E9 203 Compressed Sensing & Sparse SP

22 Mar. 2021

Last time:

- Majorization-minimization based algos.
- Reweighted l_1 and l_2 methods.

Today:

- Analysis of reweighting based methods (properties of local minima).

Recap: Separable cost fns:

$$\min_{x \in \mathbb{R}^N} \underbrace{\sum_{i=1}^{N} g(x_i)}_{G(x)} \quad \text{s.t. } Ax = y$$

Examples:
$g(x) = |x|^p, \quad 0<p<1$
$g(x) = \ln|x|$
$g(x) = |x+\beta|^p, \quad 0<p<1$
$g(x) = \ln|x+\beta|$, etc. $\Big\}\ \beta > 0.$

Reweighted l_2 methods:

Consider $g(x)$ as a fn. of $\tilde{y} = x^2$. If it is still concave in $\tilde{y}$, upper bound $g(x)$ as a linear fn. of $\tilde{y}$ via the 1st order Taylor expansion $\longrightarrow$ quadratic in x.

$$x^{(k+1)} \leftarrow \arg\min_{x \in \mathbb{R}^N} \|y - Ax\|_2^2 + \lambda \|\underbrace{W_k^{-1}}_{\text{diagonal}} x\|_2^2$$

$$\Rightarrow x^{(k+1)} = W_k (\underbrace{W_k A^T A W_k + \lambda I}_{N \times N})^{-1} W_k^T A^T y$$

Use matrix inversion lemma:

$$x^{(k+1)} = W_k^2 A^T (\underbrace{A W_k^2 A^T + \lambda I}_{m \times m})^{-1} y$$

E.g., with $g(x) = |x|^p$, $0<p<2$,

$$W_k^{-1} = \text{diag}\{|x_l^{(k)}|^{(p/2-1)}\}$$

with $g(x) = (x^2+\beta)^{p/2}$, $0<p<2$, $\beta>0$,

$$W_k^{-2} = \text{diag}\{(|x_l^{(k)}|^2 + \beta)^{(p/2-1)}\}$$

Reweighted l_1 min.:

Here we bound $g(x)$, $x \geq 0$ as a fn. of x, not x^2, and solve

$$x^{(k+1)} \leftarrow \arg\min_{x \in \mathbb{R}^N} \|Ax - y\|_2^2 + \lambda \|W_k^{-1} x\|_{\underline{\underline{1}}}.$$

E.g., with $g(x) = |x|^p$, $0<p<1$,

$$W_k^{-1} = \text{diag}\{|x_l^{(k)}|^{p-1}\} \ : N \times N$$

with $g(x) = (|x|+\beta)^p$, $0<p<1$, $\beta>0$,

$$W_k^{-1} = \text{diag}\{|x_l^{(k)} + \beta|^{p-1}\}.$$

Other $g(x)$: as long as it is concave, same approach works.
only need $g'(x)$.
E.g: $g(x) = \ln|x|$.

Weighted l_2 version:

$$g(x) = \ln|x| = \frac{1}{2}\ln x^2 \leq \frac{1}{2}\ln(\underbrace{x^2}_{\tilde{y}} + \beta), \quad \beta > 0$$

concave in y. [Check]!

$$\ln(\tilde{y}+\beta) \leq \frac{1}{\tilde{y}_0+\beta}(\tilde{y} - \tilde{y}_0) + \ln(\tilde{y}_0+\beta).$$

$$= \frac{\tilde{y}}{\tilde{y}_0+\beta} + \underbrace{\ln(\tilde{y}_0+\beta) - \frac{\tilde{y}_0}{\tilde{y}_0+\beta}}_{\text{not imp.}}$$

We thus get the iteration

$$x^{(k+1)} = \arg\min_{x \in \mathbb{R}^N} \|Ax - y\|_2^2 + \lambda \sum_{l=1}^{N} \frac{x_l^2}{(x_l^{(k)})^2 + \beta} \qquad \text{[Chartrand]}$$

(Weighted l_2 problem).

Weighted l_1 version: $x > 0$

$$g(x) = \ln x \leq \ln(x+\beta) \leq \frac{1}{x_0+\beta}(x - x_0) + \ln(x_0+\beta), \quad \beta > 0$$

And we get the iteration

$$x^{(k+1)} = \arg\min_{x \in \mathbb{R}^N} \|Ax - y\|_2^2 + \lambda \sum_{l=1}^{N} \frac{|x_l|}{|x_l^{(k)}| + \beta}. \quad \text{(Weighted } l_1 \text{ problem)}$$

[Candes, Wakin, Boyd]

Local minima of $G(x) = \sum_{l=1}^{N} |x_l|^p$ s.t. $Ax = y$.

