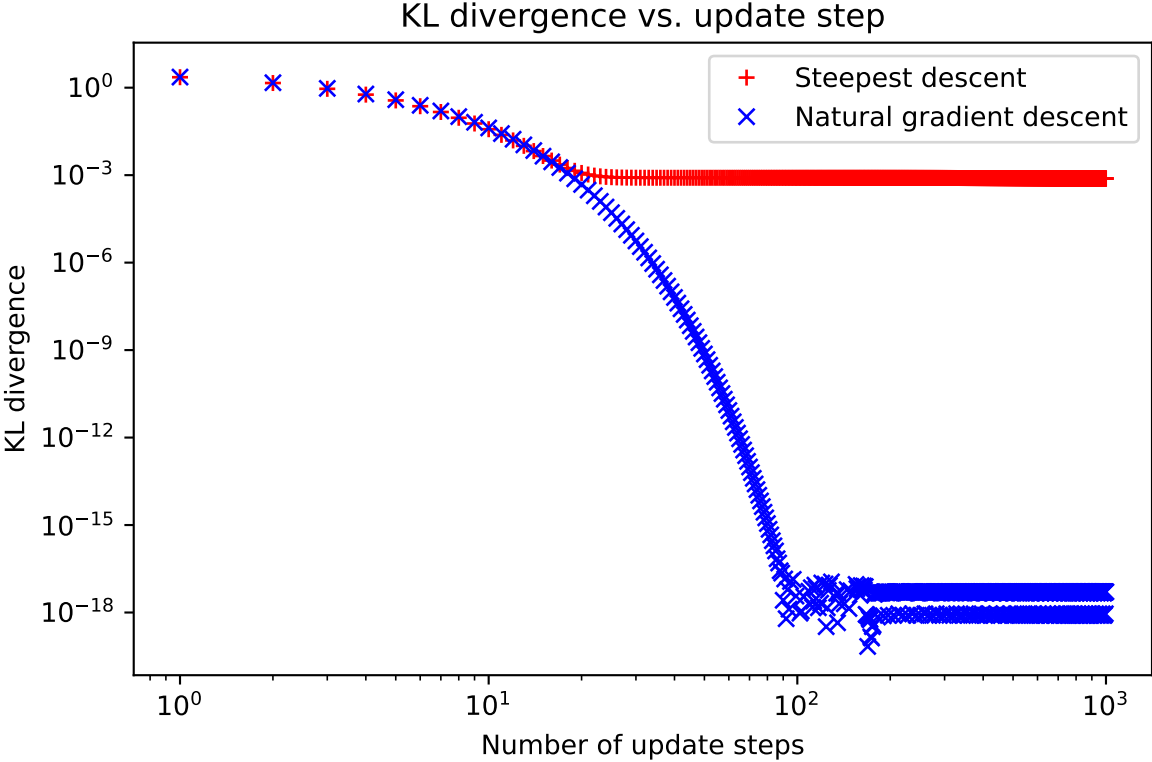
KL divergence vs. update step
Steepest descent
Natural gradient descent
KL divergence
Number of update steps
10^0
10^{-3}
10^{-6}
10^{-9}
10^{-12}
10^{-15}
10^{-18}
10^0
10^1
10^2
10^3