Proximal algorithms

let $f(x)$ be closed, proper, convex.

so epigraph $\left(= \{(x,t) : f(x) \leq t\} \right.$

$\left. = \text{shade graph} \right)$

is convex

$$\text{dom}(f) = \{ x \mid f(x) < +\infty \}$$

$$\boxed{\text{prox}_f(v) = \underset{x}{\text{argmin}} \left[f(x) + \frac{1}{2}\|x - v\|^2 \right]}$$

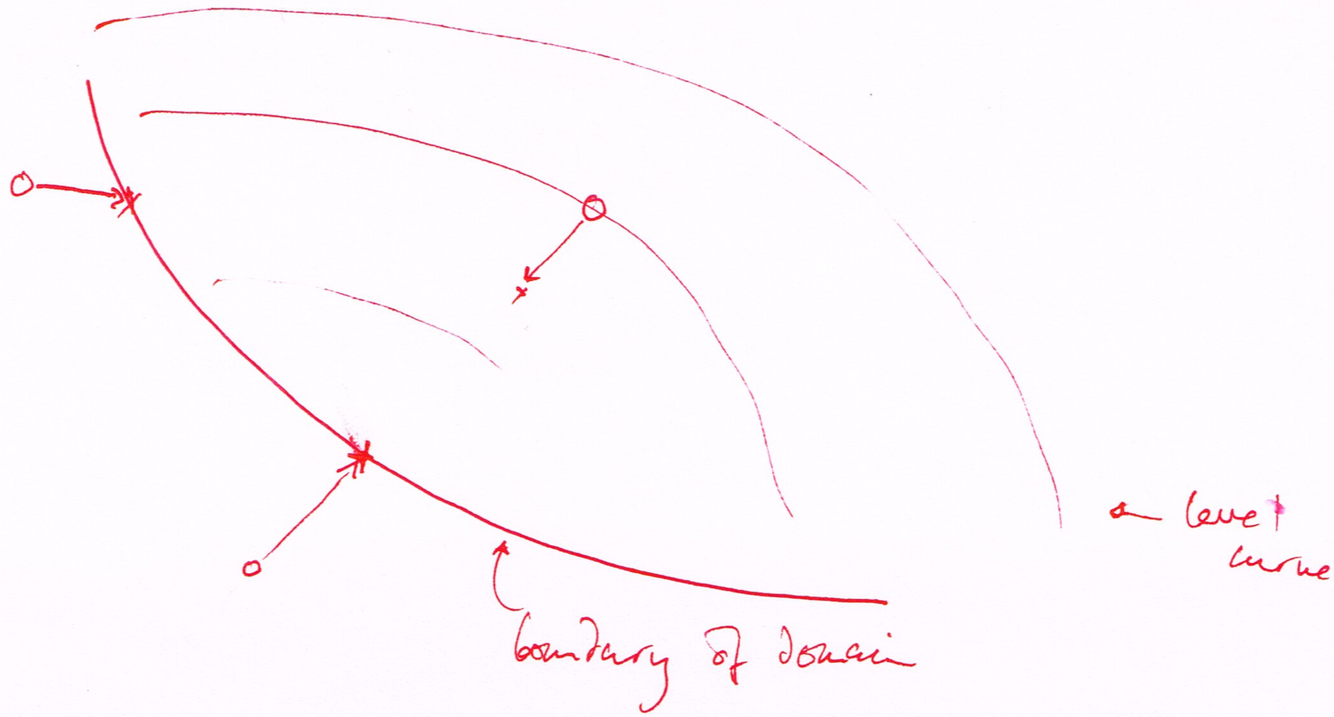
the proximal operator.

$$\text{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}} \left[f(x) + \frac{1}{2\lambda}\|x - v\|^2 \right]$$

(scaled prox operator - rescale).

Several interpretations

1



(i.e. - if you're outside, move to nearby pt on dom(f) boundary. ~~and~~)
~~gettin smaller~~
- inside - get smaller

2 let

$$f(x) = \mathbb{1}_C(x) = \begin{cases} 1 & x \in C \\ 0 & \text{ot.} \end{cases}$$

then

$\text{prox}_f(x)$ gives closest pt on C to x

Now assume f is as differentiable as I need, and λ <u>small</u>

$$\operatorname{prox}_{\lambda f}(v) = \operatorname*{argmin}_x \left[f(x) + \frac{1}{2\lambda} \| x - v \|^2 \right]$$

(λ small $\Rightarrow$ δv small)

write $x = v + \delta v$

min is "close" to v

so $f(x) \approx f(v) + \delta v^T \nabla f$

$$\operatorname{prox}_{\lambda f}(v) = \operatorname*{argmin}_{\delta v} \left[f(v) + \delta v^T \nabla f + \frac{1}{2\lambda} \| \delta v \|^2 + \cdot \right]$$

so $\nabla f + \frac{1}{\cancel{2}\lambda} \delta v = 0$

so $\operatorname{prox}_{\lambda f}(v) \approx v - \cancel{\tfrac{1}{2}}\lambda \nabla f$

fixed points:

x^* minimizes f

$\Updownarrow$

$x^* = \text{prox}_f(x^*)$.

