- Operator

Definition:

let $\mathcal{X}$ and $\mathcal{Y}$ be subsets of Hilbert spaces

an operator is a mapping from $x \in \mathcal{X}$ $\begin{cases} \text{an element } A(x) \in \mathcal{Y} \\ \text{or} \\ \text{a set of element } A(x) \subseteq \mathcal{Y} \end{cases}$

denoted $A: \mathcal{X} \rightrightarrows \mathcal{Y}$

eg. gradient $\nabla f: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$

subdifferential $\partial f: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$

Relation of A is $\{(x,y): x \in \mathcal{X}, y \in A(x)\}$

eg. $\{(x,y): x \in \mathbb{R}^n, y \in \partial f(x)\}$

$\{(x, Ax): x \in \mathbb{R}^n\}$ where $A \in \mathbb{R}^{n \times n}$

Can define generalized equation: $A(x) = 0$ or $0 \in A(x)$

- Operations on Relation

let $C \subseteq \mathcal{X} \times \mathcal{X}$, $S \subseteq \mathcal{X} \times \mathcal{X}$ be relations

Inverse: $C^{-1} = \{(y,x): (x,y) \in C\}$

Composition: $C \cdot S = \{(x,y): \exists z \text{ s.t. } (x,z) \in S, (z,y) \in C\}$

Scalar multiplication: $\alpha C = \{(x, \alpha y): (x,y) \in C\}$

addition: $C + S = \{(x, y+z): (x,y) \in C, (x,z) \in S\}$

resolvent: $(I + \lambda C)^{-1} = \{(x + \lambda y, x): (x,y) \in C\}$

eg. if f is closed convex proper, then $(\partial f)^{-1} = \partial f^*$

$$\begin{aligned}
(x,y) \in (\partial f)^{-1} &\Leftrightarrow x \in \partial f(y) \\
&\Leftrightarrow 0 \in x - \partial f(y) \\
&\Leftrightarrow y \in \arg\max_u \{u^T x - f(u)\} \\
&\Leftrightarrow y \in \partial f^*(x) \\
&\Leftrightarrow (x,y) \in \partial f^*(x)
\end{aligned}$$

• Monotone Operators

Definition: (Monotone Operator)

A relation / operator A is monotone if $\langle x-u, y-v\rangle \geq 0 \quad \forall (x,y), (u,v) \in A$

eg. Suppose f is differentiable

f convex $\Leftrightarrow \{(x, \nabla f(x))\}$ is monotone

$(\Rightarrow)$

$$\left.\begin{array}{l} f(y) \geq f(x) + \nabla f(x)^T (y-x) \\ f(x) \geq f(y) + \nabla f(y)^T (x-y) \end{array}\right] \quad [\nabla f(x) - \nabla f(y)]^T [y-x] \leq 0$$

$(\Leftarrow)$ $\forall x, y$, let $\Phi(t) = f(x+t(y-x))$ $t \in [0,1]$

$$\Phi'(t) = \nabla f(x+t(y-x))^T (y-x)$$
$$\Phi'(0) = \nabla f(x)^T (y-x)$$

$$\begin{aligned} \Phi'(t) - \Phi'(0) &= [\nabla f(x+t(y-x)) - \nabla f(x)]^T [y-x] \\ &= [\nabla f(x+t(y-x)) - \nabla f(x)]\, \tfrac{1}{t} \cdot [x+t(y-x)-x] \\ &\geq 0 \text{ since } \nabla f \text{ is monotone} \end{aligned}$$

$$f(y) - f(x) = \int_0^1 \Phi'(t)\, dt \geq \int_0^1 \Phi'(0)\, dt = \nabla f(x)^T (y-x)$$

eg. similarly $\quad f$ convex $\Rightarrow \{(x,y) : y \in \partial f(x)\}$ is monotone

Definition:

A monotone relation C is maximally monotone if there's no other monotone relation containing C

eg. ∂f is maximally monotone if f is convex and closed

- Proximal Operator

Definition: (Proximal Operator)

the proximal operator is defined as

$$\mathrm{prox}(x;th) = \arg\min_u \left\{ h(u) + \frac{1}{2t}\|u-x\|_2^2 \right\}$$

Eg. $h(u) = \|u\|_1$

$$\mathrm{prox}(x;th) = \arg\min_u \left\{ \|u\|_1 + \frac{1}{2t}\|x-u\|_2^2 \right\}$$

$$\mathrm{prox}(x;th) = \arg\min_u \left\{ |u_i| + \frac{1}{2t}(x_i-u_i)^2 \right\} \qquad 0 \in \partial|u_i^*| + \frac{1}{t}(u_i^*-x_i)$$

if $u_i^* > 0$, $0 = 1 + \frac{1}{t}(u_i^*-x_i)$ $\quad u_i^* = x_i - t > 0$ $\quad x_i > t$

if $u_i^* < 0$, $0 = -1 + \frac{1}{t}(u_i^*-x_i)$ $\quad u_i^* = x_i + t < 0$ $\quad x_i < -t$

if $u_i^* = 0$, $\frac{1}{t}(u_i^*-x_i) \in [-1,1]$ $\quad u_i^* = 0$ $\quad x_i \in [-t,t]$

$$\mathrm{prox}(x;th)_i = \begin{cases} x_i - t & \text{if } x_i > 0 \\ x_i + t & \text{if } x_i < -t \\ 0 & \text{if } x_i \in [-t,t] \end{cases}$$

Eg. $h(x) = \frac{1}{2}x^T P x + q^T x + r$

$$\mathrm{prox}(x;th) = \arg\min_u \left\{ \frac{1}{2}u^T P u + q^T u + \frac{1}{2t}\|x-u\|_2^2 \right\}$$

$$\nabla u = Pu + q + \frac{1}{t}(u-x) := 0$$

$$u^* = (tP+I)^{-1}(x-tq)$$

Eg. Indicator Function $\quad h(u) = I_C(u) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$

$$\mathrm{prox}(x;th) = \arg\min_u \left\{ h(u) + \frac{1}{2}\|x-u\|_2^2 \right\}$$

$$= \arg\min_{u\in C} \frac{1}{2}\|x-u\|_2^2$$

$$= \mathrm{proj}_C(x)$$

Eg. $h(X) = \langle C, X\rangle - \alpha \cdot \log\det(X)$

$$\mathrm{prox}(X;th) = \arg\min_P \left\{ \langle C,P\rangle - \alpha\log\det(P) + \frac{1}{2t}\|X-P\|_F^2 \right\}$$

$$\nabla P = C - \alpha P^{-1} + \frac{1}{t}(P-X) := 0$$

$$(\alpha P^{-1} - \frac{1}{t}P) = C - X := \sum_i \lambda_i u_i u_i^T \qquad \text{let } P = \sum_i \sigma_i u_i u_i^T$$

$$\alpha\frac{1}{\sigma_i} - \frac{1}{t}\sigma_i = \lambda_i$$

$$\sigma_i = \sqrt{\alpha t + \frac{(\lambda_i t)^2}{4}} - \frac{\lambda_i t}{2}$$

