04/19/18 Proximal algorithms IV.

Saddle-point problem

$$\mathcal{L}(x_{opt}, y_{opt}) = \inf_x \sup_y \mathcal{L}(x,y) = \sup_y \inf_x \mathcal{L}(x,y)$$

ex. (splitting) $\min f(x) + g(z)$ s.t. $x = z$

$\mathcal{L}(x,y,z) = f(x) + g(z) + y^T(x-z)$

$\mathcal{L}(x,y) = -f^*(y) + g(x) + y^T x$

ex. $\min f(x)$ s.t. $Ax = b$

$$\mathcal{L}(x,y) = f(x) + y^T(Ax - b)$$

Uzawa (1958) → Arrow-Hurwicz (1958)

[prox] min in $x$ $\quad x_{k+1} = \arg\min_x \mathcal{L}(x, y_k) \left[+ \frac{1}{2\lambda} \|x - x_k\|_2^2\right]$

prox max in $y$ $\quad y_{k+1} = \arg\min -\mathcal{L}(x_{k+1}, y) + \frac{1}{2\lambda} \|y - y_k\|_2^2$

(+ check Chambolle-Pock modification)

ex. (augmented Lagrangian splitting)

$\min f(x) + g(z) : x = z$

$$\mathcal{L}(x,y,z) = f(x) + g(z) + \frac{1}{\lambda} y^T(x-z) + \frac{1}{2\lambda} \|x - z\|_2^2$$

do not eliminate $z$

(useful for distributed algorithms).

Uzawa becomes (recall $\text{prox}_{\lambda f}(z) = \arg\min_x f(x) + \frac{1}{2\lambda} \|x - z\|_2^2$)

min in $x$ $\quad x_k = \underset{x}{\operatorname{argmin}}\, \mathcal{L}(x, y_k, z_k) \quad \left(= \operatorname{prox}_{\lambda f}(z_k - y_k)\right)$ (1)

min in $z$ $\quad z_{k+1} = \underset{z}{\operatorname{argmin}}\, \mathcal{L}(x_k, y_k, z) \quad \left(= \operatorname{prox}_{\lambda g}(x_k + y_k)\right)$ (2)

prox max in $y$ $\quad y_{k+1} = \underset{y}{\operatorname{argmin}} -\mathcal{L}(x_k, y, z_{k+1}) + \frac{1}{2\lambda}\|y - y_k\|_2^2$

$$\left(= \operatorname{argmin} \frac{-1}{\lambda} y^T (x_k - z_{k+1}) + \frac{1}{2\lambda}\|y - y_k\|_2^2\right.$$

$$= y_k + x_k - z_{k+1} \quad (3)$$

Def. (1), (2), (3) is called ADMM
(alternating-direction method of multipliers)
(aka alternating-split Bregman iteration)

Exercise Same as Douglas-Rachford splitting on the dual (D)
$\left(\max_y -f^*(y) - g^*(-y)\right)$

---

Case Study: matrix completion with the nuclear norm

Ref. Cai, Candès, Shen, 2008.
A singular-value thresholding algorithm for matrix completion.

$$\min \|X\|_* \quad \text{s.t.} \quad P_\Omega X = P_\Omega M$$

($\Omega$ = subset of indices,
$P_\Omega$ = mask on $\Omega$)

with $\|X\|_* = \sum \sigma_i(X)$

<u>Prop.</u> (Proximity operator of the nuclear norm).

$$\underset{X}{\operatorname{argmin}} \|X\|_* + \frac{1}{2\lambda}\|X-Y\|_F^2$$ — Frobenius

$$= S_\lambda(Y)$$

$$:= U S_\lambda(\Sigma) V^*$$

where $Y = U\Sigma V^*$

$S_\lambda(\sigma) = (\sigma-\lambda)_+$

(soft-thresholding)

<u>Pf.</u> strongly convex: unique minimizer iff

$$0 \in \partial\left[\|X\|_* + \frac{1}{2\lambda}\|X-Y\|_F^2\right]$$

$$0 \in \partial\|X\|_* + \frac{1}{\lambda}(X-Y)$$ → STP this is obeyed by $\hat{X} = S_\lambda(Y)$

Recall, for a sequence $\sigma = (\underbrace{\sigma_1, \sigma_2, \ldots, \sigma_k}_{>0}, 0, \ldots, 0)$,

$$\partial\|\sigma\|_1 = \{(1, 1, \ldots, 1, x_{k+1}, \ldots, x_n),\ |x_i| \le 1\}.$$

<u>Lemma</u> (Watson, 1992). Let $\phi(X) = \phi(\Sigma)$ when $X = U\Sigma V^*$.

