Proximal algorithms

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

So the epigraph

$E(f) = \{(x,t) : f(x) \leq t\}$

is convex.

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

prox$_f(v) = \arg\min_x \left[ f(x) + \frac{1}{2\lambda} \|x-v\|^2 \right]$

the proximal operator:

prox$_f(v) = \arg\min_x \left[ f(x) + \frac{1}{2\lambda} \|x-v\|^2 \right]$

(scaled proximal operator - rescale).
Several interpretations

1.地处 boundary of domain
2. nearby pt on
   (i.e. if you're outside, move to dom(f) boundary, and)
   get smaller
   - inside - get smaller

\[ f(x) = \mathbb{1}_C(x) = \begin{cases} 1 & x \in C \\ 0 & \text{otherwise} \end{cases} \]

then
\[ \text{prox}_f(x) \] gives closest pt on C to x
Now assume $f$ is as differentiable as 

\[\text{prox} (v) = \arg\min_x \left[ f(x) + \frac{1}{2\lambda} \|x - v\|^2 \right]\]

\[x_f \text{ s.t. } x = v + sv\]

\[\text{since } x \approx \text{ "close" to } v\]

\[f(x) \approx f(v) + sv^T \nabla f\]

\[\text{so } f(x_c) \approx f(v) + sv^T \nabla f + \frac{1}{2\lambda} \|sv\|^2\]

\[\text{prox} (v) = \arg\min_{sv} \left[ f(v) + sv^T \nabla f + \frac{1}{2\lambda} \|sv\|^2 \right]\]

\[\nabla f + \frac{1}{\lambda} sv = 0\]

\[\nabla f + \frac{1}{\lambda} sv = 0\]

\[\text{so } \text{prox}_\lambda (v) \approx v - \frac{1}{\lambda} sv\]
fixed points:

\[ x^* \text{ minimizes } f \]

\[ x^* = \text{prox}_f(x^*) \]

Which is why we care!

Proof:

\[ \Rightarrow \]: assume \( x^* \) minimizes \( f \) i.e. \( f(x) \geq f(x^*) \) all \( x \)

then

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

\[ \leq \]: assume \( x \) minimizes \( \text{prox}_f(v) \), so

\[ x^* = \text{prox}_f(v) \]

this implies

\[ 0 \in df\big|_{x^*} + (x^* - v) \]

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

\[ 0 \in df\big|_{x^*} \text{ so } x^* \text{ min } f \]
Proximal min

\[ x^{k+1} = \text{prox}_{\lambda f} (x^k) \]

- If \( f \) has a min, it is closed, proper, convex, then \( x^k \to \) set of minimizers.

- Could do

\[ x^{k+1} = \text{prox}_{\lambda f} (x^k) \]

\[ \rightarrow \text{OK if } x^k > 0 \text{ and } \sum_{k} \lambda^k = \infty \]
ADMM is a prox alg.

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

\[ h_p = f(x) + g(z) + y^T(x - z) + \frac{1}{2} \| x - z \|^2 \]

\[ x^{k+1} = \arg\min_x h_p(x, z^k, y^k) \]
\[ z^{k+1} = \arg\min_z h_p(x^{k+1}, z, y^k) \]
\[ y^{k+1} = y^k + \frac{1}{2} (x^{k+1} - z^{k+1}) \]

So
\[ x^{k+1} = \arg\min_x \left[ f(x) + y^{kT}x + \frac{1}{2} \| x - z^k \|^2 \right] \]

etc.

Now
\[ f(x) + y^{kT}x + \frac{1}{2} \| x - z^k \|^2 \]
\[ = f(x) + y^{kT}x + \frac{1}{2} \| x - z^k \|^2 + y^{kT}z^k - y^{kT}z^k \]
\[ + \frac{1}{2} y^{kT}z^k - \frac{1}{2} y^{kT}z^k \]
\[ = f(x) + \frac{1}{2} \| x - z^k + \frac{1}{\rho} y^k \|^2 + \text{const} \]
\[
\text{New:} \quad u^k = \frac{1}{\rho} y^k, \quad \lambda = \frac{1}{\rho}, \quad \text{get}
\]
\[
x^{k+1} = \text{prox}_{\lambda f}(z^k - u^k)
\]
\[
y^{k+1} = \text{prox}_{\lambda g}(x^{k+1} + u^k)
\]
\[
u^{k+1} = u^k + x^{k+1} - z^{k+1}
\]

