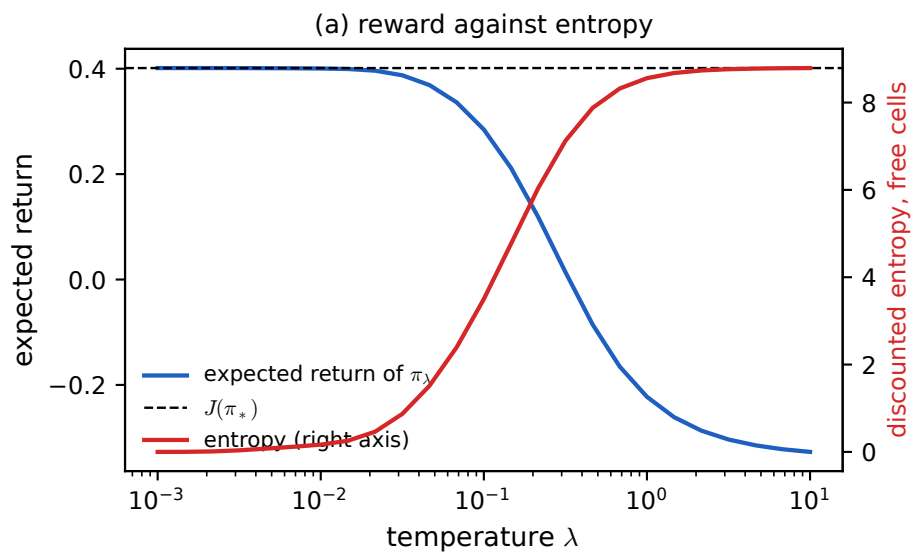
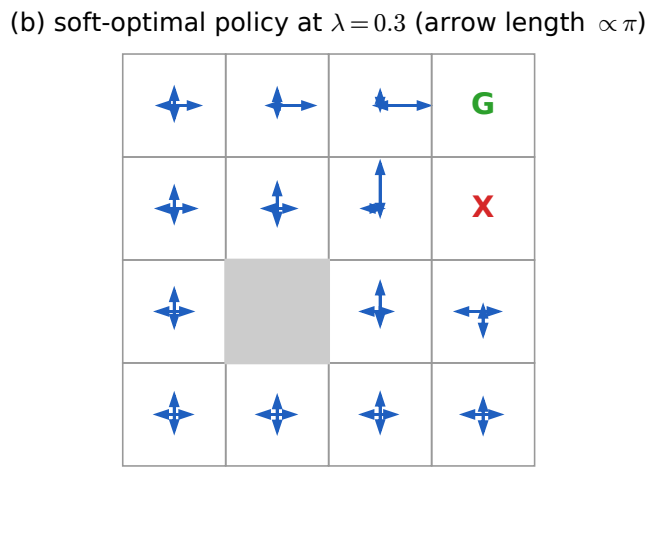

(a) reward against entropy
expected return
discounted entropy, free cells
temperature λ
expected return of π_λ
$J(\pi_*)$
entropy (right axis)
0.4
0.2
0.0
−0.2
8
6
4
2
0
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}


(b) soft-optimal policy at $\lambda = 0.3$ (arrow length $\propto \pi$)
G
X