Lec 8 ADMM cont'd

8-1 We are solving $\min_{x\in\mathbb{R}^n,\, z\in\mathbb{R}^m} f(x)+g(z)$

$$\text{s.t. } Ax+Bz=0$$

Opt conds:

primal feas: $Ax^*+Bz^*-c=0$ (1)

dual feas: $\exists y^*$ s.t. $\inf \underbrace{L_0(x,z,y^*)}_{f(x)+g(z)+y^{*T}(Ax+Bz-c)} > -\infty$

— a convex fun

i.e. $\exists x^*, z^*$ s.t.

$$\begin{bmatrix}0\\0\end{bmatrix} \in \begin{bmatrix}\partial f(x^*)+A^Ty^*\\ \partial g(z^*)+B^Ty^*\end{bmatrix}\begin{matrix}(+0+0) & (2)\\ (+0+0) & (3)\end{matrix}$$

RHS = set + vector = set.

Recall $L_\rho(x,y,z)=f(x)+g(z)+y^T(Ax+Bz-c) + \frac{\rho}{2}\|Ax+Bz-c\|_2^2$

ADMM is:

$$x^{k+1} = \arg\min_x L_\rho(x, \underline{\underline{z^k}}, y^k)$$

$$z^{k+1} = \arg\min_z L_\rho(\underline{\underline{x^{k+1}}}, z, y^k)$$

$$y^{k+1} = y^k + \rho\underbrace{(A\underline{\underline{x^{k+1}}} + B\underline{\underline{z^{k+1}}} - c)}_{r^{k+1}}$$

so, looking at $\underline{z^{k+1}}$ first,

$$0\in \partial g(z^{k+1}) + B^Ty^k + \rho B^T\underbrace{(Ax^{k+1}+Bz^{k+1}-c)}_{r^{k+1}}$$

$$0\in \partial g(z^{k+1}) + B^Ty^{k+1}.$$

So z^{k+1}, y^{k+1} always satisfy (3).

We only need to focus on (1), (2). (cf. method of multipliers)

Let's turn to x^{k+1}. By def'n.

$$0 \in \partial f(x^{k+1}) + A^T y^k + \rho A^T \underbrace{(Ax^{k+1} + Bz^k - c)}_{r^{k+1} + B(z^k - z^{k+1})}$$

$$0 \in \partial f(x^{k+1}) + A^T y^{k+1} + \rho A^T B(z^k - z^{k+1})$$

i.e. $\underbrace{\rho A^T B(z^{k+1} - z^k)}_{\text{call this } s^{k+1}} \in \partial f(x^{k+1}) + A^T y^k$

"dual residual"

as opposed to "primal residual" r^{k+1}.

In the convergence proof, we'll see $B(z^{k+1} - z^k) \to 0$, implying $s^{k+1} \to 0$.

Recall Scaled Version of ADMM.

$$x^{k+1} = \arg\min_x \left(f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + u^k\|_2^2\right)$$

$$z^{k+1} = \arg\min_z \left(g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + u^k\|_2^2\right)$$

$$u^{k+1} = u^k + r^{k+1}.$$

We'll use the notation $x \equiv x^k$, etc, $x^+ = x^{k+1}$:

$$x^+ = \arg\min_x \left(f(x) + \frac{\rho}{2}\|Ax - v\|_2^2\right)$$

where $v = -Bz + c - u$.

Specific Applications

1. $A = I$

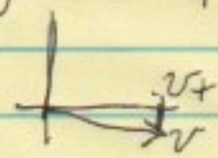
Then $x^+ = \arg\min_x f(x) + \frac{\rho}{2}\|x - v\|_2^2$

$= \text{prox}_{f,\rho}(v)$ "proximal operator"

Suppose $f(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C \end{cases}$ "indicator function" for convex set C.

Then $x^+ = \Pi_C(v)$ projection of v onto C.

