E9 203 Compressed sensing & Sparse SP

17 Mar. 2021

Last time:

- Thresholding based methods
- General diversity measures/concave optimization.

Today:

- Extreme points, BFS, concave opt.
- Properties of local optima
- Reweighting based methods.

Recall: $\boxed{\min_{z \in \mathbb{C}^N} G(z) \text{ s.t. } Az = y}$

Separable: $G(z) = \sum_{i=1}^{N} g(|z_i|)$

$g(|z|)$ monotonically $\uparrow$ in $|z|$, $g(0)$ bdd.

Examples: $g(x) = |x|^p$, $p \geq 0$

$g(x) = \ln(\varepsilon + |x|)$, $\varepsilon > 0$.

Reformulation:

(PG1) $\min_{x \in \mathbb{R}^N} J(x) = \sum_{i=1}^{N} g(|x_i|)$ s.t. $Ax = y$

(PG2) Define $w = \begin{bmatrix} x_+ \\ x_- \end{bmatrix} \in \mathbb{R}_+^{2N}$

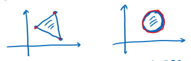
$\min_{w \in \mathbb{R}^{2N}} J_1(w) = \sum_{i=1}^{N} \left[g(|x_{+,i}|) + g(|x_{-,i}|) - g(0) \right]$

s.t. $\boxed{[A \ -A] w = y, \quad w \geq 0}$.

[HW]: There is a 1-1 mapping betn. the local minima of (PG1) and (PG2), and the value of the objective fn. is the same at corresponding local minima.

Basic feasible soln. (BFS): a soln. s.t. the cols of $[A \ -A]$ corresp. supp(w) are LI.

Extreme pt.: A point x in a convex set $\mathcal{C}$ is said to be an extreme pt. of $\mathcal{C}$ if there are no distinct $x_1, x_2 \in \mathcal{C}$ s.t. $x = \alpha x_1 + (1-\alpha) x_2$ for some $\alpha \in (0,1)$.



Equivalence of extreme pts and BFSs:

Let $\mathcal{K}$ be the convex polytope consisting of all N-vecs s.t.

$\boxed{Ax = y, \quad x \geq 0}$ — (*)

A vec. $x \in \mathbb{R}^N$ is an extreme pt. of $\mathcal{K}$ iff x is a BFS to (*).

Proof: Suppose x is a BFS, and $x_1, \ldots, x_k$ are nonzero. Then $a_1 \ldots a_k$ are LI.

If x is not an extreme pt. of $\mathcal{K}$, then $\exists$ $u, v \in \mathbb{R}^N$ s.t. $x = \alpha u + (1-\alpha) v$, $0 < \alpha < 1$, $u \neq v$, $u \geq 0$, $v \geq 0$.

$$\begin{bmatrix} x_1 \\ \vdots \\ x_k \\ \boxed{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}} \end{bmatrix} = \alpha \begin{bmatrix} u_1 \\ \vdots \\ u_k \\ \boxed{\begin{matrix} u_{k+1} \\ \vdots \\ u_N \end{matrix}} \end{bmatrix} + (1-\alpha) \begin{bmatrix} v_1 \\ \vdots \\ v_k \\ \boxed{\begin{matrix} v_{k+1} \\ \vdots \\ v_N \end{matrix}} \end{bmatrix}$$

$\Rightarrow u_{k+1} = \cdots = u_N = v_{k+1} = \cdots = v_N = 0$.

$u, v \in \mathcal{K} \Rightarrow Au = y$, $Av = y$

$a_1 (u_1 - v_1) + a_2 (u_2 - v_2) + \cdots + a_k (u_k - v_k) = 0$.

$u \neq v \Rightarrow a_1, \ldots, a_k$ are LD (contradiction).

Conversely, suppose x is an extreme pt. of $\mathcal{K}$ and $x_1, \ldots, x_k$ are nonzero. If $a_1 \ldots a_k$ are LD, then $\exists\, 0 \neq u = \begin{bmatrix} u_1 \\ \vdots \\ u_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ s.t. $Au = 0$.

Since $x_1, \ldots, x_k > 0$, can select $\varepsilon > 0$ s.t.

$\underbrace{x + \varepsilon u}_{\hat{x}_1} \geq 0$, $\underbrace{x - \varepsilon u}_{\hat{x}_2} \geq 0$

$\hat{x}_1 \neq \hat{x}_2$, $\hat{x}_1, \hat{x}_2 \in \mathcal{K}$, and

$x = \frac{1}{2} \hat{x}_1 + \frac{1}{2} \hat{x}_2$

$\Rightarrow x$ is not an extreme pt. (contradiction).

Hence, x is a BFS.

