

General Problem

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

K is a matrix

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

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

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

$$\|u\|_{TV} = \sum_i |u_{i+1} - u_i|$$
$$= \|Du\|_1$$
$$Du = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$(n-1) \times n$

Adjoint operator/matrix for a linear operator K

$$x \in \mathbb{R}^n$$

$$\langle Kx, y \rangle = \langle x, K^*y \rangle$$

$$y \in \mathbb{R}^m$$

$$K^* = K^T$$

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

K^* is simply K^T for real matrices

but we use K^* to emphasize it can

be extended to general adjoint of linear operators

$$(P) \quad \min_x f(Kx) + g(x)$$

$$f: \mathbb{R}^m \rightarrow \mathbb{R}$$

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

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

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

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

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

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

$$D = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}_{(n-1) \times n} \quad f(x) = \|x\|_1, \quad g(x) = \frac{\alpha}{2} \|x-d\|_2^2$$

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

$$g^*(y) = \max_x \langle x, y \rangle - \frac{\alpha}{2} \|x-d\|_2^2$$

$$\text{Critical point} \Rightarrow y - \alpha(x-d) = 0$$

$$\Rightarrow x_* = \frac{y}{\alpha} + d$$

$$\Rightarrow g^*(y) = \frac{y^2}{2} + dy - \frac{\alpha}{2} \frac{y^2}{\alpha^2} = \frac{1}{2\alpha} \|y + \alpha d\|_2^2 - \frac{d^2}{2\alpha}$$

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

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

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

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

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

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

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

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

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)

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

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

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

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

It may not converge without strong convexity!

Convergence can be proven if

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

↳ A good reading project (2014 paper)

Algorithm II

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

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

$\eta\tau < \frac{1}{\|K\|^2}$, 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 $\eta\tau < \frac{1}{\|K\|^2}$

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

Remark: PDHG $\Leftrightarrow$ 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)

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

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

Accelerated PDHG

g is strongly convex with $\mu > 0$

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

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

② ADMM for solving

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

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

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

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

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

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

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

$$= \max_z \left(\min_x [f(x) + \langle A^T z, x \rangle] + \min_y [g(y) + \langle B^T z, y \rangle] - \langle z, c \rangle \right)$$

$$= \max_z \left(-\max_x [\langle x, -A^T z \rangle - f(x)] - \max_y [\langle y, -B^T z \rangle - g(y)] - \langle z, c \rangle \right)$$

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

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

$$(P1) \min f(x) + g(y) \quad \text{s.t. } Ax + By = c$$

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

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

$$(D-D1) \min_u (F(u) + G(u))$$

Given a problem (P1)

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

$\downarrow$
Lagrange multiplier

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

ADMM for (P1) is defined as

$$\begin{cases} x_{k+1} = \operatorname{argmin}_x L_{\sigma}(x, y_k, z_k) \\ y_{k+1} = \operatorname{argmin}_y L_{\sigma}(x_{k+1}, y, z_k) \\ z_{k+1} = z_k + \tau (Ax_{k+1} + By_{k+1} - C) \end{cases}$$

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

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

Theorem If Total Duality holds (s.t. $(P) \Leftrightarrow (D)$)

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

Douglas Rachford Splitting $\frac{I + R_F^* R_G^*}{2}$ with $\eta = \tau$.

Example: ROF denoising for 1D problems

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

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

$$(D1) = \min_{z \in \mathbb{R}^{n-1}} \underbrace{\frac{1}{2\alpha} \|D^T z - \alpha d\|^2 - \frac{d^2}{2\alpha}}_{F^*(z)} + \underbrace{\mathcal{I}_{\{\|z\|_\infty \leq 1\}}(z)}_{G^*(z)}$$

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

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

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

$$\Leftrightarrow \begin{cases} X_{k+1} = (\underbrace{2I + \tau D^T D})^{-1} [2d - D^T z_k + \tau D^T y_k] \\ \quad \quad \quad \hookrightarrow \text{solve a discrete Poisson equation } (2I - \tau \Delta)^{-1} \\ y_{k+1} = \text{Prox}_{\frac{\tau}{2}}^{\|\cdot\|_1} (z_k + \tau D X_{k+1}) \\ z_{k+1} = z_k + \tau (D X_{k+1} - y_{k+1} - c) \end{cases}$$

