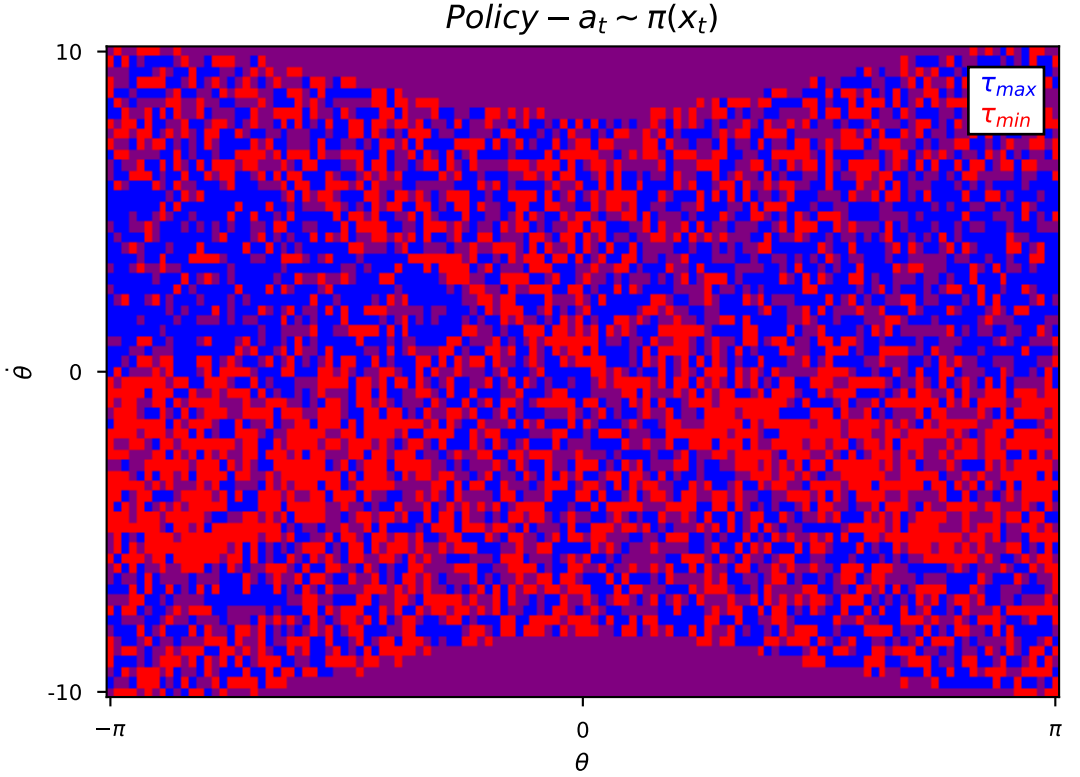
$Policy - a_t \sim \pi(x_t)$
τ_{max}
τ_{min}
10
0
-10
$\dot{\theta}$
$-\pi$
0
π
θ