Let x_0 be a feasible point of interest, i.e., $Ax_0 = y$.

In our case: local minima candidates, i.e., basic feasible solns. ($\because |x|^p$ is concave)

Consider a feasible pt. $x_0 + \varepsilon z$, $\varepsilon > 0$, $z \in \mathcal{N}(A)$. $A(\underline{\underline{x_0 + \varepsilon z}}) = y$.

Let $h(\varepsilon) \triangleq G(x_0 + \varepsilon y)$; $h(0) = G(x_0)$.

Want to examine $h(\varepsilon)$ when x_0 is a BFS.

If $h(\varepsilon) > h(0)$ for some sufficiently small ε, $\forall z \in \mathcal{N}(A)$, then x_0 is a local min.

So we compute $\frac{\partial h(\varepsilon)}{\partial \varepsilon}$ and see if it is positive.

only the nonzero entries of $x_0 + \varepsilon z$, $k > m$

$$\frac{\partial h(\varepsilon)}{\partial \varepsilon} = \sum_{l=1}^{k} \frac{\partial G(x_0 + \varepsilon z)}{\partial (x_l + \varepsilon z_l)} \cdot \frac{\partial (x_l + \varepsilon z_l)}{\partial \varepsilon}$$

$$= \sum_{l=1}^{k} p\, \text{sgn}(x_{0,l} + \varepsilon z_l)\, |x_{0,l} + \varepsilon z_l|^{p-1} z_l$$

(used $\frac{\partial |x|^p}{\partial x} = p|x|^{p-1} \underbrace{\text{sgn}(x)}_{\text{sign}}$). (For small enough ε, $x_{0,l} + \varepsilon z_l \neq 0$ for l s.t. $x_{0,l} \neq 0$)

WLOG, assume the first m entries of x_0 are nonzero.

$$\frac{\partial h(\varepsilon)}{\partial \varepsilon} = \sum_{l=1}^{m} p\, \text{sgn}(x_{0,l} + \varepsilon z_l)\, |x_{0,l} + \varepsilon z_l|^{p-1} z_l + \sum_{l=m+1}^{k} p\, \text{sgn}(\varepsilon z_l)\, \varepsilon^{p-1} |z_l|^{p-1} z_l$$

$$= T_1 + T_2$$

As $\varepsilon \to 0$, $T_1 \to$ some constant (+ve or −ve)

$$T_2 = \sum_{l=m+1}^{k} p\, \varepsilon^{p-1} |z_l|^{p-1} \cdot \underbrace{\underbrace{\text{sgn}(\varepsilon z_l)}_{\text{sgn}(z_l) \text{ since } \varepsilon > 0} z_l}_{|z_l|}$$

$$= \sum_{l=m+1}^{k} p\, \varepsilon^{p-1} |z_l|^p$$

$$= \sum_{l=m+1}^{k} \frac{p}{\varepsilon^{1-p}} |z_l|^p, \quad 0<p<1$$

$(\varepsilon^{1-p})^{-1} \uparrow$ as $\varepsilon \to 0$.

So T_2 becomes large and dominates the sum as $\varepsilon \to 0$!

$\Rightarrow$ for $\varepsilon < \varepsilon_0$, $T_1 + T_2 > 0$

$\Rightarrow h(\varepsilon) > h(0) \ \forall \varepsilon < \varepsilon_0$, $z \in \mathcal{N}(A)$.

$\Rightarrow x_0$ is a local minimum.

Hence, all <u>BFSs are local minima</u>. <u>Not desirable</u>!

Modified fn. $G(x) = \sum_{l=1}^{N} (\underline{|x_l| + \beta})^p$, $0<p<1$, $\beta>0$.

Following the same approach,

$$\frac{\partial h(\varepsilon)}{\partial \varepsilon} = \sum_{l=1}^{k} p\, (|x_{0,l} + \varepsilon z_l| + \beta)^{p-1} \text{sgn}(x_{0,l} + \varepsilon z_l)\, z_l$$

$$= \sum_{l=1}^{m} p\, (|x_{0,l} + \varepsilon z_l| + \beta)^{p-1} \text{sgn}(x_{0,l} + \varepsilon z_l)\, z_l + \sum_{l=m+1}^{k} p\, (|\varepsilon z_l| + \beta)^{p-1} \text{sgn}(\varepsilon z_l)\, z_l$$

$= T_1 + T_2$, and $T_1 \xrightarrow{\varepsilon \to 0}$ constant as before (maybe +ve or −ve)

$$T_2 = \sum_{l=m+1}^{k} p\, (\varepsilon |z_l| + \underline{\beta})^{p-1} |z_l| > 0$$

As $\varepsilon \to 0$, no guarantee that T_2 dominates T_1.

