General Problem

$$\min_x f(Kx) + g(x) \qquad K \text{ is a matrix}$$

Example: ROF (Rudin, Osher, Fatemi 1992) model

1D signal $\quad \min_{u \in \mathbb{R}^n} \|u\|_{TV} + \frac{\alpha}{2}\|u-d\|^2 \qquad \left(\frac{1}{2}\|x\|^2\right)^* = \frac{1}{2}\|x\|^2$

$$\|Du\|_1 + \frac{\alpha}{2}\|u-d\|^2$$

$$\|u\|_{TV} = \sum_i |u_{i+1} - u_i| = \|Du\|_1$$

$$Du = \underbrace{\begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}}_{(n-1)\times n} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Adjoint operator/matrix for a linear operator K

$x \in \mathbb{R}^n \qquad \langle Kx, y\rangle = \langle x, K^* y\rangle$

$y \in \mathbb{R}^m \qquad K^* = K^T$

$K \in \mathbb{R}^{m\times n}$

K^* is simply K^T for real matrices

but we use K^* to emphasize it can be extended to general adjoint of linear operators

(P) $\quad \min_x f(Kx) + g(x) \qquad f: \mathbb{R}^m \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}$, $K \in \mathbb{R}^{m\times n}$

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

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

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

Example: $\min_x \|Dx\|_1 + \frac{\alpha}{2}\|x-d\|_2^2$

$$f(x) = \|x\|_1 \quad g(x) = \frac{\alpha}{2}\|x-d\|^2$$

$$D = \underbrace{\begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}}_{(n-1)\times n}$$

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

$$g^*(y) = \max_x \langle x, y\rangle - \frac{\alpha}{2}\|x-d\|^2$$

Critical point $\Rightarrow y - \alpha(x_* - d) = 0$

$$\Rightarrow x_* = y/\alpha + d$$

$$\Rightarrow g^*(y) = \frac{y^2}{\alpha} + dy - \frac{\alpha}{2}\frac{y^2}{\alpha^2}$$

$$= \frac{1}{2\alpha}(y^2 + 2\alpha dy + d^2) - \frac{d^2}{2\alpha}$$

$$= \frac{1}{2\alpha}\|y + \alpha d\|^2 - \frac{d^2}{2\alpha}$$

(P) $\min_x F(x) + G(x) = \min_x f(Dx) + g(x)$

(PD) $\min_{x\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^{n-1}} \langle y, Dx\rangle - f^*(y) + g(x)$

ADMM on (P)
$\Updownarrow$
DR on (D)

(D) $-\min_{y\in\mathbb{R}^{n-1}} f^*(y) + g^*(-D^T y)$

Remark: In this example, we can easily use proximal gradient on (D) but not on (P)

Useful Theorems about convex conjuate f^* for convex $f(x)$:

⓪ $(f^*)^* = f$

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

Example: $G(y) = g^*(-y)$

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

② $f(x) + f^*(y) = \langle x, y\rangle \iff x = \arg\max_u \langle u, y\rangle - f(u)$

Proof: $f^*(y) = \sup_u \langle u, y\rangle - f(u) \Rightarrow f^*(y) = \langle x, y\rangle - f(x)$

③ $y \in \partial f(x) \iff x \in \partial f^*(y)$

Proof: $y \in \partial f(x) \iff x = \arg\max_u \langle u, y\rangle - f(u)$

$$\Updownarrow$$

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

$\Updownarrow$ because $(f^*)^* = f$

$x \in \partial f^*(y) \iff y = \arg\max_u \langle u, x\rangle - f^*(u) \iff f^*(y) + (f^*)^*(x) = \langle x, y\rangle$

④ $f(x)$ is strongly convex with $\mu > 0$

$\iff \nabla f^*(x)$ is L-cont. with $L = \frac{1}{\mu}$.