Which is why we care!

Proof:

$\Rightarrow$: assume x^* min f i.e $f(x) \geq f(x^*)$ all x

then

$\text{prox}_f(x) = f(x) + \frac{1}{2}\|x - x^*\|^2 \geq f(x^*) = \text{prox}_f(x^*)$

$\Leftarrow$: assume $\tilde{x}$ minimizes $\text{prox}_f(v)$, so

$\tilde{x} = \text{prox}_f(v)$

this implies

$0 \in \partial f|_{\tilde{x}} + (\tilde{x} - v)$

now if $\tilde{x} = x^*$ and $v = \hat{x}^*$ (so $\tilde{x} = x^* = \text{prox}_f(x^*) = \text{prox}_f(v)$)

$0 \in \partial f_{x^*}$ so x^* min f $\square$

Proximal min

$$x^{k+1} = \mathrm{prox}_{\lambda f}(x^k)$$

- if f has a min, is closed, proper, convex, then $x^k \to$ set of minimizers.
- Could do

$$x^{k+1} = \mathrm{prox}_{\lambda^k f}(x^k)$$

$\to$ OK if $\lambda^k > 0$ and $\sum_k \lambda^k = \infty$

ADMM is a prox alg.

ADMM

$$\min \quad f(x) + g(z)$$
$$\text{st} \quad x - z = 0$$

AL
$$L_\rho = f(x) + g(z) + y^T(x-z) + \frac{\rho}{2}\|x-z\|^2$$

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

SO
$$x^{k+1} = \operatorname{argmin}_x \left[f(x) + y^{k^T} x + \frac{\rho}{2}\|x - z^k\|^2 \right]$$

etc.

Now

$$f(x) + y^{k^T} x + \frac{\rho}{2}\|x-z^k\|^2$$
$$= f(x) + y^{k^T} x + \frac{\rho}{2}\|x - z^k\|^2 + y^{k^T} z^k - y^{k^T} z^k + \frac{1}{2\rho}\|y^k\|^2 - \frac{1}{2\rho}\|y^k\|^2$$

$$= f(x) + \frac{\rho}{2}\left\|x - z^k + \frac{1}{\rho} y^k\right\|^2 + \text{const}.$$

Now:

subst: $u^k = \frac{1}{\rho} y^k$, $\lambda = \frac{1}{\rho}$, get

$$x^{k+1} = \mathrm{prox}_{\lambda f}(z^k - u^k)$$

$$z^{k+1} = \mathrm{prox}_{\lambda g}(x^{k+1} + u^k)$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1}$$

Some examples:

Consensus

$$\min \ f(x) = \sum_i f_i(x).$$

multiple, local objectives
same x's

same as:

$$\min \ \sum_i f_i(x_i)$$

$$\text{st} \quad x_1 = x_2 \ \cdots = x_N$$

(6a)

Consensus algs

$$\min \; f(x) = \sum_i f_i(x).$$

tx to:

$$\min \; f(x) = \sum_i f_i(x_i)$$

$$\text{st} \quad x_1 = x_2 \; \cdots \; x_N$$

← Consensus constraint

tx to:

