General Problem

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

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

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

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

$$\boxed{\begin{aligned} \|u\|_{TV} &= \sum_i |u_{i+1} - u_i| \\ &= \|Du\|_1 \end{aligned} \qquad Du = \underbrace{\begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}}_{(n-1)\times n} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}}$$

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) $\quad \min_x f(Kx) + g(x) \qquad f : \mathbb{R}^m \to \mathbb{R}$

$g : \mathbb{R}^n \to \mathbb{R}$

(PD) $\quad \min_x \sup_y \langle y, Kx\rangle - f^*(y) + g(x) \qquad K \in \mathbb{R}^{m\times n}$

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

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

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

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

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

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

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

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

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

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

This problem is $\min_x f(Dx) + g(x)$

we have formula for proximal operators for $f(v)$, $g(x)$

but not for $F(x) = f(Dx)$

$$(P) \quad \min_x F(x) + g(x) = \min_{x\in\mathbb{R}^n} f(Dx) + g(x)$$

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

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

$$(P) \quad \min_{x \in \mathbb{R}^n} \|Dx\|_1 + \frac{\alpha}{2}\|x-d\|^2$$

$$(D) \quad -\min_{y \in \mathbb{R}^{n-1}} \iota_{\{\|y\|_\infty \le 1\}}(y) + \underbrace{\frac{1}{2\alpha}\|D^T y - \alpha d\|^2 - \frac{d^2}{2\alpha}}_{G^*(y)}$$

Algorithms that can be implemented:

① (Fast) Proximal Gradient on (D) but not on (P)

② PDHG, which will be explained later

③ ADMM on (P) will be explained later.

Algorithms for

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

① PDHG

Algorithm I (Arrow, Hurwitz, Uzawa 1958)

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

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

$$\Leftrightarrow \begin{cases} x_{k+1} = \arg\min_x L(x, y_k) + \dfrac{1}{2\eta}\|x - x_k\|^2 \\ y_{k+1} = \arg\max_y L(x_{k+1}, y) + \dfrac{1}{2\tau}\|y - y_k\|^2 \end{cases}$$

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

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)

Algorithm II

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

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

$\eta\tau < \frac{1}{\|K\|^2}$, no need of strong convexity

Theorem [$O(\frac{1}{k})$ Convergence Rate]

Let $\bar{x}_k = \frac{1}{k}\sum_{i=1}^{k} x_i$, for convex functions, PDHG with $\eta\tau < \frac{1}{\|K\|}$

$$L(\bar{x}_k, y) - L(x, \bar{y}_k) \leq \frac{1}{k}\left[\frac{1}{\tau}\|x-x_0\|^2 + \frac{1}{\sigma}\|y-y_0\|^2\right.$$

Remark: PDHG $\Longleftrightarrow$ DR only if $K = I$, $\tau = \frac{1}{\eta}$.

So for TV minimization, PDHG is NOT DR!

Algorithm III

Fast/Accelerated PDHG (Chambolle & Pock 2010)

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

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

Accelerated PDHG $\qquad$ g is strongly convex with $\mu > 0$

$$\begin{cases} \tau_0 \eta_0 \leq \frac{1}{\|K\|^2}, \quad \bar{x}_0 = x_0 \\ y_{k+1} = (I + \tau_n \partial f^*)^{-1}[y_k + \tau_n K \bar{x}_k] \\ x_{k+1} = (I + \eta_n \partial g)^{-1}[x_k - \eta_n K^* y_{k+1}] \\ \theta_n = \frac{1}{\sqrt{1 + 2\mu\eta_n}}, \quad \eta_{n+1} = \theta_n \eta_n, \quad \tau_{n+1} = \frac{\tau_n}{\theta_n} \\ \bar{x}_k = x_{k+1} + \theta_n (x_{k+1} - x_k) \end{cases}$$

Thm $O(\frac{1}{k^2})$ "convergence" rate

$\hookrightarrow$ a good reading project.