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

Example: Consider $\min_x \underbrace{\|x\|_1 + \frac{1}{2\alpha}\|x\|^2}_{f(x)} + \underbrace{i_{\{Ax=b\}}}_{g(x)}$, $\alpha > 0$

$f(x)$ is non-smooth and strongly convex with $\frac{1}{\alpha} > 0$

$\Rightarrow \nabla f^*$ is L-cont. with $L = \alpha$

For deriving $\nabla f^*(y)$ (no need to derive f^* first):

$$x = \nabla f^*(y) \Leftrightarrow y \in \partial f(x) = \partial\|x\|_1 + \frac{1}{\alpha}x$$

$$\Leftrightarrow \alpha y \in x + \alpha\,\partial\|x\|_1$$

$$\Leftrightarrow x = (I + \alpha\,\partial\|\cdot\|_1)^{-1}(\alpha y)$$

$$\Leftrightarrow x = \mathrm{Prox}^{\alpha}_{\|\cdot\|_1}(\alpha y)$$

$$\boxed{(P) \quad \min_x \underbrace{\|x\|_1 + \frac{\alpha}{2}\|x\|^2}_{f(x)} + \underbrace{i_{\{Ax=b\}}}_{g(x)}}$$

$$\boxed{(D) \quad -\min_y f^*(y) + g^*(-y)}$$

$$\mathrm{Prox}^{\eta}_{f}(x) = \frac{\alpha}{\alpha+\eta}\mathrm{Prox}^{\eta}_{\|\cdot\|_1}(x) \qquad \mathrm{Prox}^{\eta}_{f^*}(x) = x - \eta\,\mathrm{Prox}^{1/\eta}_{f}\left(\frac{x}{\eta}\right)$$

$$\nabla f^*(y) = \mathrm{Prox}^{\alpha}_{\|\cdot\|_1}(\alpha y)$$

$$\mathrm{Prox}^{\eta}_{g}(x) = x + A^T(AA^T)^{-1}(b - Ax) \qquad G(y) = g^*(-y)$$

$$\mathrm{Prox}^{\eta}_{G}(x) = x + \eta\,\mathrm{Prox}^{1/\eta}_{g}(-x/\eta)$$

We have discussed at least 4 kinds of algorithms and their convergence:

1) (generalized) Douglas-Rachford: $\qquad x_k \to x_*$

$$u_{k+1} = \frac{I + R_f R_g}{2} \Rightarrow \begin{cases} x_k = u_k + A^T(AA^T)^{-1}(b - Au_k) \\ u_{k+1} = \frac{\alpha}{\alpha+\eta} S_\eta(2x_k - u_k) + u_k - x_k \end{cases}$$

S_η is the shrinkage operator (Prox of $\|x\|_1$)

2) (Accelerated) Proximal-Gradient on the dual:

$$(D) \quad -\min_{y} f^*(y) + \underbrace{g^*(-y)}_{G(y)}$$

$$\eta = \frac{1}{L} = \alpha$$

$$y_{k+1} = (I + \eta\,\partial G)^{-1}(I - \eta\nabla f^*)(y_k),$$

$$= (I + \alpha\,\partial G)^{-1}(I - \alpha\nabla f^*)(y_k)$$

$$\Rightarrow \begin{cases} v_k = y_k - \alpha S_\alpha(\alpha y_k) \\ y_{k+1} = v_k + \eta\left(-\frac{v_k}{\eta} + A^T(AA^T)^{-1}\left(b - A\left(-\frac{v_k}{\eta}\right)\right)\right) \\ \phantom{y_{k+1}} = A^T(AA^T)^{-1}(\alpha b + A v_k) \end{cases}$$

$$y_k \to y_*$$

To recover x_k: $x_k = \nabla f^*(y_k)$

Recall the primal-dual relation: $x_* \in \partial f^*(y_*)$

$y_* \in \partial f(x_*)$

Remark: projection to ℓ^∞ unit ball gives $y_k - S_1(y_k)$, but this is coincidence

3) ADMM (on the primal) equivalent to $\frac{I + R_F R_G}{2}$ $\begin{cases} G(y) = f^*(y) \\ F(y) = g^*(-y) \end{cases}$

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

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

$$\Longleftrightarrow \begin{cases} z_{k+1} = (w_k - \frac{1}{\tau} y_k) + A^T(AA^T)^{-1}\left[b - A(w_k - \frac{1}{\tau} y_k)\right] \\ w_{k+1} = S_{\frac{1}{\tau}}(z_{k+1} + \frac{1}{\tau} y_k) \\ y_{k+1} = y_k - \tau(w_{k+1} - z_{k+1}) \end{cases}$$

