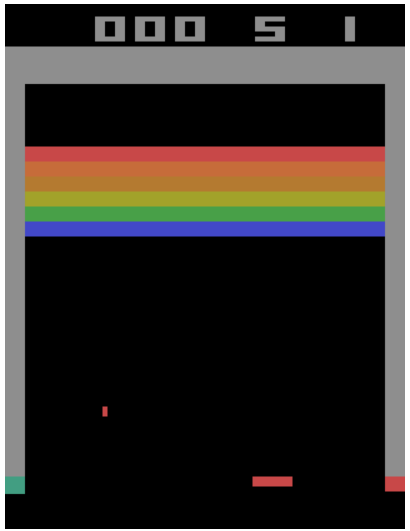
Adversarial Imitation Learning via Random Search

Authors: MyungJae Shin (Presenter) and Joongheon Kim
School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea
Emails: mjshin.cau@gmail.com , joongheon@gmail.com
Sites: github.com/170928 , https://sites.google.com/site/joongheonkim/

Reinforcement Learning

Goal : Learn policies
High-dimensional & raw observations

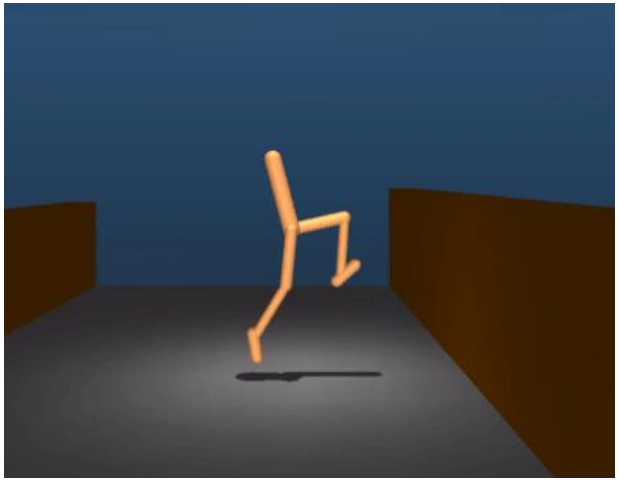


Imitation Learning

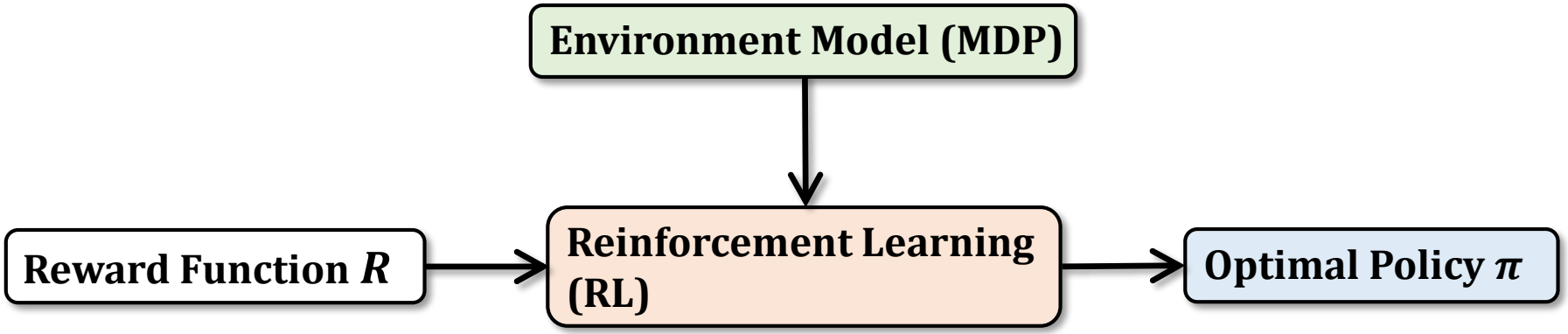
Input : expert behavior generated by expert π_E

$$\left\{\left(s_0^i, a_0^i, s_1^i, a_1^i, \ldots\right)\right\}_{i=1}^{N} \sim \pi_E$$

