7-1

In preparation for ADMM lecture:

Subgradients + Subdifferential of Convex Functions

Oddly, not in BV.

Assume f is convex + proper: $\exists x$ s.t. $f(x) < +\infty$ + closed: all sublevel sets are closed.

$f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. $\forall x,\ f(x) > -\infty$.

Def $y \in \mathbb{R}^n$ is a subgradient of f at x if

$$f(x+z) \geq f(x) + y^T z \quad \forall z \in \mathbb{R}^n$$

must include end pt if jumps up to ∞ beyond

But $f = \begin{cases} +\infty & x \leq 0 \\ -\log x & x > 0 \end{cases}$ IS CLOSED although dom f is NOT

$n = 1$: y is the slope of a line passing through $(x, f(x))$ and lying underneath the graph of x

a proper convex fun is closed iff it is LSC.

$n \geq 1$ (NOT ≥ 1) $\begin{bmatrix} y \\ -1 \end{bmatrix}$ is normal to a hyperplane in $\mathbb{R}^{n+1}$ passing through $\begin{bmatrix} x \\ f(x) \end{bmatrix}$ and lying below the graph of f.

The set of all subgradients of f at x is denoted $\partial f(x)$; the SUBDIFFERENTIAL of f at x

e.g. $f(x) = |x|$, $\partial f(0) = [-1, 1]$.

If f is differentiable at x then
$$\partial f(x) = \{\nabla f(x)\}.$$
In fact this is IFF.

<u>Note</u> For $x \in \operatorname{int} \operatorname{dom} f$, $\partial f(x)$ is always a CLOSED, CONVEX, NON-EMPTY, COMPACT set.

e.g. $f(x) = \max_{1 \le i \le n} (x_i)$ $\quad (= x_{[1]}$ in BV notation$)$

What is $\partial f(x)$ for $x = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 2 \\ 3 \end{bmatrix}$? Need

$$\max\left(\begin{bmatrix} 1+z_1 \\ 3+z_2 \\ 2+z_3 \\ 2+z_4 \\ 3+z_5 \end{bmatrix}\right) \ge 3 + y^T z \quad \forall z \in \mathbb{R}^n$$

Clearly $e_1 \notin \partial f(x)$ as RHS is $3+z_1$ (take $z=e_1$)

$e_2 \in \partial f(x)$ as RHS is $3+z_2$.

In fact $\partial f(x) = \operatorname{conv}(e_2, e_5) = \left\{ \begin{bmatrix} 0 \\ \tau \\ 0 \\ 0 \\ 1-\tau \end{bmatrix} : \tau \in [0,1] \right\}$

Does this remind you of something?

Answer: (Fenchel) conjugate.

<u>THM</u> (Fenchel-Young)
$$f(x) + f^*(y) \ge x^T y$$
with equality IFF $y \in \partial f(x)$.

Pf: HW.

Relationship to Directional Derivative

$$f'(x;d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}$$

Thm $y \in \partial f(x)$ iff $y^T d \le f'(x;d)\ \forall d \in \mathbb{R}^n$.

Pf HW.

VERY IMPORTANT

Chain Rule - simplest version.

More general versions: Borwein Lewis p. 52
Rockafellar p. 225.

Let $f : \mathbb{R}^n \to \mathbb{R}$ convex, dom $f = \mathbb{R}^n$.
Let $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$.
Let h be the convex function on $\mathbb{R}^m$ defined by

$$h(\xi) = f(A\xi + b) \quad \xi \in \mathbb{R}^m.$$

Then $\partial h(\xi) = \underbrace{A^T \partial f(A\xi + b)}_{\text{means}}$

$$\{A^T y : y \in \partial f(A\xi + b)\}$$

Works even if A does not have full rank
e.g. $A = 0$.

VERY IMPORTANT

Optimality Condition

$$0 \in \partial f(x) \iff x \text{ is global minimizer of } f.$$

Pf: immediate from def'n.
Generalizes the familiar $\nabla f(x) = 0$ in differentiable case.

7-4 From Boyd et al, ADMM paper

Sec 2 Precursors.

Dual Ascent

Consider $\min f(x)$ s.t. $Ax = b$ $\quad$ convex $f: \mathbb{R}^n \to \mathbb{R}$, A is $m \times n$ $\quad$ (P)

Lagrangian $L(x,y) = f(x) + y^T(Ax - b)$

formerly ν.

