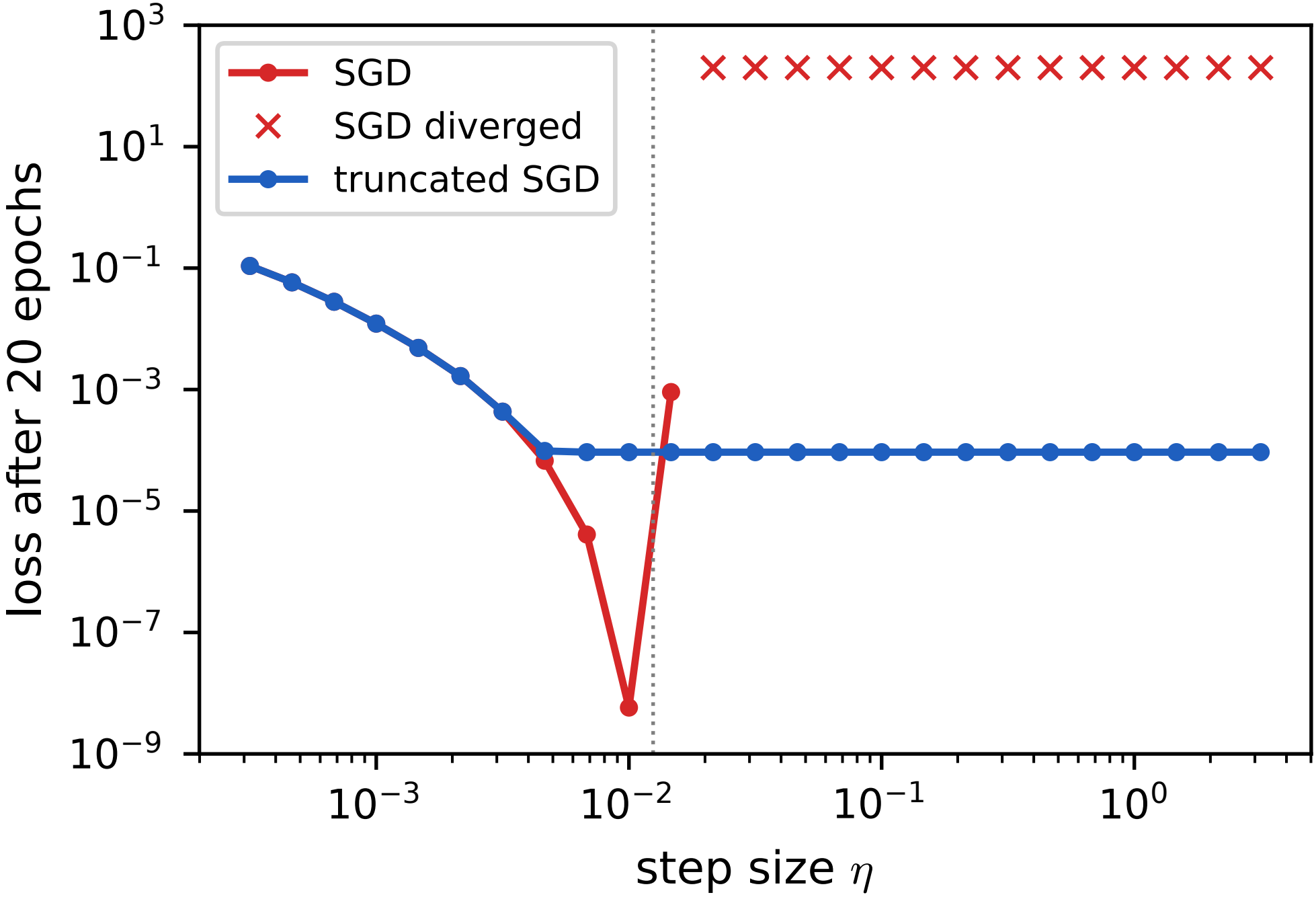

SGD
SGD diverged
truncated SGD
loss after 20 epochs
step size η
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
10^{-3}
10^{-2}
10^{-1}
10^{0}