② ADMM for solving

(P1) $\min f(x) + g(y) \quad$ s.t. $Ax + By = C$

Example: $\min\limits_{x} f(Kx) + g(x) \Leftrightarrow \min f(y) + g(x)$ s.t. $Kx - y = 0$

Lagrangian $L(x, y, z) = f(x) + g(y) + \langle z, Ax + By - C \rangle$

Assume Total Duality ($\Rightarrow$ we can switch min & max)

(convexity is not enough for total duality, though it can be shown for many convex problems need some technical conditions for total duality)

Assume some technical conditions, which are omitted

Then (P1) is equivalent to the saddle point of

Lagrangian $L(x,y,z) = f(x) + g(y) + \langle z, Ax+By-c\rangle$

$$\min_{x,y} \max_{z} \; f(x) + g(y) + \langle z, Ax+By-c\rangle$$

$$= \min_{x,y} \max_{z} \left[f(x) + \langle A^T z, x\rangle\right] + \left[g(y) + \langle B^T z, y\rangle\right] - \langle z, c\rangle$$

$$= \max_{z} \left(\min_{x} \left[f(x) + \langle A^T z, x\rangle\right] + \min_{y} \left[g(y) + \langle B^T z, y\rangle\right] - \langle z, c\rangle \right)$$

$$= \max_{z} \left(-\max_{x} \left[\langle x, -A^T z\rangle - f(x)\right] - \max_{y} \left[\langle y, -B^T z\rangle - g(y)\right] - \langle z, c\rangle \right)$$

$$= \max_{z} \left[-f^*(-A^T z) - g^*(-B^T z) - \langle z, c\rangle \right]$$

$$= -\min_{z} \left(f^*(-A^T z) + g^*(-B^T z) + \langle z, c\rangle \right)$$

(P1) $\min \; f(x) + g(y)$ s.t. $Ax + By = c$

$x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, $z \in \mathbb{R}^\ell$, $u \in \mathbb{R}^\ell$

(D1) $-\min_{z} \Big(\underbrace{f^*(A^T z)}_{F^*(z)} + \underbrace{g^*(B^T z) + \langle z, c\rangle}_{G^*(z)} \Big)$

(D-D1) $\min_{u} \big(F(u) + G(u) \big)$

Given a problem (P1)

- Lagrangian is $L(x,y,z) = f(x) + g(y) + \langle z, Ax+By-c\rangle$

  z → Lagrange multiplier

- Augmented Lagrangian is

$$\mathcal{L}_\sigma(x,y,z) = f(x) + g(y) + \langle z, Ax+By-c\rangle + \frac{\sigma}{2}\|Ax+By-c\|^2$$

ADMM for (P1) is defined as

$$\begin{cases} x_{k+1} = \underset{x}{\operatorname{argmin}}\ \mathcal{L}_\sigma(x, y_k, z_k) \\ y_{k+1} = \underset{y}{\operatorname{argmin}}\ \mathcal{L}_\sigma(x_{k+1}, y, z_k) \\ z_{k+1} = z_k + \tau(Ax_{k+1} + By_{k+1} - c) \end{cases}$$

$$\Leftrightarrow \begin{cases} x_{k+1} = \underset{x}{\operatorname{argmin}}\ f(x) + \langle z_k, Ax + By_k - c\rangle + \frac{\sigma}{2}\|Ax + By_k - c\|^2 \\ y_{k+1} = \underset{y}{\operatorname{argmin}}\ g(y) + \langle z_k, Ax_{k+1} + By - c\rangle + \frac{\sigma}{2}\|Ax_{k+1} + By - c\|^2 \\ z_{k+1} = z_k + \tau(Ax_{k+1} + By_{k+1} - c) \end{cases}$$

$$\begin{cases} \text{(P1)} \quad \min\ f(x) + g(y) \quad \text{s.t. } Ax+By=c \\ \qquad\qquad x \in \mathbb{R}^m,\ y \in \mathbb{R}^n,\ z \in \mathbb{R}^l,\ u \in \mathbb{R}^l \\ \text{(D1)} \ -\min_z \Big(\underbrace{f^*(A^T z)}_{F^*(z)} + \underbrace{g^*(B^T z) + \langle z, c\rangle}_{G^*(z)} \Big) \end{cases}$$

__Theorem__ If Total Duality holds (s.t. (P1) $\Leftrightarrow$ (D1))

then ADMM with $\tau = \sigma$ is equivalent to

Douglas Rachford Splitting $\dfrac{I + R^{\eta}_{F^*} R^{\eta}_{G^*}}{2}$ with $\eta = \tau$.