Goal : learn **cost function** or **policy**

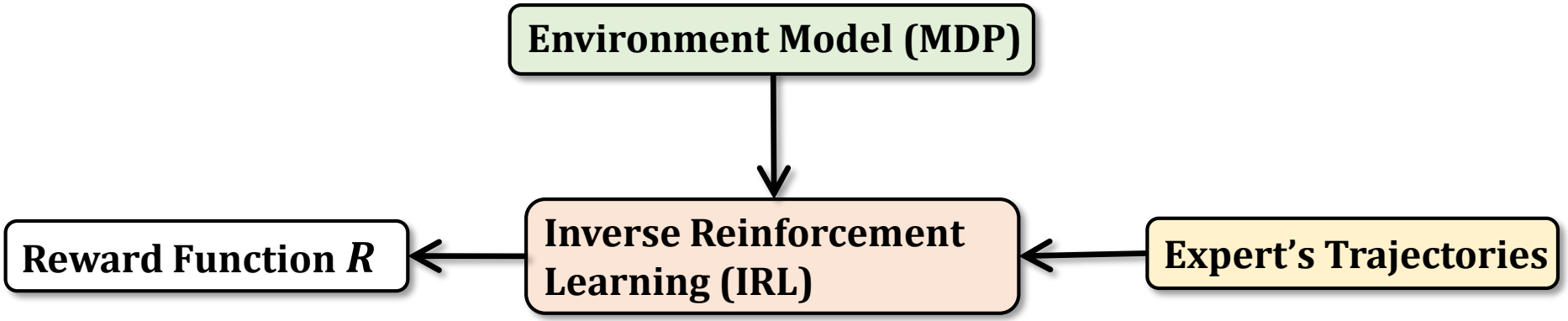


Imitation Learning



$$RL(R) = arg\min_{\pi} \mathbb{E}_{\pi}\left[R(s,a)\right] - H(\pi)$$

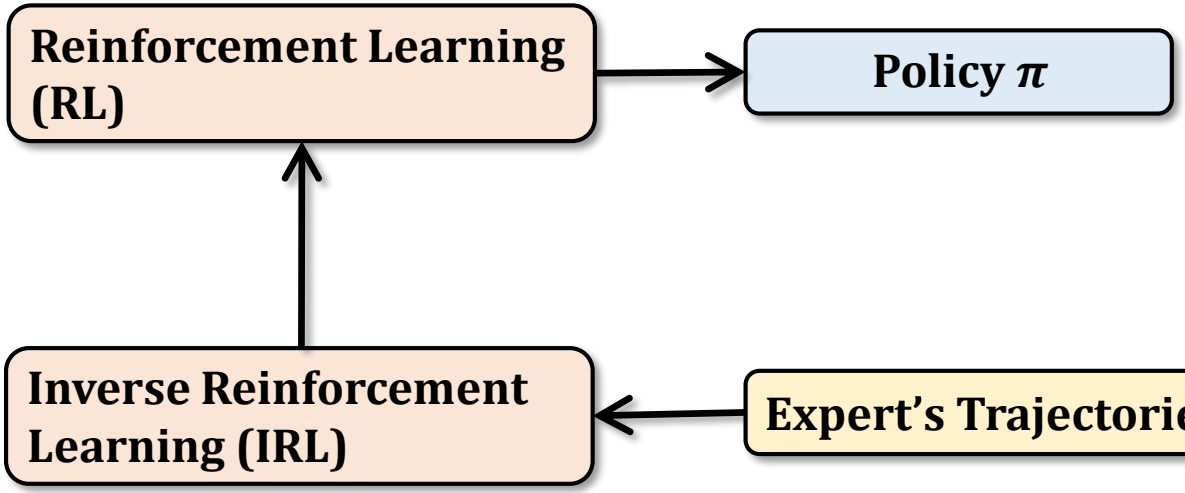
Imitation Learning



$$\max_{R}\left(\min_{\pi}\mathbb{E}_{\pi}\left[R(s,a)\right]-H(\pi)\right)-\mathbb{E}_{\pi_E}\left[R(s,a)\right]$$

Imitation Learning

$$\max_{R} -\psi(R) + \left(\min_{\pi} \mathbb{E}_{\pi}\left[R(s,a)\right] - H(\pi)\right) - \mathbb{E}_{\pi_E}\left[R(s,a)\right]$$



ρ_π = occupancy measure

➔ Distribution of state-action pairs encountered when navigating the environment with the policy

ρ_{E_π} = expert's occupancy measure

$$\min_{\pi} \psi^*(\boldsymbol{\rho_\pi} - \boldsymbol{\rho_{E_\pi}}) - H(\pi)$$

Imitation Learning

[Theorem]
ψ regularized inverse reinforcement learning implicitly, seeks a policy whose occupancy measure is close to the expert's, as measured by ψ^*

- Typical IRL finds a cost function such that the expert policy is uniquely optimal
- IRL as a procedure that tries to induce a policy that matches the expert's occupancy measure (generative model)

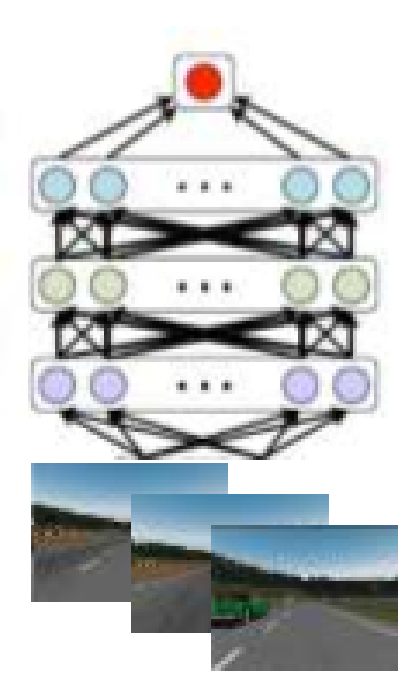
Generative Adversarial Imitation Learning (GAIL), *NIPS 2016*

