04/12/18. Proximal algorithms II.

Splitting for $\min f(x)+g(x)$.

- gradient step $x \to [I-\lambda\nabla f](x)$
- proximal step

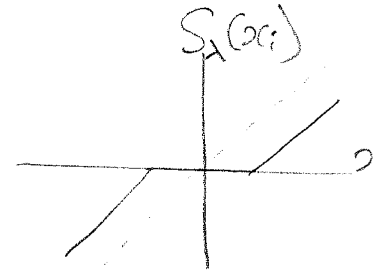
$$x \to \mathrm{prox}_{\lambda g}(x) = \operatorname*{argmin}_y g(y)+\frac{1}{2\lambda}\|y-x\|_2^2$$
$$= [I+\lambda\partial g]^{-1}(x).$$

Forward-backward splitting:

$$\begin{cases} \tilde{x}_k = x_k - \lambda\nabla f(x_k) \\ x_{k+1} = \mathrm{prox}_{\lambda g}(\tilde{x}_k) \end{cases}$$

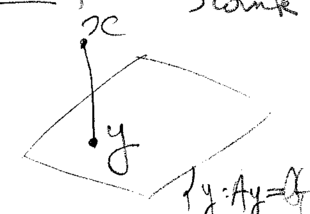
ex. $g(x)=\|x\|_1$

$\mathrm{prox}_{\lambda g}(x) = S_\lambda(x)$
(shrinkage)



ex $g(x) = \mathbb{1}_{Ax=b}(x)$ $\quad\boxed{\quad A \quad}$, full row rank

$\mathrm{prox}_{\lambda g}(x) \doteq \mathrm{Proj}_{\{Ax=b\}}(x)$



$x-y \perp \mathrm{Null}(A)$
$x-y \in \mathrm{Ran}\, A^T$
$x-y = A^T\lambda$
$Ax - \underbrace{Ay}_{b} = AA^T\lambda \implies \lambda = (AA^T)^{-1}(Ax-b)$
$y = x - A^T(AA^T)^{-1}(Ax-b)$
$y = x - A^+(Ax-b)$

<u>ex</u>. $g(x) = \frac{1}{2}\|Ax - b\|_2^2$ → seen in homework.

Can also get prox by duality:

$$f^*(y) = \sup\left(y^T x - f(x)\right)$$

<u>Thm</u> (Moreau identity)

$$x = \text{prox}_f(x) + \text{prox}_{f^*}(x)$$

or more generally for $\lambda \neq 1$,

$$x = \text{prox}_{\lambda f}(x) + \lambda\, \text{prox}_{\frac{1}{\lambda} f^*}\left(\frac{x}{\lambda}\right).$$

<u>ex</u>. $g(x) = \|x\|_1$, $\quad g^*(y) = \mathbb{1}_{\|y\|_\infty \leq 1}(y)$

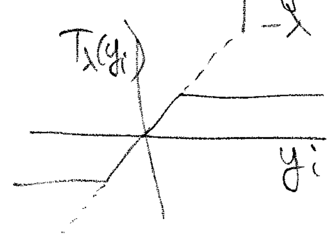
$$= \begin{cases} 0 & \text{if } \|y\|_\infty \leq 1 \\ \infty & \text{else} \end{cases}$$

$$\text{prox}_{g^*}(y) = \text{Proj}_{\|y\|_\infty \leq 1}(y)$$

$$\lambda\, \text{prox}_{\frac{g^*}{\lambda}}\left(\frac{y}{\lambda}\right) = \text{Proj}_{\|y\|_\infty \leq \lambda}(y) \to \begin{cases} \lambda & y_i > \lambda \\ y_i & |y_i| \leq \lambda \\ -\lambda & y_i < -\lambda \end{cases}$$

$$= T_\lambda(y)$$

(truncation)



$$\text{prox}_{\lambda g}(x) = x - T_\lambda(x) = S_\lambda(x)$$

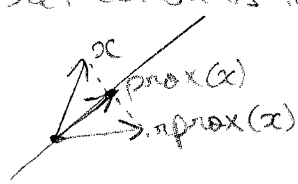
Splitting for $\min f(x)+g(x)$
when both are proximable : usually better than proximal gradient.

Rmk: $x_{k+1} = \text{prox}_{\lambda g}(\text{prox}_{\lambda f}(x_k))$

has poor convergence in general, unless
$f(x) = \mathbb{1}_A(x)$; $g(x) = \mathbb{1}_B(x)$
with A, B convex $\to$ "projection onto convex sets" (POCS).