Theorem:

let $\begin{cases} f \text{ be a closed convex function} \\ \partial f \text{ be the sub-differential} \end{cases}$

then the proximal operator is the resolvent of ∂f

i.e. $(I+\lambda\partial f)^{-1} = \text{prox}(\cdot\,;\lambda f)$

$$(u,v) \in (I+\lambda\partial f)^{-1}$$
$$(v,u) \in I+\lambda\partial f$$
$$u \in v + \lambda\partial f(v)$$
$$0 \in \partial f(v) + \tfrac{1}{\lambda}(v-u)$$
$$v = \arg\min_x \left\{ f(x) + \tfrac{1}{2\lambda}\|x-u\|_2^2 \right\} = \text{prox}(x;\lambda f)$$

- Moreau Envelop

Definition: (Moreau Envelop)

$$M(x;th) = \inf_u \left\{ h(u) + \tfrac{1}{2t}\|x-u\|_2^2 \right\}$$

(when h is convex, $h(u) + \frac{1}{2t}\|x-u\|_2^2$ is strongly convex)

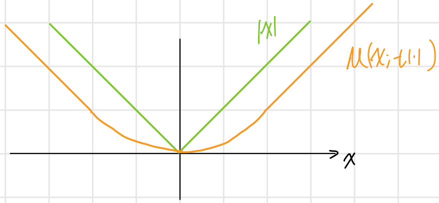
Eg $h(u) = |u|$

$$\text{prox}(x;th) = \begin{cases} x-t & \text{if } x>t \\ x+t & \text{if } x<-t \\ 0 & \text{if } x\in[-t,t] \end{cases}$$

$M(x;th) = |x-t| + \frac{1}{2t}|x-x+t|^2 = x - \frac{t}{2}$

$M(x;th) = |x+t| + \frac{1}{2t}|x-x-t|^2 = x + \frac{t}{2}$

$M(x;th) = |0| + \frac{1}{2t}|x-0|^2 = \frac{1}{2t}x^2$



Theorem:

Suppose g is convex, $x^* = \arg\min_x g(x) \Leftrightarrow x^* = \text{prox}(x^*;tg)\ \forall t>0$

$(\Rightarrow)$ $x^* = \arg\min_x g(x)$

$$g(x^*) \le g(x)\ \forall x$$
$$g(x^*) + \tfrac{1}{2t}\|x^*-x^*\|^2 \le g(x) + \tfrac{1}{2t}\|x^*-x\|^2\ \forall x, t>0$$
$$x^* = \text{prox}(x^*, tg)\ \forall t>0$$

$(\Leftarrow)$ $x^* = \arg\min_u \left\{ g(u) + \frac{1}{2t}\|u-x^*\|^2 \right\}\ \forall t>0$

$$0 \in \partial g(x^*) + \tfrac{1}{t}(x^*-x^*)\quad \forall t>0$$
$$0 \in \partial g(x^*)$$

Theorem:

$M(x;tg) = \inf_u \{ g(u) + \frac{1}{2t}\|u-x\|^2 \}$ is differentiable

$\nabla_x M(x;tg) = \frac{1}{t}(x - \mathrm{prox}(x;tg))$

$\forall x \in \mathbb{R}^n \quad u^* = \mathrm{prox}(x;tg) \quad z = \frac{1}{t}(x-u^*)$

let $r(u) = M(x+u;tg) - M(x;tg) - u^T z$ convex in u

$$= \inf_v \{ g(v) + \frac{1}{2t}\|v-x-u\|^2 \} - M(x;tg) - u^T z$$

$$\leq g(u^*) + \frac{1}{2t}\|u^*-x-u\|^2 - g(u^*) - \frac{1}{2t}\|u^*-x\|^2 - u^T z$$

$$= \frac{1}{2t}\left[\|u\|^2 - 2u^T(u^*-x)\right] - u^T z$$

$$= \frac{1}{2t}\|u\|^2$$

$r(0) = 0$

$r(0) = r(\frac{u}{2} - \frac{u}{2}) \leq \frac{1}{2}r(u) + \frac{1}{2}r(-u)$

$r(u) \geq -\frac{1}{2}r(-u) \geq -\frac{1}{2t}\|u\|^2$

$\Rightarrow -\frac{1}{2t}\|u\|^2 \leq r(u) \leq \frac{1}{2t}\|u\|^2$

$$\lim_{\theta \to 0} \frac{|M(x+\theta u;tg) - M(x;tg) - (\theta u)^T z|}{\theta} = \lim_{\theta \to 0} \frac{\frac{1}{2t}\theta^2\|u\|^2}{\theta} = 0$$

thus $M(x;tg)$ is differentiable and $\frac{1}{t}(x-u^*) = \frac{1}{t}(x - \mathrm{prox}(x;tg))$ is the gradient

Theorem:

$g(\cdot)$ and $M(\cdot;tg)$ has the same minimizer

$x^* = \arg\min_x g(x) \Leftrightarrow x^* = \mathrm{prox}(x^*;tg) \ \forall t>0 \Leftrightarrow \nabla M(x^*;tg) = 0 \ \forall t>0$

Theorem: (Moreau Decomposition)

$x = \mathrm{prox}(x;f) + \mathrm{prox}(x;f^*)$

primal viewpoint

$M(x;tf) = \inf_u \{ f(u) + \frac{1}{2t}\|x-u\|_2^2 \}$

$$= \frac{1}{2t}\|x\|_2^2 + \inf_u \{ f(u) + \frac{1}{2t}\|u\|^2 - \frac{1}{t}x^T u \} \quad \Rightarrow \nabla_x M(x;tf) = \frac{1}{t}(x - \mathrm{prox}(x;tf))$$

$$= \frac{1}{2t}\|x\|_2^2 - \frac{1}{t}\sup_u \{ x^T u - (tf(u) + \frac{1}{2}\|u\|^2) \}$$

$$= \frac{1}{2t}\|x\|_2^2 - \frac{1}{t}\left(\frac{1}{2}\|\cdot\|^2 + tf(\cdot)\right)^*(x)$$