$$\min \; \sum_i f_i(x_i) + 1_c(x_1 \; \cdots \; x_N)$$

↑ indicator fn

$= 0$ if $x_1 = x_2 \cdots = x_N$

$= \infty$ otherwise

↑ not the usual thing

Now think about partitions of vars

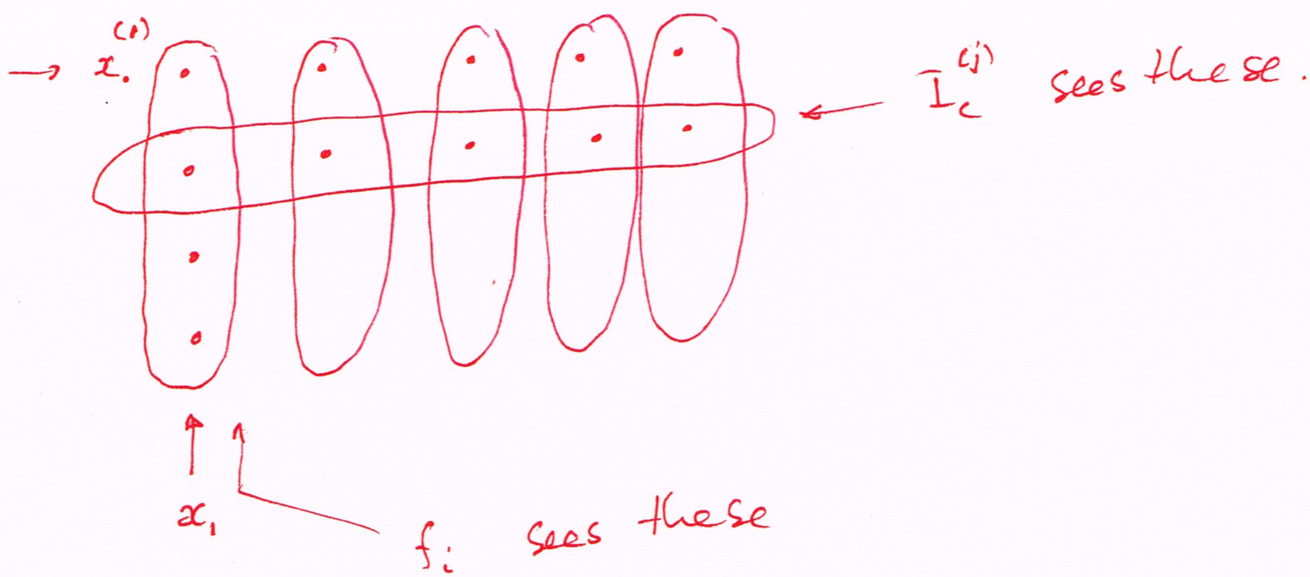
- each x_i is n Dim
- there are N

write $x_i^{(j)}$ for j'th component of i'th x

$$I_c(x_1 \quad \cdot \quad \cdot \quad x_N)$$

$$= I_c(x_1^{(1)} \cdot \cdot\ x_N^{(1)}) + I_c(x_1^{(2)}, \cdot \cdot, x_N^{(2)}) \cdot \cdot \cdot$$

So we can <u>draw</u> vars



This gets us into a quite general picture:

$$\min \quad f(x) \quad + \quad g(x)$$

depends on vars partitioned one way = P

this is truly a partition – set of disjoint sets that covers vars

depends on vars partitioned some other way = Q

as is this

Natural strategy

$$\min \quad f(x) + g(z)$$

$$\text{st} \quad x = z$$

i.e. make two copies, one for each side, impose equality

c_i = i'th component of P

d_j = j'th component of Q

so

$$\min \quad f(x) \quad + \quad g(z) \qquad \text{st} \quad x = z$$

is

$$\min \quad \sum_i f_i(x_{c_i}) \quad + \quad \sum_j g_j(z_{d_j}) \qquad \text{st} \quad \text{constraints}$$

<u>Main issue:</u> notation for L.M.'s, etc

$z^k_{c_i}$ = c_i pieces of k'th est of z

etc

u^k = all lms, k'th est

$u^k_{d_j}$ = lm's pertaining to d_j vars, k'th est

<u>ADMM</u>:

$$x^{k+1}_{c_i} = \text{prox}_{\lambda f_i}\left(z^k_{c_i} - u^k_{c_i}\right)$$

$$z^{k+1}_{d_j} = \text{prox}_{\lambda g_j}\left(x^{k+1}_{d_j} + u^k_{d_j}\right)$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1}$$

Now apply to consensus.

$\text{prox}_{\lambda g}$ is easy — this is prox for a convex set ind fun = project onto set.

— in consensus case, average.

$$z_i^{k+1} = \frac{1}{N} \sum_i z_i^k$$

or in c_i, d_j notation

$$z_{d_j}^{k+1} = \frac{1}{N} \sum_j x_{d_j}^k$$

Notice $u^0 = 0$

$$u_{d_j}^{k+1} = u_{d_j}^k + x_{d_j}^{k+1} - z_{d_j}^{k+1}$$

so $\sum_j u_{d_j}^{k+1} = 0$

so $\sum_{i=1}^{N} u_{c_i}^k = 0 = \sum_{i=1}^{N} u_i^k$

arranged to csp w/ x_i

So we can simplify to get

$$x_i^{k+1} = \text{prox}_{\lambda f_i}(\bar{x}_k - u_i^k)$$
$$u_i^{k+1} = u_i^k + x_i^{k+1} - \bar{x}^{k+1}$$

