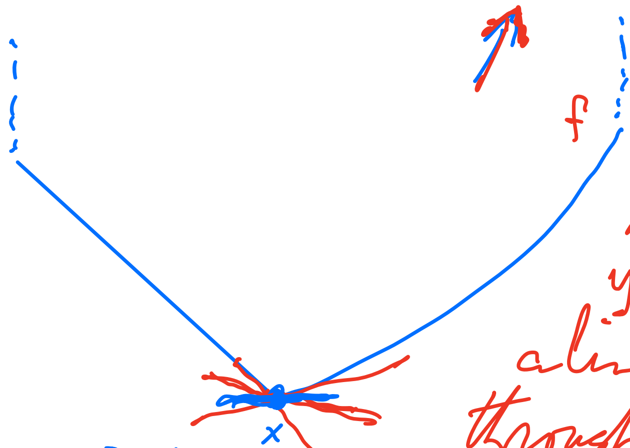
TURN ON RECORDING!

SUBGRADIENTS & SUBDIFFERENTIAL OF A CONVEX FUNCTION.

Assume f is convex, proper $\leftarrow$ $\exists x$ s.t. $f(x) < +\infty$, $\forall x$, $f(x) > -\infty$

+ closed: all sublevel sets are closed.

<u>Def</u> $y \in \mathbb{R}^n$ is a <u>subgradient</u> of f at x if $f(x+z) \geq f(x) + y^T z \quad \forall z \in \mathbb{R}^n$



$\partial f(x) = [-1, 1/2]$

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

$n \geq 1$: $\begin{bmatrix} y \\ -1 \end{bmatrix}$ is normal to a hyperplane in $\mathbb{R}^{n+1}$ passing through $(x, f(x))$ and lying below the graph of f.

The set of all subgradients of f at x is denoted $\partial f(x)$: the <u>subdifferential</u> of f at x.

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

If f is differentiable at x then

$$\partial f(x) = \{\nabla f(x)\}$$

For $x \in \text{int dom} f$, $\partial f(x)$ is always a closed, convex, compact, non-empty set.

e.g. $f(x) = \max_{1 \le i \le n} x_i$ $\qquad x \in \mathbb{R}^n$

$(f = x_{[1]})$ in BV

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

For $y \in \partial f(x)$ need

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

Is $e_1 \in \partial f(x)$? Let $y = e_1$, then RHS $= 3 + z_1$

$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

Is true $\forall z$?

<u>NO</u> e.g. $z = e_1$

Is $e_2 \in \partial f(x)$? Let $y = e_2$ then RHS $= 3 + z_2$

<u>YES</u> Also for e_5.

Fenchel-Young Thm

$$f(x) + f^*(y) \geq x^T y$$

with equality IFF $y \in \partial f(x)$

Directional Derivative

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

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

CHAIN RULE

Simplest case. | More general cases.

Borwein + Lewis p.52

Rockafellar p.225

Let $f : \mathbb{R}^n \to \mathbb{R}$ convex, $\operatorname{dom} f = \mathbb{R}^n$.

Let $A \in \mathbb{R}^{n \times m}$ $b \in \mathbb{R}^n$

Define h by

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

$$h : \mathbb{R}^m \to \mathbb{R}$$

Then $\partial h(\xi) = A^T \partial f(A\xi + b)$

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

OPTIMALITY CONDITION

$0 \in \partial f(x) \iff x$ is a global minimizer of f

known. from def'n.

Precursors of ADMM

(from ADMM paper by Boyd + al.)

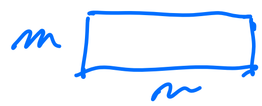
DUAL ASCENT

$$\min f(x)$$
$$\text{s.t. } Ax = b$$

$f : \mathbb{R}^n \to \mathbb{R}$

convex, differentiable

A is $m \times n$



Lagrangian

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

formerly ν

Lagr. Dual Fun.

$$g(y) = \inf_x L(x, y)$$

Lagr. Dual Prob:

$$\max_y g(y).$$

Assuming strong duality, can recover primal solution x^* from a dual sol'n y^* as
$$x^* = \arg\min_x L(x, y^*)$$

DUAL ASCENT METHOD

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

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

where $x^+ = \arg\min L(x, y)$

<u>Pf</u> see book by BAZAARA Ch. 6.

Then iterate: start with x^0, y^0, then for $k = 0, 1, \ldots$

$$x^{k+1} = \arg\min_x L(x, y^k)$$
$$y^{k+1} = y^k + \underbrace{\alpha_k}_{\text{Stepsize}} \underbrace{(Ax^{k+1} - b)}_{\nabla g(y^k)}$$

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

DUAL DECOMPOSITION

Suppose f is separable
$$f(x) = \sum_{i=1}^{n} f_i(x_i)$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \leftarrow \text{sub vectors}$$

Partition A ...