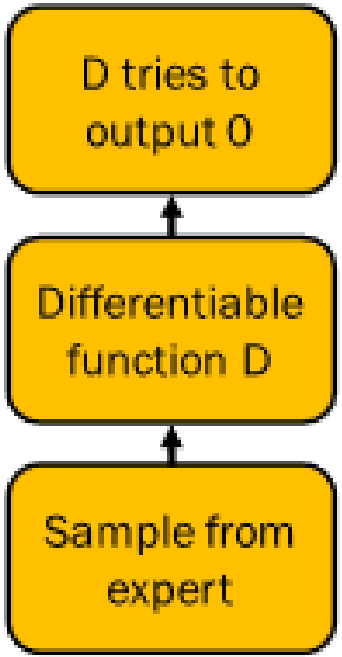
Use this regularizer $$\psi_{GA}(R) = \begin{cases} \mathbb{E}_{\pi_E}\left[g\left(R(s,a)\right)\right] & if\ R < 0 \\ +\infty & otherwise \end{cases}$$

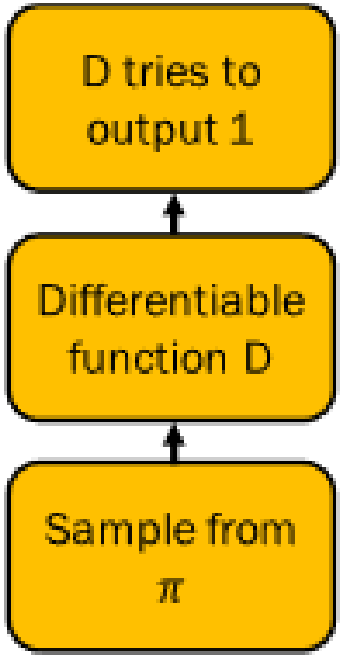
Generative Adversarial Networks, Ian J. Goodfellow, *NIPS 2014*

Adversarial Imitation Learning via Random Search

Generative Adversarial Imitation Learning (GAIL), *NIPS 2016*







Adversarial Imitation Learning via Random Search

Generative Adversarial Imitation Learning (GAIL), *NIPS 2016*

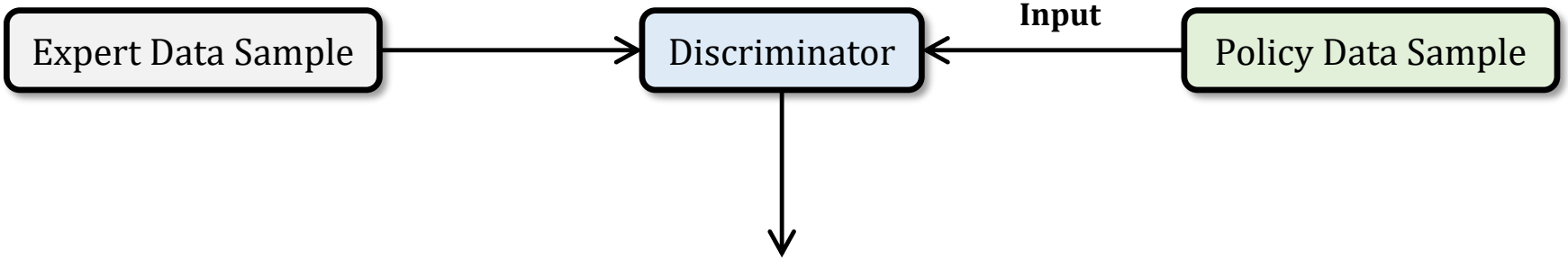
➔ State-Action Distribution of Training Agent

➔ State-Action Distribution of Expert

Based on the output of the discriminator (Generative Adversarial Networks, Ian J. Goodfellow, 2014), we could know the difference between the distribution of expert data and that of agent.