Some examples:

**Consensus**

\[
\min \ f(x) = \sum_i f_i(x)
\]
\text{multiple, local objectives}
\text{same } x \text{'s}

Same as:

\[
\min \ \sum_i f_i(x) \quad \text{subject to} \quad x_1 = x_2 = \ldots = x_N
\]
Consensus algs

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

\[\text{subject to:}\]

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

$$\text{subject to:} x_1 = x_2 = \ldots = x_N$$

\[\text{consensus constraint}\]

\[\text{tx to:}\]

$$\min \sum_i f_i(x_i) + I_c(x_1, \ldots, x_N)$$

\[\text{INVALID}\]

\[\text{tx to:}\]

$$\min \sum f_i(x_i) + I_c(x_1, x_2, \ldots, x_N)$$

\[\text{rev:}\]

$$I_c = \begin{cases} 
0 & \text{if } x_1 = x_2 = \ldots = x_N \\
\infty & \text{otherwise}
\end{cases}$$

\[\text{not the usual thing}\]

Now think about partitions of vars

- each $x_i$ is in $D_i$

- there are $N$
Write $x_i^{(j)}$ for $j$th component of $i$th $x$

$$I_c(x_1, \ldots, x_N) = I_c(x_1^{(u)}, \ldots, x_N^{(u)}) + I_c(x_1^{(v)}, \ldots, x_N^{(v)})$$

so we can draw vars

$\rightarrow x_i^{(u)}$ sees these.

$\rightarrow I_c^{(u)}$ sees these.

$\rightarrow x_i$.

$\rightarrow f_i$ sees these.
This gets us into a quite general picture:

\[ \min f(x) + g(y) \]

depends on vars partitioned some other way

\[ = 0 \]

as is this

This is truly a partition - set of disjoint sets that covers vars

Natural strategy

\[ \min f(x) + g(z) \]

**st** \[ x = z \]

i.e. make two copies, one for each side, impose equality
\[ c_i = \text{ith component of } P \]
\[ d_j = \text{jth component of } Q \]

So
\[
\min \quad f(x) + g(z) \quad \text{st} \quad x = z
\]
is
\[
\min \quad \sum_i f_i(x_i) + \sum_j g_j(z_j) \quad \text{st constraints}
\]

**Main Issue:** notation for \( L \)-\( M \)'s, etc

\[ z_{ci}^k = \text{ci pieces of } k\text{th est of } z \]
\[ z_{di}^k = \text{all } l\text{'s, } k\text{th est} \]
\[ u_k^k = \text{l's pertaining to } d\text{'s vars, } k\text{th est} \]

**ADMM:**
\[
x_{ci}^{k+1} = \text{prox} \quad \lambda f_i \quad (z_{ci}^k - u_{ci}^k)
\]
\[
z_{ci}^{k+1} = \text{prox} \quad g_i \quad (x_{ci}^{k+1} + u_{ci}^k)
\]
\[
z_{di}^{k+1} = \text{prox} \quad g_i \quad (x_{di}^{k+1} + u_{di}^k)
\]
\[
u^{k+1} = u^k + x - z^{k+1}
\]
Now apply to consensus.

\[ \lim_{g \to 0} \text{prox}_{g \Phi} = \text{project onto set}. \]

- In consensus case, average.

\[ z_{i}^{k+1} = \frac{1}{N} \sum_{i}^{k} \]

or in \( c_i, d_j \) notation

\[ z_{d_j}^{k+1} = \frac{1}{N} \sum_{j}^{k} \]

