Projection onto unit L_2 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 23, 2020

Projection

- Given $\mathbf{y} \in \mathbb{R}^n$, the Euclidean projection of $\mathbf{y}$ onto a (non-empty and compact)set $S \subseteq \mathbb{R}^n$, denoted as $P_S(\mathbf{y})$, is a function $\mathbb{R}^n \to \mathbb{R}^n$ that output a point $\hat{\mathbf{x}}$ by solving the following optimization problem

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

  Such optimization always has a unique solution, details here.

- **Question** What if S is the unit L_2 ball?

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

  where

$$\text{Unit } \; L_2 \; \text{ ball } S = \{\mathbf{x} \in \mathbf{R}^n : \|\mathbf{x}\|_2 \leq 1\}.$$

Remarks

- The projection onto unit L_2 ball is a convex optimization problem
  - The objective function is (strongly) convex.
  - The constraint set $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$ is convex, non-empty and compact.

  The problem has a unique global minimizer.

- Note. The inequality in $\|\mathbf{x}\|_2 \leq 1$ is important, without the inequality $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}$ which is non-convex.

- *Proving S is non-convex by counterexample*
  Consider 2-dimensional case, the point $\mathbf{x}_1 = [1, 0]$ and $\mathbf{x}_2 = [0, 1]$ both have unit norm so $\mathbf{x}_1, \mathbf{x}_2 \in S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1\}$ but the point $0.5\mathbf{x}_1 + 0.5\mathbf{x}_2 = [0.5, 0.5] \notin S$.

Close form solution of $\underset{\|\mathbf{x}\|_2 \leq 1}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{y}\|_2$

- A theorem

$$\frac{\mathbf{y}}{\max\{1, \|\mathbf{y}\|_2\}} = \underset{\|\mathbf{x}\|_2 \leq 1}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{y}\|_2.$$

- A "geometric" proof: let $S = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$.
  - If $\mathbf{y} \in S$, then $\|\mathbf{y}\|_2 \leq 1$ and $\mathbf{y}$ itself is the closest to $\mathbf{y}$.
  - If $\mathbf{y} \notin S$, then $\|\mathbf{y}\|_2 > 1$ and the closest point $\mathbf{x} \in S$ to $\mathbf{y}$ will be simply $\frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ as the norm of $\frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ is 1 :

$$\left\| \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2 = \sqrt{\frac{\mathbf{y}^\top \mathbf{y}}{\|\mathbf{y}\|_2^2}} = \sqrt{\frac{\|\mathbf{y}\|_2^2}{\|\mathbf{y}\|_2^2}} = 1.$$

  - Combine both cases gives $\frac{\mathbf{y}}{\max\{1, \|\mathbf{y}\|_2\}}$.

- If we project to L_2 ball with radius a instead of 1, just replace the 1 in the equation by a.

Second proof, using Lagrangian and KKT conditions

- Instead of proving

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

  prove

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

  The squares are for simplicity of taking derivatives. As the square does not affect the structure of the proof, so proving the squared-problem proves the original problem.

- Proof idea: use Lagrangian method by constructing the Lagrangian and KKT system so we have 2 variables $(\mathbf{x}, \lambda)$. Find λ such that KKT system is satisfied. With such λ, the KKT system reduces to a one variable problem in $\mathbf{x}$. Solve for $\mathbf{x}$ and we are done.

The proof of $\underset{\|\mathbf{x}\|_2^2 \le 1}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{y}\|_2^2 = \dfrac{\mathbf{y}}{\max\{1, \|\mathbf{y}\|_2^2\}}$

- Lagrangian : $L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda(\|\mathbf{x}\|_2^2 - 1)$.
- KKT conditions

$$\begin{aligned} \text{Stationarity} \quad & \nabla_{\mathbf{x}} L = 2(\mathbf{x} - \mathbf{y}) + 2\lambda \mathbf{x} = 0 \\ \text{Slackness} \quad & \lambda(\|\mathbf{x}\|_2^2 - 1) = 0 \\ \text{Primal feasibility} \quad & \|\mathbf{x}\|_2^2 - 1 \le 0 \\ \text{Dual feasibility} \quad & \lambda \ge 0 \end{aligned}$$

- Simplifying the KKT conditions gives

$$\begin{array}{llll} (1) & (1 + \lambda)\mathbf{x} = \mathbf{y} & (2) & \lambda(\|\mathbf{x}\|_2^2 - 1) = 0 \\ (3) & \|\mathbf{x}\|_2^2 - 1 \le 0 & (4) & \lambda \ge 0 \end{array}$$

As we want to eliminate λ, we start with (4). We have two cases: $\lambda = 0$ or $\lambda > 0$.

The proof

$$(1)\ (1+\lambda)\mathbf{x} = \mathbf{y} \quad (2)\ \lambda(\|\mathbf{x}\|_2^2 - 1) = 0 \quad (3)\ \|\mathbf{x}\|_2^2 - 1 \leq 0.$$

- **Cases** $\lambda = 0$
  (1) gives $\mathbf{x} = \mathbf{y}$ and (3) gives $\|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2 \leq 1$.
  This corresponds to the case $\mathbf{y}$ inside S.
- **Cases** $\lambda > 0$
  (2) gives $\|\mathbf{x}\|_2^2 = 1$, multiply (1) by $\mathbf{x}^\top$, use (2) we have
  $$(1+\lambda)\underbrace{\mathbf{x}^\top\mathbf{x}}_{1} = \mathbf{x}^\top\mathbf{y} \implies \lambda = \mathbf{x}^\top\mathbf{y} - 1 > 0$$
  And $\mathbf{x}^\mathbf{t}\mathbf{y} - 1 > 0 \iff \mathbf{x}^\top\mathbf{y} > 1 \iff \underbrace{\|\mathbf{x}\|_2^2}_{1}\|\mathbf{y}\|_2^2 > 1$ so $\|\mathbf{y}\|_2^2 > 1$, this is the case $\mathbf{y}$ outside S.
  Furthermore, put $\lambda = \mathbf{x}^\top\mathbf{y} - 1$ to (1) gives $(\mathbf{x}^\top\mathbf{y})\mathbf{x} = \mathbf{y}$ meaning $\mathbf{x}$, $\mathbf{y}$ are parallel so $\mathbf{x} = k\mathbf{y}$ for some scalar k, put $\mathbf{x} = k\mathbf{y}$ into $(\mathbf{x}^\top\mathbf{y})\mathbf{x} = \mathbf{y}$ get $k = \dfrac{1}{\|\mathbf{y}\|_2^2}$ so $\mathbf{x} = \dfrac{\mathbf{y}}{\|\mathbf{y}\|_2^2}$.
- Combine the 2 cases gives $\mathbf{x} = \dfrac{\mathbf{y}}{\max\{1, \|\mathbf{y}\|_2^2\}}$.