$\boxed{\text{Convex Prog}}$ 5232 Part3 Lec3

Algorithms & Models

① Sub-gradient calculus ⟷ convex f^{n}

$\xi \in \partial f(\bar{x})$ $\boxed{\text{set-value mapping}}$

$\Leftrightarrow \quad f(x) \geq f(\bar{x}) + \xi^T (x - \bar{x}) \quad \forall x$

$\rightarrow \quad \partial f(x) := \{ g : f(z) - f(x) \geq g^T (z - x), \text{ for all } z \}$

① [prop] x is local minima $\Rightarrow$ $0 \in \partial f(x)$ → optimality condition

f is convex $\Leftrightarrow$ $0 \in \partial f(x)$

② [prop] $\partial f(x)$ singleton

$\Leftrightarrow$ f is differentiable at x and $\{\nabla f(x)\} = \partial f(x)$

$\boxed{\text{Norm-2 Subgradient}}$

$f(\underline{x}) = \|\underline{x}\|_1$

$\Downarrow$

$$\partial f(\underline{x}) = \begin{pmatrix} * \\ \vdots \\ * \end{pmatrix} \qquad * = \begin{cases} +1 & \text{if } x_i > 0 \\ [-1, +1] & \text{if } x_i = 0 \\ -1 & \text{if } x_i < 0 \end{cases}$$

Norm-2 Subgradient

$f(\underline{x}) = \|x\|_2^2 \Rightarrow \partial f(\underline{x}) = 2\underline{x}$

$$f(\underline{x}) = \|x\|_2 \Rightarrow \partial f(\underline{x}) = \begin{cases} \dfrac{x}{\|x\|_2} & \text{if } x \neq 0 \\ 0 & \text{if } \underline{x} = 0 \end{cases}$$

② Proximal Algorithm

$P_f(y) = \arg\min_x \{ f(x) + \frac{1}{2}\|y-x\|^2 \}$

for $f = \lambda\|\cdot\|_1$

generalization

$\text{prox}_{\|\cdot\|_1}(y) = \arg\min_x \frac{1}{2}\|y-x\|_2^2 + \lambda\|x\|_1$

Soft-threshold operator ⟹ stable in de-noising

$$\text{prox}(y)_i = \begin{cases} y_i - \lambda & y_i \geq \lambda \\ 0 & \lambda \geq y_i \geq -\lambda \\ y_i + \lambda, & -\lambda \geq y_i \end{cases}$$

⟹ Shrinkage

P12

$P_f: \quad \hat{x} \in \arg\min_x \frac{1}{2}\|y-x\|_2^2 + \lambda\|x\|_1 := F(x)$

$\Leftrightarrow \quad 0 \in \partial F(\hat{x})$

$\Leftrightarrow \quad 0 \in \hat{x} - y + \lambda\partial\|\hat{x}\|_1$

$\Leftrightarrow \quad \hat{x} \in y - \lambda\partial\|\hat{x}\|_1$

$\Leftrightarrow \quad \hat{x} \in \text{Prox}_{\|\cdot\|_1}(y)$

Q: De-noising

$y = x^* + \epsilon$

observation noise

if x^* is very sparse

(it has 1-nonzero entry)

hard threshold operator → unstable

$$\text{thresh}(y)_i = \begin{cases} y_i & y_i \geq \lambda \\ 0 & \lambda \geq y_i \geq -\lambda \\ y_i & -\lambda \geq y_i \end{cases}$$

$\|P_f(x) - P_f(y)\| \leq \|x-y\| \longrightarrow$ stability of proximal operator

Algorithm part:

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

$$\begin{cases} f(\cdot) \text{ differentiable} \\ g(\cdot) \text{ convex but non-smooth} \end{cases}$$

Proximal Gradient Method

$$x \leftarrow \text{prox}_{\eta g}(\underbrace{x - \eta \nabla f(x)}_{\text{gradient step}})$$

Application $\boxed{\text{LASSO}}$

$$\boxed{\underset{x}{\operatorname{argmin}} \ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1}$$

$$\begin{cases} f(x) := \frac{1}{2}\|y - Ax\|_2^2 \\ g(x) = \lambda\|x\|_1 \end{cases}$$

$\Rightarrow$ algorithm :

step 1 : $(x - \eta A^T(Ax - y)) \to x$

step 2 : $\text{prox}_{\eta g(\cdot)}(x) \to x$

proximal operator with respect to nuclear norm

$$\boxed{P_{nuc}(Y) := \underset{X}{\operatorname{argmin}} \ \frac{1}{2}\|Y - X\|_F^2 + \lambda\|X\|_{nuc}}$$

if $Y = U\Sigma V^T$

$$= U \boxed{P_{nuc}(\Sigma)} V^T$$

easy shrink $\boxed{\min\{\sigma_i - \lambda, 0\}}$ soft threshold

$\boxed{\text{ADMM}}$

$$\begin{cases} \min & f(x) \\ \text{s.t} & Ax=b \end{cases}$$

$$\Rightarrow L(x;y) = f(x) + y^T(Ax-b)$$

Gradient Ascent $\rightarrow$ Saddle $L(x;y)$

$$\begin{cases} x \leftarrow \arg\min L(x,y) & \text{fix } y \rightarrow \text{convex} \\ y \leftarrow y + \eta(Ax-b) \end{cases}$$

$\boxed{\text{Augmented Lagrangian}}$

$$L_A(x;y) = L(x;y) + \frac{\rho}{2}\|Ax-b\|_2^2$$

Large Scale

First-Order : Cheap each + More iteration

$\boxed{\text{ML}}$, we prefer this, we do not need very accurate solution

Second-Order: Expensive + fewer iteration

$\boxed{\text{Optimal Power Flow}}$ $\boxed{\text{2 hour}}$