Adversarial Imitation Learning via Random Search

$$minimize \;\; \mathbb{E}_{\pi}[\log(D(s,a)] + \mathbb{E}_{\pi_E}[\log(1 - D(s,a)]$$



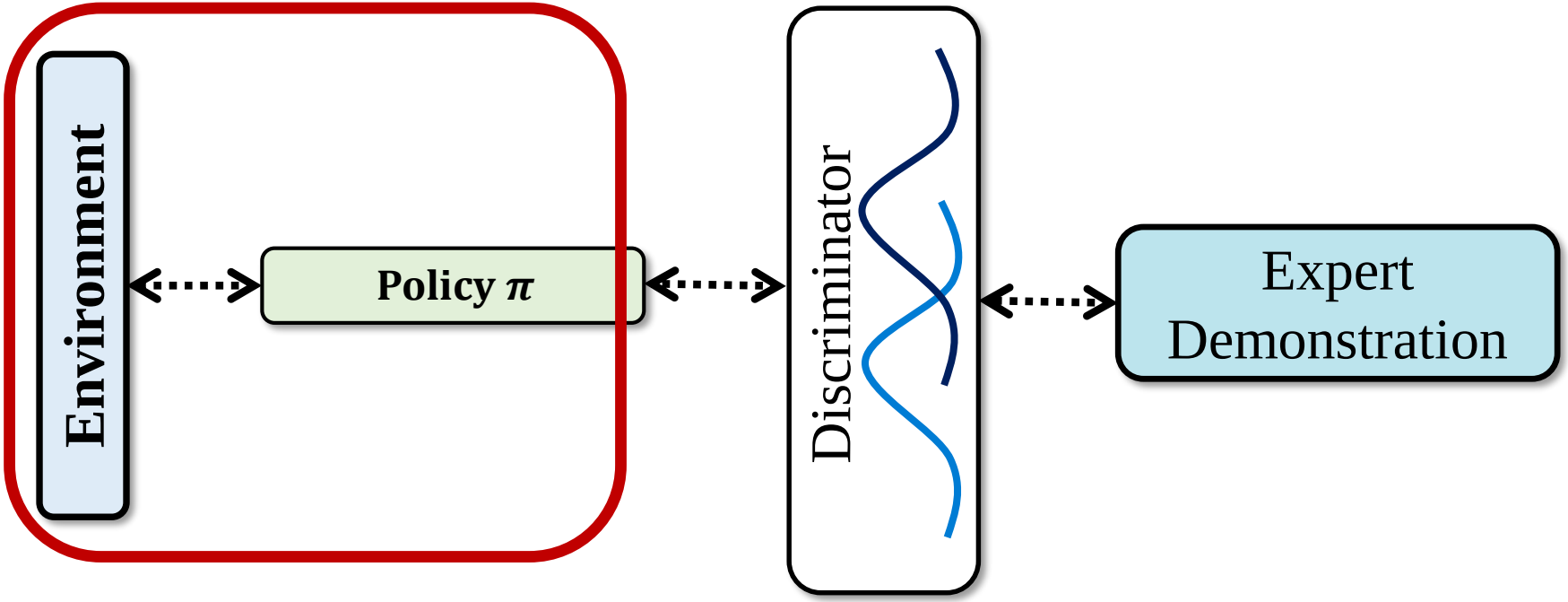
$\boldsymbol{D(s,a)}$ **:** Probability between 0 and 1

The probability that the input data sample is the expert data sample

Adversarial Imitation Learning via Random Search

Challenge

A lot of interaction with the environment is required to optimize the policy through GAIL framework



Adversarial Imitation Learning via Random Search

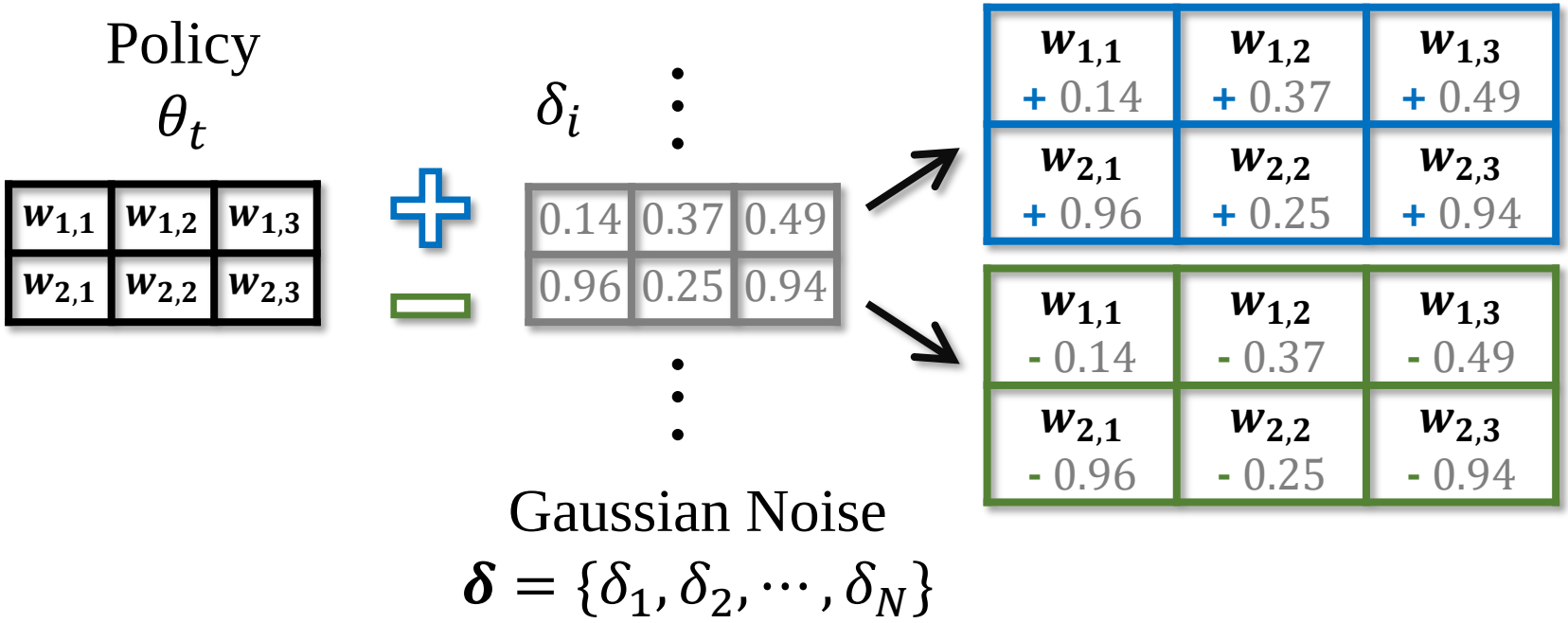
Usually in AI:

$$f'(x) = \frac{df}{dx}$$

Proposed method

$$f'(a) = \frac{f(a+h) - f(a)}{h}$$

Adversarial Imitation Learning via Random Search



Adversarial Imitation Learning via Random Search

R_{d-pos}

$w_{1,1} + d_{1,1}$	$w_{1,2} + d_{1,2}$	$w_{1,3} + d_{1,3}$
$w_{2,1} + d_{2,1}$	$w_{2,2} + d_{2,2}$	$w_{2,3} + d_{2,3}$

R_{e-pos}

$w_{1,1} + e_{1,1}$	$w_{1,2} + e_{1,2}$	$w_{1,3} + e_{1,3}$
$w_{2,1} + e_{2,1}$	$w_{2,2} + e_{2,2}$	$w_{2,3} + e_{2,3}$

R_{f-pos}

$w_{1,1} + f_{1,1}$	$w_{1,2} + f_{1,2}$	$w_{1,3} + f_{1,3}$
$w_{2,1} + f_{2,1}$	$w_{2,2} + f_{2,2}$	$w_{2,3} + f_{2,3}$

R_{g-pos}

$w_{1,1} + g_{1,1}$	$w_{1,2} + g_{1,2}$	$w_{1,3} + g_{1,3}$
$w_{2,1} + g_{2,1}$	$w_{2,2} + g_{2,2}$	$w_{2,3} + g_{2,3}$

R_{d-neg}

$w_{1,1} - d_{1,1}$	$w_{1,2} - d_{1,2}$	$w_{1,3} - d_{1,3}$
$w_{2,1} - d_{2,1}$	$w_{2,2} - d_{2,2}$	$w_{2,3} - d_{2,3}$

R_{e-neg}

$w_{1,1} - e_{1,1}$	$w_{1,2} - e_{1,2}$	$w_{1,3} - e_{1,3}$
$w_{2,1} - e_{2,1}$	$w_{2,2} - e_{2,2}$	$w_{2,3} - e_{2,3}$