Dual viewpoint

$$M(x;tf) = \inf_u \{ f(u) + \frac{1}{2t}\|x-u\|^2 \}$$

$$= \sup_v \{ -f^*(v) - \frac{t}{2}\|v\|^2 + v^T x \}$$

$$= (f^* + \frac{t}{2}\|\cdot\|^2)^*(x)$$

$$\Downarrow$$

$$\begin{cases} M(x;tf) = \frac{1}{2t}\|x\|_2^2 - \frac{1}{t}(\|\cdot\|^2 + tf(\cdot))^*(x) \\ M(x;tf) = (f^* + \frac{t}{2}\|\cdot\|^2)^*(x) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} M(x;f) = \frac{1}{2}\|x\|_2^2 - (\|\cdot\|^2 + f(\cdot))^*(x) \\ M(x;f) = (f^* + \frac{1}{2}\|\cdot\|^2)^*(x) \end{cases}$$

$$\frac{1}{2}\|x\|_2^2 = (f^* + \frac{1}{2}\|\cdot\|^2)^*(x) + (\|\cdot\|^2 + f(\cdot))^*(x)$$

$$= \sup_u \{ u^T x - f^*(u) - \frac{1}{2}\|u\|^2 \} + \sup_u \{ u^T x - f(u) - \frac{1}{2}\|u\|^2 \}$$

$$= -\inf_u \{ f^*(u) + \frac{1}{2}\|u\|^2 - u^T x \} - \inf_u \{ f(u) + \frac{1}{2}\|u\|^2 - u^T x \}$$

$$\Downarrow \nabla_x$$

$$x = -(x - \mathrm{prox}(x;f^*)) - (x - \mathrm{prox}(x;f))$$

$$x = \mathrm{prox}(x;f^*) + \mathrm{prox}(x;f)$$

$$\min_{u,z} \; f(u) + \frac{1}{2t}\|z\|^2$$

$$\text{s.t.} \quad x - u = z \quad ; v$$

$$\mathcal{L}(u,z,v) = f(u) + \frac{1}{2t}\|z\|^2 + v^T(x-u-z)$$

$$= f(u) - v^T u + \frac{1}{2t}\|z\|^2 - v^T z + v^T x$$

$$g(v) = \inf_{u,z} \mathcal{L}(u,z,v)$$

$$= \inf_u \{ f(u) - v^T u \} + \inf_z \{ \frac{1}{2t}\|z\|^2 - v^T z \} + v^T x$$

$$= -f^*(v) - \frac{t}{2}\|v\|_2^2 + v^T x$$

- (Firmly) Non-Expansive Operator

Definition: (Non-Expansive)

An operator T is non-expansive if $\forall\ (x,y)\ (u,v) \in T \quad \|y-v\| \le \|x-u\|$

Definition: (Firmly Non-Expansive)

An operator T is firmly non-expansive if $\forall (x,y)\ (u,v) \in T \quad \langle x-u, y-v\rangle \ge \|v-y\|^2$

Theorem:

Firmly non-expansive $\Rightarrow$ non-expansive

$$\|x-u\|\,\|y-v\| \ge \langle x-u, y-v\rangle \ge \|y-v\|^2 \quad \Rightarrow \quad \|x-u\| \ge \|v-y\|$$

Theorem:

If F is monotone, $\lambda > 0$

the resolvent $(I+\lambda F)^{-1}$ is single-valued and firmly non-expansive

let $(x,y)\ (u,v) \in (I+\lambda F)^{-1}$

$(y,x), (v,u) \in I+\lambda F$

$x \in y+\lambda F(y) \quad u \in v+\lambda F(v)$

$\langle x-u, y-v\rangle \in \langle y-v+\lambda(F(y)-F(v)), y-v\rangle$

$= \langle y-v, y-v\rangle + \lambda\langle F(y)-F(v), y-v\rangle$

F monotone $\Rightarrow \langle F(y)-F(v), y-v\rangle \ge 0$

$\langle x-u, y-v\rangle \ge \|y-v\|^2$

Eg: proximal operator

∂f is monotone $\Rightarrow (I+\lambda\partial f)^{-1}$ is single-valued

$\text{prox}(x;\lambda f) = \arg\min_u \{f(u) + \frac{1}{2\lambda}\|x-u\|^2\}$ is single-valued since $f(u)+\frac{1}{2\lambda}\|x-u\|^2$ is strongly convex

Theorem: (Cayley Operator)

let F be a monotone operator, $R_F = (I+\lambda F)^{-1}$ be the resolvent

the Cayley operator $C_A = 2R_A - I$ is non-expansive

let $(x,y)\ (u,v) \in (I+\lambda A)^{-1} = R_A$

$(x, 2y-x)\ (u, 2v-u) \in C_A$

$$\|2y-x-2v+u\|^2 = \|2(y-v)-(x-u)\|^2$$

$$= \underbrace{4\|y-v\|^2 - 4\langle y-v, x-u\rangle}_{\le 0} + \|x-u\|^2$$

≤ 0 since R_A is non-expansive.

Theorem: (Averaged operator)

let F be a non-expansive operator, $0<\theta<1$

$$\breve{F} = \theta F + (1-\theta) I$$

then

① F and $\breve{F}$ have the same fixed point

② the fixed point iteration

$$x^+ = \breve{F}(x)$$

converges with $\|Fx^{(k)} - x^{(k)}\|^2 \le \frac{1}{(k+1)\theta(1-\theta)} \|x^{(0)} - x^*\|^2$

