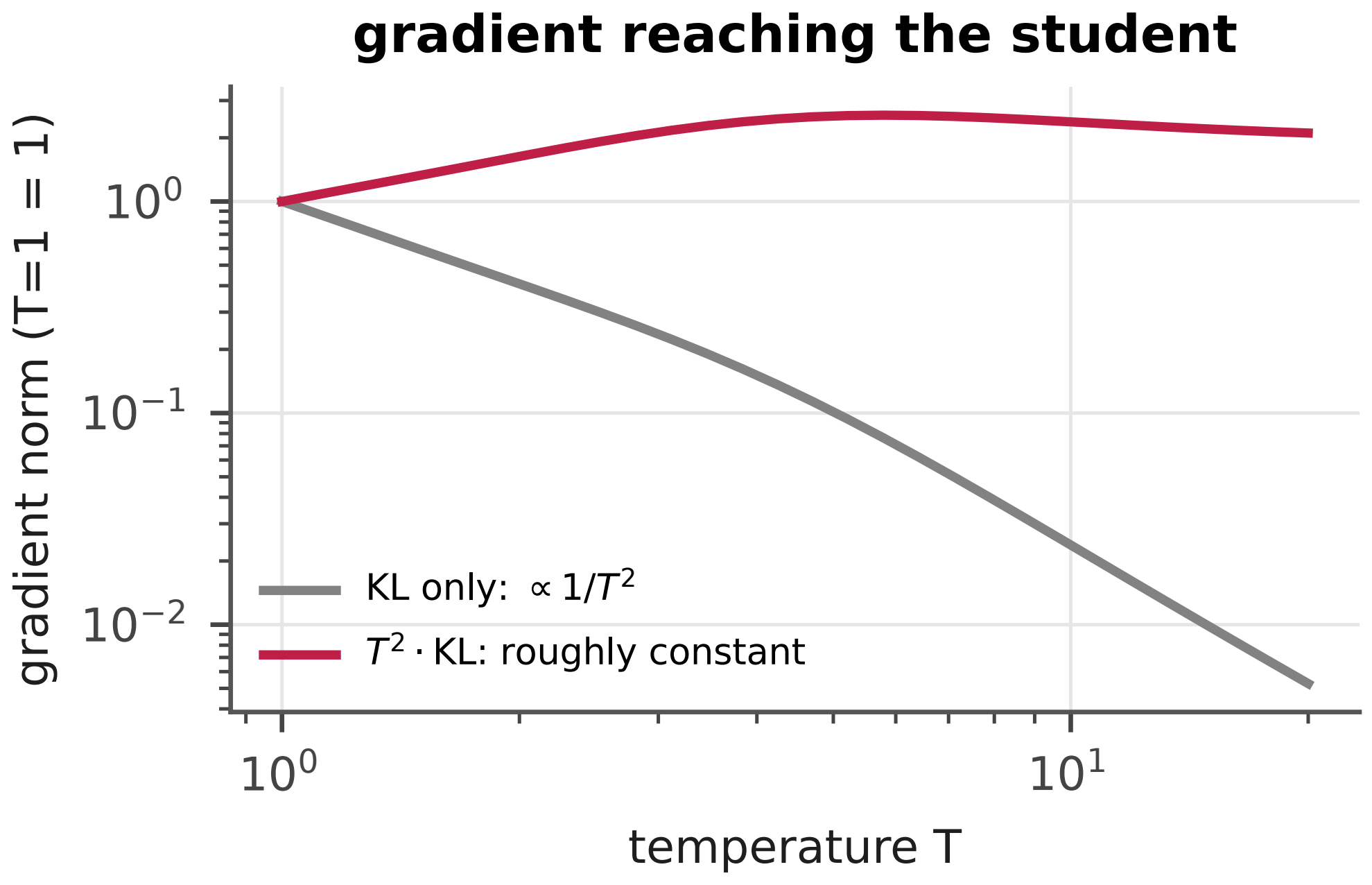

gradient reaching the student
gradient norm (T=1 = 1)
10^0
10^{-1}
10^{-2}
KL only: $\propto 1/T^2$
$T^2 \cdot$ KL: roughly constant
10^0
10^1
temperature T