L. Dual fun:
$$g(y) = \inf_x L(x,y)$$

can omit

$$\left[= -\sup_x \left(-f(x) - (A^T y)^T x\right) - b^T y \right.$$
$$\left. = -f^*(-A^T y) - b^T y \right]$$

Fenchel conjugate (see Lec 2).

L. dual prob:
$$\max g(y) \quad (D)$$

Assuming strong duality holds, we can recover a primal optimal point x^* from a dual optimal point y^* as
$$x^* = \arg\min_x L(x, y^*).$$

Dual Ascent Method

Need $\nabla g(y)$, assuming g is differentiable.

First evaluate $g(y) = \min_x L(x,y)$ (assuming inf is attained)

giving $x^+ = \arg\min_x L(x,y)$

Then $\nabla g(y) = Ax^+ - b$

- intuitively, because of def of $L(x,y)$

Pf see book by BAZAARA, Ch. 6

Then iterate:

$$\begin{cases} x^{k+1} = \arg\min_x L(x, y^k) \\ y^{k+1} = y^k + \underset{\text{stepsize}}{\alpha_k}(Ax^{k+1} - b) \end{cases}$$

If α_k is small enough, $g(y^{k+1}) < g(y^k)$.

Convergence results are limited.
Fails if f is a nonzero affine function of x, since then typically L is unbounded below in x.

Dual decomposition.

Suppose f is separable:

$$f(x) = \sum_{i=1}^{N} f_i(x_i)$$

where x_i are SUBVECTORS of x.

Partition A conformally:

$$A = [A_1 \, A_2 \cdots A_N]$$

so $Ax = \sum_{i=1}^{N} A_i x_i$, then

$$L(x,y) = \sum_{i=1}^{N} L_i(x_i, y) \equiv \sum_{i=1}^{N} \left(f_i(x_i) + y^T A_i x_i - \frac{1}{N} y^T b \right)$$

So D.A. alg becomes

$$\begin{cases} x_i^{k+1} = \underset{x_i}{\arg\min}\, L_i(x_i, y^k), \quad i = 1, \dots, N \\ y^{k+1} = y^k + \alpha_k (Ax^{k+1} - b) \end{cases}$$

can all be solved in parallel

Goes back to early 1960s.

7-7

Augmented Lagrangians & Method of Multipliers

$$L_\rho(x,y) = f(x) + y^T(Ax-b) + \frac{\rho}{2}\|Ax-b\|_2^2$$

Penalty Parameter

Equivalent to the usual Lagrangian for the problem

$$\min f(x) + \frac{\rho}{2}\|Ax-b\|_2^2 \qquad (*)$$
$$\text{s.t. } Ax=b$$

which is equivalent to the original problem.

Applying dual ascent to $(*)$ gives

$$x^{k+1} = \operatorname*{argmin}_x L_\rho(x, y^k)$$
$$y^{k+1} = y^k + \rho(Ax^{k+1} - b)$$

NOTE BD. BELOW

"Method of Multipliers" or "Augmented Lagrangian Method"

But why $\alpha_k = \rho$? To justify this, assume f is differentiable. Opt conds for (P) (p. 7-4) are

$$Ax^* - b = 0 \quad \text{(primal feas)}$$
$$\nabla f(x^*) + A^T y^* = 0 \quad \text{(KKT, or just dual feas, i.e. } g(y^*) > -\infty, \text{ since } f \text{ is smooth)}$$

By def'n, x^{k+1} minimizes $L_\rho(x,y^k)$, so

$$0 = \nabla_x L_\rho(x^{k+1}, y^k)$$
$$= \nabla_x f(x^{k+1}) + A^T y^k + \rho A^T(Ax^{k+1} - b)$$
$$= \nabla_x f(x^{k+1}) + A^T y^{k+1}$$

So (x^{k+1}, y^{k+1}) is dual feasible: means $g(y^{k+1}) = \inf_x L(x, y^{k+1}) > -\infty$: AS $\nabla f(x) + A^T y^{k+1} = 0$ for $x = x^{k+1}$.

As method proceeds, primal infeasibility → 0.

Much better convergence properties but no longer separable.

ALTERNATING DIRECTION METHOD OF MULTIPLIERS

Combine decomposability of Dual Ascent with better convergence properties of Method of Multipliers

$$\min_{x\in\mathbb{R}^n,\, z\in\mathbb{R}^m} f(x) + g(z)$$
$$\text{S.T. } Ax + Bz = c$$

TOTALLY NEW USE OF g! NOT THE DUAL!