(ind. of ρ). E.g. $C = \mathbb{R}^n_+$, then $x^+ = \underbrace{v_+}_{\text{proj of } v \text{ onto } \mathbb{R}^n}$



2. $f(x) = \frac{1}{2}x^T P x + q^T x \qquad P \in S^n_{++}$

Then x^+ minimizes $\frac{1}{2}x^T P x + q^T x + \frac{\rho}{2}(Ax - v)^T(Ax - v)$

$= \frac{1}{2}x^T(P + \rho A^T A)x + (q - \rho A^T v)^T x + \text{const.}$

Differentiate: get

$$x^+ = (P + \rho A^T A)^{-1}(\rho A^T v - q)$$

3. "Soft Thresholding"

$f(x) = \lambda\|x\|_1 \qquad \lambda > 0$

$A = I$

$x^+ = \arg\min \lambda\|x\|_1 + \frac{\rho}{2}\|x - v\|_2^2$

separable:

$$x_i^+ = \arg\min \underbrace{\lambda|x_i| + \frac{\rho}{2}(x_i - v_i)^2}_{h^{(i)}(x)}$$

Need $0 \in \partial h^{(i)}(x_i) = \begin{cases} [-\lambda, \lambda] - \rho v_i & \text{if } x_i = 0 \\ \lambda \, \mathrm{sgn}(x_i) + \rho x - \rho v_i & x_i \neq 0 \end{cases}$

We see $0 \in \partial h^{(i)}(0)$ iff $\rho |v_i| \leq \lambda$

so $x_i^+ = 0$ if $|v_i| \leq \frac{\lambda}{\rho}$.

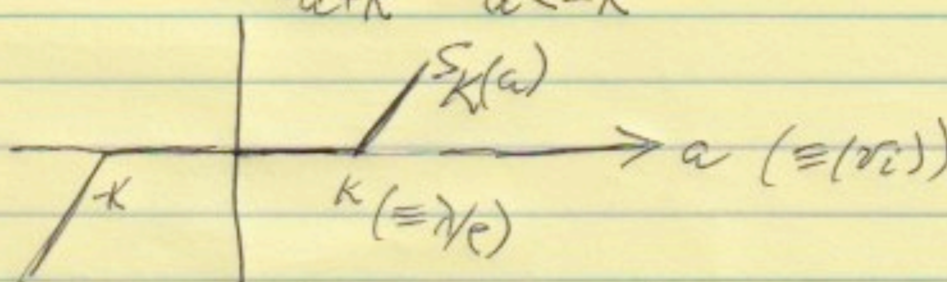
Otherwise need $\frac{\lambda}{\rho} \mathrm{sgn}(x_i^+) + x_i^+ - v_i = 0$

so if $v_i > \frac{\lambda}{\rho}$, set $x_i^+ = v_i - \frac{\lambda}{\rho} > 0$

and if $v_i < \frac{\lambda}{\rho}$, set $x_i^+ = v_i + \frac{\lambda}{\rho} < 0$ } satisfy

So $x_i^+ = S_{\lambda/\rho}(v_i)$

where $S_K(a) = \begin{cases} a - K & a > K \\ 0 & |a| \leq K \\ a + K & a < -K \end{cases}$



"prox operator for the l_1 norm"

Write in vector form

$x^+ = S_{\lambda/\rho}(v)$.

4. LASSO "Least Absolute Shrinkage + Selection Operator"

NOT always used to mean the same thing:

$$\min \frac{1}{2}\|Ax-b\|_2^2 + \underbrace{\lambda\|x\|_1}$$

"encourage" "sparsity" in x^*

larger λ: more components of x^* will be 0.

ADMM formulation:

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

$$g(z) = \lambda\|z\|_1$$

Constraint: $x = z$ or $A = I$, $B = -I$, $c = 0$

Then ADMM becomes

$$x^{k+1} = \arg\min_x \frac{1}{2}\|Ax-b\|_2^2 + \frac{\rho}{2}\|x - z^k + u^k\|_2^2$$

Diff: $0 = (A^TA + \rho I)x - A^Tb - \rho(z^k - u^k)$

Solve using Cholesky factor of $A^TA + \rho I$

$$z^{k+1} = \arg\min \lambda\|z\|_1 + \frac{\rho}{2}\|x^{k+1} - z + u^k\|_2^2$$

$$= S_{\lambda/\rho}(x^{k+1} + u^k)$$

$$u^{k+1} = u^k + \underbrace{x^{k+1} - z^{k+1}}_{r^{k+1}}$$

8-6 Convergence Analysis

Recall assumptions

1: f, g are convex, closed + proper

2. L_0 has a saddle pt (x^*, z^*, y^*) (not nec. unique.)

WTS $r^k \to 0$, $p^k = f(x^k) + g(z^k) \to p^*$,

and $s^k = \rho A^T \underbrace{B(z^k - z^{k+1})}_{\text{in fact } \to 0} \to 0$

Let $V^k = \frac{1}{\rho}\|y^k - y^*\|_2^2 + \rho\|B(z^k - z^*)\|_2^2$

Want to show

$$V^{k+1} \le V^k - \rho\|r^{k+1}\|_2^2 - \rho\|B(z^{k+1} - z^k)\|_2^2. \quad (A.1)$$