$$\partial\phi(X) = \operatorname{conv}\{UDV^* : \operatorname{diag}(D) \in \partial\phi(\sigma)\}$$

Here, $\phi(X) = \|X\|_*$, $U = [U_1 | U_2]$, $V = [V_1 | V_2]$

$X = U_1\Sigma_1 V_1^*$ — $\sigma_i = 0$

$$UDV^* = U_1V_1^* + U_2 W V_2^*$$

fixed

$W$ diag with |elements| $\le 1$

arbitrary orthocomplements to $U_1$, $V_1$. (floating)

$$\sum \lambda_i = 1, \quad \lambda_i \geq 0$$

$$U_1 V_1^* + \sum_i \lambda_i U_{2,i} W_i V_{2,i}^*$$

$$= U_1 V_1^* + U^{(2)} \Big( \sum_i \lambda_i Y_i W_i Z_i^T \Big) V^{(2)*}$$

fixed bases of $(\operatorname{Ran} U_1)^\perp$, $(\operatorname{Ran} V_1)^\perp$.

obeys $\|\cdot\| \leq 1$ (spectral norm)

$$\Rightarrow \partial \|X\|_* = \{ U_1 V_1^* + W \;;\; U^* W = 0,\ W V = 0,\ \|W\| \leq 1 \}.$$

---

Let $Y = U_1 \Sigma_1 V_1^* + U_2 \Sigma_2 V_2^*$

$\Sigma_1$: s.v. $> \lambda$ $\qquad$ $\Sigma_2$: s.v. $\leq \lambda$

$$\hat{X} = U_1 (\Sigma_1 - \lambda I) V_1^*$$

$$Y - \hat{X} = \lambda \big( U_1 V_1^* + \underbrace{U_2 D V_2^*}_{W} \big)$$ with $D$ diagonal, $|D_{ii}| \leq 1$.

$\Rightarrow \|W\| \leq 1$.

and $U_1^* W = 0$ because $U_1^* U_2 = 0$

$W V_1 = 0$ because $V_2^* V_1 = 0$

$$\Rightarrow \frac{Y - \hat{X}}{\lambda} \in \partial \|\hat{X}\|_*$$

$$\Rightarrow 0 \in \partial \|\hat{X}\|_* + \frac{1}{\lambda}(\hat{X} - Y)$$

$\square$

Try $\mathcal{L}(X,Y) = \|X\|_* + \langle Y, P_\Omega(M-X)\rangle$

$\rightarrow X=0$ if $\|P_\Omega Y\| \le 1$
or $\inf \mathcal{L} = -\infty$ otherwise.

Idea. $\mathcal{L}_\lambda(X,Y) = \|X\|_* + \frac{1}{2\lambda}\|X\|_F^2 + \frac{1}{\lambda}\langle Y, P_\Omega(M-X)\rangle$

(i.e, augment with a regularization).

$= \|X\|_* + \frac{1}{2\lambda}\|X - P_\Omega Y\|_F^2 + \text{stuff}(Y,M)$

(1) $X_{k+1} = \underset{X}{\operatorname{argmin}}\ \mathcal{L}_\lambda(X, Y_k) = \operatorname{prox}_{\lambda\|X\|_*}(P_\Omega Y_k)$

$= S_\lambda(P_\Omega Y_k)$ use (2)

$\Rightarrow \boxed{X_{k+1} = S_\lambda(Y_k)}$

(2) $Y_{k+1} = \underset{Y}{\operatorname{argmin}}\ -\mathcal{L}_\lambda(X_{k+1}, Y) + \frac{1}{2\delta\lambda}\|Y - Y_k\|_F^2$

$= \underset{Y}{\operatorname{argmin}}\ -\delta\langle Y, P_\Omega(M - X_{k+1})\rangle + \frac{1}{2}\|Y - Y_k\|_F^2$

$= \underset{Y}{\operatorname{argmin}}\ \frac{1}{2}\|Y - (Y_k + \delta P_\Omega(M - X_{k+1}))\|_F^2$

$\Rightarrow \boxed{Y_{k+1} = Y_k + \delta P_\Omega(M - X_{k+1})}$

Compare with FB splitting for $\|X\|_* + \frac{1}{2\lambda}\|P_\Omega(M-X)\|_F^2$

$X_{k+1} = S_\lambda(X_k + \delta P_\Omega(M - X_k))$

$$\begin{cases} X_{k+1} = S_\lambda(Y_k) \\ Y_k = X_k + \delta P_\Omega(M - X_k) \end{cases}$$

Very different! Approaches solution from high ranks ($\lambda \to 0$) rather than low rank ($\lambda \to \infty$)