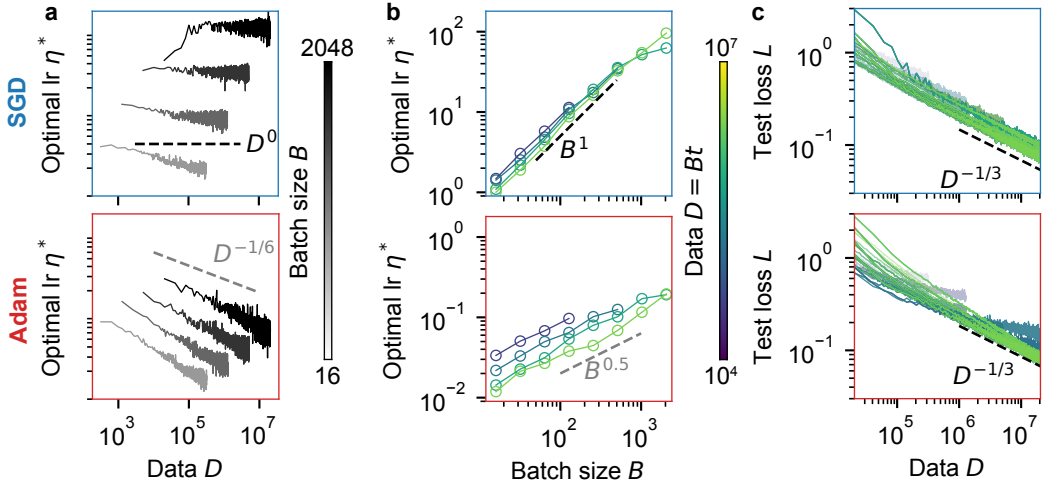

a
SGD
Adam
Optimal lr η^*
D^0
$D^{-1/6}$
2048
Batch size B
16
10^3
10^5
10^7
Data D
b
Optimal lr η^*
B^1
$B^{0.5}$
10^7
Data $D = Bt$
10^4
10^2
10^3
Batch size B
c
Test loss L
$D^{-1/3}$
10^5
10^6
10^7
Data D