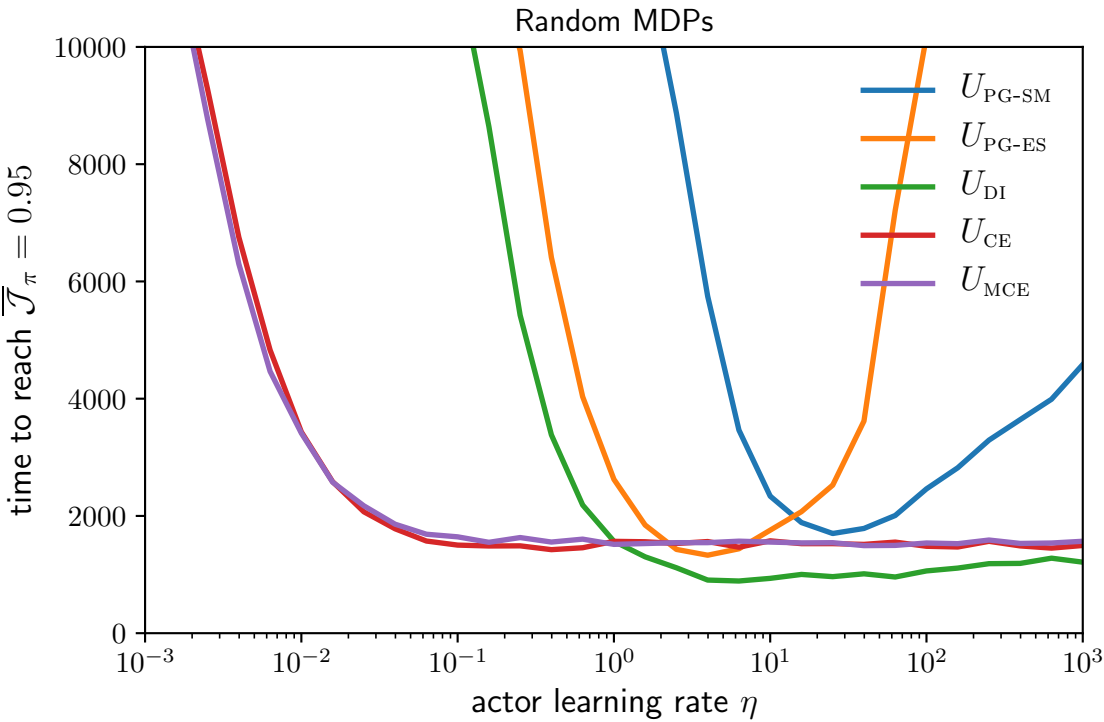
Random MDPs
time to reach $\overline{\mathcal{J}}_\pi = 0.95$
actor learning rate η
$U_{\text{PG-SM}}$
$U_{\text{PG-ES}}$
U_{DI}
U_{CE}
U_{MCE}
0
2000
4000
6000
8000
10000
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}