Partition A conformably

$$A = [A_1 \; A_2 \; \ldots \; A_N]$$

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

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

So Dual Ascent becomes:

$$x_i^{k+1} = \operatorname{argmin} L_i(x_i, y^k) \qquad i = 1, \ldots, N$$

$$y^{k+1} = y^k + \alpha_k \underbrace{(A x^{k+1} - b)}_{\nabla g(y^k)}$$

AUGMENTED LAGRANGIANS + METHOD OF MULTIPLIERS

$$L_\rho(x,y) = f(x) + \underbrace{y^T(Ax-b)}_{\text{Lagrangian term}} + \underbrace{\frac{\rho}{2} \|Ax-b\|_2^2}_{\text{Penalty term}}$$

$\boxed{\rho > 0}$

Equivalent to usual Lagrangian for

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

$$\text{s.t. } Ax = b \qquad \equiv (P)$$

Apply Dual Ascent to $(*)$

METHOD OF MULTIPLIERS $\Leftrightarrow$ AUGMENTED LAGR. METHOD.

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

$$y^{k+1} = y^k + \underbrace{\rho}_{\alpha_k}(Ax^{k+1} - b)$$

NOTE: THIS IS BD BELOW

Opt conds for (P) are

$$Ax^* - b = 0 \quad \text{(primal feas)}$$

$$\nabla f(x^*) + A^T y^* = 0 \quad \text{(KKT)}$$

(dual feas:

for dual feas, need $g(y^*) > -\infty$

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

$$0 = \nabla_x L_\rho(x^{k+1}, y^k)$$

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

as $\frac{\rho}{2}\|Ax-b\|_2^2 = (Ax-b)^T(Ax-b)\cdot\frac{\rho}{2}$

$$A^T(\underbrace{y^k + \rho(Ax^{k+1} - b)}_{y^{k+1}})$$

$$0 = \nabla f(x^{k+1}) + A^T y^{k+1}$$

so (x^{k+1}, y^{k+1}) is dual feasible: means

$g(y^{k+1}) = \inf_x L(x, y^{k+1}) > -\infty$: yes,

as $\nabla f(x) + A^T y^{k+1} = 0$ for $x = x^{k+1}$

(x^{k+1}, y^{k+1}) always dual feasible

and as method proceeds,

primal infeasibility $\to 0$

Much better convergence properties

than dual descent, but NO LONGER SEPARABLE

ALTERNATING DIRECTION METHOD OF MULTIPLIERS (ADMM)

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} \quad f(x) + g(z)$$

NOT DUAL FCN

$$\text{S.T.} \quad Ax + Bz = c$$

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = c$$

OPT. VAL. p^*.

Assume f, g are convex.

Define

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

ADMM

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

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

$$y^{k+1} = y^k + \rho(Ax^{k+1}+Bz^{k+1}-c)$$

(In method of multipliers, get

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

$$y^{k+1} = y^k + \rho(Ax^{k+1}+Bz^{k+1}-c)$$

SCALED FORM of ADMM.

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

Then

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

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

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

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

$$\arg\min_x f(x) + y^T r + \frac{\rho}{2}\|r\|^2$$

$$= \arg\min_x f(x) + \frac{\rho}{2}\left\|r + \frac{1}{\rho}y\right\|^2 - \cancel{\frac{1}{2\rho}\|y\|^2}$$

RHS of def'n of z^{k+1} is similar

RHS of def'n of y^{k+1}

$$\underbrace{y^{k+1}}_{\rho u^{k+1}} = \underbrace{y^k}_{\rho u^k} + \rho r$$

with $r = Ax^{k+1} + Bz^{k+1} - c$

SO ADMM BECOMES

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

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

$$u^{k+1} = u^k + \underbrace{Ax^{k+1} + Bz^{k+1} - c}_{r^{k+1}},$$

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

$$= (u^{k-1} + r^k) + r^{k+1}$$

$$\vdots$$

$$u^{k+1} = u^0 + \sum_{j=1}^{k+1} r^j$$

CONVERGENCE THEORY

<u>Assumption 1</u> f, g are convex, closed & proper

equivalently $\operatorname{epi} f = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^{n+1} : f(x) \le t \right\}$

is a closed convex nonempty set.

<u>Assumption 2</u> L_0 has a saddle pt,

i.e. $\exists\, (x^*, z^*, y^*)$, not nec. unique,

s.t. $L_0(x^*, z^*, y) \le L_0(x^*, z^*, y^*) \le L(x, z, y^*)$

holds for all x, z, y.

Then ADMM has following properties

- PRIMAL RESIDUAL $\to 0$

  $r^k \to 0$ as $k \to \infty$

- PRIMAL OBJECTIVE $\to p^*$

  $f(x^k) + g(z^k) \to p^*$

- DUAL VARIABLE $\to y^*$

  $y^k \to y^*$