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

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

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

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

<u>Lemma</u> $\quad L(x,y) = \langle Kx, y\rangle - f^*(y) + g(x)$

If (x_*, y_*) is a saddle point of (PD),

and B_1 & B_2 are sets s.t. $x_* \in B_1$, $y_* \in B_2$

then the partial duality gap

$$\max_{y \in B_2} L(\hat{x}, y) - \min_{x \in B_1} L(x, \hat{y}) \geq 0$$

and it's $0 \iff (\hat{x}, \hat{y})$ is a saddle point.

Proof: $\quad \max_{y \in B_2} L(\hat{x}, y) - \min_{x \in B_1} L(x, \hat{y})$

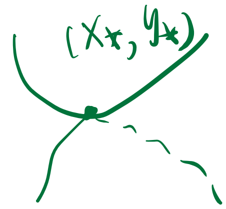
$$\geq L(\hat{x}, y_*) - L(x_*, \hat{y})$$

$$\geq L(\hat{x}, y_*) - L(x_*, y_*)$$

$$+ L(x_*, y_*) - L(x_*, \hat{y})$$

$$\geqslant 0$$

$$\begin{cases} L(\hat{x}, y_*) - L(x_*, y_*) \geqslant 0 \\ L(x_*, y_*) - L(x_*, \hat{y}) \geqslant 0 \end{cases}$$



$\Rightarrow$

It's equal to $0 \Leftrightarrow \begin{cases} L(\hat{x}, y_*) - L(x_*, y_*) = 0 \\ L(x_*, y_*) - L(x_*, \hat{y}) = 0 \end{cases}$

Theorem If $\tau\eta < \rho(K^*K)$, PDHG satisfies

① For closed bounded sets B_1, B_2 containing a saddle

$$0 \leq \max_{y \in B_2} L(\tilde{x}_n, y) - \min_{x \in B_1} L(x, \hat{y}_n) \leq \frac{1}{n} \sup_{x \in B_1} \sup_{y \in B_2} \left(\frac{\|y - y_0\|^2}{\tau} + \frac{\|x - x_0\|^2}{\eta} \right)$$

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

$$\tilde{x}_n = \frac{1}{n} \sum_{k=1}^{n} x_k , \quad \hat{y}_n = \frac{1}{n} \sum_{k=1}^{n} y_k$$

② $$\frac{\|y_k - y_*\|^2}{2\tau} + \frac{\|x_k - x_*\|^2}{2\eta} \leq C \left(\frac{\|y_* - y_0\|^2}{2\tau} + \frac{\|x_* - x_0\|^2}{2\eta} \right)$$

③ There is one saddle point (x_*, y_*) s.t.

$$x_k \to x_*, \quad y_k \to x_*$$

Proof: ① Following last lecture,

$$L(x_{k+1}, y) - L(x, y_{k+1})$$

$$+ \underbrace{\frac{\|y-y_{k+1}\|^2}{2\tau} + \frac{\|x-x_{k+1}\|^2}{2\eta} - \langle K(x-x_{k+1}), y-y_{k+1}\rangle}_{(1)}$$

$$+ \underbrace{\frac{\|y_k-y_{k+1}\|^2}{2\tau} + \frac{\|x_k-x_{k+1}\|^2}{2\eta} - \langle K(x_{k+1}-x_k), y_{k+1}-y_k\rangle}_{(2)}$$

$$\leq \frac{\|y-y_k\|^2}{2\tau} + \frac{\|x-x_k\|^2}{2\eta} - \langle K(x-x_k), y-y_k\rangle$$

(1), (2) have the same form:

$$\frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \langle Ka, b\rangle$$

Cauchy-Schwartz $\geq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \|Ka\|\cdot\|b\|$

$\|Ka\| \leq \|K\|\cdot\|a\|$ $\geq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \|K\|\cdot\|a\|\cdot\|b\|$

$\tau\eta < \frac{1}{\rho(K^*K)} = \frac{1}{\|K\|^2}$ $\geq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \frac{1}{\sqrt{\tau\eta}}\|a\|\cdot\|b\|$

$$= \left\|\frac{a}{\sqrt{2\tau}} - \frac{b}{\sqrt{2\eta}}\right\|^2$$

$(2) \geq 0 \Rightarrow$

$$L(x_{k+1}, y) - L(x, y_{k+1})$$

$$+ \frac{\|y-y_{k+1}\|^2}{2\tau} + \frac{\|x-x_{k+1}\|^2}{2\eta} - \langle K(x-x_{k+1}), y-y_{k+1}\rangle$$

$$\leq \frac{\|y-y_k\|^2}{2\tau} + \frac{\|x-x_k\|^2}{2\eta} - \langle K(x-x_k), y-y_k \rangle$$

Sum over k

$$\Rightarrow \sum_{k=0}^{n-1} \left[L(x_{k+1}, y) - L(x, y_{k+1}) \right]$$

$$+ \underbrace{\frac{\|y-y_n\|^2}{2\tau} + \frac{\|x-x_n\|^2}{2\eta} - \langle K(x-x_n), y-y_n \rangle}_{\geq 0}$$

$$\leq \underbrace{\frac{\|y-y_0\|^2}{2\tau} + \frac{\|x-x_0\|^2}{2\eta} - \langle K(x-x_0), y-y_0 \rangle}_{(3)}$$

$$(3): \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \langle Ka, b \rangle$$

$$\leq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} + \|Ka\| \cdot \|b\|$$

$$\leq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} + \|K\| \cdot \|a\| \cdot \|b\|$$

$$\leq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} + \frac{1}{\sqrt{\tau\eta}} \|a\| \cdot \|b\|$$

$$\leq \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} + \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta}$$

$$\Rightarrow \sum_{k=0}^{n-1} \left[L(x_{k+1}, y) - L(x, y_{k+1}) \right] \leq \frac{\|y-y_0\|^2}{\tau} + \frac{\|x-x_0\|^2}{\eta}$$

