Moreau - Yosida.

$$\begin{cases} \psi_{f,t}(x) = \min_y f(y) + \frac{1}{2t}\|y-x\|_2^2 \\ \downarrow \\ P_{tf}(x) = \operatorname{argmin}_y f(y) + \frac{1}{2t}\|y-x\|_2^2 \end{cases}$$

$$\begin{cases} f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x)\} \\ \psi_{f,t}(x) = \min_y \{f(y) + \frac{1}{2t}\|y-x\|_2^2\} \\ P_{tf}(x) = \operatorname{argmin}\{tf(y) + \frac{1}{2}\|y-x\|_2^2\} \end{cases}$$

$\Downarrow$

$$\begin{cases} \min f(x) \Leftrightarrow \min \psi_{f,t}(x) \\ \operatorname{argmin} f(x) \Leftrightarrow \operatorname{argmin} \psi_{f,t}(x) \end{cases}$$

① For closed, proper, convex f,

Moreau decomposition

$\Downarrow$ $x = P_f(x) + P_{f^*}(x)$

$\boxed{\text{More}}$

$\boxed{Q_f(x) = P_{f^*}(x)}$

$\Rightarrow x - Q_f(x) \in \partial f^*(Q_f(x))$

$\Rightarrow P_f(x) \in \partial f^*(Q_f(x))$

$Q_f(x) \in \partial f(P_f(x))$

Since $x - P_f(x) \in \partial f(P_f(x))$

② use Moreau Decomposition, we can find.

$$g(x) = \min_s f^*(-s) + \frac{\lambda}{2}\|s-x\|^2$$

$$= \lambda\, \psi_{f^*(-\cdot),\lambda^{-1}}(x)$$

i)

$$\Rightarrow \nabla g = \lambda \nabla \psi_{\lambda^{-1} f^*(-\cdot)}$$

$$= \lambda\,(x - P_{\lambda^{-1} f^*(-\cdot)}(x)) = -P_{\lambda f}(-\lambda x)$$

ii) $P_{\lambda^{-1}f^*(-\cdot)}(x)$

$= -P_{\lambda^{-1}f^*(\cdot)}(-x)$

$= -\left(-x - P_{(\lambda^{-1}f^*(\cdot))^*}(-x)\right)$

$= x + P_{\lambda^{-1}f(\lambda x)}(-x)$

$= x + \lambda^{-1} P_{\lambda f}(-\lambda x)$

3, Some Simple results. (for $P_h(x)$)

$\downarrow$ derive by $\underline{P_h(x) = x - P_{h^*}(x)}$

① $h = \|\cdot\|_{\#}$

$h^* \to$ indicator function $\delta(x \mid \partial h(0))$

$\downarrow$

B_*'

$\Rightarrow h^*(x) = \delta(x \mid B_*')$

$\Rightarrow P_{h^*}(x) = \Pi_{B_*'}(x)$

$\Rightarrow P_h(x) = x - \Pi_{B_*'}(x)$

② $f = \max\{x_1, \dots, x_n\}$

$\Rightarrow \partial f(0) = \{\xi : \xi_1 + \xi_2 + \dots + \xi_n = 1,\ \xi \geq 0\}$

$\Rightarrow f^* = \delta(x \mid S)$

$\Rightarrow P_{f^*} = \Pi_S(x)$

$\Rightarrow P_f(x) = x - \Pi_S(x)$ $\quad$ $\xi \in \underline{\partial \delta_{\mathbb{R}_+^n}(x)}$

$\delta_{\mathbb{R}_+^n}(y) \geq \quad + \xi^T(y - x)$

③ $f = \max\{x_1, \dots, x_n\} + \delta_{\mathbb{R}_+^n}(x)$

$\Rightarrow \partial f(0) = \{x \in \mathbb{R}^n : e^T x^+ \leq 1\}$

$\Rightarrow f^* = \delta(x \mid \partial f(0))$ $\quad$ Some simplified form

$\Rightarrow P_{f^*}(x) = \Pi_{\partial f(0)}(x)$

$$\Rightarrow \quad P_f(x) = x - P_{f^*}(x)$$

PPA → Proximal Point Alg.

$$\min f(x)$$
$$\downarrow$$
$$\min \varphi_{f,t}(x)$$
$$\Downarrow \text{ "steepest descent"}$$
$$\boxed{x_{k+1}} \approx x_k - t_k \nabla \varphi_{f,t_k}(x_k)$$
$$\approx P_{t_k f}(x_k)$$
$$\approx \boxed{\arg\min_x \; t_k f(x) + \frac{1}{2}\|x - x_k\|_2^2}$$

