FDR2 for Quartic toy model

October 9, 2019

We consider the one-dimensional quartic potentials

$$f^-(\theta) = (\theta - 1)^4 \tag{1}$$

and

$$f^+(\theta) = (\theta + 1)^4 \tag{2}$$

and assume that they are sampled each time with probability 1/2.

Total loss function

The full-batch loss function is the average of the two mini-batch loss functions above, that is:

$$\begin{aligned} f(\theta) &= \frac{1}{2}\Big(f^-(\theta) + f^+(\theta)\Big) && (3) \\ &= \frac{1}{2}\Big((\theta - 1)^4 + (\theta + 1)^4\Big) && (4) \\ &= \theta^4 + 6\theta^2 + 1 && (5) \end{aligned}$$

So the total gradient w.r.t. the parameter θ is

$$\partial_\theta f(\theta) = 4\theta^3 + 12\theta \tag{6}$$

And thus the left hand side of Sho's FDR2 should be the expectation of this quantity squared, namely:

$$\begin{aligned} \langle(\partial_\theta f(\theta))^2\rangle &= \langle[4(\theta^3 + 3\theta)^2]\rangle && (7) \\ &= 16\langle\theta^6 + 6\theta^4 + 9\theta^2\rangle && (8) \end{aligned}$$