Convexity

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

$$k \left[\max_{y \in B_2} L(\tilde{x}_n, y) - \min_{x \in B_1} L(x, \tilde{y}_n) \right] \leq \max_{x \in B_1,\, y \in B_2} \left[\frac{\|y-y_0\|^2}{\tau} + \frac{\|x-x_0\|^2}{\eta} \right]$$

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

(2) $$\underbrace{\sum_{k=0}^{n-1}\Big[L(x_{k+1}, y) - L(x, y_{k+1})\Big]}_{\geq 0 \text{ if } (x,y) \text{ is saddle}}$$

$$+ \frac{\|y-y_n\|^2}{2\tau} + \frac{\|x-x_n\|^2}{2\eta} - \langle K(x-x_n), y-y_n\rangle$$

$$\leq \underbrace{\frac{\|y-y_0\|^2}{2\tau} + \frac{\|x-x_0\|^2}{2\eta} - \langle K(x-x_0), y-y_0\rangle}_{\leq \frac{\|y-y_0\|^2}{\tau} + \frac{\|x-x_0\|^2}{\eta}}$$

$$-\langle Ka, b\rangle \geq -\|K\|\cdot\|a\|\cdot\|b\|$$

$$\geq -\frac{\|K\|\cdot\sqrt{\tau\eta}}{2\tau}\|a\|^2 - \frac{\|K\|\sqrt{\tau\eta}}{2\eta}\|b\|^2$$

$$-2A\cdot B \geq -\alpha A^2 - \frac{1}{\alpha}B^2$$

$$\Rightarrow \frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta} - \langle Ka, b\rangle$$

$$\geq (1-\|K\|\cdot\sqrt{\tau\eta})\left[\frac{\|a\|^2}{2\tau} + \frac{\|b\|^2}{2\eta}\right]$$

$$\Rightarrow \frac{\|y_*-y_n\|^2}{2\tau} + \frac{\|x_*-x_n\|^2}{2\eta} \leq \frac{1}{1-\|K\|\cdot\sqrt{\tau\eta}}\left[\frac{\|y_*-y_0\|^2}{\tau} + \frac{\|x_*-x_0\|^2}{\eta}\right]$$

$$\tau\eta < \frac{1}{\rho(K^*K)} = \frac{1}{\|K\|^2}$$

③ ② $\Rightarrow$ boundedness of $\{x_k\}$ & $\{y_k\}$

$\Rightarrow$ convergence of subsequence

$$\Rightarrow \begin{cases} x_{k_i} \longrightarrow \hat{x} \\ y_{k_j} \longrightarrow \hat{y} \end{cases}$$

keep term (2) before summing over k,

use similar inequality

we can get

$$\frac{\|y_* - y_n\|^2}{2\tau} + \frac{\|x_* - x_n\|^2}{2\eta} + \sum_{k=1}^{n} \frac{\|y_k - y_{k-1}\|^2}{2\tau} + \sum_{k=1}^{n} \frac{\|x_k - x_{k-1}\|^2}{2\sigma}$$

$$\leq \frac{1}{1 - \|K\| \cdot \sqrt{\tau\eta}} \left[\frac{\|y_* - y_0\|^2}{\tau} + \frac{\|x_* - x_0\|^2}{\eta} \right]$$

$$\Rightarrow \begin{cases} x_n \longrightarrow \hat{x} \\ y_n \longrightarrow \hat{y} \end{cases}$$

$\Rightarrow$ $(\hat{x}, \hat{y})$ is a fixed point of PDHG

$\Rightarrow$ a saddle point.

Remark: The $O(\frac{1}{k})$-rate is for the partial duality gap, which is a weak result.

Fast/Accelerated PDHG (Chambolle & Pock 2010)

PDHG is

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

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

Accelerated Version $O\left(\frac{1}{k^2}\right)$ $\quad$ g is strongly convex with $\mu > 0$

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

Compare this with Accelerated Proximal Gradient

Example: 1 $\quad \min_x \|x\|_1 + \frac{1}{2}\|Ax-b\|^2$

Proximal gradient for $\min_x f(x) + g(x)$

$$x_{k+1} = (I+\eta\partial f)^{-1}(I-\eta\partial g)(x_k)$$

$$= S_\eta(x_k - \eta \nabla g(x_k))$$

$$= S_\eta(x_k - \eta A^T(Ax_k - b))$$

PDHG for $\min_x \|x\|_1 + \frac{1}{2}\|Ax-b\|^2$

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

$$f(y) = \frac{1}{2}\|y-b\|^2$$

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

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

Critical Point $\Rightarrow$ $z - \nabla f(y_*) = 0$

$\Rightarrow$ $z - (y_* - b) = 0 \Rightarrow y_* = z + b$

$$\Rightarrow f^*(z) = \langle z, z+b\rangle - f(z+b) = \|z\|^2 + \langle z, b\rangle - \frac{1}{2}\|z\|^2$$

$$= \frac{1}{2}\|z\|^2 + \langle z, b\rangle$$

$$= \frac{1}{2}\|z+b\|^2 - \frac{1}{2}\|b\|^2$$