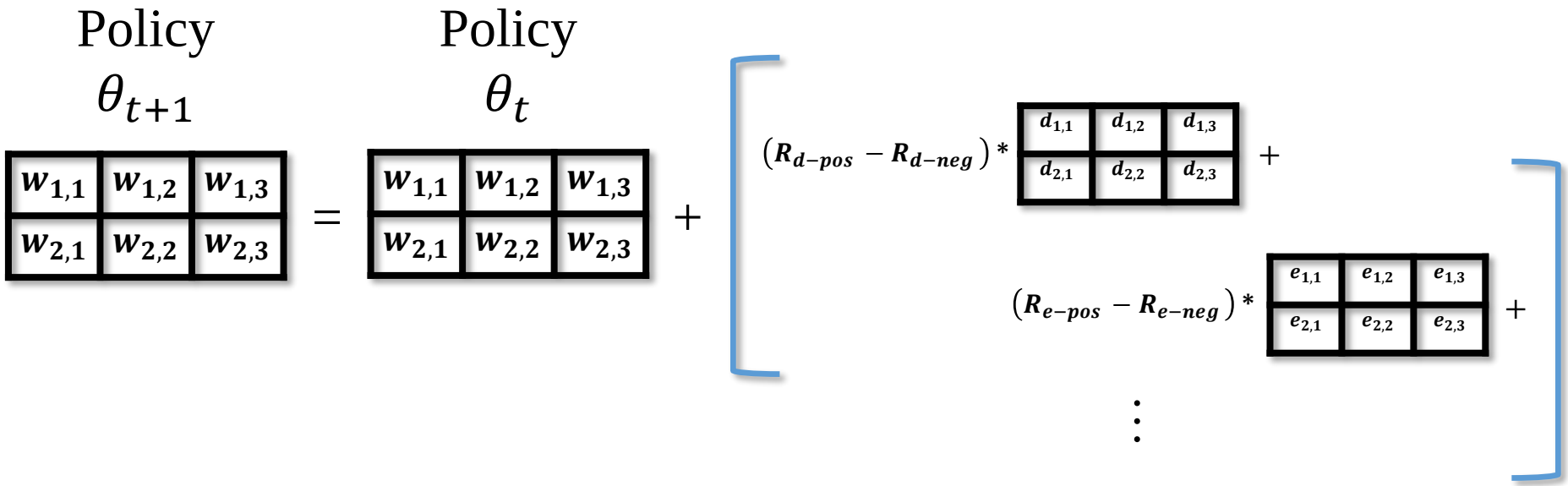
R_{f-neg}

$w_{1,1} - f_{1,1}$	$w_{1,2} - f_{1,2}$	$w_{1,3} - f_{1,3}$
$w_{2,1} - f_{2,1}$	$w_{2,2} - f_{2,2}$	$w_{2,3} - f_{2,3}$

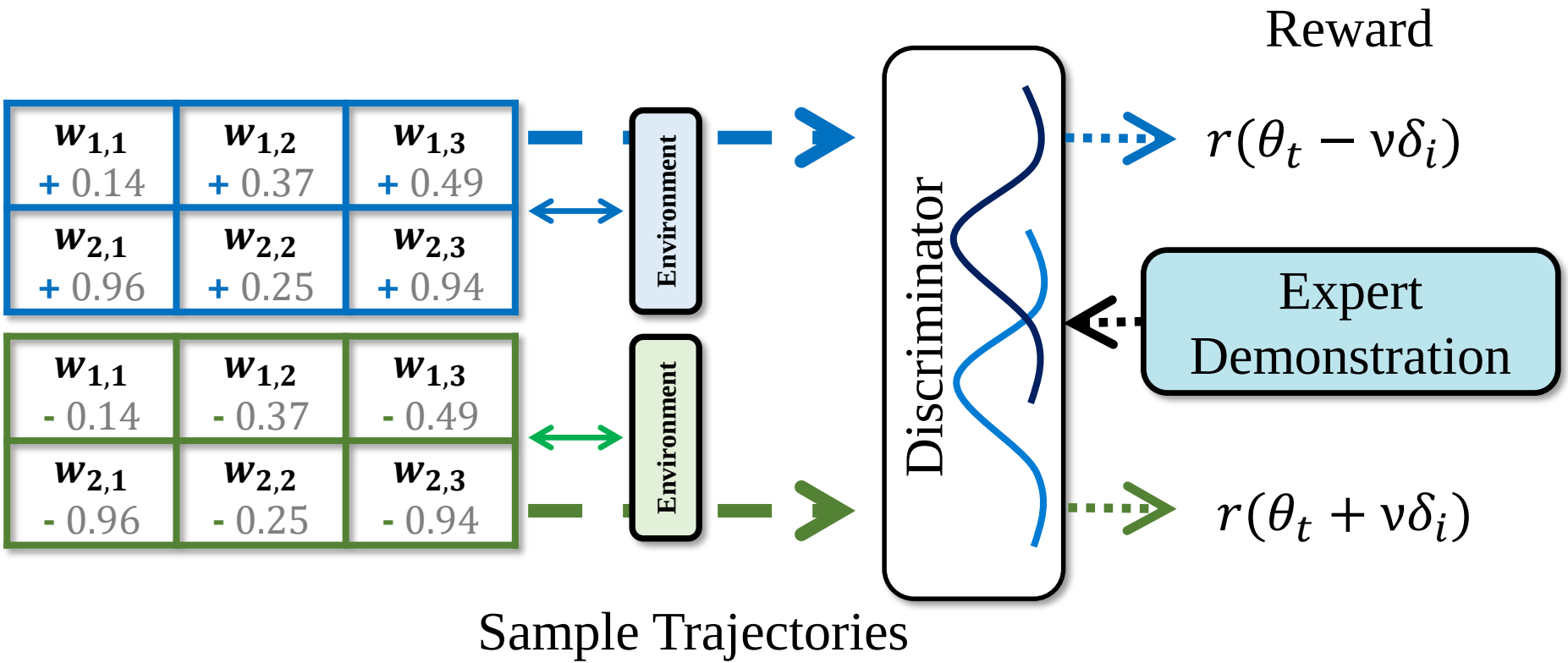
R_{g-neg}

$w_{1,1} - g_{1,1}$	$w_{1,2} - g_{1,2}$	$w_{1,3} - g_{1,3}$
$w_{2,1} - g_{2,1}$	$w_{2,2} - g_{2,2}$	$w_{2,3} - g_{2,3}$