① $x^* = \breve{F}(x^*) = \theta F(x^*) + (1-\theta) x^*$

$$\Updownarrow$$

$x^* = F(x^*)$

② $\|(1-\theta)a + \theta b\|^2 = (1-\theta)\|a\| + \theta\|b\|^2 - \theta(1-\theta)\|a-b\|^2$

$$\|x^{(k+1)} - x^*\|^2 = \|\breve{F}(x^{(k)}) - x^*\|^2$$

$$= \|\theta F(x^{(k)}) + (1-\theta)x^{(k)} - x^*\|^2$$

$$= \|\theta(F(x^{(k)}) - x^*) + (1-\theta)(x^{(k)} - x^*)\|^2$$

$$= \theta\|F(x^{(k)}) - x^*\|^2 + (1-\theta)\|x^{(k)} - x^*\|^2 - \theta(1-\theta)\|F(x^{(k)}) - x^{(k)}\|^2$$

$$\le \|x^{(k)} - x^*\|^2 - \theta(1-\theta)\|F(x^{(k)}) - x^{(k)}\|^2$$ decaying

$$\|F(x^{(k)}) - x^{(k)}\|^2 = \left\|\frac{1}{\theta}\breve{F}(x^{(k)}) - \frac{1-\theta}{\theta}x^{(k)} - x^{(k)}\right\|^2$$

$$= \frac{1}{\theta^2}\|\breve{F}(x^{(k)}) - x^{(k)}\|^2$$

$$= \frac{1}{\theta^2}\|x^{(k+1)} - x^{(k)}\|^2$$ decaying

$$\|x^{(k+1)}-x^*\|^2 - \|x^{(k)}-x^*\| \le -\theta(1-\theta)\,\|F(x^{(k)})-x^{(k)}\|^2$$
$$\|x^{(k)}-x^*\|^2 - \|x^{(0)}-x^*\| \le -\theta(1-\theta)\sum_{i=0}^{k-1}\|F(x^{(i)})-x^{(i)}\|^2$$
$$\sum_{i=0}^{k-1}\|F(x^{(i)})-x^*\|^2 \le \frac{1}{\theta(1-\theta)}\|x^{(0)}-x^*\|^2$$
$$k\cdot \min_{i=1\cdots k}\|F(x^{(i)})-x^*\|^2 \le \frac{1}{\theta(1-\theta)}\|x^{(0)}-x^*\|^2$$
$$\|F(x^{(k)})-x^*\|^2 \le \frac{1}{k\theta(1-\theta)}\|x^{(0)}-x^*\|^2$$

Eg. $\mathrm{prox}(x;\lambda f)$ converges to a fixed point for convex function

∂f is monotone $\Rightarrow$ $(I+\lambda\partial f)^{-1} = \frac{1}{2}\underbrace{C_{\partial f}}_{\text{non-expansive}} + \frac{1}{2}I$

- Proximal Point Method

$$\min_x f(x)$$

optimality condition: $0 \in \lambda\,\partial f(x)$

$$\Updownarrow$$

$$x^* \in x^* + \lambda\,\partial f(x) = (I+\lambda\partial f)(x^*)$$
$$x^* = (I+\lambda\partial f)^{-1}(x^*)$$

$$\Updownarrow$$

x^* is the fixed point of $(I+\lambda\partial f)^{-1}$

$$\Downarrow$$

$$x^{(k+1)} = (I+\lambda\partial f)^{-1}(x^{(k)}) = \mathrm{prox}(x^{(k)};\lambda f)$$

- Proximal Gradient Method

$\min_x f(x)+g(x)$ $\quad$ f differentiable, g non-differentiable

optimality condition: $0 \in \lambda(\nabla f(x^*) + \partial g(x^*))$

$$x^* - \lambda\nabla f(x^*) \in x^* + \lambda\,\partial g(x^*)$$
$$(I-\lambda\nabla f)(x^*) \in (I+\lambda\partial g)(x^*)$$
$$x^* = (I+\lambda\partial g)^{-1}\,(I-\lambda\nabla f)(x^*)$$

$I-\lambda\nabla f$ might not be non-expansive. need line search

$$\Downarrow$$

$$x^{(k+1)} = (I+\lambda\partial g)^{-1}(I-\lambda\nabla f)(x^{(k)}) = \mathrm{prox}(x^{(k)}-\lambda\nabla f(x^{(k)});\lambda g)$$

$$x^{(k+1)} = \text{prox}(x^{(k)} - t\nabla f(x^{(k)}); \lambda g)$$
$$= x^{(k)} - t \cdot \frac{1}{t}\left[x^{(k)} - \text{prox}(x^{(k)} - t\nabla f(x^{(k)}); \lambda g)\right]$$
$$= x^{(k)} - t \cdot G_t(x^{(k)})$$

Assume:

f is convex, differentiable, and L-smooth
g is convex, non-differentiable
minimum of $f+g$ is finite and attained at x^*

L-smooth: $f(y) \le \nabla f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|x-y\|_2^2$

$$f(x^+) = f(x - tG_t(x))$$
$$\le f(x) - t\nabla f(x)^T G_t(x) + \frac{Lt^2}{2}\|G_t(x)\|_2^2$$
$$:\le f(x) - t\nabla f(x)^T G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2$$

when $t < 1/L$, the inequality is satisfied

$$x^+ = \text{prox}(x - t\nabla f(x); tg)$$
$$= \arg\min_u \left\{ g(u) + \frac{1}{2t}\|x - t\nabla f(x) - u\|_2^2 \right\}$$

$$0 \in \partial g(x^+) + \frac{1}{t}(x^+ - x + t\nabla f(x)) = \partial g(x^+) + \nabla f(x) - G_t(x)$$

$$f(x^+) + g(x^+) \le f(x) - t\nabla f(x)^T G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2 + g(x - tG_t(x))$$

$$f(z) \ge f(x) + \nabla f(x)^T(z-x) \quad \forall z$$
$$g(z) \ge g(x - tG_t(x)) + u^T(z - x + tG_t(x)) \quad \forall z, \quad u \in \partial g(x - tG_t(x))$$

$$\le f(z) - \nabla f(x)^T(z-x) - t\nabla f(x)^T G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2$$
$$+ g(z) - [G_t(x) - \nabla f(x)]^T[z - x + tG_t(x)]$$

$$= f(z) + g(z) + \nabla f(x)^T[(x-z) - tG_t(x)] - [G_t(x) - \nabla f(x)]^T[z - x + tG_t(x)] + \frac{t}{2}\|G_t(x)\|_2^2$$

$$= f(z) + g(z) - G_t(x)^T(z-x) - \frac{t}{2}\|G_t(x)\|_2^2$$

$$f(x^+) + g(x^+) \le f(x) + g(x) - \frac{t}{2}\|G_t(x)\|_2^2$$

$$p(x^+) - p(z) \le G_t(x)^T(x-z) - \frac{t}{2}\|G_t(x)\|_2^2 = \frac{1}{2t}\left(\|x-z\|_2^2 - \|x - z - tG_t(x)\|_2^2\right)$$
$$p(x^+) - p^* \le \frac{1}{2t}\left(\|x - x^*\| - \|x^+ - x^*\|_2^2\right)$$

