Convex Optimization — Operator Splitting Methods

Problem 1

Compute the proximal mapping $\mathrm{prox}_f(x)$ of the following functions.

a). $f(x)=\frac{1}{2}x^TAx+b^Tx+c$, $A\in S_+^n$. $b\in\mathbb{R}^n$.

$$\mathrm{prox}_f(x)=\underset{y}{\arg\min}\ g(y):=\frac{1}{2}y^TAy+b^Ty+C+\frac{1}{2}\|y-x\|_2^2.$$

$$\partial_y g(y)=Ay+b+y-x=0$$

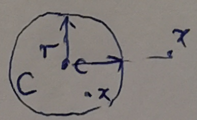
$$\Rightarrow\quad y=(I+A)^{-1}(x-b).$$

$$\Rightarrow\quad \mathrm{prox}_f(x)=(I+A)^{-1}(x-b)$$

b) $f(x)=I_C(x)$, $C=\{x\in\mathbb{R}^n:\ \|x-c\|_2\le r\}$. $c\in\mathbb{R}^n$. $r\in\mathbb{R}_{++}$

$$\mathrm{prox}_f(x)=\underset{y}{\arg\min}\ I_C(y)+\frac{1}{2}\|y-x\|_2^2$$

$$=\underset{y\in C}{\arg\min}\ \frac{1}{2}\|y-x\|_2^2=\begin{cases} x & x\in C.\\ c+\frac{x-c}{\|x-c\|}\cdot r & x\notin C.\end{cases}$$



c) $f(x)=I_C(x)$, $C=\{x\in\mathbb{R}^n:\ Px=q\}$. $P\in\mathbb{R}^{m\times n}$ has full row rank. $q\in\mathbb{R}^m$. $m\le n$

$$\mathrm{prox}_f(x)=\underset{y}{\arg\min}\ I_C(y)+\frac{1}{2}\|y-x\|_2^2$$

$$=\underset{Py=q}{\arg\min}\ \frac{1}{2}\|y-x\|_2^2.$$

$$d(\mu):=\inf_y \frac{1}{2}\|y-x\|_2^2+\mu^T(Py-q).$$

$$\Rightarrow\quad y-x+P^T\mu=0\quad\Rightarrow\quad y=x-P^T\mu.$$

Hence, $d(\mu)=\frac{1}{2}\|P^T\mu\|_2^2+\mu^TP(x-P^T\mu)-\mu^Tq$

$$=-\frac{1}{2}\|P^T\mu\|_2^2+\mu^TPx-\mu^Tq.$$

$\Rightarrow$ optimal $\bar\mu$. $\quad -PP^T\bar\mu+Px=0$

$$\bar\mu=(PP^T)^{-1}P\bar{x}..$$

therefore. $\bar{x}$. $\bar\mu$ satisfies KKT condition, and. $I_C(y)+\frac{1}{2}\|y-x\|_2^2$ is convex.

$\Rightarrow$ $\bar{x}$. $\bar\mu$ are optimal.

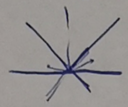
$\Rightarrow \quad \bar{y} = x - P^T(PP^T)^{-1}Px$

$\text{prox}_f(x) = x - P^T(PP^T)^{-1}Px.$

(d). $f(x) = \lambda|x|, \ x \in \mathbb{R}, \ \lambda > 0.$

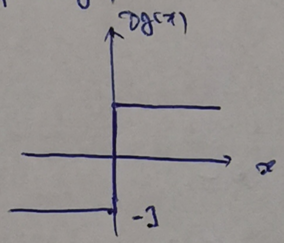
$\text{prox}_{\lambda g}(x) = (I_d + \lambda \partial g)^{-1}(x).$

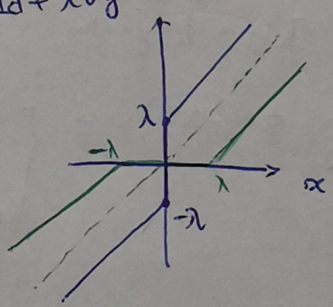
Let $g = |x|.$



$\Rightarrow \quad \partial g = \begin{cases} \text{sgn}(x) & x \neq 0 \\ [-1, 1]. & x = 0. \end{cases}$

graphically, therefore, Inverse of $I_d + \lambda\partial g$.





