quad \mu \|x_1 - x_2\|_2^2 \leq \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle$ ①

L-smooth. $f(x_1) \leq f(x_2) + \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{L}{2} \|x_1 - x_2\|_2^2$

$f(x_2) \leq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \frac{L}{2} \|x_1 - x_2\|_2^2$

相加有 $\quad L \|x_1 - x_2\|_2^2 \geq \langle \nabla f(x_1) - \nabla f(x_2) \rangle.$

由 Cauchy 不等式和① $\quad \mu \|x_1 - x_2\|_2^2 \leq \|\nabla f(x_1) - \nabla f(x_2)\| \|x_1 - x_2\|$

$\Rightarrow \mu \|x_1 - x_2\| \leq \|\nabla f(x_1) - \nabla f(x_2)\|.$