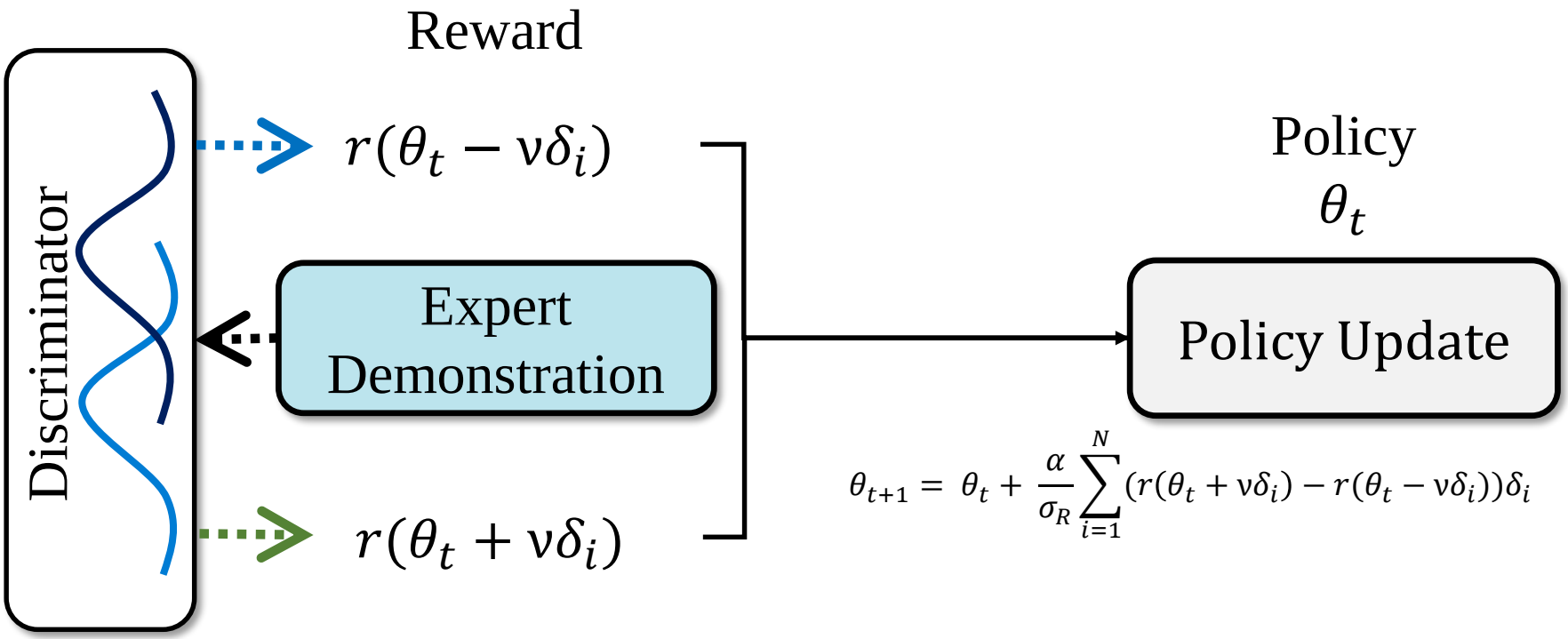
Adversarial Imitation Learning via Random Search



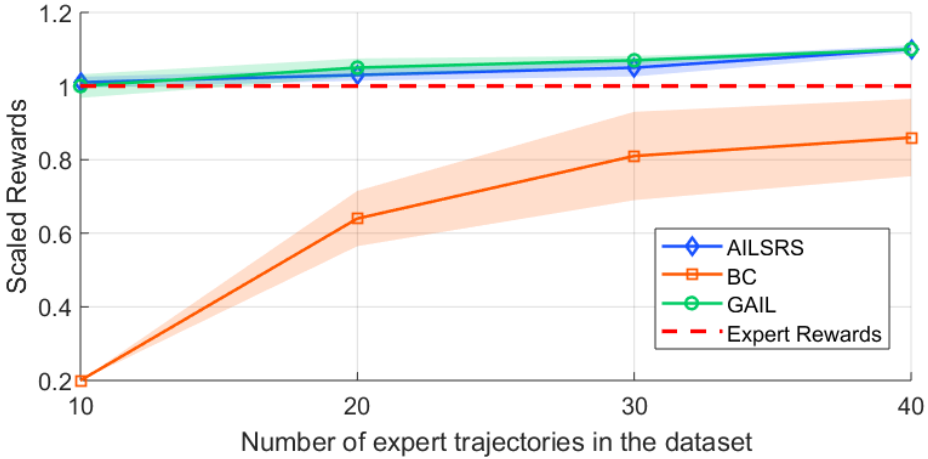
Adversarial Imitation Learning via Random Search