So some BFSs may not be local minima. ☺

Some remarks on reweighted l_1 vs. reweighted l_2:

1. Each iteration of l_1 reweighting is more computationally expensive than reweighted l_2 (closed form update)
   - But # iterations w/ rewt. l_1 typically << # itns w/ rewt. l_2.
   - With rewt. l_1, even 1 itn $\longrightarrow$ sparse soln. (unlike rewt. l_2).
2. Typically, much easier to incorporate additional constraints (non-negativity, box constraints, etc) w/ rewt. l_1
   Rewt. $l_2 \Rightarrow$ no longer have closed-form solns.
3. Can consider other ways to bound besides l_1 or l_2; as long as the bounding fn. is convex, each itn solves a convex opt. pb.
   $l_2 \Rightarrow$ closed form updates
   $l_1 \Rightarrow$ Each iteration promotes sparsity
4. The choice of $W^{(k)}$ is determined by the surrogate cost fn. for promoting sparsity that is being minimized.
   - Different surrogate fns. lead to different algos with different performance-complexity tradeoffs.

[Chartrand and Yin] Reweighted ℓ_2

$$g(x) = (x^2+\beta)^{p/2}$$

$$h(\tilde{y}) = (\tilde{y}+\beta)^{p/2}, \quad \tilde{y} \geq 0$$

$$h'(\tilde{y}) = \frac{p}{2}(\tilde{y}+\beta)^{\frac{p}{2}-1}, \quad h''(\tilde{y}) = \underbrace{\frac{p}{2}\left(\frac{p}{2}-1\right)}_{<0}(\tilde{y}+\beta)^{\frac{p}{2}-2} < 0 \quad \text{concave.}$$

$$h(\tilde{y}) \leq h'(\tilde{y}_0)(\tilde{y}-\tilde{y}_0) + h(\tilde{y}_0)$$

$$x^{(k+1)} = \arg\min_{x \in \mathbb{R}^N} \|Ax-y\|_2^2 + \lambda \sum_{\ell=1}^{N} \left(h'(\tilde{y})\Big|_{\tilde{y}=(x_\ell^{(k)})^2} \right) x_\ell^2$$

$$= \arg\min_{x \in \mathbb{R}^N} \|Ax-y\|_2^2 + \lambda \sum_{\ell=1}^{N} \underbrace{\left(|x_\ell^{(k)}|+\beta\right)^{\frac{p}{2}-1}}_{w_\ell^{(k)}} x_\ell^2$$

$$W_k^{-2} = \operatorname{diag}\left\{ \left(|x_\ell^{(k)}|^2+\beta\right)^{\frac{p}{2}-1} \right\}$$

$\Rightarrow$ Reweighted l_2 based method.

let $w_\ell^{(k)} = \left(|x_\ell^{(k)}|+\beta\right)^{\frac{p}{2}-1}$

When $p \to 0$, Chartrand-Yin update:

$$w_\ell^{(k)} = \left(|x_\ell^{(k)}|+\beta\right)^{-1}.$$

When $\beta \to 0$, FOCUSS:

$$w_\ell^{(k)} = |x_\ell^{(k)}|^{p-2}.$$

When $p = 1$,

(Recall: $\min_{x \in \mathbb{R}^N} \underline{\sum_{i=1}^{N} (x_i^2+\beta)^{p/2}, \ \text{s.t. } Ax=y}$)

(As $\beta \to 0$ and $p \to 1$, $\min_{x \in \mathbb{R}^N} \sum_{i=1}^{N} |x_i|$ s.t. $Ax=y$. l_1 min.)

$$w_\ell^{(k)} = \left(|x_\ell^{(k)}|^2+\beta\right)^{\frac{-1}{2}}.$$

$\Rightarrow$ A more robust, iterative soln. to (P_1).

As $p \to 0$, we are trying to solve

$$\min_{x \in \mathbb{R}^N} \|Ax-y\|_2^2 + \lambda \sum_{\ell=1}^{N} |x_\ell^2+\beta|^0$$

$\Rightarrow$ Intuitive explanation of why the Chartrand & Yin updates yield sparse solns.