

Convex Conjugate (Fenchel Dual Function)

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

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

$$\left\{ \begin{array}{l} \text{Primal Problem (P)} \\ \min_x f(x) + g(x) \end{array} \right. \quad f(x) \text{ \& } g(x) \text{ are convex}$$

$$\Rightarrow \left\{ \begin{array}{l} f^*(y) \text{ is convex} \\ g^*(-y) \text{ is convex} \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{l} \text{Dual Problem (D)} \\ -\min_y f^*(y) + g^*(-y) \end{array} \right.$$

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

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

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

Primal Dual Minimizers Relation

$$\left\{ \begin{array}{l} x^* \in \partial f^*(y^*) \\ y^* \in -\partial g(x^*) \end{array} \right.$$

In (PD), the cost function is

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

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

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

Primal Dual (PD)

$$\min_x \sup_y \langle x, y \rangle - f^*(y) + g(x)$$

For $\min_x \max_y L(x, y)$, a simple algorithm is

to use gradient descent/ascent for approaching saddle point

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

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

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

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

$\eta > 0, \tau > 0$

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

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

$\eta > 0, \tau > 0$

$\eta \tau \leq 1$

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

③ PDHG with $\tau = \frac{1}{\eta}$

Moreau-Decomposition

$f(x)$ is convex

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

$$\text{Let } v_k = x_k - \eta y_k \Leftrightarrow y_k = \frac{x_k - v_k}{\eta}$$

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

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

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

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

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

$$\Rightarrow \begin{cases}
 x_{k+1} = \text{Prox}_g^\eta (V_k) & \boxed{V_k = x_k - \eta y_k} \\
 V_{k+1} = V_k - x_{k+1} + \text{Prox}_f^\eta (2x_{k+1} - V_k)
 \end{cases}$$

$$\begin{aligned}
 \frac{I + R_f R_g}{2} [V_k] &= \frac{V_k + R_f [R_g(V_k)]}{2} = \frac{1}{2} V_k + \frac{1}{2} R_f [2x_{k+1} - V_k] \\
 &= \frac{1}{2} V_k + \frac{1}{2} [2 \text{Prox}_f^\eta (2x_{k+1} - V_k) - (2x_{k+1} - V_k)]
 \end{aligned}$$

So PDHG ($\eta\tau \leq 1$) 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

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

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

The Lagrangian for constrained minimization:

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

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 :

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

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

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

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

Let $\sigma = \tau$ $\frac{w_{k+1} - y_k}{\tau} = z_{k+1}$ & $\frac{y_k - y_k}{\tau} = w_k$

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

Fact : (ADMM) $\Leftrightarrow$ Douglas-Rankford on (D)

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

$$\Leftrightarrow \begin{cases} z_{k+1} = \text{Prox}_g^{\frac{1}{2}}(w_k - \frac{1}{2}y_k) \\ w_{k+1} = \text{Prox}_f^{\frac{1}{2}}(z_{k+1} + \frac{1}{2}y_k) \\ y_{k+1} = y_k - \tau(w_{k+1} - z_{k+1}) \end{cases}$$

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

Moreau Decomposition

$$\text{Prox}_{f^*}^{\eta}(x) + \eta \text{Prox}_f^{\frac{1}{\eta}}(\frac{x}{\eta}) = x$$

$$\text{Prox}_G^{\eta}(x) = \text{Prox}_{f^*}^{\eta}(x) = x - \eta \text{Prox}_f^{\frac{1}{\eta}}(\frac{x}{\eta}) \quad F(x) = g^*(-x)$$

$$\text{Prox}_F^{\eta}(x) = \arg\min_u [g^*(-u) + \frac{1}{2\eta} \|u - x\|^2] \quad u = -v$$

$$\begin{aligned} u^* = -v^* &\Leftarrow (\Leftrightarrow) - \arg\min_v [g^*(v) + \frac{1}{2\eta} \|v - x\|^2] \\ &= - \arg\min_v [g^*(v) + \frac{1}{2\eta} \|v - (-x)\|^2] \\ &= - \text{Prox}_{g^*}^{\eta}(-x) \\ &= \eta \text{Prox}_g^{\frac{1}{\eta}}(-\frac{x}{\eta}) - (-x) \end{aligned}$$

