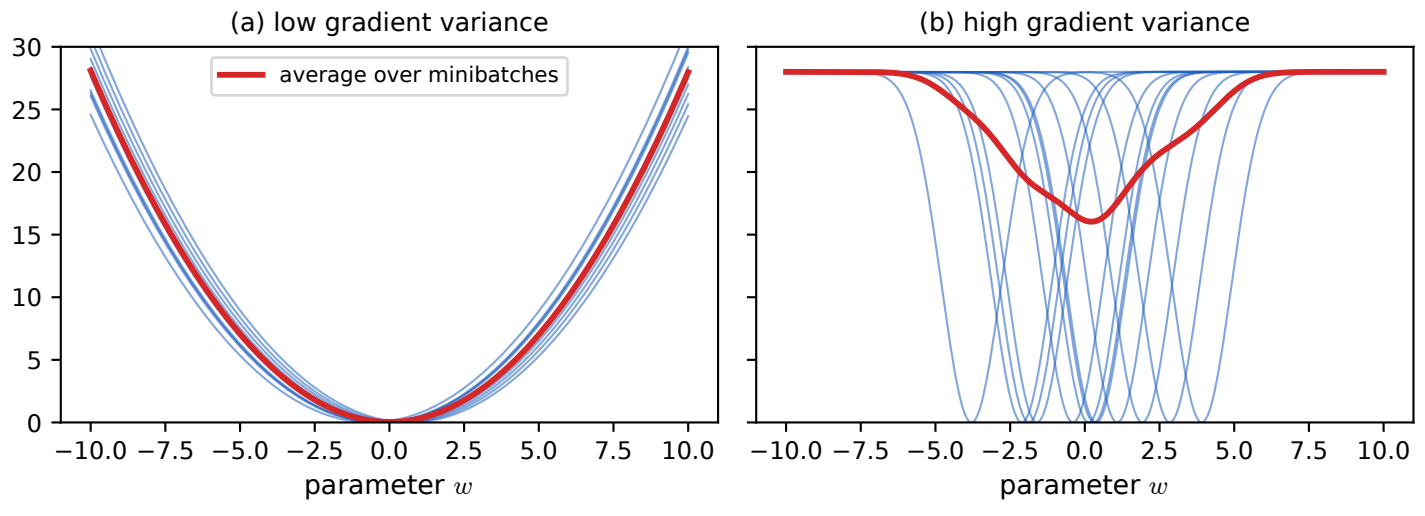

(a) low gradient variance
average over minibatches
30
25
20
15
10
5
0
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
parameter w
(b) high gradient variance
−10.0 −7.5 −5.0 −2.5 0.0 2.5 5.0 7.5 10.0
parameter w