$y_k \to y_*$ $\quad w_k \to x_*$ $\quad z_k \to x_*$

4) PDHG $\quad$ with $\sigma = \frac{1}{\eta}$, it's $\frac{I + R_f R_g}{2}$

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

$\mathrm{Prox}_f^{\eta}(x) = \frac{\alpha}{\alpha + \eta} \mathrm{Prox}_{\|\cdot\|_1}^{\eta}(x) \qquad \mathrm{Prox}_{f^*}^{\eta}(x) = x - \eta\, \mathrm{Prox}_f^{1/\eta}(\frac{x}{\eta})$

$$\begin{cases} x_{k+1} = x_k - \eta y_k + A^T(AA^T)^{-1}\left[b - A(x_k - \eta y_k)\right] \\ y_{k+1} = y_k + \sigma(2x_{k+1} - x_k) - \sigma \frac{\alpha}{\frac{1}{\sigma} + \alpha} S_{\frac{1}{\sigma}}\left[\frac{1}{\sigma} y_k + 2x_{k+1} - x_k\right] \end{cases}$$

$$\begin{cases} x_k \to x_* \\ y_k \to y_* \end{cases}$$

Convergence of PDHG

$$\min_x f(Kx) + g(y)$$

Algorithm I (Arrow, Hurwitz, Uzawa 1958)

$$\text{(PD)} \quad L(x,y) = \langle y, Kx\rangle - f^*(y) + g(x)$$

$$\begin{cases} \dfrac{x_{k+1}-x_k}{\eta} = -\dfrac{\partial L}{\partial x}(x_{k+1}, y_k) \\ \dfrac{y_{k+1}-y_k}{\tau} = \dfrac{\partial L}{\partial y}(x_{k+1}, y_{k+1}) \end{cases} \Rightarrow \begin{cases} x_{k+1} \in x_k - \eta K^* y_k - \eta\, \partial g(x_{k+1}) \\ y_{k+1} \in y_k + \tau K x_{k+1} - \tau\, \partial f^*(y_{k+1}) \end{cases}$$

$$\Longleftrightarrow \begin{cases} x_{k+1} = \underset{x}{\operatorname{argmin}}\ L(x, y_k) + \frac{1}{2\eta}\|x - x_k\|^2 \\ y_{k+1} = \underset{y}{\operatorname{argmax}}\ L(x_{k+1}, y) + \frac{1}{2\tau}\|y - y_k\|^2 \end{cases}$$

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

It may not converge without strong convexity!

Convergence can be proven if

$$\begin{cases} g(x) \text{ is strongly convex with } \mu > 0 \\ \tau < \dfrac{\mu}{\|K\|} \end{cases}$$

↳ A good reading project (2014 paper)

Convergence of Primal-Dual-Hybrid-Gradient (PDHG, 2010)

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

$\eta\tau < \frac{1}{\|K\|}$, no need of strong convexity

The saddle point of $L(x,y) = \langle y, Kx\rangle - f^*(y) + g(x)$

satisfies

$$\begin{cases} 0 \in \frac{\partial L}{\partial x}(x_*, y_*) = K^* y_* + \partial g(x_*) \\ 0 \in -\frac{\partial L}{\partial y}(x_*, y_*) = -K x_* + \partial f^*(x_*) \end{cases}$$

$$0 \in T\begin{pmatrix} x_* \\ y_* \end{pmatrix}$$

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} K^* y + \partial g(x) \\ -Kx + \partial f^*(y) \end{pmatrix} = \begin{pmatrix} \partial g(x) \\ \partial f^*(y) \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

$$0 \in T\begin{pmatrix} x_* \\ y_* \end{pmatrix} \qquad \vec{z} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad M = \begin{pmatrix} \frac{1}{\eta} I & -K^* \\ -K & \frac{1}{\tau} I \end{pmatrix}$$