Example: ROF denoising for 1D problems

(P) $\min\limits_{x \in \mathbb{R}^n} \|Dx\|_1 + \frac{\alpha}{2}\|x - d\|^2$ $\quad x \in \mathbb{R}^n, D \in \mathbb{R}^{(n-1)\times n}$, $y \in \mathbb{R}^{n-1}$

(P1) $\min \frac{\alpha}{2}\|x-d\|^2 + \|y\|_1 \quad$ s.t. $Dx - y = 0$

(D1) $-\min\limits_{z \in \mathbb{R}^{n-1}} \underbrace{\frac{1}{2\alpha}\|D^T z - \alpha d\|^2 - \frac{d^2}{2\alpha}}_{F^*(z)} + \underbrace{\iota_{\{\|z\|_\infty \le 1\}}(z)}_{G^*(z)}$

ADMM with $\tau = \sigma$ for (P) is

$$\begin{cases} x_{k+1} = \arg\min\limits_{x} \frac{\alpha}{2}\|x-d\|^2 + \langle z_k, Dx - y_k\rangle + \frac{\tau}{2}\|Dx - y_k\|^2 \\ y_{k+1} = \arg\min\limits_{y} \|y\|_1 + \langle z_k, Dx_{k+1} - y\rangle + \frac{\tau}{2}\|Dx_{k+1} - y\|^2 \\ z_{k+1} = z_k + \tau(Dx_{k+1} - y_{k+1} - c) \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha(x_{k+1} - d) + D^T z_k + \tau D^T D x_{k+1} - \tau D^T y_k = 0 \\ 0 \in \partial\|y_{k+1}\|_1 - z_k + \tau(y_{k+1} - Dx_{k+1}) \end{cases}$$

$$\Longleftrightarrow \begin{cases} x_{k+1} = (\alpha I + \tau D^T D)^{-1}[\alpha d - D^T z_k + \tau D^T y_k] \\ y_{k+1} = \mathrm{Prox}_{\|\cdot\|_1}^{\frac{1}{\tau}}(z_k + \tau D x_{k+1}) \\ z_{k+1} = z_k + \tau(D x_{k+1} - y_{k+1} - c) \end{cases}$$

↳ solve a discrete Poisson equation $(\alpha I - \tau\Delta)^{-1}$

So three different algorithm on ROF denoising

① (Fast) Proximal Gradient on (D1)

② (fast) PDHG

$$\min_x f(Dx) + g(x) \quad f(y) = \|y\|_1, \; g(x) = \frac{\alpha}{2}\|x-d\|^2$$

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

③ ADMM on (P1) $\Longleftrightarrow$ DR on (D1)

$\Updownarrow$

DR on (D-D1)

↳ need to compute a Poisson equation, much more expensive than ① & ② in each iteration.

Remark: For a problem where Prox_f and Prox_g are available,

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

$$\begin{cases} \text{Primal Problem (P)} \\ \min\limits_x f(x)+g(x) \end{cases} \iff \begin{cases} \text{Dual Problem (D)} \\ -\min\limits_y f^*(y)+g^*(-y) \end{cases}$$

$$\boxed{\begin{matrix}\text{PDHG} \\ \text{with } \tau=\frac{1}{\eta}\end{matrix}} \iff \boxed{\begin{matrix}\text{DR on (P)} \\ \text{with } \eta\end{matrix}} \iff \boxed{\begin{matrix}\text{Simple} \\ \text{ADMM} \\ \text{on (D)} \\ \text{with } \tau=\sigma=\eta\end{matrix}} \iff \boxed{\begin{matrix}\text{Split Bregman} \\ \text{on (D)} \\ \tau=\eta\end{matrix}}$$

$$\Updownarrow$$

$$\boxed{\begin{matrix}\text{DR on (D)} \\ \text{with } \frac{1}{\eta}>0\end{matrix}} \iff \boxed{\begin{matrix}\text{Simple} \\ \text{ADMM} \\ \text{on (P)} \\ \text{with } \tau=\sigma=\frac{1}{\eta}\end{matrix}} \iff \boxed{\begin{matrix}\text{Split Bregman} \\ \text{on (P)} \\ \tau=\frac{1}{\eta}\end{matrix}}$$