Stochastic Spectral Descent for 1-layer Recurrent Neural Network

BIG FRAMEWORK

- Lipchitz relation: Functions that satisfies the following relation$(p^{-1} + q^{-1} = 1)$

$$\|f'(x_1) - f'(x_2)\|_q \leq L_p \|x_1 - x_2\|_p$$

has an upper bound

$$f(x_2) \leq f(x_1) + \langle f'(x_1), x_2 - x_1 \rangle + \frac{L_p}{2} \|x_2 - x_1\|_p^2$$

- Maximize the right part leads to a MM optimization method

$$x_{k+1} = x_k - [f'(x_k)]^\#$$

where

$$x^\# = argmax_s\{\langle x, s \rangle - \frac{1}{2}\|s\|_p^2\}$$

- When $p = q = 2$, it reduces to SGD. In our method, $p = \infty$, which leads to a more tight upper bound of log-of-sum function.