Projection onto unit L_1 ball

Andersen Ang

Mathématique et recherche opérationnelle, UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: August 6, 2018
Last update: December 24, 2020

Projection

- The Euclidean projected of a given vector $\mathbf{y} \in \mathbb{R}^n$, denoted as $P_S(\mathbf{y})$ is a function $\mathbb{R}^n \to \mathbb{R}^n$ that output a point $\mathbf{x}$ by solving the following optimization problem

$$\hat{\mathbf{x}} = P_S(\mathbf{y}) = \underset{\mathbf{x} \in S}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{y}\|_2,$$

- **Question**: what if S is the unit L_1 ball?

$$\text{Unit } L_1 \text{ ball} \quad S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq 1\}$$

- Remark about the projection problem.
  - The projection onto unit L_1 ball is a convex optimization problem.
    - The objective function is (strongly) convex.
    - The constraint set $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq 1\}$ is convex and compact.
  - There exists a unique sol. to this projection problem. Details here.

Remarks

- The inequality in $\|\mathbf{x}\|_1 \leq 1$ is important, without it S is non-convex.
- *Showing* $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 = 1\}$ *is non-convex by example*
  Consider 2-dimensional case, the point $\mathbf{x}_1 = [1, 0]$ and $\mathbf{x}_2 = [-1, 0]$ both have unit L_1 norm but for the point $0.5\mathbf{x}_1 + 0.5\mathbf{x}_2 = [0, 0]$, its L_1 norm is 0.
- Unlike the case of L_2 norm, there is no close form solution for projection onto unit L_1 norm ball.

Lagrangian and KKT system

- Instead of considering

$$\underset{\|\mathbf{x}\|_1 \leq 1}{\operatorname{argmin}} \|\mathbf{x}-\mathbf{y}\|_2,$$

we consider

$$\underset{\|\mathbf{x}\|_1 \leq 1}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2{}_2,$$

the squares are for simplicity of taking derivatives and it does not affect the structure of the proof, so proving the squared-problem also proves the original problem.

- We now study the problem by Lagrangian method:
  - Build the Lagrangian and KKT system, we get 2 variables $(\mathbf{x}, \lambda)$.
  - Find λ such that KKT system is satisfied. Such λ reduces the KKT system to a single variable in x.
  - Solve for $\mathbf{x}$ and we are done.

Lagrangian of $\underset{\|\mathbf{x}\|_1 \leq 1}{\operatorname{argmin}} \|\mathbf{x}-\mathbf{y}\|_2^2$

$$\mathcal{L}(\mathbf{x}, \lambda) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda(\|\mathbf{x}\|_1 - 1) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1 - \lambda.$$

- L_1, L_2 norms are separable, so $\mathcal{L}$ can be expressed as

$$\mathcal{L}(\mathbf{x}, \lambda) = \sum_{i=1}^{n} \Big(\underbrace{\frac{1}{2}(x_i - y_i)^2 + \lambda|x_i|}_{l_i(x_i,\lambda)} \Big) - \lambda,$$

- Recall the stationarity condition in KKT is $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \lambda) = 0$, which is

$$\left[\frac{\partial l_1(x_1, \lambda)}{\partial x_1}, \frac{\partial l_2(x_2, \lambda)}{\partial x_2}, \dots, \frac{\partial l_n(x_n, \lambda)}{\partial x_n}\right] = 0.$$

  As L_1 norm is not differentiable so we have to use subgradient

- Using subgradient, $\partial_{\mathbf{x}} l_i = 0$ becomes $0 \in x_i - y_i + \lambda \mathsf{sgn}(x_i)$, which is

$$y_i = x_i + \lambda \mathsf{sgn}(x_i)$$

  Swaps the subject (express x_i as a function of y_i) gives the soft-thresholding operator

$$x_i = \mathsf{sgn}(y_i)(|y_i| - \lambda)_+, \text{ where } (\cdot)_+ = \max\{\cdot, 0\}.$$

KKT conditions

$$\mathcal{L}(\mathbf{x}, \lambda) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda(\|\mathbf{x}\|_1 - 1) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1 - \lambda.$$

- KKT conditions

$$\begin{array}{rl} \text{Stationarity} & 0 \in \partial\mathcal{L}(\mathbf{x}, \lambda) \\ \text{Slackness} & \lambda(\|\mathbf{x}\|_1 - 1) = 0 \\ \text{Primal feasibility} & \|\mathbf{x}\|_1 - 1 \leq 0 \\ \text{Dual feasibility} & \lambda \geq 0 \end{array} \implies \begin{array}{rl} i & x_i = \mathsf{sgn}(y_i)(|y_i| - \lambda)_+ \\ ii & \lambda(\|\mathbf{x}\|_1 - 1) = 0 \\ iii & \|\mathbf{x}\|_1 - 1 \leq 0 \\ iv & \lambda \geq 0 \end{array}$$

- As we want to find λ that the KKT system holds so we start with (iv) with two possibilities : $\lambda = 0$ or $\lambda > 0$.