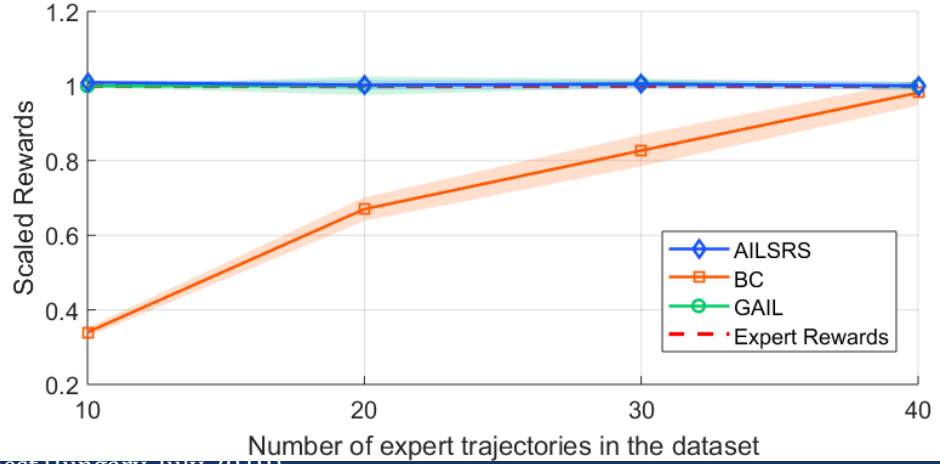


Adversarial Imitation Learning via Random Search

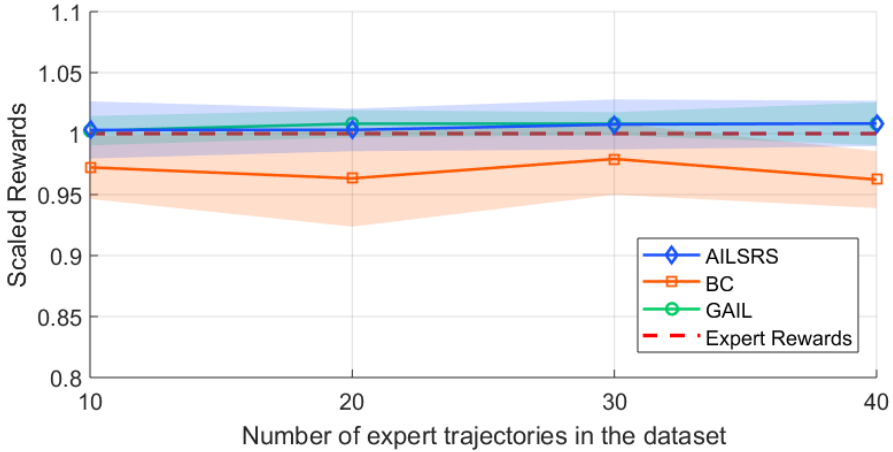


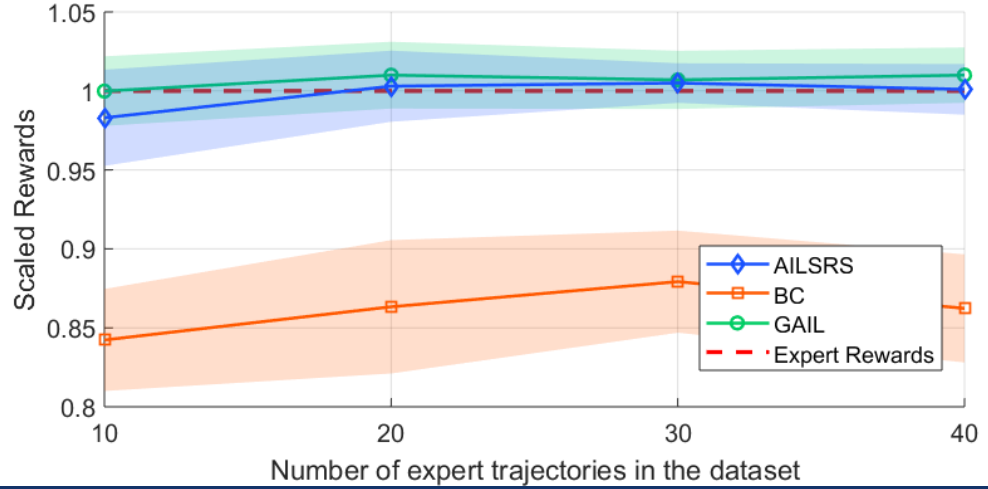
Adversarial Imitation Learning via Random Search





Adversarial Imitation Learning via Random Search





Thank You