Remarks:

1. If the convex set $\mathcal{K}$ corresp. to (*) $(Ax = y, x \geq 0)$ is nonempty, it has at least one extreme pt.
2. If the LP has a finite optimal soln., it has a finite optimal soln. which is an extreme pt.
3. $\mathcal{K}$ has at most a finite # extreme pts.
4. If the convex polytope $\mathcal{K}$ is bdd, then $\mathcal{K}$ is a convex polyhedron, i.e., $\mathcal{K}$ consists of points that are a convex comb. of a finite # pts.

Local minima of concave fns.

Let $g(x) = \sum_{i=1}^{N} g(|x_i|)$,

strictly concave for $x \geq 0$. (*)

(PG2) $\min_{x \in \mathbb{R}^N} \sum_{i=1}^{N} g(|x_i|)$ s.t. $\underbrace{Ax = y, \ x \geq 0}_{\mathcal{K}}$.

Claim: When $g(\cdot)$ is strictly concave, all local minima of (PG2) occur at extreme pts of $\mathcal{K}$.

$\Rightarrow$ all local minima occur at BFSs to (*).

Proof: Let x_0 be a local min.

$Ax_0 = y$, $x_0 \geq 0$, $\exists\, \varepsilon > 0$ s.t.

$g(x_0) \leq g(x')$ $\forall\, x'$ s.t. $\|x_0 - x'\|_2 \leq \varepsilon$.

If x_0 is not an extreme pt., then $\exists\, x_1, x_2 \in \mathcal{K}$, $x_1 \neq x_2$, $\lambda \in (0,1)$ s.t. $x_0 = \lambda x_1 + (1-\lambda) x_2$.

By the strict concavity of $g(\cdot)$,

$g(x_0) > \lambda g(x_1) + (1-\lambda) g(x_2) \geq \min(g(x_1), g(x_2))$.

Now, pick any $0 \neq z \in \mathcal{N}(A)$, $\text{supp}(z) \subseteq \text{supp}(x_0)$.

Then, for small enough δ, $x_0 \pm \delta z$ is feasible.

$x_0 = \frac{1}{2}(x_0 + \delta z) + \frac{1}{2}(x_0 - \delta z)$

$\Rightarrow g(x_0) > \frac{1}{2} g(x_0 + \delta z) + \frac{1}{2} g(x_0 - \delta z)$

$\geq \min\left(g(x_0 + \delta z), g(x_0 - \delta z)\right)$

$\Rightarrow g(x_0) > \min\left(g(x_0 + \delta z), g(x_0 - \delta z)\right)$

$\Rightarrow (x_0 \pm \delta z)$ is a better pt. than x_0 (contradiction). ∎

General result on maximizing convex fns:

Thm. Let f be a convex fn. defined on the bounded, closed convex set Ω. If f has a maximum over Ω, it is achieved at an extreme pt. of Ω.

Cor. The global min. of (PG1) or (PG2) is also a BFS.

$$\min_{x \in \mathbb{R}^N} \sum_{i=1}^{N} g(|x_i|) \quad \text{s.t.} \quad \underline{Ax = y}, \ x \geq 0.$$

1. $g(x) = g(-x) = g(|x|)$
2. $g(x)$ is monotone $\uparrow$ for $x \in \mathbb{R}_+$
3. $g(x)$ is strictly concave $x \in \mathbb{R}_+$.

Noisy data:

$$\min_{\substack{x \in \mathbb{R}^N \\ x \geq 0}} \lambda \sum_{i=1}^{N} g(|x_i|) + \|Ax - y\|_2$$

$$\min_{\substack{x \in \mathbb{R}^N \\ x \geq 0}} \sum_{i=1}^{N} g(|x_i|) \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \varepsilon.$$

A general approach: Majorization-minimization (MM):

Let $f(\theta)$ be a fn. to be minimized.

Let $\boxed{g(\theta \mid \theta^{(m)}) \geq f(\theta)}$ $\forall \theta$

$(\underline{g_{\theta^{(m)}}(\theta)})$ $g(\theta^{(m)} \mid \theta^{(m)}) = f(\theta^{(m)})$.

$g(\theta \mid \theta^{(m)})$ an upper bd. on $f(\theta)$, eq. at $\theta^{(m)}$.

MM algo:

Init. $\theta^{(0)}$ = something convenient.

Iterate

$\theta^{(m+1)} = \arg\min_{\theta} g(\theta \mid \theta^{(m)})$ $\leftarrow$ Convenient for opt. e.g. linear, quadratic, etc

determine $g(\theta \mid \theta^{(m+1)})$.

Until convergence

$$\underline{f(\theta^{(m+1)})} \leq \underline{g(\theta^{(m+1)} \mid \theta^{(m)})} \leq g(\theta^{(m)} \mid \theta^{(m)}) = \underline{f(\theta^{(m)})}$$