Convex Conjugate (Fenchel Dual Function)

$$f^*(x) = \max_y \langle x, y \rangle - f(y)$$

$f(x)$ is convex $\Rightarrow (f^*)^* = f$

$$\begin{cases} \text{Primal Problem (P)} \\ \min_x f(x) + g(x) \end{cases}$$

$f(x)$ & $g(x)$ are convex

$$\Rightarrow \begin{cases} f^*(y) \text{ is convex} \\ g^*(-y) \text{ is convex} \end{cases}$$

$$\Leftrightarrow \begin{cases} \text{Dual Problem (D)} \\ -\min_y f^*(y) + g^*(-y) \end{cases}$$

Ex: $f(x) = \|x\|_1$

$$f^*(x) = \begin{cases} 0, & \|x\|_\infty \leq 1 \\ +\infty, & \|x\|_\infty > 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} \text{Primal Dual (PD)} \\ \min_x \sup_y \langle x, y \rangle - f^*(y) + g(x) \end{cases}$$

Primal Dual Minizers Relation

$$\begin{cases} x^* \in \partial f^*(y^*) \\ y^* \in -\partial g(x^*) \end{cases}$$

In (PD), the cost function is

$$L(x, y) = \langle x, y \rangle - f^*(y) + g(x)$$

$$\frac{\partial L}{\partial x}(x, y) = y + \partial g(x)$$

$$\frac{\partial L}{\partial y}(x, y) = x - \partial f^*(y)$$

$$\boxed{\begin{array}{l} \text{Primal Dual (PD)} \\ \min_x \sup_y \langle x, y \rangle - f^*(y) + g(x) \end{array}}$$

For $\min_x \max_y L(x, y)$, a simple algorithm is to use gradient descent/ascent for approaching saddle point

$$\begin{cases} \dfrac{x_{k+1}-x_k}{\eta} = -\dfrac{\partial L}{\partial x}(x_{k+1}, y_k) \\ \dfrac{y_{k+1}-y_k}{\tau} = \dfrac{\partial L}{\partial y}(x_{k+1}, y_{k+1}) \end{cases}$$

$$\Leftrightarrow \begin{cases} x_{k+1} = x_k - \eta y_k - \eta \partial g(x_{k+1}) \\ y_{k+1} = y_k + \tau x_{k+1} - \tau \partial f^*(y_{k+1}) \end{cases}$$

$\Leftrightarrow$ ① Arrow-Hurwitz Algorithm (1958)

$$\begin{cases} x_{k+1} = (I+\eta\partial g)^{-1}[x_k - \eta y_k] \\ y_{k+1} = (I+\tau\partial f^*)^{-1}[y_k + \tau x_{k+1}] \end{cases}, \quad \eta>0, \tau>0$$

② Primal-Dual-Hybrid-Gradient (PDHG, 2010)

$$\begin{cases} x_{k+1} = (I+\eta\partial g)^{-1}[x_k - \eta y_k] \\ y_{k+1} = (I+\tau\partial f^*)^{-1}[y_k + \tau(2x_{k+1} - x_k)] \end{cases} \quad \eta>0, \tau>0, \quad \eta\tau \le 1$$

$y_{k+1} = \mathrm{Prox}_{f^*}^{1/\eta}[\frac{1}{\eta}(\eta y_k + 2x_{k+1} - x_k)]$

③ PDHG with $\tau = \frac{1}{\eta}$

$\boxed{\text{Moreau-Decomposition}}$ $f(x)$ is convex

$$\mathrm{Prox}_f^{\eta}(x) + \eta\,\mathrm{Prox}_{f^*}^{1/\eta}(x/\eta) = x$$

Let $V_k = x_k - \eta y_k \Leftrightarrow y_k = \dfrac{x_k - V_k}{\eta}$

$$\eta\left(I+\frac{1}{\eta}\partial f^*\right)^{-1}\left(\eta\left[y_k + \frac{1}{\eta}(2x_{k+1}-x_k)\right]/\eta\right)$$

$$= \eta\left[y_k + \frac{1}{\eta}(2x_{k+1} - x_k)\right] - \mathrm{Prox}_f^{\eta}(2x_{k+1} - x_k + \eta y_k)$$