Given $x \in \text{dom} f \cap \text{dom} g$

for $k = 1, \cdots$

compute $\nabla f(x)$

$t = 1$

$x^+ = \text{prox}(x - t\nabla f(x); tg)$

while $f(x^+) \geq f(x) - t\nabla f(x)^T G_t(x) + \frac{t}{2}\|G_t(x)\|_2^2$

$= f(x) - \nabla f(x)^T (x^+ - x) + \frac{1}{2t}\|x^+ - x\|_2^2$

$t := \beta t$

$x^+ = \text{prox}(x - t\nabla f(x); tg)$

Quit when $\|G_t(x)\| = \frac{1}{t}\|x^+ - x\| \leq \epsilon$

- Augmented Lagrangian

$$\min_x f(x)$$
$$\text{st } Ax = b$$

let $F = \{(v, b - Ax) : x \in \arg\min_u f(u) + v^T(Au - b)\}$

optimality condition : $\begin{cases} Ax^* = b \\ x^* = \arg\min_x f(x) + v^{*T}(Ax - b) \end{cases}$

$\Updownarrow$

$0 \in F(v^*)$

$\Updownarrow$

$v^* \in v^* + \lambda F(v^*) = (I + \lambda F)(v^*)$

$\Updownarrow$

$v^* = (I + \lambda F)^{-1}(v^*)$

$\Downarrow$

$v^{(k+1)} = (I + \lambda F)^{-1}(v^{k})$

let $(v, z) \in (I + \lambda F)^{-1}$ $\quad (z, v) \in I + \lambda F$

$v \in z + \lambda F(z)$

$v = z + \lambda(b - Ax)$ for some $x \in \arg\min_u f(u) + z^T(Au - b)$

$\Updownarrow$

$\begin{cases} v = z + \lambda(b - Ax) \\ 0 \in \partial f(x) + A^T z = \partial f(x) + A^T[v + \lambda(Ax - b)] \end{cases}$

$\Downarrow$

$\begin{cases} x = \arg\min_u f(u) + v^T(Au - b) + \frac{\lambda}{2}\|Au - b\|_2^2 \\ z = v + \lambda(Ax - b) \end{cases}$ $\Rightarrow$

$x^+ = \arg\min_u f(u) + v^T(Au - b) + \frac{\lambda}{2}\|Au - b\|_2^2$

$v^+ = v + \lambda(Ax^+ - b)$

F is a monotone operator

let $(v_1, b - Ax_1)$ $(v_2, b - Ax_2) \in F$

$0 \in \partial f(x_1) + A^T v_1 \quad 0 \in \partial f(x_2) + A^T v_2$

$(v_1 - v_2)^T(b - Ax_1 - b + Ax_2) = A^T(v_1 - v_2)^T(x_2 - x_1)$

$\in [\partial f(x_1) - \partial f(x_2)]^T(x_2 - x_1)$

≥ 0

since f is convex, ∂f is monotone

Operator Splitting

Theorem:

let A, B be maximally monotone, $A+B$ be maximally monotone

$$0 \in A(x) + B(x) \iff \begin{cases} (2R_A - I)(2R_B - I)z = z \\ x = R_B(z) \end{cases}$$

where $R_A = (I+\lambda A)^{-1}$ $R_B = (I+\lambda B)^{-1}$ are resolvents of A, B

let $x = R_B(z)$ $\quad \tilde{z} = (2R_B - I)z = 2x - z$

$\tilde{x} = R_A(\tilde{z})$ $\quad z = (2R_A - I)(2R_B - I)z = 2\tilde{x} - \tilde{z}$

$$\left.\begin{array}{l} \tilde{z} = 2x - z \\ z = 2\tilde{x} - \tilde{z} \end{array}\right] \begin{array}{l} x = \tilde{x} \\ z + \tilde{z} = 2x \end{array}$$

$$\left.\begin{array}{l} x = R_B(z) = (I+\lambda B)^{-1}(z) \Rightarrow z \in (I+\lambda B)(x) = x + \lambda B(x) \\ \tilde{x} = R_A(\tilde{z}) = (I+\lambda A)^{-1}(\tilde{z}) \Rightarrow \tilde{z} \in (I+\lambda A)(\tilde{x}) = \tilde{x} + \lambda A(\tilde{x}) \end{array}\right] \begin{array}{l} z + \tilde{z} \in 2x + \lambda(A(\tilde{x}) + B(x)) \\ 0 \in A(x) + B(x) \end{array}$$

we saw that $C_A = (2R_A - I)$ $C_B = (2R_B - I)$ are non-expansive

Peaceman-Rachford Splitting: $\quad z^+ = (2R_A - I)(2R_B - I)(z)$

Douglas-Rachford Splitting: $\quad z^+ = \left[\frac{1}{2}I + \frac{1}{2}(C_A \circ C_B)\right](z)$

- Douglas-Rachford Splitting & Update Rules

$$\min_x f(x) + g(x) \iff 0 \in \partial f(x) + \partial g(x)$$

the Douglas-Rachford Splitting is $\quad z^+ = \left[\frac{1}{2}I + \frac{1}{2}(2R_{\lambda f} - I)(2R_{\lambda g} - I)\right](z)$

$$\begin{cases} y_1^{(k)} = 2\,\mathrm{prox}(z^{(k)}; \lambda g) - z^{(k)} \\ y_2^{(k)} = 2\,\mathrm{prox}(y_1^{(k)}; \lambda f) - y_1^{(k)} \\ z^{(k+1)} = \frac{1}{2}z^{(k)} + \frac{1}{2}\left[2\,\mathrm{prox}(y_1^{(k)}; \lambda f) - 2\,\mathrm{prox}(z^{(k)}; \lambda g) + z^{(k)}\right] \\ \quad = z^{(k)} + \mathrm{prox}(y_1^{(k)}; \lambda f) - \mathrm{prox}(z^{(k)}; \lambda g) \end{cases}$$

$\Downarrow$ let $y^{(k)} = \mathrm{prox}(z^{(k)}; \lambda g)$

then $y_1^{(k)} = 2y^{(k)} - z^{(k)}$

$$
\begin{cases}
y^{(k)} = \mathrm{prox}(z^{(k)}; \lambda g) \\
z^{(k+1)} = z^{(k)} + \mathrm{prox}(2y^{(k)} - z^{(k)}; \lambda g) - y^{(k)}
\end{cases}
$$

this also work as the "post-processing" $x = R_B(z)$ in the operator splitting

thus y is the result we should return

$\Downarrow$ shift iteration

$$
\begin{cases}
y^{(0)} = \mathrm{prox}(z^{(0)}; \lambda g) \\
z^{(k+1)} = \overbrace{\boxed{z^{(k)} - y^{(k)}}}^{-w^{(k)}} + \underbrace{\boxed{\mathrm{prox}(2y^{(k)} - z^{(k)}; \lambda g)}}_{u^{(k+1)}} \\
\boxed{y^{(k+1)} = \mathrm{prox}(z^{(k+1)}; \lambda g)}
\end{cases}
$$

