The Fastest $\ell_{\infty,1}$ Prox in the West

Béjar, Benjamín and Dokmanić, Ivan and Vidal, René

Speaker: Apple Zhang

Shenzhen University

November 29, 2022

Table of Contents

Table of Contents

Sparse regression with $L_{\infty,1}$ constraints

Consider the following optimization model [1]:

$$\min_{\boldsymbol{W}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_2^2, \quad \text{s.t. } \|\boldsymbol{W}\|_{\infty,1} \leq \tau. \tag{1}$$

Recall $L_{p,q}$-norm

Define $L_{p,q}$-norm

$$\|\boldsymbol{W}\|_{p,q} = \left(\sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_p^q \right)^{1/q} \tag{2}$$

And we usually use $L_{2,1}$-norm:

$$\|\boldsymbol{W}\|_{2,1} = \sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_2 \tag{3}$$

In this paper, we will use $L_{\infty,1}$-norm and $L_{1,\infty}$-norm:

$$\|\boldsymbol{W}\|_{\infty,1} = \sum_{i=1}^{m} \|\boldsymbol{w}^{(i)}\|_\infty = \sum_{i=1}^{m} \max_{j=1,2,\cdots,c} |w_{ij}|, \tag{4}$$

$$\|\boldsymbol{W}\|_{1,\infty} = \max_{i=1,2,\cdots,m} \|\boldsymbol{w}^{(i)}\|_1. \tag{5}$$

Table of Contents

KKT conditions for convex problems

For the following convex problem, where $f_i(x)$ are convex functions.

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}), \quad \text{s.t. } f_i(\boldsymbol{x}) \leq 0 \; (i = 1, 2, \cdots, m), \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{6}$$

We can write Lagrange function:

$$L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{x}) + \boldsymbol{\beta}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}).$$

And its KKT conditions are

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*), \tag{7}$$

$$f_i(\boldsymbol{x}^*) \leq 0, \quad \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}, \quad \alpha_i^* \geq 0, \quad \alpha_i^* f_i(\boldsymbol{x}^*) = 0. \tag{8}$$

Note that if $f_i(\boldsymbol{x}),\ i = 0, 1, \cdots, m$ are differentiable, then we have

$$\nabla_{\boldsymbol{x}} L = \nabla_{\boldsymbol{x}} f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \alpha_i \nabla_{\boldsymbol{x}} f_i(\boldsymbol{x}) - \boldsymbol{A}^\top \boldsymbol{\beta} = \boldsymbol{0}. \tag{9}$$

Proximal operator

For a convex function $f(\cdot)$, define its proximal operator as

$$\text{prox}_f(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2,$$

which means minimize $f(x)$ and make the solution is close to $\boldsymbol{x}$.

Projection: a special case of proximal operator

For non-empty convex set $\mathcal{C}$, define

$$\pi_{\mathcal{C}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}\in\mathcal{C}} \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2. \tag{10}$$

Projection is a special proximal operator since

$$\pi_{\mathcal{C}}(\boldsymbol{x}) = \mathrm{prox}_{\mathbb{I}(\cdot\in\mathcal{C})}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} \mathbb{I}(\boldsymbol{z}\in\mathcal{C}) + \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2. \tag{11}$$

where $\mathbb{I}(\cdot \in \mathcal{C})$ is indicator function:

$$\mathbb{I}(\boldsymbol{z}\in\mathcal{C}) = \begin{cases} 0, & \boldsymbol{z}\in\mathcal{C}, \\ +\infty, & \text{otherwise}. \end{cases} \tag{12}$$

Example: projection onto of L_1-ball

Consider L_1-ball projection:

$$\pi_{\|\cdot\|_1 \le \lambda}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2, \quad \text{s.t. } \|\boldsymbol{z}\|_1 \le \lambda. \tag{13}$$

Lagrange:

$$\begin{aligned} L(\boldsymbol{z}, \alpha) &= \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + \alpha(\|\boldsymbol{z}\|_1 - \lambda) \qquad (14) \\ &= \left[\frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|_2^2 + \alpha\|\boldsymbol{z}\|_1\right] - \alpha\lambda. \qquad (15) \end{aligned}$$