If (A.1) holds, then $\{y^k\}$ and $\{Bz^k\}$ are bounded as $V^k \le V^0$

Also $V^0 - V^1 \ge \rho\|r^1\|_2^2 + \rho\|B(z^1 - z^0)\|_2^2$

$V^1 - V^2 \ge \rho\|r^2\|_2^2 + \rho\|B(z^2 - z^1)\|_2^2$

$\vdots$

Sum up:

$$V^0 \ge \rho\sum_{k=0}^{\infty}\|r^k\|^2 + \rho\sum_{k=0}^{\infty}\|B(z^k - z^{k-1})\|_2^2$$

so $r^k \to 0$ and $B(z^k - z^{k-1}) \to 0$ as $k \to \infty$.

so $s^k \to 0$ also.

Also want to show

$$p^{k+1} - p^* \le -(y^{k+1})^T \underbrace{r^{k+1}}_{\to 0} - \rho\left(\overbrace{B(z^{k+1} - z^k)}^{\to 0}\right)^T \left(\underbrace{-r^{k+1}}_{\to 0} + \underbrace{B(z^{k+1} - z^*)}_{BD}\right) \quad (A.2)$$

and

$$p^* - p^{k+1} \le (y^*)^T \underbrace{r^{k+1}}_{\to 0} \quad (A.3)$$

So (A.1), (A.2), (A.3) $\Longrightarrow$

$$\underbrace{|p^{k+1} - p^*|}_{\text{could have + or − sign}} \to 0$$

as iterates are not feasible.

<u>Proof of (A.3)</u>

We have $p^* \le L_0(x^{k+1}, z^{k+1}, y^*)$,

$$\overset{\Vert}{L_0(x^*, z^*, y^*)} \qquad \underbrace{f(x^{k+1}) + g(z^{k+1})}_{p^{k+1}} + y^{*}r^{k+1}$$

✓

<u>Proof of (A.2)</u>

x^{k+1} minimizes $L_\rho(x, \underline{z^k}, y^k)$, so

$$0 \in \partial_x L_\rho(x^{k+1}, z^k, y^k) = \partial f(x^{k+1}) + A^T y^k + \rho A^T(Ax^{k+1} + Bz^k - c)$$

so

$$0 \in \partial f(x^{k+1}) + A^T\big(\underbrace{y^k + \rho r^{k+1}}_{y^{k+1}} - \rho B(z^{k+1} - z^k)\big)$$

(already stated on p. 8-2)

so x^{k+1} minimizes $f(x) + (y^{k+1} - \underline{\rho B(z^{k+1} - z^k)})^T A x$

Also

z^{k+1} minimizes $L_\rho(\underline{x^{k+1}}, z, y^k)$, so

$$0 \in \partial g(z^{k+1}) + B^T(\underbrace{y^k + \rho r^{k+1}}_{y^{k+1}})$$

(already stated on p. 8-1)

so z^{k+1} minimizes $g(z) + (y^{k+1})^T B z$.

It follows that

$$f(x^{k+1}) + \left(y^{k+1} - \rho B(z^{k+1} - z^k)\right)^T A x^{k+1}$$
$$\leq f(x^*) + \left(y^{k+1} - \rho B(z^{k+1} - z^k)\right)^T A x^*$$

and

$$g(z^{k+1}) + (y^{k+1})^T B z^{k+1}$$
$$\leq g(z^*) + (y^{k+1})^T B z^*$$

Now add these inequalities:

$$p^{k+1} + (y^{k+1})^T (A x^{k+1} + B z^{k+1}) - \rho\left(B(z^{k+1} - z^k)\right)^T A x^{k+1}$$
$$\leq p^* + (y^{k+1})^T (\underbrace{A x^* + B z^*}_{c}) - \rho\left(B(z^{k+1} - z^k)\right)^T A x^*$$

$$p^{k+1} - p^* \leq (y^{k+1})^T (-r^{k+1}) - \rho\left(B(z^{k+1} - z^k)\right)^T \underbrace{A(x^* - x^{k+1})}_{\overbrace{(\underbrace{Ax^* + Bz^*}_{c}) - (Ax^{k+1} + Bz^{k+1} + B(z^* - z^{k+1}))}^{-r^{k+1} + B(z^{k+1} - z^*)}}$$

which is (A.2) ✓

<u>Pf of (A.1)</u>: Takes another 1.5 pages! omit!