$\Downarrow$ let $w^{(k)} = y^{(k)} - z^{(k)}$

then $z^{(k+1)} = u^{(k+1)} - w^{(k)}$

$$
\begin{cases}
y^{(0)} = \mathrm{prox}(z^{(0)}; \lambda g) \\
w^{(0)} = y^{(0)} - z^{(0)} \\
\boxed{u^{(k+1)} = \mathrm{prox}(y^{(k)} + w^{(k)}; \lambda f)} \\
\boxed{y^{(k+1)} = \mathrm{prox}(u^{(k+1)} - w^{(k)}; \lambda g)} \\
w^{(k+1)} = y^{(k+1)} - z^{(k+1)} = y^{(k+1)} + w^{(k)} - u^{(k+1)}
\end{cases}
$$

the update rule for Douglas-Rachford Splitting

$$
\begin{cases}
y^{(0)} = \mathrm{prox}(z^{(0)}; \lambda g) \\
w^{(0)} = y^{(0)} - z^{(0)} \\
u^{(k+1)} = \mathrm{prox}(y^{(k)} + w^{(k)}; \lambda f) \\
y^{(k+1)} = \mathrm{prox}(u^{(k+1)} - w^{(k)}; \lambda g) \\
w^{(k+1)} = y^{(k+1)} - u^{(k+1)} + w^{(k)}
\end{cases}
$$

- ADMM: Douglas-Rachford Splitting in Dual

Theorem:

let $\tilde{f}(\nu) = f^*(-A^T\nu) + b^T\nu$

then the solution to $\text{prox}(z; \lambda\tilde{f})$ is given by

$$z + \lambda(Ax^* - b)$$

where x^* is the solution to

$$\min_x \; f(x) + z^T(Ax - b) + \frac{\lambda}{2}\|Ax - b\|_2^2$$

$$\min_x \; f(x) + z^T(Ax-b) + \frac{\lambda}{2}\|Ax-b\|_2^2$$

$$\Downarrow$$

$$\min_x \; f(x) + \frac{\lambda}{2}\|Ax - b + \frac{1}{\lambda}z\|_2^2$$

$$\Downarrow$$

$$\min_{x,y} \; f(x) + \frac{\lambda}{2}\|y\|_2^2$$

$$\text{s.t.} \quad Ax - b + \frac{1}{\lambda}z = y \; : \; \nu$$

$$\mathcal{L}(x,y,\nu) = f(x) + \frac{\lambda}{2}\|y\|_2^2 + \nu^T\left[Ax - b + \frac{1}{\lambda}z - y\right]$$

$$= f(x) + (A^T\nu)^T x + \frac{\lambda}{2}\|y\|_2^2 - \nu^T y + \frac{1}{\lambda}\nu^T z$$

$$\begin{cases} 0 \in \partial f(x^*) + A^T\nu^* \Rightarrow x^* \in \arg\max_x \{(-A^T\nu^*)^T x - f(x)\} \Rightarrow x^* \in \partial f^*(-A^T\nu^*) \\ \left.\begin{array}{l} \lambda y^* - \nu^* = 0 \\ Ax^* - b + \frac{1}{\lambda}z = y^* \end{array}\right\} \Rightarrow Ax^* - b + \frac{1}{\lambda}(z - \nu^*) = 0 \end{cases}$$

$$0 \in A\,\partial f^*(-A^T\nu^*) - b + \frac{1}{\lambda}(z - \nu^*) = 0$$

$$0 \in -A\,\partial f^*(-A^T\nu^*) + b + \frac{1}{\lambda}(\nu^* - z) = 0$$

$$\nu^* \in \arg\min_\nu \left\{ f^*(-A^T\nu) + b^T\nu + \frac{1}{2\lambda}\|\nu - z\|_2^2 \right\}$$

$$= \arg\min_\nu \left\{ \tilde{f}(\nu) + \frac{1}{2\lambda}\|\nu - z\|_2^2 \right\}$$

$$= \text{prox}(z; \lambda\tilde{f})$$

$$\nu^* = \lambda y^* = z + \lambda(Ax^* - b)$$

$$\min_{x_1, x_2} \ f_1(x_1) + f_2(x_2)$$

$$\text{s.t.} \quad A_1 x_1 + A_2 x_2 = b \ : v$$

$$\mathcal{L}(x_1, x_2, v) = f_1(x_1) + f_2(x_2) + v^T(A_1 x_1 + A_2 x_2 - b)$$
$$= f_1(x_1) + (A_1^T v)^T x_1 + f_2(x_2) + (A_2^T v)^T x_2 - b^T v$$

$$g(v) = \inf_{x_1} \{ f_1(x_1) + (A_1^T v)^T x_1 \} + \inf_{x_2} \{ f_2(x_2) + (A_2^T v)^T x_2 \} - b^T v$$
$$= -f_1^*(-A_1^T v) - b^T v - f_2^*(-A_2^T v)$$

dual problem:

$$\min_v \ \underbrace{f_1^*(-A_1^T v) + b^T v}_{\tilde{f}_1(v)} + \underbrace{f_2^*(-A_2^T v)}_{\tilde{f}_2(v)}$$