OPT. VALUE p^*.

Assume f, g are convex, but not nec. differentiable. Define augmented Lagrangian

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

ADMM:

$$\left[\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x L_\rho(x, \underline{\underline{z^k}}, y^k) \\ z^{k+1} &= \operatorname*{argmin}_z L_\rho(\underline{\underline{x^{k+1}}}, z, y^k) \\ y^{k+1} &= y^k + \rho(A\underline{\underline{x^{k+1}}} + B\underline{\underline{z^{k+1}}} - c) \end{aligned}\right.$$

cf. Method of Multipliers:

$$\left[\begin{aligned} (x^{k+1}, z^{k+1}) &= \operatorname*{argmin}_{(x,z)} L_\rho(x, z, y^k) \\ y^{k+1} &= y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \end{aligned}\right.$$

Scaled Form of ADMM

Let residual $r = Ax + Bz - c$. Then, with $\|\cdot\| = \|\cdot\|_2$,

$$\|r + \tfrac{1}{\rho} y\|^2 = \|r\|^2 + \tfrac{2}{\rho} r^T y + \tfrac{1}{\rho^2}\|y\|^2$$

$$\tfrac{\rho}{2}\|r + \tfrac{1}{\rho} y\|^2 = \tfrac{\rho}{2}\|r\|^2 + r^T y + \tfrac{1}{2\rho}\|y\|^2$$

Let $u = \frac{1}{\rho} y$ "scaled dual variable". Then

RHS of x^{k+1} is, with $r = Ax + Bz^k - c$,

$$\arg\min_x f(x) + y^T r + \tfrac{\rho}{2}\|r\|^2 \quad \text{(as } g(z^k) \text{ does not depend on } x\text{)}$$

$$= \arg\min_x f(x) + \tfrac{\rho}{2}\|r+u\|^2 \underbrace{- \tfrac{1}{2\rho}\rho^2\|u\|^2}_{\text{does not depend on } x}$$

RHS of z^{k+1} is similar

RHS of y^{k+1} is $\underbrace{y^k}_{\rho u^k} + \rho r$ (with $r = Ax^{k+1} + Bz^{k+1} - c$)

So ADMM becomes

$$\left[\begin{array}{l} x^{k+1} = \arg\min_x \left(f(x) + \frac{\rho}{2}\|Ax + B\underline{\underline{z^k}} - c + u^k\|^2 \right) \\ z^{k+1} = \arg\min_z \left(g(z) + \frac{\rho}{2}\|A\underline{\underline{x^{k+1}}} + Bz - c + u^k\|^2 \right) \\ u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c \end{array}\right.$$

Define $r^k = Ax^k + Bz^k - c$

Then
$$u^{k+1} = u^k + r^{k+1} = u^{k-1} + r^k + r^{k+1} = \dots = u^0 + r^1 + r^2 + \dots + r^{k+1}$$

Convergence Theory:

Assumption 1: $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, $g : \mathbb{R}^{\hat{}} \to \mathbb{R} \cup \{\infty\}$ are convex, closed + proper (see p. 7-1)

Equivalently
$$\text{epi } f = \{(x,t) \in \mathbb{R}^{n+1} : f(x) \le t\}$$
is a closed nonempty convex set.

Note: f, g may be nondifferentiable at some points + may take the value $+\infty$ at some pts.

Ass'n 1 implies that subproblems defining x^{k+1}, z^{k+1} are solvable (not nec. uniquely).

Assumption 2 L_0 has a saddle pt, i.e., $\exists\ (x^*, z^*, y^*)$, not nec. unique, s.t.

$$L_0(x^*, z^*, y) \le L_0(x^*, z^*, y^*) \le L_0(x, z, y^*)$$

holds for all x, y, z. This implies that strong duality holds with primal + dual opt values $L_0(x^*, z^*, y^*)$ which is finite.

Note that we don't assume that A or B is full rank.

Then...

Under assumptions 1 & 2, ADMM satisfies

- Residual convergence: $r^k \to 0$ as $k \to \infty$
  (primal feasibility in limit)
- Objective convergence:
  (primal) $f(x^k) + g(z^k) \to p^*$
- Dual variable convergence:
  $y^k \to y^*$, where
  y^* is dual optimal.

Note that x^k, z^k do not nec. converge.

Details (not all) in next lecture.

Also, applications in next lecture.