Amath 515, March 8th

Recap:

$$\min_x \; g(x) + h(Ax - b) - \langle c, x \rangle$$

$$\max_z \; -\langle z, b \rangle - h^*(z) - g^*(-A^T z + c)$$

Problem without red additions will lead us to the Chambolle-Pock algorithm for $\min_x g(x) + h(Ax)$

$\rightarrow$ match the algo in their paper.

* Add red terms, will get a slightly modified version we need for Lasso on hw.

Opt conditions:

$$\begin{bmatrix}0\\0\end{bmatrix} \in \overbrace{\left(\begin{bmatrix}\partial g & 0\\ 0 & \partial h^*\end{bmatrix} + \begin{bmatrix}0 & A^T\\ -A & 0\end{bmatrix}\right)}^{T} \overset{\bar{w}}{\begin{bmatrix}\bar{x}\\ \bar{z}\end{bmatrix}}$$

<u>Iterations:</u>

* ① Basic: $w^+ = (I+T)^{-1} w$

** ② 'Tricky': $w^+ = (I+\underline{\underline{P}}^{-1}T)\, w$

Chambolle-Pock iteration

Shown T is <u>monotone</u>

$$\langle x-y,\ u-v\rangle \geq 0 \quad \forall x,y \quad \begin{matrix} u\in Tx \\ v \in Ty\end{matrix}$$

$$\langle x-y,\ Tx-Ty\rangle \geq 0$$

<u>Plan:</u>

① Show $(I+T)^{-1}$ is <u>FNE</u>

② pf sketch for convergence of *, **

③ Understand 'Trick', and P

Thm: T monotone $\Rightarrow$ $S = (I+T)^{-1}$ FNE

Pf: pick x, $u \in Sx$, y, $v \in Sy$

$u \in Sx$ means $u \in (I+T)^{-1}x$

multiply by $I+T$, get

$$x \in \underbrace{(I+T)u}_{u+Tu}$$

$$x-u \in Tu$$

$v \in Sy$ means $y - v \in Tv$

T monotone means that

$$\langle u - v, \underbrace{(x-u)}_{\substack{\uparrow \\ Tu \\ x - Sx}} - \underbrace{(y-v)}_{\substack{\uparrow \\ Tv \\ y - Sy}} \rangle \geq 0 \quad \checkmark$$

$\star$

$$\langle Sx - Sy, (I-S)x - (I-S)y \rangle \geq 0$$

Choice of x, y arbitrary, so concl. holds for all x, y.

$\star$ holding for all x, y means S is FNE by hw 3b.

<u>Thm:</u> Krasnoselskii-Mann

<u>Informal:</u> Any Lip-1 (NE) operator can be <u>tweaked</u> so repeated application of it converges to its fixed pts,

<u>Cor:</u> If S is <u>FNE</u>, then don't need 'tweak', so

$w^+ = Sw$ conv to fixed points of S.

Formal KM statement:

Let R be NE, $\|Rx - Ry\| \le \|x-y\|$.

Suppose Fix R, $\{x : Rx = x\} \notin \phi$.

For $\lambda \in (0,1)$, define 'tweak'

$R_\lambda w = (1-\lambda) w + \lambda R w.$

① $\text{Fix } R = \text{Fix } R_\lambda$ (same fixed pts)

② $w^+ = R_\lambda w$ conv. to Fix R at a sublinear rate.

Pf details: By hw 2b, we have for any $\bar{w} \in \text{Fix } R$,

$$\|R_\lambda w - \bar{w}\|^2 \le \|w - \bar{w}\|^2 - \lambda(1-\lambda)\|Rw - w\|^2$$

$w^{k+1} = R_\lambda w^k$ our iteration, can get

$$\lambda(1-\lambda)\|w^k - Rw^k\|^2 \le \|w^k - \bar{w}\|^2 - \|w^{k+1} - \bar{w}\|^2$$

Telescopes

add up across $i = 0, \dots, k$, get

$$\min_{0 \le i \le k} \|Rw^i - w^i\|^2 \le \frac{1}{k}\sum_{i=1}^{k} \|Rw^i - w^i\|^2$$

$$\| \quad \frac{\|w_0 - \bar{w}\|^2}{k(\lambda)(1-\lambda)}$$

Big picture:

some work we do not show here.

1. $\{w^k\}$ stay bounded
2. $\{w^k\}$ have unique limit point
3. Our detail above gives a 'sublinear conv' rate for this limit point.

Cor: If S is FNE, RM simplifies to $w^+ = S w$.

