Def $f: \mathbb{R}^n \to \mathbb{R}$ is a function

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

the convex conjugate of $f(x)$.

a.k.a. Legendre Transform

Fenchel Transform

Fenchel dual

Theorem $f^*(x)$ is convex on its domain $\{x : f^*(x) < +\infty\}$

even if $f(x)$ is not convex

Example: $f(x) = e^x$

$$f^*(x) = \sup_y (xy - e^y)$$

sup attains at critical point $\Rightarrow x - e^{y_*} = 0$

$$\Rightarrow y_* = \log x$$

$$\Rightarrow f^*(x) = \begin{cases} x\log x - x, & x > 0 \\ 0, & x = 0 \\ +\infty, & x < 0 \end{cases}$$

Example: $f(x) = ax + b$

$$f^*(x) = \sup_y (xy - f(y)) = \sup_y (xy - ay - b)$$

$$= \begin{cases} -b, & x = a \\ +\infty, & x \neq a \end{cases}$$

<u>Theorem</u> $f^*(x) + f(y) \geq \langle x, y \rangle$ if $f^*(x) \in \mathbb{R}$

Proof: $f^*(x) = \sup_y \langle x, y \rangle - f(y) \geq \langle x, y \rangle - f(y)$

<u>Theorem</u> ① $f^{**}(x) \leq f(x)$

② $f(x)$ is convex $\Rightarrow f^{**}(x) = f(x)$

Example: ① $f(x) = \|x\|$ for some norm $\|\cdot\|$

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

indicator function of unit ball

$\|x\|_*$ is the dual norm

② $f(x) = \frac{1}{2}\|x\|^2 \Rightarrow f^*(x) = \frac{1}{2}\|x\|_*^2$

Examples of dual norms for $x \in \mathbb{R}^n$

1) The dual norm of $\|x\|$ is $\|x\|$

vector 2-norm

2) The dual norm of $\|x\|_1$ is $\|x\|_\infty = \max_i |x_i|$

3) The dual norm of $\|x\|_\infty$ is $\|x\|_1$

4) The dual norm of $\|x\|_p$ is $\|x\|_q$ $\quad \frac{1}{p}+\frac{1}{q}=1$

$$\|x\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{\frac{1}{p}}$$

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

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

Example:

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

$$f^*(x) = \|x\|_1$$

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

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

Example: Find $\mathrm{Prox}_f^{\eta}(x)$ for $f(x) = \begin{cases} 0 & , \|x\|_\infty \le 1 \\ +\infty & , \|x\|_\infty > 1 \end{cases}$

Solution I : $f^*(x) = \|x\|_1$ $\quad f^*(x) = \|x\|_1$

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

$$= x - \eta S_{\frac{1}{\eta}}\left(\frac{x}{\eta}\right)$$

$$S_\gamma(X)_i = \begin{cases} X_i - \gamma & , \; X_i > \gamma \\ X_i + \gamma & , \; X_i < -\gamma \\ 0 & , \; X_i \in [-\gamma, \gamma] \end{cases}$$

$$S_{\frac{1}{\eta}}\left(\frac{X}{\eta}\right)_i = \begin{cases} X_i/\eta - 1/\eta & , \; X_i > 1 \\ X_i/\eta + 1/\eta & , \; X_i < -1 \\ 0 & , \; X_i \in [-1, 1] \end{cases}$$

$$\text{Prox}_f^\eta(X)_i = X_i - \eta \, S_{\frac{1}{\eta}}\left(\frac{X}{\eta}\right)_i = \begin{cases} 1 & , \; X_i > 1 \\ -1 & , \; X_i < -1 \\ X_i & , \; X_i \in [-1, 1] \end{cases}$$

Solution II: $f(X) = \begin{cases} 0 & , \; \|X\|_\infty \leq 1 \\ +\infty & , \; \|X\|_\infty > 1 \end{cases}$

is the indicator function of $\|\cdot\|_\infty$-ball

So $\text{Prox}_f^\eta(X)$ should be projection to this ball.