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

Conclusion: Convergence of Douglas-Rachford

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

Convergence rate of either $F(X_k) - F(X_*)$
or $\|X_k - X_*\|^2$

$\min_x f(x)$ ∇f

Gradient Descent

Accelerated GD

Convexity

$$O\left(\frac{1}{k}\right)$$

$$O\left(\frac{1}{k^2}\right)$$

Strong Convexity

$$O\left(\left[\frac{L-\mu}{L+\mu}\right]^2\right)^k \quad \eta = \frac{2}{L+\mu}$$

$$O\left((1-\sqrt{\frac{\mu}{L}})^k\right) \quad \eta = \frac{1}{L}$$

$\min_x f(x)$ ∂f

Subgradient Method

Proximal Point Method

$$O\left(\frac{1}{\sqrt{k}}\right)$$

$$O\left(\frac{1}{k}\right)$$

$$O\left(\frac{1}{k}\right)$$

$$O\left(\left[\left(\frac{1}{1+\eta\mu}\right)^2\right]^k\right) \quad \forall \eta > 0$$

$\min_x f(x) + g(x)$ $\partial f \quad \nabla g$

Proximal Gradient

Accelerated Prox Grad

$$O\left(\frac{1}{k}\right)$$

$$O\left(\frac{1}{k^2}\right)$$

$$O\left((1-\frac{\mu}{L})^k\right) \quad \eta = \frac{1}{L}$$

$$O\left((1-\sqrt{\frac{\mu}{L}})^k\right) \quad \eta = \frac{1}{L}$$

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

Generalized Douglas-Rachford $\lambda \in (0, 2)$ $O\left(\frac{1}{k}\right)$

$$O\left(\frac{1}{k}\right)$$

$\forall \eta > 0$

Peaceman-Rachford $\lambda = 2$ divergent

$$O\left(\frac{1}{k}\right)$$

$\forall \eta > 0$

$\min_x f(x) + g(x)$ $\partial f \quad \nabla g$

$$C = \sqrt{1 - \frac{4\mu\eta}{(1+\eta L)^2}} \quad \forall \eta > 0$$

Generalized Douglas-Rachford $\lambda \in (0, 2)$

$$O\left(\frac{1}{k}\right) \quad O\left(\left[1 - \frac{\lambda}{2} + \frac{\lambda}{2}C\right]^{2k}\right)$$

Split Bregman (2009)

$$\min f(\mathbf{w}) + g(\mathbf{z}), \quad \mathbf{w} = \mathbf{z}.$$

Consider an unconstrained problem

$$\min_{\mathbf{w}, \mathbf{z}} f(\mathbf{w}) + g(\mathbf{z}) + \frac{\tau}{2} \|\mathbf{w} - \mathbf{z}\|^2,$$

for which the Bregman iteration is given as

$$\mathbf{z}_{k+1} = \operatorname{argmin}_{\mathbf{z}} g(\mathbf{z}) + \frac{\tau}{2} \|\mathbf{w}_k - \mathbf{z} - \mathbf{u}_k\|^2$$

$$(\text{Split Bregman}) : \quad \mathbf{w}_{k+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) + \frac{\tau}{2} \|\mathbf{w} - \mathbf{z}_{k+1} - \mathbf{u}_k\|^2$$