Def. $\text{rprox}_f(x) = 2\,\text{prox}_f(x) - x$

(is to prox what a reflection is to a projection :



)

Def. (Peaceman-Rachford)

$$x_{k+1} = \text{rprox}_{\lambda g}(\text{rprox}_{\lambda f}(x_k))$$

$\to$ converges $(\forall \lambda)$

Def. (Douglas-Rachford)

$$\begin{cases} \tilde{x}_{k+1} = \frac{1}{2}\left(I + \text{rprox}_{\lambda g} \circ \text{rprox}_{\lambda f}\right)(\tilde{x}_k) \\ x_{k+1} = \text{prox}_{\lambda f}(\tilde{x}_{k+1}) \end{cases}$$

$\to$ converges as well $(\forall \lambda)$

(Recall $f^*(y) = \sup_x (y^T x - f(x))$)

Primal-dual splitting for $\min f(x) + g(x)$:

(P) $\left[\min_{x=z} f(z) + g(x) \right.$

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

(PD) $\left[\min_x \max_y -f^*(y) + g(x) + y^T x \right.$

$$\min_x \min_z \mathcal{L}(x,y,z) = -f^*(y) + \underbrace{\min_x (g(x) + y^T x)}_{-g^*(-y)}$$

(D) $\left[\max_y -f^*(y) - g^*(-y) \right.$

In (PD), proximal maximization for y, from (x_k, y_k)

$$\min_y \; f^*(y) - \underbrace{y^T x_k + \frac{1}{2\lambda}\|y - y_k\|_2^2}_{\frac{1}{2\lambda} y^T y - \frac{1}{\lambda} y^T (y_k + \lambda x_k) + \frac{1}{2\lambda} y_k^T y_k}$$

$$= \frac{1}{2\lambda}\|y - (y_k + \lambda x_k)\|_2^2 + \text{stuff}(x_k, y_k)$$

$$\Rightarrow y_{k+1} = \text{prox}_{\lambda f^*}(y_k + \lambda x_k) \qquad (1)$$

Proximal minimization for x, from (x_k, y_{k+1})

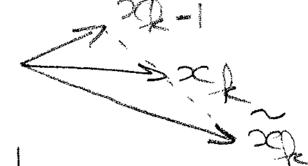
$$\min_x \; g(x) + \underbrace{y_{k+1}^T x + \frac{1}{2\lambda}\|x - x_k\|_2^2}_{\frac{1}{2\lambda}\|x - (x_k - \lambda y_{k+1})\|_2^2 + \text{stuff}(x_k, y_{k+1})}$$

$$\Rightarrow x_{k+1} = \text{prox}_{\lambda g}(x_k - \lambda y_{k+1}) \qquad (2)$$

<u>Def</u>. (1),(2) is the Arrow-Hurwicz method

<u>Def</u>. (Chambolle-Pock, 2011).

$$\begin{cases} y_{k+1} = \text{prox}_{\lambda f^*}(y_k + \lambda \bar{x}_k) \\ x_{k+1} = \text{prox}_{\lambda g}(x_k - \lambda y_{k+1}) \\ \bar{x}_{k+1} = 2x_{k+1} - x_k \end{cases}$$ (linear extrapolat°)



<u>Exercise</u> Same as Douglas-Rachford, if $\tilde{x}_{k+1} = x_k - \lambda y_{k+1}$

Augmented-Lagrangian splitting for $\min f(x)+g(x)$

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

Proximal minimization for x, from (y_k, z_k)

$$x_k = \text{prox}_{\lambda f}(y_k - z_k) \quad (3)$$

Proximal minimization for y, from (x_k, z_k)

$$y_{k+1} = \text{prox}_{\lambda g}(x_k + z_k) \quad (4)$$

Proximal maximization for z,

$$\min -\frac{1}{\lambda} z^T(x-y) + \frac{1}{2\lambda}\|z - z_k\|_2^2$$

$$\rightarrow z_{k+1} = z_k + x_k - y_{k+1} \quad (5)$$

<u>Def</u>. (3),(4),(5) is the Alternating-direction method of multipliers (ADMM).
AKA alternating split Bregman (in the context of ℓ_1)

~~Exercise: Same as Douglas-Rachford splitting on the dual (D).~~

Also: variants for $\min f(x) + g(Ax)$ (in the (PD) formulation), so only $\text{prox}_{\gamma g}$ needs to be computed.

Ref.

- Combettes, Pesquet, Proximal splitting methods in signal processing, 2010.
- Chambolle, Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, 2010.
- Demanet, Zhang, Eventual linear convergence of the DR iteration for basis pursuit, 2013
- Peyré, A review of proximal methods, with a new one, slides, 2012 (see also numerical tours)