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} K^* y + \partial g(x) \\ -Kx + \partial f^*(y) \end{pmatrix} = \begin{pmatrix} \partial g(x) \\ \partial f^*(y) \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

$$0 \in M(\vec{z}_{k+1} - \vec{z}_k) + T\vec{z}_{k+1}$$

$$\Leftrightarrow \begin{cases} 0 \in \frac{1}{\eta}(x_{k+1} - x_k) - K^*(y_{k+1} - y_k) + \partial g(x_{k+1}) + K^* y_{k+1} \\ 0 \in \frac{1}{\tau}(y_{k+1} - y_k) - K(x_{k+1} - x_k) + \partial f^*(y_{k+1}) - K x_{k+1} \end{cases}$$

$$\Leftrightarrow \begin{cases} x_k - \eta K^* y_k \in x_{k+1} + \eta \partial g(x_{k+1}) \\ y_k + \tau K(2x_{k+1} - x_k) \in y_{k+1} + \tau f^*(y_{k+1}) \end{cases}$$

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

$$\vec{z} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad M = \begin{pmatrix} \frac{1}{\eta} I & -K^* \\ -K & \frac{1}{\tau} I \end{pmatrix}$$

So PDHG $\Leftrightarrow$ $M z_{k+1} + T z_{k+1} \ni M z_k$

$$\Leftrightarrow \quad z_{k+1} + M^{-1} T z_{k+1} \ni z_k$$

$$\Leftrightarrow \quad z_{k+1} = (I + M^{-1} T)^{-1} z_k$$

M is positive definite $\Leftrightarrow$ $\eta \tau < \|K\| = \max\limits_i \lambda_i (K K^*)$

$$|\lambda I - M| = \begin{vmatrix} (\lambda - \frac{1}{\eta}) I & +K^* \\ +K & (\lambda - \frac{1}{\tau}) I \end{vmatrix}$$

$$= (\lambda - \tfrac{1}{\eta})^n \left| (\lambda - \tfrac{1}{\tau}) I - (\lambda - \tfrac{1}{\eta})^{-1} K K^* \right|$$

$$\Rightarrow \qquad \eta \tau < \|K\|$$

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A) \det(D - CA^{-1}B)$$

Define $\|z\|_M^2 = z^T M z = (x \;\; y) \begin{pmatrix} \frac{1}{\eta} I & -K^* \\ -K & \frac{1}{\tau} I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$

Then can prove

$$\|z_{k+1} - z_*\|_M \leq \|z_k - z_*\|_M - c \|z_{k+1} - z_k\|_M$$

→ A good reading project (2012 paper)

Some details on Arrow-Hurwitz.

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

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

$$\Rightarrow \begin{cases} K^* y_k + \partial g(x_{k+1}) + \frac{1}{\eta}(x_{k+1} - x_k) \ni 0 \\ -K x_{k+1} + \partial f^*(x_{k+1}) + \frac{1}{\tau}(y_{k+1} - y_k) \ni 0 \end{cases}$$

$$\Rightarrow \begin{cases} x_{k+1} = \underset{x}{\arg\min}\ g(x) + \langle K^* y_k + \frac{1}{\eta}(x_{k+1} - x_k), x \rangle \\ y_{k+1} = \underset{y}{\arg\min}\ f^*(y) - \langle K x_{k+1} + \frac{1}{\tau}(y_{k+1} - y_k), y \rangle \end{cases}$$

$$\Rightarrow \begin{cases} g(x) - g(x_{k+1}) + \langle x - x_{k+1}, K^* y_k + \frac{1}{\eta}(x_{k+1} - x_k) \rangle \geq 0 \\ f^*(y) - f^*(y_{k+1}) + \langle y - y_{k+1}, -K x_{k+1} + \frac{1}{\tau}(y_{k+1} - y_k) \rangle \geq 0 \end{cases}$$

$$u = \begin{pmatrix} x \\ y \end{pmatrix} \quad F(u) = g(x) + f^*(y)$$

$$\Rightarrow F(u) - F(u_{k+1}) + \left\langle \begin{pmatrix} x - x_{k+1} \\ y - y_{k+1} \end{pmatrix}, \begin{pmatrix} K^* y_{k+1} \\ -K x_{k+1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\eta}(x_{k+1} - x_k) - K^*(y_{k+1} - y_k) \\ \frac{1}{\tau}(y_{k+1} - y_k) \end{pmatrix} \right\rangle$$

$$\Rightarrow F(u) - F(u_{k+1}) + (u - u_{k+1})^T M u_{k+1} \geq (u - u_{k+1})^T Q (u_k - u_{k+1})$$

$$M = \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} \frac{1}{\eta} I & -K^* \\ 0 & \frac{1}{\tau} I \end{pmatrix}$$