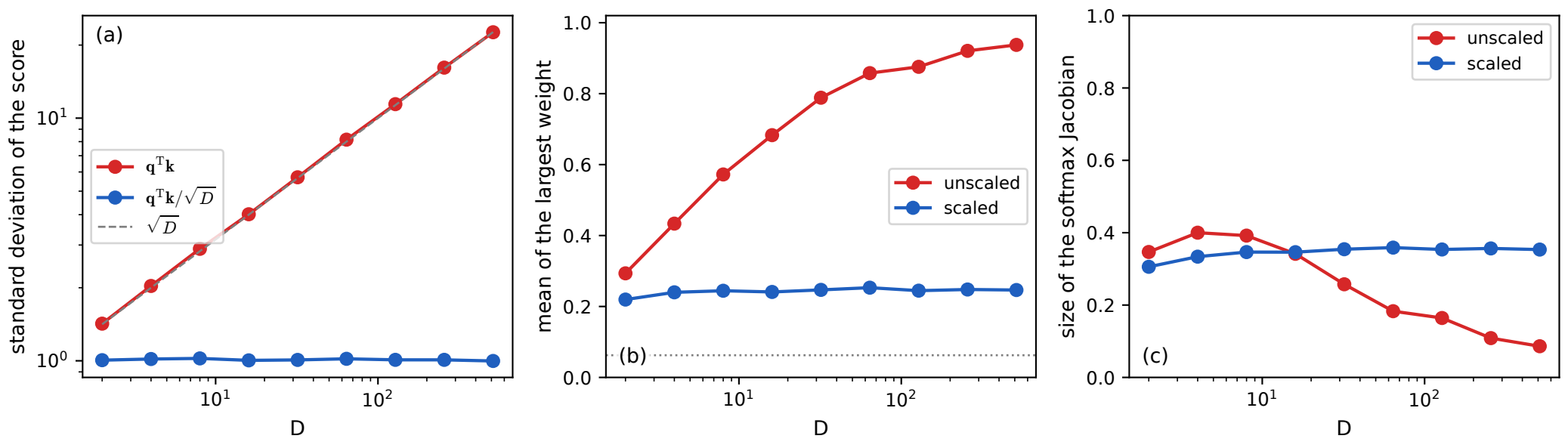

(a)
standard deviation of the score
10^1
10^0
$\mathbf{q}^{\mathrm{T}}\mathbf{k}$
$\mathbf{q}^{\mathrm{T}}\mathbf{k}/\sqrt{D}$
$\sqrt{D}$
10^1
10^2
D
(b)
mean of the largest weight
1.0
0.8
0.6
0.4
0.2
0.0
unscaled
scaled
10^1
10^2
D
(c)
size of the softmax Jacobian
1.0
0.8
0.6
0.4
0.2
0.0
unscaled
scaled
10^1
10^2
D