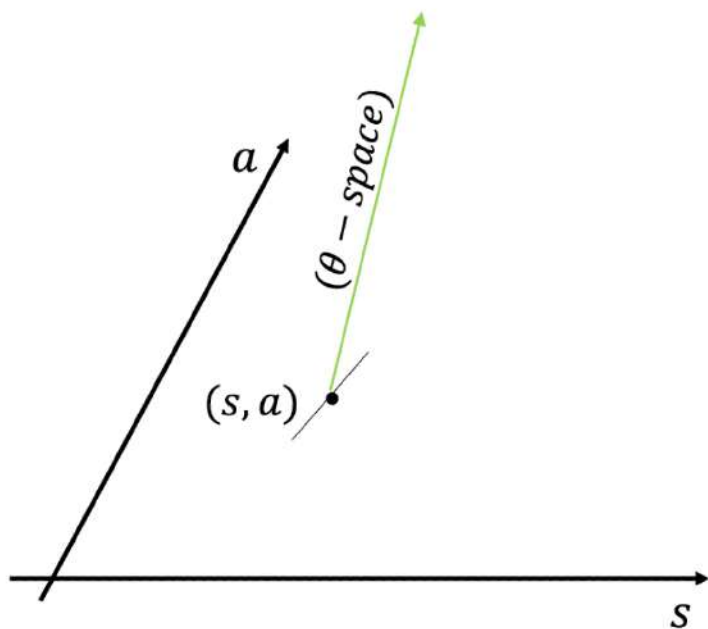


TD-learning

$$\Delta\theta \propto \delta_\theta(s, a) \nabla_\theta Q_\theta(s, a)$$



Taylor TD-learning

$$E_{\xi_a, \xi_s}[\Delta\theta] \approx E_{\xi_a, \xi_s}[\delta_\theta(s + \xi_s, a + \xi_a) \nabla_\theta Q_\theta(s + \xi_s, a + \xi_a)]$$

where, $\xi_a \sim \mathcal{N}(0, \lambda_a)$

$\xi_s \sim \mathcal{N}(0, \lambda_s)$