$$= 2x_{k+1} - (x_k - \eta y_k) - \mathrm{Prox}_f^{\eta}(2x_{k+1} - (x_k - \eta y_k))$$

$$\begin{cases} x_{k+1} = (I + \eta\partial g)^{-1}[x_k - \eta y_k] \\ y_{k+1} = (I + \frac{1}{\eta}\partial f^*)^{-1}[y_k + \frac{1}{\eta}(2x_{k+1} - x_k)] \end{cases}$$

$$\Rightarrow \begin{cases} x_{k+1} = \mathrm{Prox}_g^{\eta}(v_k) \\ y_{k+1} = \dfrac{2x_{k+1} - v_k - \mathrm{Prox}_f^{\eta}(2x_{k+1} - v_k)}{\eta} \end{cases}$$

$$\Rightarrow \begin{cases} x_{k+1} = \mathrm{Prox}_g^{\eta}(v_k) \\ \mathrm{Prox}_f^{\eta}(2x_{k+1} - v_k) = x_{k+1} - v_k + (x_{k+1} - \eta y_{k+1}) \end{cases}$$

$$\Rightarrow \begin{cases} x_{k+1} = \mathrm{Prox}_g^{\eta}(v_k) \quad \boxed{v_k = x_k - \eta y_k} \\ v_{k+1} = v_k - x_{k+1} + \mathrm{Prox}_f^{\eta}(2x_{k+1} - v_k) \end{cases}$$

$$\frac{I + R_f R_g}{2}[v_k] = \frac{v_k + R_f[R_g(v_k)]}{2} = \frac{1}{2}v_k + \frac{1}{2}R_f[2x_{k+1} - v_k]$$

$$= \frac{1}{2}v_k + \frac{1}{2}\left[2\mathrm{Prox}_f^{\eta}(2x_{k+1} - v_k) - (2x_{k+1} - v_k)\right]$$

So PDHG ($\eta\tau \le 1$) with $\tau = \frac{1}{\eta}$

is Douglas-Rachford on (P)

thus converges with $\tau = \frac{1}{\eta}$, $\forall \eta > 0$.

④ ADMM (Alternating Direction Method of Multipliers)

since 1975

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

$$\Longleftrightarrow \min f(w) + g(z) \quad \text{s.t. } w = z$$

The Lagrangian for constrained minimization:

$$L(w, z, y) = f(w) + g(z) - \langle \underbrace{y}_{\text{Lagrangian multiplier}}, w - z \rangle$$

Lagrangian multiplier

If (w_*, z_*, y_*) is a solution (saddle point) to

$$\min_{w,z} \max_{y} L(w, z, y)$$

then $w_* = z_*$ is a solution to

$$\min_{w,z} f(w) + g(z) \quad \text{s.t. } w = z$$

Augmented Lagrangian is

$$\mathcal{L}(w, z, y) = f(w) + g(z) - \langle y, w - z \rangle + \frac{\tau}{2} \|w - z\|^2$$

The saddle point of Augmented Lagrangian still gives minimizer to (P)

ADMM:

$$\min_{w,z}\max_{y} \mathcal{L}(w,z,y)$$

$$\begin{cases} z_{k+1} = \underset{z}{\operatorname{argmin}}\ \mathcal{L}(w_k, z, y_k) \\ w_{k+1} = \underset{w}{\operatorname{argmin}}\ \mathcal{L}(w, z_{k+1}, y_k) \\ y_{k+1} = y_k + \sigma \frac{\partial}{\partial y}\mathcal{L}(w_{k+1}, z_{k+1}, y_k) \end{cases}$$

$$\mathcal{L}(w,z,y) = f(w) + g(z) - \langle y, w-z\rangle + \frac{\tau}{2}\|w-z\|^2$$

$$\Leftrightarrow \begin{cases} z_{k+1} = \underset{z}{\operatorname{argmin}}\ g(z) - \langle y_k, w_k - z\rangle + \frac{\tau}{2}\|w_k - z\|^2 \\ w_{k+1} = \underset{w}{\operatorname{argmin}}\ f(w) - \langle y_k, w - z_{k+1}\rangle + \frac{\tau}{2}\|w - z_{k+1}\|^2 \\ y_{k+1} = y_k - \sigma(w_{k+1} - z_{k+1}) \end{cases}$$

$$\Leftrightarrow \begin{cases} z_{k+1} = \underset{z}{\operatorname{argmin}}\ g(z) + \langle y_k, z\rangle + \frac{\tau}{2}\|z\|^2 - \tau\langle w_k, z\rangle \\ w_{k+1} = \underset{w}{\operatorname{argmin}}\ f(w) - \langle y_k, w\rangle + \frac{\tau}{2}\|w\|^2 - \tau\langle w, z_{k+1}\rangle \\ y_{k+1} = y_k - \sigma(w_{k+1} - z_{k+1}) \end{cases}$$

