E9 203 Compressed Sensing & Sparse Signal Processing

21 May 2021

Last time:
- Algos for l_1 min.
- Coordinate descent ("shooting algo")
- Homotopy based methods : LARS
- Proximal and gradient projection methods

Today:
- Complete the discussion on proximal & gradient proj$^{\underline{n}}$ methods
- Gelfand m-widths and <u>lower bounds</u> on the # meas. needed to satisfy RIP.

<u>Proximal and Gradient Projection Methods</u>

Objective fn: $f(x) = \underbrace{L(x)}_{\substack{\frac{1}{2}\|y-Ax\|_2^2 \\ \text{Loss}}} + \underbrace{R(x)}_{\substack{\lambda\|x\|_1 \\ \text{Regularizer}}}$

To develop intuition, suppose $A = I$. Then,

$$f(x) = \frac{1}{2}\|y - x\|_2^2 + R(x).$$

Soln.: $x_{opt} = \underbrace{\text{prox}_R(y)}_{\text{proximal operator}} = \arg\min_z \underbrace{\left\{ R(z) + \frac{1}{2}\|y - z\|_2^2 \right\}}_{\text{Minimize } R\text{, but remain close to } y}$

Will use the above in an iterative algo:

$$\text{prox}_R(\underbrace{x_k}_{\text{prev. iterate.}}) = \arg\min_z \left\{ R(z) + \frac{1}{2}\|x_k - z\|_2^2 \right\} \quad \text{—}(*)$$

(a) How to solve the above pb;
(b) How to extend to a general A.

(a) Proximal operator

If $R(x) = \lambda\|x\|_1$,

$\text{prox}_R(x) = \text{soft}(x, \lambda)$ (component-wise)

where $\text{soft}(a, \delta) = \text{sign}(a)(|a| - \delta)_+$

$x_+ = \max\{x, 0\}$.

If $R(x) = \lambda\|x\|_0$,

$\text{prox}_R(x) = \text{hard}(x, \sqrt{2\lambda})$

where $\text{hard}(u, a) = u\, \mathbb{1}_{\{|u| > a\}}$.

(b) Proximal gradient method

$$x_{k+1} = \arg\min_z \left[R(z) + L(x_k) + g_k^T(z - x_k) + \underbrace{\frac{1}{2t_k}\|z - x_k\|_2^2}_{(*)} \right]$$

(*) Alternatives to $\|\cdot\|_2^2$ exist.
For e.g., if we replace $\|\cdot\|_2^2$ with a Bregman divergence, we get "mirror method".

(*) We approximate $\nabla^2 L(x_k) \approx \frac{1}{t_k} I$.

Thus,
$$\boxed{\begin{aligned} x_{k+1} &= \arg\min_z \left[t_k R(z) + \frac{1}{2}\|z - u_k\|_2^2 \right] \\ &= \text{prox}_{t_k R}(u_k) \end{aligned}}$$

where

$$u_k = x_k - t_k g_k, \quad g_k = \nabla L(x_k).$$

If $R(x) = 0$, it is the gradient descent method.
If $R(x) = \lambda\|x\|_1$, get <u>iterative soft thresholding</u>.

Choice of t_k: natural choice $\left[\frac{1}{t_k} I \approx \underbrace{\nabla^2 L}_{\approx g(x_k) - g(x_{k-1})} \right.$

$$\frac{(x_k - x_{k-1})}{t_k} \approx g_k - g_{k-1}$$

Hence, we compute

$$\alpha_k = \arg\min_\alpha \|\alpha(x_k - x_{k-1}) - (g_k - g_{k-1})\|_2^2$$

$$= \frac{(x_k - x_{k-1})^T(g_k - g_{k-1})}{(x_k - x_{k-1})^T(x_k - x_{k-1})}$$

and set $t_k = \frac{1}{\alpha_k}$.

This is known as the <u>Barzilai-Browein (BB)</u> or <u>spectral stepsize</u> method.

BB stepsize + Iterative soft thresholding + continuation method (gradually reduce λ) $\Rightarrow$ <u>SpaRSA</u> [Wright 2009]

Sparse reconstruction by separable approx.

<u>Nesterov's Method</u>

$$x_{k+1} = \text{prox}_{t_k R}(\phi_k - t_k g_k)$$

(Earlier: $x_{k+1} = \text{prox}_{t_k R}(x_k - t_k g_k)$)

$$\phi_k = x_k + \left(\frac{k-1}{k+2}\right)(x_k - x_{k-1})$$

$$g_k = \nabla L(\phi_k)$$

This + IST ($R(x) = \lambda\|x\|_1$) + Continuation
$\Rightarrow$ <u>FISTA</u> (fast iterative shrinkage & thresholding algo).