Utilize **KKT conditions,**

$$\begin{aligned} \boldsymbol{z} = \arg\min_{\boldsymbol{z}} L(\boldsymbol{z}, \alpha) = \operatorname{prox}_{\alpha\|\cdot\|_1}(\boldsymbol{x}), \qquad (16) \\ \alpha \ge 0, \quad \alpha(\|\boldsymbol{z}\|_1 - \lambda) = 0. \qquad (17) \end{aligned}$$

Example: projection onto of L_1-ball

Algorithm 1 $O(n \log n)$ Algorithm for $\pi_{\|\cdot\|_1 \le \lambda}$

Require: $\boldsymbol{x} \in \mathbb{R}^n, \lambda > 0$.

Ensure: $\boldsymbol{z} = \pi_{\|\cdot\|_1 \le \lambda}(\boldsymbol{x})$.

```
if ‖x‖₁ ≤ λ then
    z ← x.
else
    u ← sort(x, 'descend'),
    ρ ← max{j = 1, 2, ⋯ , n | u_j − (∑_{r=1}^{j} u_r − λ)/j > 0},
    α ← (∑_{r=1}^{ρ} u_i − λ)/ρ,
    z_i ← sign(x_i)[u_i − α]_+,   ∀i = 1, 2, ⋯ , n.
end if
```

if $\|\boldsymbol{x}\|_1 \le \lambda$ **then**

$\quad \boldsymbol{z} \leftarrow \boldsymbol{x}$.

else

$\quad \boldsymbol{u} \leftarrow \mathrm{sort}(\boldsymbol{x}, \texttt{'descend'})$,

$\quad \rho \leftarrow \max\{j = 1, 2, \cdots, n \mid u_j - (\sum_{r=1}^{j} u_r - \lambda)/j > 0\}$,

$\quad \alpha \leftarrow (\sum_{r=1}^{\rho} u_i - \lambda)/\rho$,

$\quad z_i \leftarrow \mathrm{sign}(x_i)[u_i - \alpha]_+, \quad \forall i = 1, 2, \cdots, n$.

end if

Table of Contents

Projected gradient descent

$$\min_{\boldsymbol{W}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_2^2, \quad \text{s.t. } \|\boldsymbol{W}\|_{\infty,1} \leq \tau. \tag{18}$$

Update

$$\boldsymbol{V} \leftarrow \boldsymbol{W} - \eta \nabla_{\boldsymbol{W}} J(\boldsymbol{W}), \tag{19}$$

$$\boldsymbol{W} \leftarrow \pi_{\|\cdot\|_{\infty,1} \leq \tau}(\boldsymbol{V}). \tag{20}$$

Projection and proximal operator

Lemma

For all $\tau > 0$,

$$\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V}) + \pi_{\|\cdot\|_{\infty,1}\leq\tau}(\boldsymbol{V}) = \boldsymbol{V} \tag{21}$$

Therefore, we should first compute $\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V})$, i.e.,

$$\mathrm{prox}_{\tau\|\cdot\|_{1,\infty}}(\boldsymbol{V}) = \arg\min_{\boldsymbol{W}} \tau\|\boldsymbol{W}\|_{1,\infty} + \frac{1}{2}\|\boldsymbol{W} - \boldsymbol{V}\|_2^2. \tag{22}$$

Just for reminder, the definition of $L_{1,\infty}$ is

$$\|\boldsymbol{W}\|_{1,\infty} = \max_{i=1,2,\cdots,m} \|\boldsymbol{w}^{(i)}\|_1. \tag{23}$$

Table of Contents

$L_{\infty,1}$ v.s. $L_{2,1}$

Datasets	RFS ($L_{2,1}$)	$L_{\infty,1}$-regression	SVM
DNA	91.69±0.97	93.60±0.59	90.51±1.31
Binaryalpha	55.27±2.56	57.53±1.37	62.81±1.81
USPS	88.10±0.51	88.26±0.38	92.10±0.59

Thank you!

[1] B. Béjar, I. Dokmanić, and R. Vidal, "The Fastest $\ell_{\infty,1}$ Prox in the West," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3858–3869, 2022.