Let $\sigma = \tau$ $\quad \frac{v_{k+1} - y_k}{\eta} = z_{k+1}$ & $\frac{v_k - y_k}{\eta} = w_k$

$$\Leftrightarrow \begin{cases} z_{k+1} = \underset{z}{\operatorname{argmin}}\ g(z) + \frac{\tau}{2}\|z - w_k + \frac{1}{\tau}y_k\|^2 \\ w_{k+1} = \underset{w}{\operatorname{argmin}}\ f(w) + \frac{\tau}{2}\|w - z_{k+1} - \frac{1}{\tau}y_k\|^2 \\ y_{k+1} = y_k - \tau(w_{k+1} - z_{k+1}) \quad \text{(ADMM)} \end{cases}$$

Claim: (ADMM) $\Leftrightarrow$ Douglas-Rachford on (D)

(ADMM):
$$\begin{cases} z_{k+1} = \underset{z}{\operatorname{argmin}}\ g(z) + \frac{\tau}{2}\|z - (w_k - \frac{1}{\tau}y_k)\|^2 \\ w_{k+1} = \underset{w}{\operatorname{argmin}}\ f(w) + \frac{\tau}{2}\|w - (z_{k+1} + \frac{1}{\tau}y_k)\|^2 \\ y_{k+1} = y_k - \tau(w_{k+1} - z_{k+1}) \end{cases}$$

$$\Longleftrightarrow \begin{cases} z_{k+1} = \mathrm{Prox}_g^{1/\tau}(w_k - \frac{1}{\tau} y_k) \\ w_{k+1} = \mathrm{Prox}_f^{1/\tau}(z_{k+1} + \frac{1}{\tau} y_k) \\ y_{k+1} = y_k - \tau(w_{k+1} - z_{k+1}) \end{cases}$$

Consider $\min_y \underbrace{f^*(y)}_{G(y)} + \underbrace{g^*(-y)}_{F(Y)}$ $\qquad \frac{I + R_F R_G}{2}$

$$\boxed{\begin{array}{l}\text{Moreau Decomposition} \\ \qquad \mathrm{Prox}_{f^*}^{\eta}(x) + \eta\, \mathrm{Prox}_f^{1/\eta}\left(\frac{x}{\eta}\right) = x\end{array}}$$

$$\mathrm{Prox}_G^{\eta}(x) = \mathrm{Prox}_{f^*}^{\eta}(x) = x - \eta\, \mathrm{Prox}_f^{1/\eta}\left(\frac{x}{\eta}\right) \qquad F(x) = g^*(-x)$$

$$\mathrm{Prox}_F^{\eta}(x) = \operatorname*{argmin}_u \left[g^*(-u) + \frac{1}{2\eta}\|u - x\|^2 \right] \qquad u = -v$$

$$u^* = -v^* \Leftarrow (=) - \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|-v - x\|^2 \right]$$

$$= - \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|v - (-x)\|^2 \right]$$

$$= - \mathrm{Prox}_{g^*}^{\eta}(-x)$$

$$= \eta\, \mathrm{Prox}_g^{1/\eta}\left(\frac{-x}{\eta}\right) - (-x)$$

$$= \eta\, \mathrm{Prox}_g^{1/\eta}\left(-\frac{x}{\eta}\right) + x$$

DR $\begin{cases} v_{k+1} = \frac{I + R_F R_G}{2}(v_k) \\ y_k = \mathrm{Prox}_G(v_k) \end{cases}$

$$\begin{cases} y = \mathrm{Prox}^{\eta}_{G}(v) = v - \eta\, \mathrm{Prox}^{1/\eta}_{f}(v/\eta) \\ \dfrac{I + R_F R_G}{2}(v) = v - y + \mathrm{Prox}_F[2y - v] \end{cases}$$