Douglas-Rachford Splitting

$$\begin{cases} y^{(0)} = \mathrm{prox}(z^{(0)}; \lambda g) \\ w^{(0)} = y^{(0)} - z^{(0)} \\ u^{(k+1)} = \mathrm{prox}(y^{(k)} + w^{(k)}; \lambda f) \\ y^{(k+1)} = \mathrm{prox}(u^{(k+1)} - w^{(k)}; \lambda g) \\ w^{(k+1)} = y^{(k+1)} - u^{(k+1)} + w^{(k)} \end{cases}$$

$$y^{(0)} = \mathrm{prox}(z^{(0)}; \lambda \tilde{f}_2)$$

Solve $x_2^{(0)} = f_2(x) + z^{(0)T} A_2 x_2 + \frac{\lambda}{2} \| A_2 x_2 \|_2^2$

$$y^{(0)} = z^{(0)} + \lambda (A_2 x_2^{(0)})$$

$$w^{(0)} = y^{(0)} - z^{(0)} = \lambda A_2 x_2^{(0)}$$

$$u^{(k+1)} = \mathrm{prox}(y^{(k)} + w^{(k)}; \lambda \tilde{f}_1)$$

Solve $x_1^{(k+1)} = \arg\min_{x_1} f_1(x_1) + (y^{(k)} + w^{(k)})^T (A_1 x_1 - b) + \frac{\lambda}{2} \| A_1 x_1 - b \|_2^2$

$$= \arg\min_{x_1} f_1(x_1) + y^{(k)T}(A_1 x_1 + A_2 x_2^{(k)} - b) + \frac{\lambda}{2} \| A_1 x_1 + \tfrac{1}{\lambda} w^{(k)} - b \|_2^2$$

$$u^{(k+1)} = y^{(k)} + w^{(k)} + \lambda (A_1 x_1^{(k+1)} - b)$$

$$y^{(k+1)} = \mathrm{prox}(u^{(k+1)} - w^{(k)}; \lambda \tilde{f}_2)$$
$$= \mathrm{prox}(y^{(k)} + \lambda (A_1 x_1^{(k+1)} - b); \lambda \tilde{f}_2)$$

Solve $x_2^{(k+1)} = \arg\min_{x_2} f_2(x_2) + (y^{(k)} + \lambda (A_1 x_1^{(k+1)} - b))^T A_2 x_2 + \frac{\lambda}{2} \| A_2 x_2 \|_2^2$

$$= \arg\min_{x_2} f_2(x_2) + y^{(k)T} A_2 x_2 + \frac{\lambda}{2} \| A_1 x_1^{(k+1)} + A_2 x_2 - b \|_2^2$$

$$y^{(k+1)} = y^{(k)} + \lambda (A_1 x_1^{(k+1)} - b) + \lambda A_2 x_2^{(k+1)}$$
$$= y^{(k)} + \lambda (A_1 x_1^{(k+1)} + A_2 x_2^{(k+1)} - b)$$

$$w^{(k+1)} = y^{(k+1)} + w^{(k)} - u^{(k+1)}$$
$$= y^{(k)} + \lambda (A_1 x_1^{(k+1)} + A_2 x_2^{(k+1)} - b) + w^{(k)} - (y^{(k)} + w^{(k)} + \lambda (A_1 x_1^{(k+1)} - b))$$
$$= \lambda A_2 x_2^{(k+1)}$$

$$w^{(k)} = \lambda A_2 x_2^{(k)}$$

$$A(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^T(A_1x_1 + A_2x_2 - b) + \frac{\lambda}{2}\|A_1x_1 + A_2x_2 - b\|_2^2$$

$$= f_1(x_1) + f_2(x_2) + \frac{\lambda}{2}\|A_1x_1 + A_2x_2 - b - \frac{y}{\lambda}\|_2^2 - \frac{1}{2\lambda}\|y\|_2^2$$

randomly initialize $y^{(0)}, x_1^{(0)}, x_2^{(0)}$

$$\left[\begin{aligned} x_1^{(k+1)} &= \arg\min_{x_1} A(x_1, x_2^{(k)}, y^{(k)}) \\ x_2^{(k+1)} &= \arg\min_{x_2} A(x_1^{(k+1)}, x_2, y^{(k)}) \\ y^{(k+1)} &= y^{(k)} + \lambda(A_1x_1^{(k+1)} + A_2x_2^{(k+1)} - b) \end{aligned}\right.$$

- Stopping Criterion

$$\min_{x_1, x_2} f_1(x_1) + f_2(x_2)$$

$$\text{s.t. } A_1x_1 + A_2x_2 = b$$

$$\mathcal{L}(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^T(A_1x_1 + A_2x_2 - b)$$

$$A(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^T(A_1x_1 + A_2x_2 - b) + \frac{\rho}{2}\|A_1x_1 + A_2x_2 - b\|_2^2$$

update rule

$$\left\{\begin{aligned} x_1^{(k+1)} &= \arg\min_{x_1} A(x_1, x_2^{(k)}, y^{(k)}) \\ x_2^{(k+1)} &= \arg\min_{x_2} A(x_1^{(k+1)}, x_2, y^{(k)}) \\ y^{(k+1)} &= y^{(k)} + \lambda(A_1x_1^{(k+1)} + A_2x_2^{(k+1)} - b) \end{aligned}\right.$$

optimality condition

$$\left\{\begin{aligned} & A_1x_1^* + A_2x_2^* = b \\ & 0 \in \partial f_1(x_1^*) + A_1^Ty^* \\ & 0 \in \partial f_2(x_2^*) + A_2^Ty^* \end{aligned}\right.$$

$$\left.\begin{aligned} 0 &\in \partial f_2(x_2^{(k+1)}) + A_2^Ty^{(k)} + \rho A_2^T(A_1x_1^{(k+1)} + A_2x_2^{(k+1)} - b) \\ &= \partial f_2(x_2^{(k+1)}) + A_2^T(y^{(k)} + \rho(A_1x_1^{(k+1)} + A_2x_2^{(k+1)} - b)) \\ &= \partial f_2(x_2^{(k+1)}) + A_2^Ty^{(k+1)} \end{aligned}\right\} \text{ 3rd KKT condition is always satisfied}$$

$$\begin{aligned} 0 &\in \partial f_1(x_1^{(k+1)}) + A_1^Ty^{(k)} + \rho A_1^T(A_1x_1^{(k+1)} + A_2x_2^{(k)} - b) \\ &= \partial f_1(x_1^{(k+1)}) + A_1^T(y^{(k)} + \rho(A_1x_1^{(k+1)} + A_2x_2^{(k)} - b)) \\ &= \partial f_1(x_1^{(k+1)}) + A_1^T(y^{(k)} + \rho(A_1x_1^{(k+1)} + A_2x_2^{(k+1)} - b + A_2x_2^{(k)} - A_2x_2^{(k+1)})) \\ &= \partial f_1(x_1^{(k+1)}) + A_1^Ty^{(k+1)} + \underbrace{\rho A_1^TA_2(x_2^{(k)} - x_2^{(k+1)})}_{\text{dual residual}} \end{aligned}$$

need to check:

