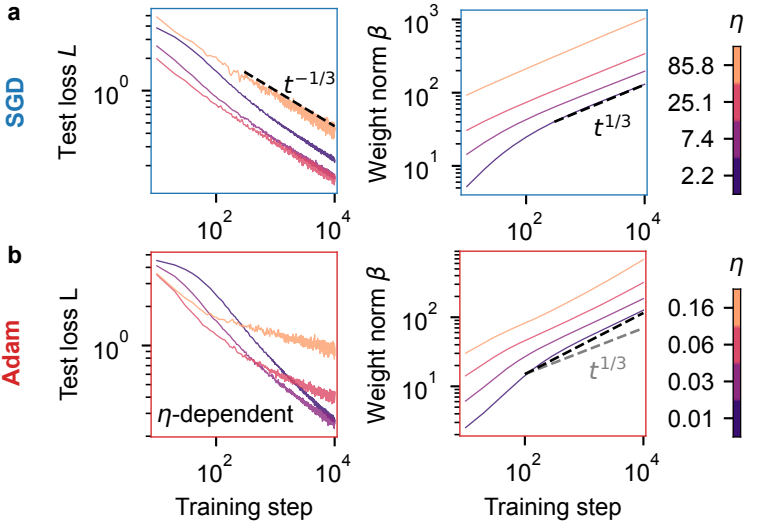

a
SGD
Test loss L
$10^0$
$t^{-1/3}$
$10^2$
$10^4$
Weight norm β
$10^3$
$10^2$
$10^1$
$t^{1/3}$
$10^2$
$10^4$
η
85.8
25.1
7.4
2.2
b
Adam
Test loss L
$10^0$
η-dependent
$10^2$
$10^4$
Training step
Weight norm β
$10^2$
$10^1$
$t^{1/3}$
$10^2$
$10^4$
Training step
η
0.16
0.06
0.03
0.01