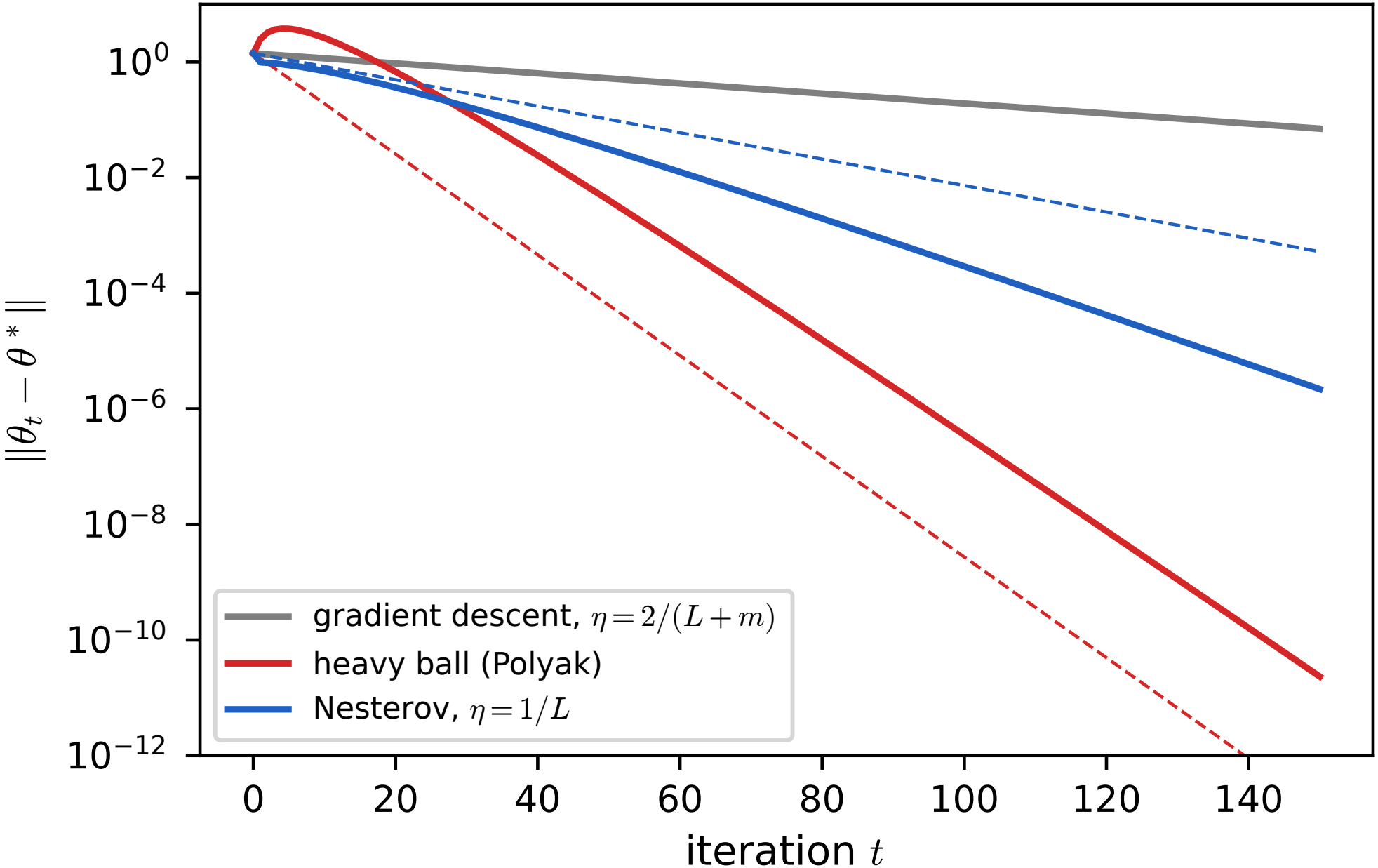

10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
$\|\theta_t - \theta^*\|$
0
20
40
60
80
100
120
140
iteration t
gradient descent, $\eta = 2/(L+m)$
heavy ball (Polyak)
Nesterov, $\eta = 1/L$