So three different algorithm on ROF denoising

① (Fast) Proximal Gradient on (D1)

② (Fast) PDHG

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

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

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



DR on $(D - D1)$

$\hookrightarrow$ need to compute

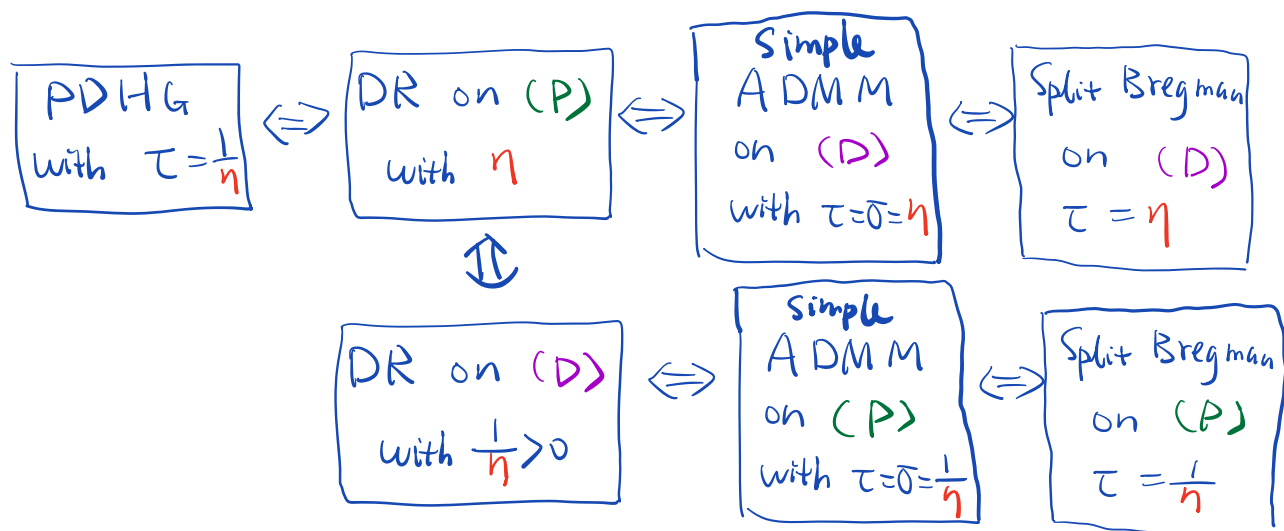
a Poisson equation, much more expensive than

① & ② in each iteration.

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

Such as $\min_x \|x\|_1 + \mathcal{I}_{\{x: Ax=b\}}(x)$

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



Last Topic for Part II

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

Two-operator Douglas-Rachford splitting (1979):

$$\begin{cases} y_{k+1} = \frac{I + R_f^n R_g^n}{2} (y_k) = y_k - x_k + \text{Prox}_f^n(2x_k - y_k) \\ x_k = \text{Prox}_g^n(y_k) \end{cases}$$

Convexity of f & g , $\forall \eta > 0 \Rightarrow$ Convergence

How about $\min_x f(x) + g(x) + h(x)$ (P)

$$F^*(x) + G^*(y) + H^*(z) \quad Ax + By + Cz = d \quad (1)$$

Three-operator splitting Davis-Yin (2016):

$$\begin{cases} y_{k+1} = \text{Prox}_g^n(z_k) \\ x_{k+1} = \text{Prox}_f^n(2y_{k+1} - z_k - \eta \nabla h(y_{k+1})) \\ z_{k+1} = z_k + x_{k+1} - y_{k+1} \end{cases}$$

1) Convexity of f, g, h
2) ∇h is L -continuous
3) $\eta \leq \frac{1}{L}$

$\Rightarrow$ Convergence