Pf: S FNE $\iff$ $2S - I$ NE by HW 3C

Apply RM to $2S - I$, $\lambda = 1/2$

$$\underbrace{(2S-I)}_{R}\underbrace{{}_{1/2}}_{\lambda} w = \frac{1}{2} w + \frac{1}{2}(2S-I)w$$

$$= S w$$

Summary: T monotone, then

$$w^+ = (I+T)^{-1} w$$ will

converge to fixed pts of $(I+T)^{-1}$, which satisfy $0 \in T\bar{w}$

Aside: $\min_{x} f(x)$, f convex, closed, proper, ...

$$0 \in \underbrace{\partial f}_{T} (\bar{x})$$

Iteration: $\underline{x^+ \in (I + \partial f)^{-1} x}$

what is this ??

$$x \in (I + \partial f) x^+$$

$$x \in x^+ + \partial f(x^+)$$

$$\underline{x - x^+ \in \partial f(x^+)}$$

opt for prox

$$\underline{x^+ = \mathrm{prox}_f (x)}$$

* Proximal point algorithm

Why is this not as satisfying as it seems?

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

$$x^+ = (I+T)^{-1} x$$

$$\downarrow$$

$$\partial f$$

$$x^+ = \overbrace{\text{prox}}_{g+h(A\cdot)}(x)$$

$$= \underset{y}{\text{argmin}} \ \underline{\tfrac{1}{2}\|x-y\|^2} + g(y) + h(Ay)$$

Our T was

$$\begin{bmatrix} \partial f & \\ & \partial h^* \end{bmatrix} + \begin{bmatrix} 0 & -A^T \\ -A & 0 \end{bmatrix}$$

If we try $(I+T)^{-1}$, we will find we can't easily implement it.

Trick: want to use a 'preconditioner' P.

Let P be any pos def matrix, we can define a new inner product:

$$\langle x, y\rangle_P : x^T P y$$

Fact: $P^{-1}T$ is monotone with respect to $\langle x, y\rangle_P$ $\iff$ T is monotone.

Allows us to use iteration

* $w^+ = (I + P^{-1}T)^{-1} w$ for any P pos. def.

$$w \in (I + P^{-1}T) w^+$$

$$w - w^+ \in P^{-1} T w^+$$

* $P(w - w^+) \in T w^+$

Our optimality conditions:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} A^T z + \partial g(x) \\ -Ax + \partial h^*(z) \end{bmatrix}}_{Tw}, \quad w = \begin{bmatrix} x \\ z \end{bmatrix}$$

Clever P:

$$\begin{bmatrix} \frac{1}{\gamma} I & -A^T \\ -A & \frac{1}{\gamma} I \end{bmatrix}$$

① Can pick γ to make sure P is pos def

② P was designed to make the iteration $w^+ = (I + P^{-1}T)^{-1} w$ implementable.

Let's now compute w^+ via

$$P(w - w^+) \in T w^+$$

$$\underbrace{\begin{bmatrix} \frac{1}{\gamma} I & -A^T \\ -A & \frac{1}{\gamma} I \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} x - x^+ \\ z - z^+ \end{bmatrix}}_{w - w^+} \in \underbrace{\begin{bmatrix} A^T z^+ + \partial g(x^+) - c \\ -A x^+ + \partial h^*(z^+) + b \end{bmatrix}}_{T w^+}$$

Row 1:

$$\frac{1}{\gamma}(x - x^+) - A^T(z - \cancel{z^+}) \in \cancel{A^T z^+} + \partial g(x^+) - c$$

$$\frac{1}{\gamma}(x - x^+) - A^T z \in \partial g(x^+)$$

$$\frac{1}{\gamma}(x - \gamma A^T z - x^+) \in \partial g(x^+)$$

$$x^+ = \text{prox}_{\gamma g}(x - \gamma A^T z)$$

$$\text{prox}_{\gamma g}(x - \gamma(A^T z - c))$$

Row z:

$$-A(x-x^+) + \frac{1}{\gamma}(z-z^+) \in -Ax^+ + \partial h^*(z^+)$$

$$\frac{1}{\gamma}(z-z^+) \underline{- Ax + 2Ax^+} \in \cancel{-Ax^+} + \partial h^*(z^+)$$

$$\underline{\frac{1}{\gamma}(z - \gamma A(x-2x^+) - z^+)} \in \partial h^*(z^+)$$

$$z^+ = \mathrm{prox}_{\gamma h^*}\left(z - \gamma A(x-2x^+)\right)$$

$$z^+ = \mathrm{prox}_{\gamma h^*}\left(z - \gamma A(x-2x^+) \underline{- \gamma b}\right)$$

Derived the Chambolle-Pock iteration (see their paper)

Try to make things easier in hw

$r_x = b - Ax, \quad r_x^+ = b - Ax^+$

$r_z = c - A^T z$

Iteration in the code:

$$x^+ = \text{prox}_{\gamma g}(x - \gamma(r_z))$$

$$z^+ = \text{prox}_{\gamma h^*}(z + \gamma(2r_x^+ - r_x))$$

$z \longleftrightarrow v$ in code

in lecture

$g \longleftrightarrow \underline{\underline{k}}$ in code