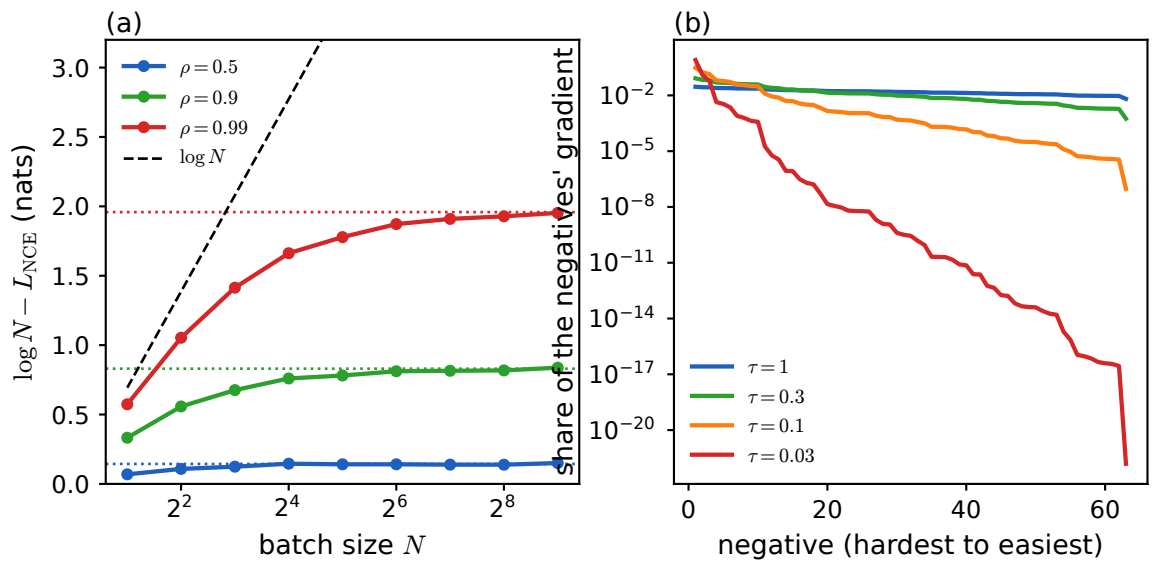

(a)
$\rho = 0.5$
$\rho = 0.9$
$\rho = 0.99$
$\log N$
$\log N - L_{\mathrm{NCE}}$ (nats)
batch size N
(b)
share of the negatives' gradient
$\tau = 1$
$\tau = 0.3$
$\tau = 0.1$
$\tau = 0.03$
negative (hardest to easiest)