More general form of consensus

the c_i are no longer a partition

eg $\quad f_1(x_1, x_2) + f_2(x_2, x_3) + \ldots f_N(x_N, x_{N+1})$

x_{c_i} = set of x's for each function

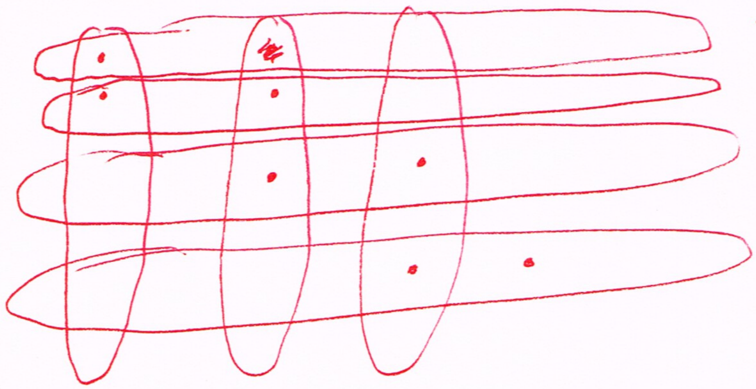
Strategy
Make one copy of relevant vars for each term

eg: $\quad f_1(x_1, x_2) + \ldots$

becomes $\quad f_1(z_1) + f_2(z_2) + \ldots$

AND set equalities $\quad$ so

$I_q^{(j)}$ sees these



for our example

f_1 sees these

ADMM then becomes

$$x_i^{k+1} = \operatorname{prox}_{\lambda f_i}(\bar{x}_i^k + u_i^k)$$

$$u_i^{k+1} = u_i^k + x_i^{k+1} - \bar{x}_i^{k+1}$$

$(\bar{x}^k)_i$ = average of x's relevant to this (ith) fn.

Exchange:

$$\min \sum_i f_i(x_i)$$
$$\text{st} \quad \sum_i x_i = 0$$

rewrite:

$$\min \sum_i f_i(x_i) + I_C(x_1 \cdots x_N)$$

$$\begin{cases} = 0, & x_1 + \cdots x_N = 0 \\ \infty & \text{otherwise.} \end{cases}$$

same procedure, partition

– project onto C ? $\equiv$ subtract mean

$$\therefore \quad x_i^{k+1} = \text{prox}_{\lambda f_i}^{\frac{1}{2}} (x_i^k - \bar{x}^k - u^k)$$
$$u^{k+1} = u^k + \bar{x}^{k+1}$$

Some example proximal operators

$$\text{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}}\left[f(x) + \frac{1}{2\lambda}\|x - v\|^2 \right]$$

st $x \in C$ ← some convex set

1) $f = \frac{x^T A x}{2} + b^T x + c$

then $\text{prox}_{\lambda f}(v)$ obtained by solving

$$\left(A + \frac{I}{\lambda}\right) x = \frac{v}{\lambda} - b$$

or $(\lambda A + I)\, x = v - b\lambda$

→ either factor $(\lambda A + I)$ or warm start an iterative method (C.G.)

2) Sum of scalar fns.

$$f = \sum_i f_i(x_i)$$

then

$$\text{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}} \sum_i f(x_i) + \frac{1}{2\lambda} \sum_i \| x_i - v_i \|^2$$

$$= \begin{matrix} \underset{x_1}{\text{argmin}} \; f(x_1) + \frac{1}{2\lambda}(x_1 - v_1)^2 \\ \vdots \\ \underset{x_N}{\text{argmin}} \; f(x_N) + \frac{1}{2\lambda}(x_N - v_N)^2 \end{matrix}$$

→ 2 cases of particular interest

$f = -\log(x)$; $\text{prox}_{\lambda f}(v) = \dfrac{v + \sqrt{v^2 + 4\lambda}}{2}$

$f = |x|$; $\text{prox}_{\lambda f}(v) = \begin{cases} v - \lambda, & v \geq \lambda \\ 0 & |v| \leq \lambda \\ v + \lambda & v \leq -\lambda \end{cases}$

Polyhedron:

$$P = \{x \mid Ax = b,\ Cx \le d\}$$

a) project w to P

$$\operatorname*{argmin}_x \ \tfrac{1}{2}\|x - w\|^2$$
$$\text{st } Ax = b,\ Cx \le d.$$

b) $\mathrm{prox}_{\lambda f}(v)$ for f some convex fn defined on P

$$\operatorname{argmin} \ f + \tfrac{1}{2\lambda}\|x - v\|^2$$
$$\text{st } Ax = b,\ Cx \le d.$$

IP method ?