(a) loss vs iter

(b) acc vs iter

Figure 1. The convergence performance with different penalty parameters ρ.
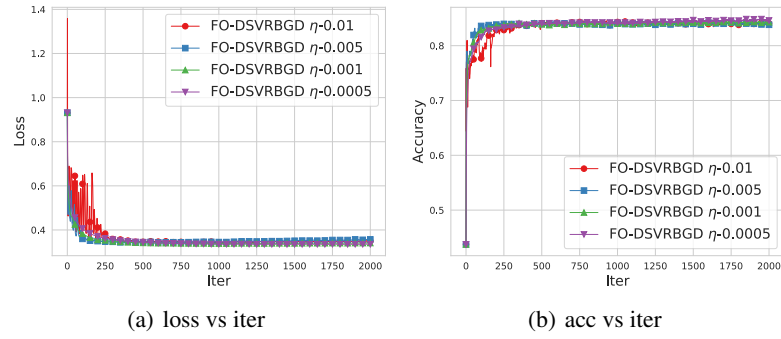

(a) loss vs iter

(b) acc vs iter

Figure 2. The convergence performance with different learning rates η.
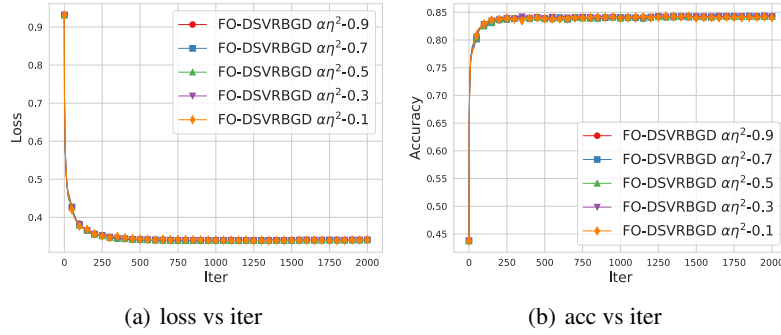

(a) loss vs iter

(b) acc vs iter

Figure 3. The convergence performance with different momentum coefficients $\alpha\eta^2$.