Projection onto L-Lipschitz matrix

Andersen Ang

Mathématique et recherche opérationnelle, UMONS, Belgium

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

First draft: November 21, 2018
Last update: December 23, 2020

Definition of Lipschitz

- **Definition of Lipschitz of function** f
  Given a function $f : \mathcal{Q} \subset \mathbb{R}^n \to \mathbb{R}$. We say f is L-Lipschitz if the following is true for all $\mathbf{x}, \mathbf{y} \in \mathcal{Q}$

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2.$$

- **Definition of Lipschitz matrix (linear map)**
  Suppose f is a linear map (i.e., f is a matrix). We say a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is $L-$Lipschitz if the following is true for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2.$$

Projection problem

- Let $\mathcal{L}_L$ denotes the set of matrices that are L-Lipschitz :

$$\mathcal{L}_L = \left\{ \mathbf{A} \in \mathbb{R}^{m \times n} \;\middle|\; \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2 \;\forall\; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \right\}.$$

- **A projection problem**
  Given a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$, find the matrix $\mathbf{X} \in \mathcal{L}_L$ that is closest to $\mathbf{M}$ under a norm:

$$\mathcal{P}_{(L,\zeta)} \;:\; \underset{\mathbf{X} \in \mathcal{L}_L}{\operatorname{argmin}} \; \frac{1}{2}\|\mathbf{X} - \mathbf{M}\|_\zeta^2.$$

- In fact, this problem has simple solution when $L = 1$ and $\zeta = F$: take SVD of $\mathbf{M}$, set all singular values larger than 1 to 1, the resulting matrix is the solution.

The set $\mathcal{L}_L$ is in fact simple

$$\mathcal{L}_L = \left\{ \mathbf{A} \in \mathbb{R}^{m \times n} \;\middle|\; \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2 \;\forall\; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \right\}.$$

- Let set $\mathcal{L}_L$ is just the set of matrices with singular values smaller than L. To see this : consider

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_2 &\leq \|\mathbf{A}(\mathbf{x} - \mathbf{y})\|_2 \\ &\leq \|\mathbf{A}\|_2 \cdot \|\mathbf{x} - \mathbf{y}\|_2 \quad \text{(operator-norm inequality)} \end{aligned}$$

- Hence $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2$ is equivalent to $\|\mathbf{A}\|_2 \leq L$.

- Denoting the largest singular value(s) of the matrix be σ_1. Then $\mathcal{L}_L$ is

$$\mathcal{L}_L = \left\{ \mathbf{A} \in \mathbb{R}^{m \times n} \;\middle|\; \sigma_1(\mathbf{A}) \leq L \right\}.$$

A projected gradient method

- In compact notation, we can re-write the problem $\mathcal{P}$ as

$$\mathcal{P}_{(L,\zeta)} : \underset{\mathbf{X}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{M}\|_\zeta^2 \ \text{ s.t. } \mathbf{X} \in \left\{ \mathbf{A} \in \mathbb{R}^{m\times n} \ \middle| \sigma_1(\mathbf{A}) \leq L \right\}.$$

- Projected gradient method to solve this problem:
  - Step 0: make an initial guess of $\mathbf{X} \in \mathcal{L}_L$.
  - Step 1: perform a gradient update

    $$\mathbf{X} = \mathbf{X} - t\nabla_{\mathbf{X}}\left(\frac{1}{2}\|\mathbf{X}-\mathbf{M}\|_\zeta^2\right).$$

    If $\|\mathbf{X}-\mathbf{M}\|_\zeta^2$ is not differentiable, use a sub-gradient.
  - Step 2: perform a projection on the singular value of the updated $\mathbf{X}$
    - Step 2a: run a SVD : $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top$
    - Step 2b: replace all singular values in Σ larger than L by L :

      $$\sigma_i = \min\{\sigma_i, L\}.$$

  - Repeat steps 1-2 until converge.

A singular value thresholding algorithm

- The main loop of the projected gradient has 2 lines:
  - 1. (gradient update)

$$\mathbf{X} = \mathbf{X} - t\nabla_{\mathbf{X}}\Big(\frac{1}{2}\|\mathbf{X} - \mathbf{M}\|_\zeta^2\Big).$$

  - 2. (singular value thresholding)

$$\mathbf{X} = \texttt{SVT}_L(\mathbf{X}),\ \text{where } \texttt{SVT}_L : \sigma_i = \min\{\sigma_i, L\}.$$

  Therefore, the projected gradient framework is a *singular value thresholding* (SVT) algorithm.

- Writing the 2 steps in 1 line :

$$\mathbf{X} = \texttt{SVT}_L\left[\mathbf{X} - t\nabla_{\mathbf{X}}\Big(\frac{1}{2}\|\mathbf{X} - \mathbf{M}\|_\zeta^2\Big)\right].$$

- Next slide: a special case where $L = 1$ and $\zeta = F$.
  Why this is useful: when a mapping is 1-Lipschitz, that function is a *non-expansive* mapping.

SVT for projection onto set of 1-Lipschitz under F-norm

- Put $L = 1$ and $\zeta = F$, we have the iteration

$$\mathbf{X} = \texttt{SVT}_1 \left[\mathbf{X} - t \nabla_{\mathbf{X}} \Big(\frac{1}{2} \|\mathbf{X} - \mathbf{M}\|_F^2 \Big) \right].$$

- The gradient $g(\mathbf{X}) = \nabla_{\mathbf{X}} \Big(\frac{1}{2} \|\mathbf{X} - \mathbf{M}\|_F^2 \Big) = \mathbf{X} - \mathbf{M}$, the stepsize t can be set as L^{-1}, which is 1. Hence

$$\mathbf{X} = \texttt{SVT}_1[\mathbf{X} - 1(\mathbf{X} - \mathbf{M})] = \texttt{SVT}_1[\mathbf{M}].$$

  i.e. to solve the 1-Lipschitz matrix projection problem in F-norm, just shrink those singular values (of the input matrix) larger than 1 to 1.

- Such mapping $\mathbf{X}$ is *non-expansive*: let the distance between $\mathbf{x}$ and $\mathbf{y}$ in the original space in $\mathbb{R}^n$ be d, the linear map $\mathbf{X}$ transforms $(\mathbf{x}, \mathbf{y})$ from $\mathbb{R}^n$ to a space in $\mathbb{R}^m$, such mapping either shrinks d, or keep d unchanged.

End of document