$$= v - y + (2y - v) + \eta\, \mathrm{Prox}^{1/\eta}_{g}\left(-\frac{2y}{\eta} + \frac{v}{\eta}\right)$$

$$= y + \eta\, \mathrm{Prox}^{1/\eta}_{g}\left(-\frac{2y}{\eta} + \frac{v}{\eta}\right)$$

$$\begin{cases} v_{k+1} = y_k + \eta\, \mathrm{Prox}^{1/\eta}_{g}\left(-\frac{2y_k}{\eta} + \frac{v_k}{\eta}\right) \\ y_{k+1} = v_{k+1} - \eta\, \mathrm{Prox}^{1/\eta}_{f}(v_{k+1}/\eta) \end{cases}$$

$y_k \to y_*$

$\min_y f^*(y) + g^*(-y)$

Let $\dfrac{v_{k+1} - y_k}{\eta} = z_{k+1}$ & $\dfrac{v_k - y_k}{\eta} = w_k$

Then

$$\begin{cases} z_{k+1} = \mathrm{Prox}^{1/\eta}_{g}\left(w_k - \frac{1}{\eta} y_k\right) \\ w_{k+1} = \mathrm{Prox}^{1/\eta}_{f}\left(z_{k+1} + \frac{1}{\eta} y_k\right) \\ y_{k+1} = y_k - \tau\,(z_{k+1} - w_{k+1}) \end{cases}$$

DR

$\min_y \underbrace{f^*(y)}_{G(y)} + \underbrace{g^*(-y)}_{F(y)}$

$(I + R_F R_G)/2$

$\Updownarrow$

(ADMM)

$\Updownarrow$

$$\begin{cases} z_{k+1} = \mathrm{Prox}^{1/\tau}_{g}\left(w_k - \frac{1}{\tau} y_k\right) \\ w_{k+1} = \mathrm{Prox}^{1/\tau}_{f}\left(z_{k+1} + \frac{1}{\tau} y_k\right) \\ y_{k+1} = y_k - \tau\,(w_{k+1} - z_{k+1}) \end{cases}$$

$$\begin{cases} w_{k+1} = \mathrm{Prox}_f^{1/\tau}\left(z_k + \frac{1}{\tau} y_k\right) \\ z_{k+1} = \mathrm{Prox}_g^{1/\tau}\left(w_{k+1} - \frac{1}{\tau} y_k\right) \\ y_{k+1} = y_k + \tau\left(w_{k+1} - z_{k+1}\right) \end{cases}$$

(ADMM 2)

$\Updownarrow$

$$\min_y \underbrace{f^*(y)}_{F(y)} + \underbrace{g^*(-y)}_{G(y)}$$

$(I + R_F R_G)/2$

$\Updownarrow$

Conclusion: Convergence of Douglas-Rachford

$\Rightarrow$ Convergence of $\begin{cases} \text{1) ADMM} \\ \text{2) PDHG with } \tau = \frac{1}{\eta} \end{cases}$

$$g \qquad \eta$$

$$\mathrm{Prox}_F^{\eta}(x) = x - \eta\, \mathrm{Prox}_{F^*}^{1/\eta}\left(\frac{x}{\eta}\right) = x - \eta\, \mathrm{Prox}_g^{1/\eta}\left(-\frac{x}{\eta}\right)$$

$$\begin{aligned}
F^*(x) &= \max_y \langle x, y\rangle - F(y) \\
&= \max_y \langle x, y\rangle - g^*(-y) \\
&= \max_y \langle -x, -y\rangle - g^*(-y) \\
&= \max_v \langle -x, v\rangle - g^*(v) \\
&= g^{**}(-x) = g(-x)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Prox}_{F^*}^{\tau}(x) &= \operatorname*{argmin}_u F^*(u) + \frac{1}{2\tau}\|u - x\|^2 \\
&= \operatorname*{argmin}_u g(-u) + \frac{1}{2\tau}\|u - x\|^2 \\
&= \operatorname*{argmin}_v g(v) + \frac{1}{2\tau}\|v + x\|^2 = \mathrm{Prox}_g^{\tau}(-x)
\end{aligned}$$