See. gree line.

$\Rightarrow \quad \text{prox}_{\lambda g}(x) = \text{prox}_f(x) = \begin{cases} x + \lambda & x \leq -\lambda \\ 0 & -\lambda < x \leq \lambda \\ x - \lambda. & x > \lambda \end{cases}$

e). l_1-norm: $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|.$

$\text{prox}_f(x) = \underset{y}{\text{argmin}} \left\{ \sum_{i=1}^n |x_i| + \frac{1}{2}\|y - x\|_2^2 \right\}$

$= \begin{bmatrix} \underset{y_1}{\text{argmin}} \ |y_1| + \frac{1}{2}(y_1 - x_1)^2 \\ \vdots \\ \underset{y_n}{\text{argmin}} \ |y_n| + \frac{1}{2}(y_n - x_n)^2 \end{bmatrix}.$

using result of (d). (Let $\lambda = 1$).

$[\text{prox}_f(x)]_i = \begin{cases} x_i + 1 & x_i \leq -1 \\ 0 & -1 < x_i \leq 1 \\ x_i - 1 & x_i > 1. \end{cases} \quad i = 1, \dots, n.$

f) l_2-norm. $f(x) = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$

Moreau's Identity shows

$$\mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) = Id(x)$$

The Conjugate of l_2-norm is

$$f^*(x) = \begin{cases} 0 & \|x\|_2 \le 1 \\ \infty & \|x\|_2 > 1 \end{cases}$$

$$= I_C(x), \text{ where } C = \{x \in \mathbb{R}^n, \|x\|_2 \le 1\}.$$

By (b) we know

$$\mathrm{prox}_{f^*}(x) = \begin{cases} x & x \in C \\ \frac{x}{\|x\|_2} & x \notin C \end{cases}$$

therefore. $\mathrm{prox}_f(x) = Id(x) - \mathrm{prox}_{f^*}(x)$

$$= \begin{cases} 0 & \|x\|_2 \le 1 \\ x - \frac{x}{\|x\|_2} & \|x\|_2 > 1 \end{cases}$$

g) in subtask 1.a).

$$\mathrm{prox}_f(x) = (I+A)^{-1}(x-b)$$

therefore. $\mathrm{prox}_f(x) - \mathrm{prox}_f(y) = (I+A)^{-1}(x-y)$.

Let $\|\mathrm{prox}_f(x) - \mathrm{prox}_f(y)\|_2^2 = (x-y)^T.(I+A)^{-2}(x-y) \le \beta^2 \|x-y\|_2^2$.

Denote $u = x-y$.

$$\Rightarrow \quad u^T(I+A)^{-2}u \le \beta^2 u^T u.$$

Let $u = (I+A).\tilde{u}$.

$$\Rightarrow \quad \tilde{u}^T\tilde{u} \le \tilde{u}^T\beta^2.(I+A)^2.\tilde{u}.$$

$$\Rightarrow \quad (I+A)^2 \succeq \frac{1}{\beta^2}.I.$$

$$\Rightarrow \quad A^2 + 2A \succeq \left(\frac{1}{\beta^2} - 1\right).I.$$

In order $\mathrm{prox}_f(x)$ Contractive, i.e. $\beta < 1$.

we need $A^2 + 2A \succ 0$.

Because $A \in S^n_+$.

Hence we need $A \in S^n_{++}$.

and $\frac{1}{\beta^2} \le \lambda_{min}[(I+A)^2]$.

$\Rightarrow \quad \beta \ge \sqrt{\lambda_{min}(I+A)^2}$.

therefore.

Contraction factor. is $\sqrt{\lambda_{min}[(I+A)^2]}$..