In practice, if we have Prox for $f(x)$

then we also have Prox for $f^*(x)$

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

$\Rightarrow$

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

$$= \min_x \left(\max_y \left[\langle x, y \rangle - f^*(y) \right] + g(x) \right)$$

min & max can be switched under some assumptions

$$= \min_x \max_y \left(\langle x, y \rangle - f^*(y) + g(x) \right)$$

$$= \max_y \min_x \left(\langle x, y \rangle - f^*(y) + g(x) \right)$$

$$= \max_y \left[\min_x \left(\langle x, y \rangle + g(x) \right) - f^*(y) \right]$$

$$= \max_y \left[-\max_x \left(\langle x, -y \rangle - g(x) \right) - f^*(y) \right]$$

$$= \max_y \left[-g^*(-y) - f^*(y) \right]$$

$$= -\min_y \left[f^*(y) + g^*(-y) \right]$$

Fencel's Duality Theorem

$f(x): \mathbb{R}^n \to \mathbb{R}$, $g(x): \mathbb{R}^n \to \mathbb{R}$ are convex

$$\min_x \left[f(x) + g(x) \right] = -\min_y \left[f^*(y) + g^*(-y) \right]$$

Primal-Dual Relation

$$\min \underbrace{\|x\|_1 + \alpha \|x\|_2^2}_{f(x)} + \underbrace{i_{\{Ax=b\}}}_{g(x)}$$

$$x^* = \arg\min_x \langle x, y^* \rangle + g(x) \Leftrightarrow 0 \in y^* + \partial g(x^*) \Leftrightarrow -y^* \in \partial g(x^*)$$

Similarly $y^* = \arg\min_y \left[f^*(y) - \langle x^*, y \rangle \right] \Leftrightarrow x^* \in \partial f^*(y^*)$

Strong Convexity – Smoothness Correspondance

① $f(x)$ is convex, $\nabla f(x)$ is L-cont. with $L = \sigma > 0$ $\Big\} \Leftrightarrow f^*(x)$ is strongly convex with $\mu = \frac{1}{\sigma}$.

② $f(x)$ is strongly convex with $\mu = \sigma$ $\Leftrightarrow \begin{cases} f^*(x) \text{ is convex} \\ \nabla f^*(x) \text{ is L-cont. with } L = \frac{1}{\sigma}. \end{cases}$

Remark: strong convexity in $f(x) \Leftrightarrow$ Smoothness in dual

Consider $\min_x \|x\|_1$ s.t. $Ax=b$

$$\Updownarrow$$

$$\min_x \|x\|_1 + \iota_{\{x:Ax=b\}}(x)$$

$$f(x) + g(x)$$

$$\Updownarrow$$

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

We have
$$\begin{cases} \mathrm{Prox}_f^{\eta}(x) = S_\eta(x) & \text{shrinkage} \\ \mathrm{Prox}_g^{\eta}(x) = P(x) & \text{projection} \\ \qquad\qquad\;\; = x + A^T(AA^T)^{-1}(b-Ax) \end{cases}$$

By Moreau-Decomposition, we get

$$\mathrm{Prox}_{f^*}^{\eta}(x) = x - \eta\,\mathrm{Prox}_f^{1/\eta}\left(\frac{x}{\eta}\right) = P_1(x) = \begin{cases} 1 & , \; x_i > 1 \\ -1 & , \; x_i < -1 \\ x_i & , \; x_i \in [-1,1] \end{cases}$$

$$\begin{aligned} \mathrm{Prox}_{g^*}^{\eta}(x) &= x - \eta\,\mathrm{Prox}_g^{\frac{1}{\eta}}\left(\frac{x}{\eta}\right) \\ &= x - \eta\left[\frac{x}{\eta} + A^T(AA^T)^{-1}\left(b - A\frac{x}{\eta}\right)\right] \\ &= -A^T(AA^T)^{-1}(\eta b - Ax) \\ &= A^T(AA^T)^{-1}(Ax - \eta b) \end{aligned}$$