$$\mathbf{u}_{k+1} = \mathbf{u}_k + (\mathbf{w}_{k+1} - \mathbf{z}_{k+1})$$

Plug in $\mathbf{u} = \frac{1}{\tau} \mathbf{y}$, then it can be written as

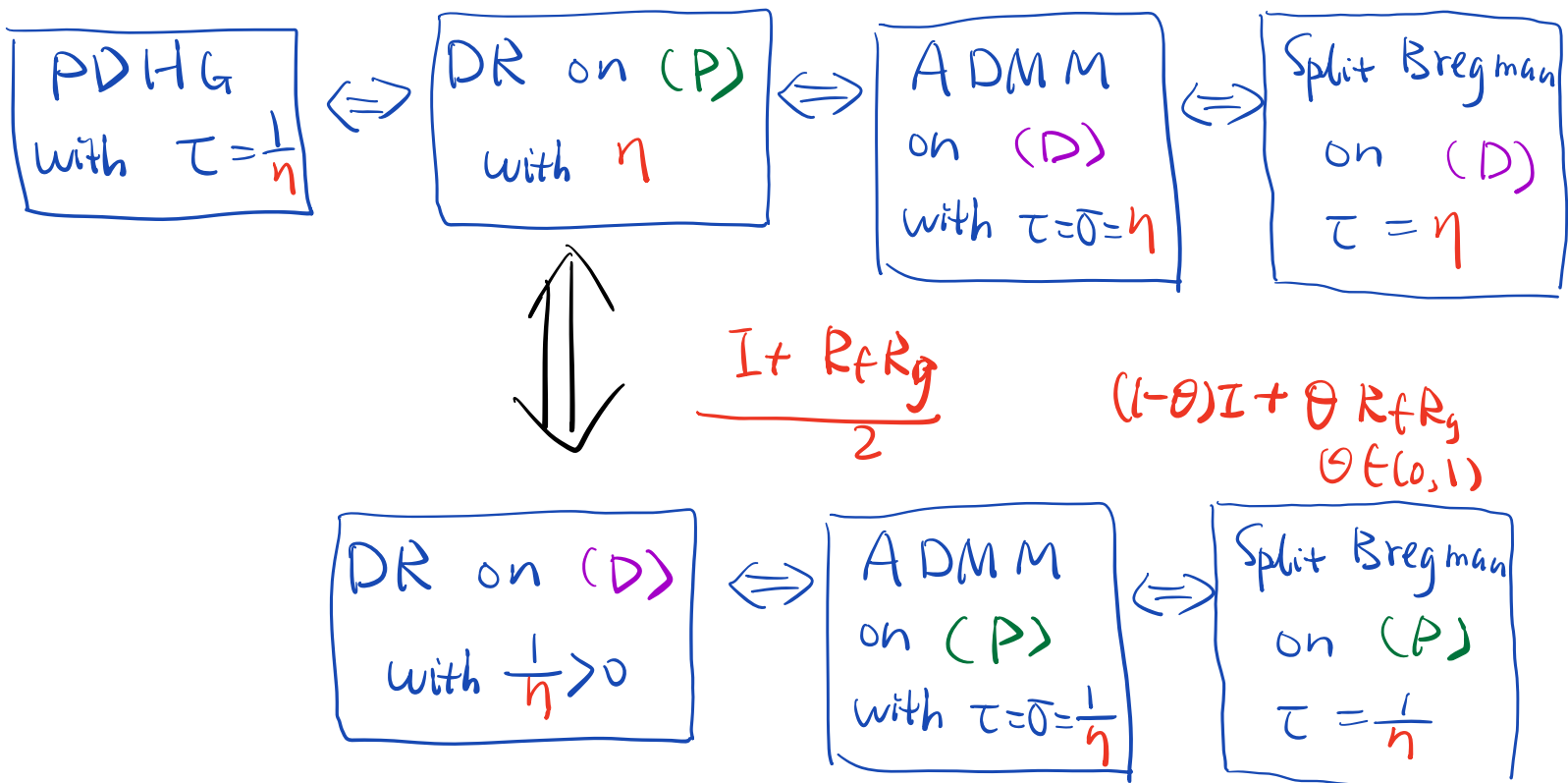
$$\mathbf{z}_{k+1} = \operatorname{argmin}_{\mathbf{z}} g(\mathbf{z}) - \langle \mathbf{y}_k, \mathbf{w}_k - \mathbf{z} \rangle + \frac{\tau}{2} \|\mathbf{w}_k - \mathbf{z}\|^2$$

$$(\text{Split Bregman}) : \quad \mathbf{w}_{k+1} = \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) - \langle \mathbf{y}_k, \mathbf{w} - \mathbf{z}_{k+1} \rangle + \frac{\tau}{2} \|\mathbf{w} - \mathbf{z}_{k+1}\|^2$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \tau(\mathbf{w}_{k+1} - \mathbf{z}_{k+1})$$

So this version of split Bregman method is equivalent to ADMM with $\sigma = \tau$, thus also equivalent to the DR splitting on the dual.

$$\left\{ \begin{array}{l} \text{Primal Problem (P)} \\ \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \text{Dual Problem (D)} \\ -\min_{\mathbf{y}} f^*(\mathbf{y}) + g^*(-\mathbf{y}) \end{array} \right.$$



Exercise:

$$\text{For } \min_x \underbrace{\|x\|_1}_{f(x)} + \underbrace{\tilde{\mathcal{I}}_{\{x: Ax=b\}}(x)}_{g(x)} + \frac{1}{2d} \|x\|_2^2$$

how many different algorithms we have learned?

Which one is the best?