Training set trajectory sc $f_{\phi}$  $f_{\phi}$  $f_{\phi}$ loss loss loss  $\tilde{c}$  $\pi_{\theta}$  $\pi_{\theta}$ 

Stochastic policy network (optimized)

loss

$$= \|s - \tilde{s}\|_2 + \lambda \tilde{c}$$