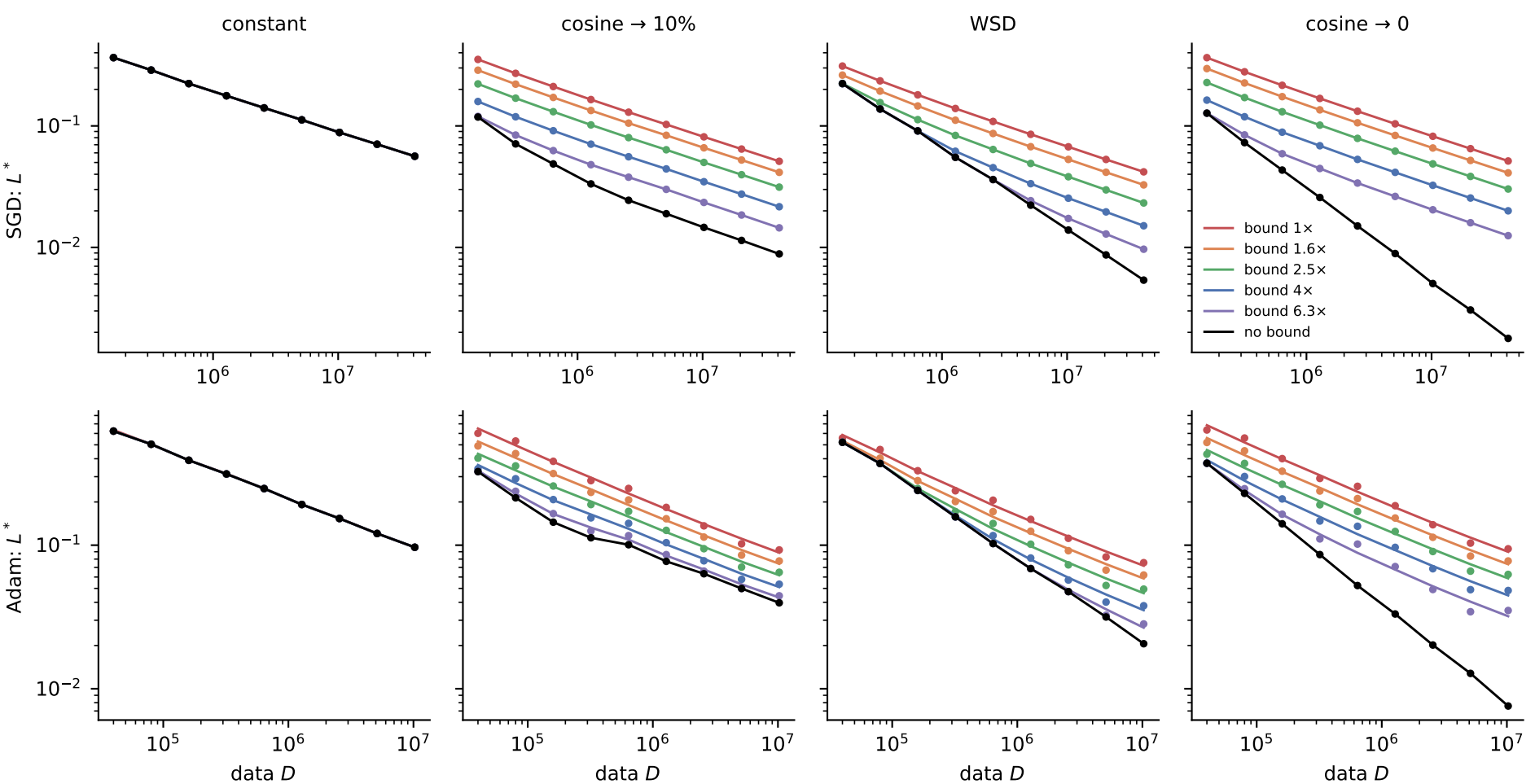

constant
cosine → 10%
WSD
cosine → 0
SGD: $L^*$
Adam: $L^*$
$10^{-1}$
$10^{-2}$
$10^{5}$
$10^{6}$
$10^{7}$
data $D$
bound 1×
bound 1.6×
bound 2.5×
bound 4×
bound 6.3×
no bound