Variant 1:

① Proximal Gradient alg.

$$f(x) = g(x) + \underset{\sim}{h(x)}$$

$\downarrow$ closed form $P_h(x)$ → like $h = \|\cdot\|_1$

$$x_{k+1} = P_{t_k f}(x_k)$$

$x_k - t_k \nabla \varphi_{\tilde{f}_{x_k}, t_k}(x_k)$

$$\approx P_{t_k \tilde{f}_{x_k}}(x_k)$$

解析 $P_{t_k \tilde{f}_{x_k}}(x_k)$

$$= P_{t_k h(\cdot)}(x_k - t_k \nabla g(x_k))$$

$$f(x) \approx \tilde{f}_{x_k} = g(x_k) + \nabla g(x_k)^T (x - x_k) + h(x)$$

$\boxed{\text{Note}}$: if $h(x) \to P_h(x)$ has closed form.

then $f(x) = \alpha h(x) + \beta$

$f(x) = h(\alpha x + b)$ also has closed form

$$f(x) = h(x) + a^T x + \beta$$

$$f(x) = h(x) + \frac{\alpha}{2}\|x-b\|_2^2$$

Variant 2:

Dual PPA.

$$\min \; g(x) + h(x)$$

$h(x) = \|Bx\|_1$ → no closed form

$$\Leftrightarrow \min_x \; g(x) + \|Bx\|_1$$

$$\Leftrightarrow \min_{x,y} \; g(x) + \|y\|_1$$

$$\text{s.t.} \quad y = Bx$$

$$(B, -I)\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

⇒ generalize

$$\boxed{\begin{aligned} &\min \; f(z) \\ &\text{s.t.} \; Az = b \end{aligned}}$$

$$\min \; \underline{f(z)} \to \text{convex}$$

$$\text{s.t.} \; \underline{Az = b} \to \text{affine}$$

⇒ strong Duality

$$\Updownarrow$$

$$\min_u \; d(u)$$

$$d(u) = -\min_x L(x, u)$$

$$= -\min \; f(x) + u^T(Ax - b)$$

$$\partial d(u) = \text{conv}\{ b - A\bar{x} ;\; \bar{x} \in X(u) \}$$

Apply PPA Framework

$$u_{k+1} = u_k - t_k \nabla \varphi_{d, t_k}(u_k)$$

$$= P_{t_k d}(u_k) = u_k - t_k(b - Ax^+)$$

考虑 $L_A(x, u_k, t_k) = f(x) + u^T(Ax - b)$

$$+\frac{t_k}{2}\|Ax-b\|_2^2$$

$$P_{t_k d}(u_k) = u_k - t_k(b - Ax^+) \quad x^+ \in \underbrace{L_A(x, u_k, t_k)}_{\substack{\text{argmin} \\ x}}$$

$$\text{arg min } L_A(x, u_k, t_k)$$

$$\Leftrightarrow \underset{x}{\text{argmin}}\ f(x) + u^T(Ax-b) + \frac{\cancel{t_k}}{2t_k}\|Ax-b\|_2^2$$

$$\Leftrightarrow \underset{x,y}{\text{argmin}}\ g(x) + \|y\|_1 + \lambda^T(Bx-y) + \frac{1}{2t_k}\|Bx-y\|_2^2$$

$$\Leftrightarrow \underset{x,y}{\text{argmin}}\ g(x) + \|y\|_1 + \cancel{0}\ \frac{1}{2t_k}\|Bx-y+t_k\lambda\|_2^2$$

$\boxed{\text{APMM}}$

$$x^{k+1} = \text{argmin}\ g(x) + \frac{1}{2t_k}\|Bx + t_k\lambda^k - y^k\|_2^2$$

$$y^{k+1} = \text{argmin}\ \|y\|_1 + \frac{1}{2t_k}\|(Bx^{k+1} + t_k\lambda^k - y\|_2^2$$

$$= P_{t_k\|\cdot\|_1}(Bx^{k+1} + t_k\lambda^k) \Rightarrow \boxed{\text{closed form}}$$

$$z - u \in cT(u)$$

$$\Rightarrow u = (I + cT)^{-1}(z)$$

$$= P_c(z)$$

Two ways of define $\underline{P_k}$

① $P_k = (I + c_kT)^{-1}$

② $z = (I + c_kT)\cdot P_k(z)$

$$\Rightarrow z - P_k(z) \in c_k T(P_k(z))$$