Problem 2.

a). 1. $u = prox_f(x) \iff u = \arg\min_y \{ f(y) + \frac{1}{2}\|y-x\|_2^2 \}$

By Fermat's rule

$\iff 0 \in \partial f(u) + u - x$

$\iff x - u \in \partial f(u)$ $\square$

2. x is a minimizer of f $\iff 0 \in \partial f(x)$

$\iff x - x \in \partial f(x)$

use first statement

$\iff x = prox_f(x)$ $\square$

b) let $u = prox_f(x)$. $v = prox_f(y)$.

From a).1. we know.

$u = prox_f(x) \iff x - u \in \partial f(u)$.

multiply by $(v-u)^T$.

$\Rightarrow \quad (v-u)^T(x-u) = \partial f^T(u)(v-u) \le f(v) - f(u)$ ①

$v = prof(y) \iff y - v \in \partial f(v)$

$\Rightarrow \quad (u-v)^T(y-v) = \partial f^T(v) \cdot (u-v) \le f(u) - f(v)$ ②

① + ② $\Rightarrow$

$$(v-u)^T(x-u) + (u-v)^T(y-v) \le 0.$$

$\Leftrightarrow$

$$\|u-v\|_2^2 \le (u-v)^T(x-y)$$

$$\le \|x-y\| \cdot \|u-v\|$$ by Cauchy-Schwarz inequality.

If $u \ne v$,

$\Rightarrow$ $\|u-v\| \le \|x-y\|$.

If $u = v$.

then $\|u-v\| = 0 \le \|x-y\|$.

therefore, $\forall x, y \in \mathbb{R}^n$.

we have $\|u-v\| \le \|x-y\|$. $\square$

Problem 3

$$\min_{x\in\mathbb{R}^n} f(x) + g(x) = \min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \|x\|_1.$$

Here. $f(x) = \|Ax-b\|_2^2$ is β-smooth. ($\beta = \lambda_{max}(A^TA)$)

$g(x) = \|x\|_1$ is a proper closed function, convex.

$$J_{\gamma\partial f} = (Id + \gamma\partial f)^{-1} = [Id + \gamma \cdot A^T(Ax-b)]^{-1} = prox_{\gamma f}^{x}$$

$$J_{\gamma\partial g} = (Id + \gamma\partial g)^{-1} = prox_{\gamma g}(x) = \sum_{i=1}^{n} [prox_{\gamma g}(x)]_i e_i.$$

where $[prox_{\gamma g}(x)]_i = \begin{cases} x_i + \gamma & x_i \le -\lambda \\ 0 & -\lambda < x_i \le \lambda \\ x_i - \lambda & x_i > \lambda \end{cases} := h(x_i).$ (*).

$$R_{\gamma\partial f} = 2prox_{\gamma f} - Id = 2[Id + \gamma A^T(Ax-b)]^{-1} - Id$$

$$R_{\gamma\partial g} = 2prox_{\gamma g} - Id = \sum_{i=1}^{n} \{2[prox_{\gamma g}(x)]_i - x_i\} e_i$$

Therefore, for

a). Douglas-Rachford splitting

$$z^{k+1} = ((1-\lambda)\cdot I_d + \lambda \cdot R_{\gamma g} R_{\gamma f}) z^k.$$

$$x^k = \text{prox}_{\gamma f}(z^k)$$

$$= [I_d + \gamma A^T(Ax-b)]^{-1} z^k.$$

~~And.~~

$$\cancel{z^{k+1} = [(1-\lambda) I_d z^k + \lambda.}$$

or More explicitly.

$$x^k = [I_d + \gamma A^T(Ax-b)]^{-1} z^k.$$

$$y^k = \sum_{i=1}^{n} h(2x_i^k - z_i^k) e_i. \quad \text{where } h(x_i) \text{ is defined in } (*)$$

$$z^{k+1} = z^k + 2\lambda(y^k - x^k).$$

b). Forward-backward splitting.

$$x^{k+1} = \text{prox}_{\gamma g}(I_d - \gamma \nabla f) x^k$$

$$= \text{prox}_{\gamma g}.[I_d - \gamma A^T(Ax-b)] x^k$$

$$= \sum_{i=1}^{n} h(x_i^k - \gamma a_i^T(Ax-b)x^k), \quad \text{Let. } A = [a_1. a_2 \cdots a_n]$$

$\square$