For using generalized Douglas Rachford, we get four families of algorithms:

$$x_k = \mathrm{Prox}_g(y_k)$$

$$R_g(y_k) = 2\mathrm{Prox}_g(y_k) - y_k = 2x_k - y_k$$

$$R_f[R_g(y_k)] = 2\mathrm{Prox}_f(2x_k - y_k) - (2x_k - y_k)$$

$$\frac{I + R_f R_g}{2}(y_k) = \mathrm{Prox}_f(2x_k - y_k) - x_k + y_k$$

① $$\begin{cases} x_k = \mathrm{Prox}_g^{\eta}(y_k) \\ y_{k+1} = (1-\lambda)y_k + \lambda\left[\mathrm{Prox}_f^{\eta}(2x_k - y_k) + y_k - x_k\right] \end{cases}$$

$x_k \to x_*$

$\lambda \in (0,2)$

② $$\begin{cases} x_k = \mathrm{Prox}_f^{\eta}(y_k) \\ y_{k+1} = (1-\lambda)y_k + \lambda\left[\mathrm{Prox}_g^{\eta}(2x_k - y_k) + y_k - x_k\right] \end{cases}$$

$x_k \to x_*$

$\lambda \in (0,2)$

$$\min_z f^*(z) + g^*(-z) \qquad F(z) = f^*(z)$$

$$\Leftrightarrow \min_z F(z) + G(z) \qquad G(z) = g^*(-z)$$

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

$$\begin{aligned} \mathrm{Prox}_G^{\eta}(x) &= \operatorname*{argmin}_u \left[g^*(-u) + \frac{1}{2\eta}\|u - x\|^2\right] \\ &= \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|-v - x\|^2\right] \\ &= \operatorname*{argmin}_v \left[g^*(v) + \frac{1}{2\eta}\|v - (-x)\|^2\right] \\ &= \mathrm{Prox}^{\eta}_{g^*}(-x) \\ &= -x - \eta\,\mathrm{Prox}_g^{1/\eta}\left(-\frac{x}{\eta}\right) \end{aligned}$$

Apply $\lambda I + (1-\lambda)\frac{I + R_F R_G}{2}$

③ $\begin{cases} z_k = \mathrm{Prox}^{\eta}_{G}(w_k) \\ w_{k+1} = (1-\lambda) w_k + \lambda\left[\mathrm{Prox}^{\eta}_{F}(2z_k - w_k) + w_k - z_k\right] \end{cases}$ $z_k \to z_*$, $\lambda \in (0,2)$

$$\begin{cases} z_k = -w_k - \eta\, \mathrm{Prox}^{1/\eta}_{g}\left(-\frac{w_k}{\eta}\right) \\ w_{k+1} = (1-\lambda) w_k + \lambda\left[z_k - \eta\, \mathrm{Prox}^{\frac{1}{\eta}}_{f}\left(\frac{2z_k - w_k}{\eta}\right)\right] \end{cases}$$

④ $\begin{cases} z_k = \mathrm{Prox}^{\eta}_{G}(w_k) \\ w_{k+1} = (1-\lambda) w_k + \lambda\left[\mathrm{Prox}^{\eta}_{F}(2z_k - w_k) + w_k - z_k\right] \end{cases}$ $z_k \to z_*$, $\lambda \in (0,2)$

Assume $g(x)$ is strongly convex with $\mu > 0$

Then $\begin{cases} g^*(-y) \text{ is convex} \\ \nabla g^*(y) \text{ is L-cont. with } L = -\frac{1}{\mu} \end{cases}$

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

can also use Forward-Backward splitting

Fast Proximal-Gradient $\frac{1}{k^2}$

worst case rate.

$$\min_{z} f^*(z) + g^*(-z) \qquad x_* \in \partial f^*(z_*)$$

③

$$\begin{cases} z_k = -w_k - \eta \, \mathrm{Prox}_g^{1/\eta}\left(-\frac{w_k}{\eta}\right) \\ w_{k+1} = (1-\lambda) w_k + \lambda \left[z_k - \eta \, \mathrm{Prox}_f^{\frac{1}{\eta}}\left(\frac{2z_k - w_k}{\eta}\right) \right] \end{cases}$$

$$\min_{x} f(x) + g(x)$$

①

$$\begin{cases} x_k = \mathrm{Prox}_g^{\eta}(y_k) \qquad x_k \to x_* \\ y_{k+1} = (1-\lambda) y_k + \lambda \left[\mathrm{Prox}_f^{\eta}(2x_k - y_k) + y_k - x_k \right] \end{cases} \quad \lambda \in (0,2)$$