Notice \( u = 0 \)

\[ u_{d_j}^{k+1} = u_{d_j}^{k} + \alpha z_{d_j}^{k+1} - z_{d_j}^{k+1} \]

so \( \sum u_{d_j}^{k+1} = 0 \)

so \( \sum_{i=1}^{N} u_{i}^{k} = 0 = \sum_{i=1}^{N} u_{i}^{k} \]

arranged to cop w/ \( \alpha \).
So we can simplify to get

\[ x_{i}^{K+1} = \text{prox}_{\lambda f_{i}} (x_{k} - u_{i}^{K}) \]

\[ u_{i}^{K+1} = u_{i}^{K} + x_{i}^{K+1} - \bar{x}^{K+1} \]

More general form of consensus

the \( c_{i} \) are no longer a partition

eg \[ f(x_{1}, x_{2}) + f(x_{2}, x_{3}) + \cdots + f(x_{N}, c_{N+1}) \]

\( x_{ci} \) = set of \( x \)'s for each function

strategy: Make one copy of relevant vars for each term

eg:

\[ f(x_{1}, x_{2}) + \cdots \]

becomes \[ f_{1}(z_{1}) + f_{2}(z_{2}) + \cdots \]

AND set equalities: so
I see these

for our example

ADM then becomes

\[ x_i^{k+1} = \text{prox}_{\lambda f_i} (x_i^k + u_i^k) \]

\[ u_{i}^{k+1} = u_{i}^{k} + x_{i}^{k+1} - x_{i}^{k} \]

\[ (\bar{x}_i^k)_i = \text{average of } x_i^k \text{ relevant to this channel}. \]
Exchange:

\[ \begin{align*}
\min & \quad \sum_i f_i(x_i) \\
\text{s.t.} & \quad \sum_i x_i = 0
\end{align*} \]

Rewrite:

\[ \begin{align*}
\min & \quad \sum_i f_i(x_i) + I_C(x, \ldots, x_N) \\
\text{s.t.} & \quad \sum_i x_i = 0, x_1 + \ldots + x_N = 0, \text{ otherwise}
\end{align*} \]

Same procedure, partition
- Project onto \( C \) ? = subtract mean

\[ \begin{align*}
x_i^{k+1} &= \text{prox}_{\frac{1}{2}} \left( x_i^k - \bar{x}^k + u^k \right) \\
x_i^k &= \lambda f_i, \bar{x}^k + u^{k+1} \\
u^{k+1} &= u^k + \bar{x}^{k+1}
\end{align*} \]
Some example proximal operators

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

\[ \text{s.t. } x \in C \subseteq \text{some convex set} \]

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

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

\[ (A + I) \alpha = \frac{v - b}{\lambda} \]

or \[ (\lambda A + I) \alpha = v - b \lambda \]

\( \Rightarrow \) either factor \((\lambda A + I)\) or warm start an iterative method (e.g.)
2) **Sum of scalar fun.**

\[ f = \sum_i f_i(x_i) \]

Then

\[ \text{prox}_f(v) = \arg\min_x f(x) + \frac{1}{2\lambda} \| x - v \|^2 \]

\[ = \arg\min_x f(x_i) + \frac{1}{2\lambda} (x_i - v_i)^2 \]

\[ = \arg\min_{x_N} f(x_N) + \frac{1}{2\lambda} (x_N - v_N)^2 \]

\[ \rightarrow \text{2 cases of particular interest} \]

\[ f = \log(x) \quad \text{prox}_f(v) = \frac{v + \sqrt{v^2 + 4\lambda}}{2} \]

\[ f = |x| \quad \text{prox}_f(v) = \begin{cases} 
-\lambda & v \geq \lambda \\
0 & |v| \leq \lambda \\
v + \lambda & v \leq -\lambda 
\end{cases} \]
Polyhedron:
\[ P = \{ x \mid Ax = b, \ Cx \leq d \} \]

a) project \( w \) to \( P \)
\[ \text{argmin}_x \frac{1}{2} \| x - w \|^2 \]
\[ \text{s.t.} \ Ax = b, \ Cx \leq d. \]

b) \( \text{prox}_{\lambda f} (v) \)
for \( f \) some convex fn defined on \( P \)
\[ \text{argmin}_x f + \frac{1}{2\lambda} \| x - v \|^2 \]
\[ \text{s.t.} \ Ax = b, \ Cx \leq d. \]

\[ \text{IP method?} \]