Chapter 10: <u>Gelfand Widths of l_1 balls</u>

<u>Defn.</u> The <u>Gelfand m-width</u> of a subset K of a <u>normed space</u> $\mathcal{X}$ is

$$d^m(K, \mathcal{X}) = \inf_{A \in \mathbb{R}^{m \times N}} \sup_{x \in K \cap \mathcal{N}(A)} \|x\|$$

$\{d^m(K, \mathcal{X})\}_{m \geq 0}$ is nonincreasing.

$$d^0(K, \mathcal{X}) = \sup_{x \in K} \|x\|$$

If K contains the all zero vector,

$$d^m(K, \mathcal{X}) = 0 \quad \forall\ m \geq N.$$

<u>Defn.</u> The compressive m-width of a subset K of a normed space $\mathcal{X}$ is defined as

$$E^m(K, \mathcal{X}) = \inf\left\{ \sup_{x \in K} \|x - \Delta(Ax)\|, \begin{array}{l} A : \mathcal{X} \to \mathbb{R}^m \text{ linear} \\ \Delta : \mathbb{R}^m \to \mathcal{X} \end{array} \right\}.$$

Δ : Reconstruction map $\mathbb{R}^m \to \mathcal{X}$, arbitrary
(Could even be the NP-hard l_0 min. soln.)

A : nonadaptive $\quad Ax = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix}$ chosen once and for all.

In the context of CS, K = set of compressible vecs.

<u>Prop. 2.3</u> For any $q > p > 0$, and any $x \in \mathbb{C}^N$,

$$\sigma_s(x)_q \leq \frac{1}{s^{\frac{1}{p} - \frac{1}{q}}}\|x\|_p.$$

$\Rightarrow$ the nonconvex unit balls B_p^N with $p < 1$ are good models for compressible vecs.

E.g. if $p = 0.1$, $\|x\|_p \leq 1$, then with $q = 1$,

[E.g.

$$\sigma_s(x)_1 \leq \frac{1}{s^q} . \text{]}$$

For simplicity, we consider $p = 1$ here.

We have the foll. result:

Thm. 10.5 For $1 < p \leq 2$ and $m < N$, $\exists$ consts. $c_1, c_2 > 0$ depending only on p s.t.

$$c_1 \min\left\{1, \frac{\ln\left(\frac{eN}{m}\right)}{m}\right\}^{1-\frac{1}{p}} \leq d^m(\underbrace{B_1^N}_{K}, \underbrace{\ell_p^N}_{X}) \leq c_2 \min\left\{1, \frac{\ln\left(\frac{eN}{m}\right)}{m}\right\}^{1-\frac{1}{p}}$$

Matching upper and lower bounds.

Leads to this result on compressive widths:

Cor. 10.6 For $1 < p \leq 2$ and $m < N$, the compressive m-width satisfies

$$c_1' \min\left\{1, \frac{\ln\left(\frac{eN}{m}\right)}{m}\right\}^{1-\frac{1}{p}} \leq E^m(B_1^N, \ell_p^N) \leq c_2' \min\left\{1, \frac{\ln\left(\frac{eN}{m}\right)}{m}\right\}^{1-\frac{1}{p}}$$

for some consts. $c_1', c_2' > 0$ that dep. only on p.

Remark: We have seen that, if $m \geq c\, s \ln\left(\frac{eN}{s}\right)$, there are matrices $A \in \mathbb{R}^{m \times N}$ with small RIC and reconstruction maps Δ (e.g. BP, IHT, OMP, etc) s.t.

$$\| x - \Delta(Ax) \|_p \leq \frac{C}{s^{1-\frac{1}{p}}} \sigma_s(x)_1$$

[$m \geq c\, s \ln\left(\frac{eN}{s}\right)$ is sufficient]. Conversely, using the above corollary, we can s.t. the existence of A & Δ providing such a stability result in the reconstruction error forces $m \geq c'\, s \ln\left(\frac{eN}{s}\right)$ to be satisfied:

Prop. 10.7 Let $1 < p \leq 2$. Suppose $\exists\, A \in \mathbb{R}^{m \times N}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^N$ s.t. $\forall\, x \in \mathbb{R}^N$

$$\| x - \Delta(Ax) \|_p \leq \frac{C}{s^{1-\frac{1}{p}}} \sigma_s(x)_1$$

Then, for some consts. $c_1, c_2 > 0$ dep. only on C,

$$m \geq c_1\, s \ln\left(\frac{eN}{s}\right), \text{ provided } s > c_2 .$$