① primal residual $A_1 x_1^{(k+1)} + A_2 x_2^{(k+1)} - b$ is small

② dual residual $\rho A_1^T A_2 (x_2^{(k)} - x_2^{(k+1)})$ is small

Check:

$$\| A_1 x_1^{(k+1)} + A_2 x_2^{(k+1)} - b \|_2 \le \epsilon_p = \sqrt{p}\, \epsilon^{abs} + \epsilon^{rel} \max \{ \| A_1 x_1^{(k+1)} \|_2, \| A_2 x_2^{(k+1)} \|, \|b\| \}$$

$$\| \rho A_1^T A_2 (x_2^{(k)} - x_2^{(k+1)}) \|_2 \le \epsilon_d = \sqrt{n}\, \epsilon^{abs} + \epsilon^{rel} \| A_1^T y^{(k+1)} \|_2$$

p is the number of constraints i.e. $A_1 \in \mathbb{R}^{p \times n}$ $\epsilon^{abs} > 0$

n is the dimension of x_1 $\epsilon^{rel} = 10^{-3}$ or 10^{-4}

Eg. ADMM Lasso

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1$$

$$\Downarrow$$

$$\min_{x_1, x_2} \frac{1}{2} \|Ax_1 - b\|_2^2 + \lambda \|x_2\|_1$$

$$\text{s.t. } x_1 - x_2 = 0$$

$$\begin{aligned} A(x_1, x_2, y) &= \frac{1}{2} \|Ax_1 - b\|_2^2 + \lambda \|x_2\|_1 + y^T(x_1 - x_2) + \frac{\rho}{2} \|x_1 - x_2\|_2^2 \\ &= \frac{1}{2} \|Ax_1 - b\|_2^2 + \lambda \|x_2\|_1 + \frac{\rho}{2} \|x_1 - x_2 + \frac{y}{\rho}\|_2^2 + \frac{1}{2\rho} \|y\|_2^2 \end{aligned}$$

randomly initialize $x_1^{(0)}$ $x_2^{(0)}$ $y^{(0)}$

$$x_1^{(k+1)} = \arg\min_{x_1} \frac{1}{2} \|Ax_1 - b\|_2^2 + \frac{\rho}{2} \|x_1 - (x_2^{(k)} - \frac{y^{(k)}}{\rho})\|_2^2$$

$$\nabla_{x_1} = A^T(Ax_1 - b) + \rho(x_1 - x_2^{(k)} + \frac{y^{(k)}}{\rho}) := 0$$

$$(A^TA + \rho I) x = \rho x_2^{(k)} - y^{(k)} + A^T b$$

$$x_1^{(k+1)} = \underbrace{(A^TA + \rho I)^{-1}}_{\text{rank-1 update}} (\rho x_2^{(k)} - y^{(k)} + A^T b)$$

$$\begin{aligned} x_2^{(k+1)} &= \arg\min_{x_2} \|x_2\|_1 + \frac{\rho}{2\lambda} \|x_2 - (x_1^{(k)} + \frac{y^{(k)}}{\rho})\|_2^2 \\ &= \text{prox}(x_1^{(k)} + y^{(k)}/\rho ; \frac{\lambda}{\rho} \|\cdot\|_1) \end{aligned}$$

$$y^{(k+1)} = y^{(k)} + \rho(x_1^{(k+1)} - x_2^{(k+1)})$$

Eg. Constrained optimization

$$\min_x f(x) \quad \text{s.t. } x \in C$$

$$\Downarrow$$

$$\min_{x_1, x_2} f(x_1) + I_C(x_2) \quad \text{s.t. } x_1 - x_2 = 0$$

$$\begin{aligned} A(x_1, x_2, y) &= f(x_1) + I_C(x_2) + y^T(x_1 - x_2) + \frac{\rho}{2}\|x_1 - x_2\|_2^2 \\ &= f(x_1) + I_C(x_2) + \frac{\rho}{2}\|x_1 - x_2 + \frac{y}{\rho}\|_2^2 - \frac{1}{2\rho}\|y\|_2^2 \end{aligned}$$

randomly initialize $x_1^{(0)}$ $x_2^{(0)}$ $y^{(0)}$

$$\begin{aligned} x_1^{(k+1)} &= \arg\min_{x_1} f(x_1) + \frac{\rho}{2}\|x_1 - (x_2^{(k)} - y^{(k)}/\rho)\|_2^2 \\ &= \text{prox}(x_2^{(k)} - y^{(k)}/\rho \,;\, \rho f) \end{aligned}$$

$$\begin{aligned} x_2^{(k+1)} &= \arg\min_{x_2} I_C(x_2) + \frac{\rho}{2}\|x_1^{(k+1)} + y^{(k)}/\rho - x_2\|_2^2 \\ &= \text{proj}_C(x_1^{(k+1)} + y^{(k)}/\rho) \end{aligned}$$

$$y^{(k+1)} = y^{(k)} + \rho(x_1^{(k+1)} - x_2^{(k+1)})$$

Eg. $\min_x f_1(x) + f_2(Ax)$ $\quad$ $\text{prox}(\cdot\,; f_1)$ $\text{prox}(\cdot\,; f_2)$ are easy

$$\Downarrow$$

$$\min_{x_1, x_2, x_3} f_1(x_3) + f_2(x_2)$$

$$\text{s.t. } \begin{bmatrix} A \\ I \end{bmatrix} x_1 = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \quad ; \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$A(x_1, x_2, x_3) = f_1(x_3) + f_2(x_2) + \frac{\rho}{2}\left\| \begin{bmatrix} A \\ I \end{bmatrix} x_1 - \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \frac{1}{\rho}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\|_2^2 + \frac{1}{2\rho}\|y\|_2^2$$

$$x_1^{(k+1)} = \arg\min_{x_1} \frac{\rho}{2}\left\| \begin{bmatrix} A \\ I \end{bmatrix} x_1 - \begin{bmatrix} x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \frac{1}{\rho}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}^{(k)} \right\|_2^2 \qquad \text{least square}$$

$$x_2^{(k+1)}, x_3^{(k+1)} = \underbrace{\arg\min_{x_2, x_3} f_1(x_3) + f(x_2) + \frac{\rho}{2}\left\| \begin{bmatrix} A \\ I \end{bmatrix} x_1^{(k+1)} - \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - \frac{1}{\rho}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}^{(k)} \right\|_2^2}_{\text{separable in } x_2 \text{ and } x_3\text{, just prox operators}}$$

$$y^{(k+1)} = y^{(k)} + \lambda\left(\begin{bmatrix} A \\ I \end{bmatrix} x_1^{(k+1)} - \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}^{(k+1)} \right)$$