$(i)\ x_i = \mathsf{sgn}(y_i)(|y_i| - \lambda)_+, \quad (ii)\ \lambda(\|\mathbf{x}\|_1 - 1) = 0 \quad (iii)\ \|\mathbf{x}\|_1 - 1 \leq 0.$

- **Cases** $\lambda = 0$
  (i) gives $x_i = \mathsf{sgn}(y_i)|y_i| = y_i$, together with (iii) gives $\|\mathbf{x}\|_1 = \|\mathbf{y}\|_1 \leq 1$, this corresponds to the case $\mathbf{y} \in S$.
- **Cases** $\lambda > 0$
  (ii) gives $\|\mathbf{x}\|_1 = 1$, that is

  $$|x_1| + |x_2| + ... + |x_i| + ... + |x_n| = 1 \tag{1}$$

  Apply (i) into (1) we get $\sum_{i=1}^{n} |\mathsf{sgn}(y_i)(|y_i| - \lambda)_+| = 1$, note that $\mathsf{sgn}(y_i)$ and the absolute sign can be removed and we have

  $$\sum_{i=1}^{n} (|y_i| - \lambda)_+ = 1. \tag{2}$$

  So in this case the solution is

  $$x_i = \mathsf{sgn}(y_i)(|y_i| - \lambda)_+,$$

  where λ is the root of (2), a piece-wise linear equation.

$x_i = \text{sgn}(y_i)(|y_i| - \lambda)_+$

- The expression $x_i = \text{sgn}(y_i)(|y_i| - \lambda)_+$ has explicit form by considering 3 intervals

$$(1)\ y_i > \lambda, \quad (2)\ y_i \in [-\lambda, \lambda], \quad (3)\ y_i < -\lambda$$

- Case (1) : as $\lambda > 0$ so
  - $y_i > \lambda > 0$ is positive, so $\text{sgn}(y_i) = 1$
  - $y_i > \lambda$, so $(.)_+$ can be removed

  and we have $x_i = y_i - \lambda$.

- Case (2) : as $|y_i| \leq \lambda$ so $|y_i| - \lambda \leq 0$ so $(|y_i| - \lambda)_+ = 0$ thus $x_i = 0$

- Case (3) : as $\lambda > 0$ so
  - $y_i < -\lambda < 0$ is negative, so $\text{sgn}(y_i) = -1$ and $y_i = -|y_i|$
  - $y_i < -\lambda \implies |y_i| > \lambda \implies |y_i| - \lambda > 0$ so $(.)_+$ can be removed

    $x_i = \text{sgn}(y_i)(|y_i| - \lambda)_+ = -(|y_i| - \lambda) = -|y_i| + \lambda = y_i + \lambda.$

Summary

- If $\|\mathbf{y}\|_1 \leq 1$, $\mathbf{x} = \mathbf{y}$. Otherwise we have component-wise solution as
$$x_i = \mathrm{sgn}(y_i)(|y_i| - \lambda)_+ = \begin{cases} y_i - \lambda & y_i > \lambda \\ 0 & y_i \in [-\lambda, \lambda] \\ y_i + \lambda & y_i < -\lambda \end{cases},$$
where λ is the root of the piecewise linear equation to establish KKT:
$$\sum_{i=1}^{n} (|y_i| - \lambda)_+ = 1.$$
- The piecewise linear equation can be solved by sorting. Details here. This method has the cost $\mathcal{O}(n \log n)$.
- The best method so far: Laurent Condat, Fast projection onto the simplex and the l_1 ball, Mathematical Programming, 2016.
- The paper also reviewed several methods (including the highly cited Duchi2008), and listed out the errors of these methods.
- For small n, all methods are OK, can simply use the sorting-based method. For big n ($\geq 10^6$), one should use Laurent Condat's method.

End of document