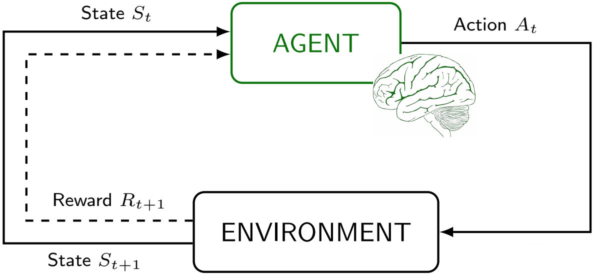
Policy Space Identification in Configurable Environments

Alberto Maria Metelli   Guglielmo Manneschi   Marcello Restelli

Politecnico di Milano
ECML PKDD 2021 - Journal Track

September 2021

Reinforcement Learning

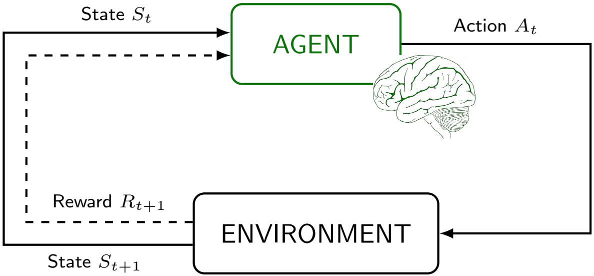


- Markov Decision Process (MDP, Puterman, 2014)
  1. Observe the state S_t
  2. Perform an action $A_t \sim \pi^{\text{Ag}}(\cdot|S_t)$
  3. Transition to the next state $S_{t+1} \sim P(\cdot|S_t, A_t)$
  4. Obtain reward $R_{t+1} = r(S_t, A_t, S_{t+1})$

- **Goal**: maximize the expected cumulative discounted reward (Sutton and Barto, 2018):

$$\pi^{\text{Ag}} \in \arg\max_{\pi \in \Pi_\Theta} J^\pi = \mathbb{E}^\pi \left[\sum_{t \in \mathbb{N}} \gamma^t R_{t+1} \right] \qquad \Pi \text{ policy space}$$

Reinforcement Learning



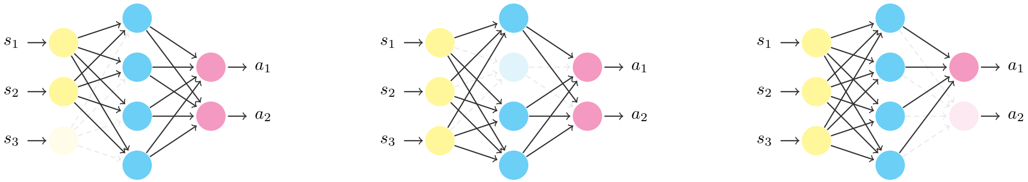
- Markov Decision Process (MDP, Puterman, 2014)
  1. Observe the state S_t
  2. Perform an action $A_t \sim \pi^{\text{Ag}}(\cdot|S_t)$
  3. Transition to the next state $S_{t+1} \sim P(\cdot|S_t, A_t)$
  4. Obtain reward $R_{t+1} = r(S_t, A_t, S_{t+1})$

- **Goal**: maximize the expected cumulative discounted reward (Sutton and Barto, 2018):

$$\pi^{\text{Ag}} \in \arg\max_{\pi \in \Pi_\Theta} J^\pi = \mathbb{E}^\pi \left[\sum_{t \in \mathbb{N}} \gamma^t R_{t+1} \right] \qquad \Pi \text{ policy space}$$

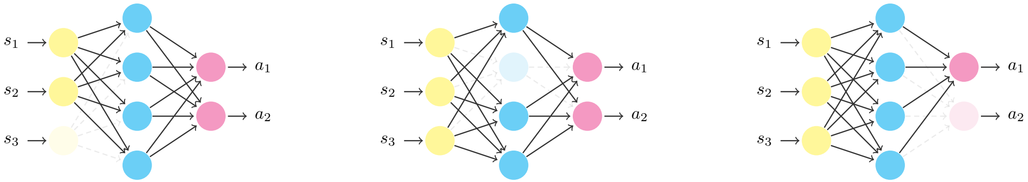
Motivations and Problem

- The **policy space** Π defines the **perception**, **actuation**, and **mapping** capabilities of an agent
- **Research Question**: How to identify the **policy space** of an agent by observing its behavior π^{Ag}? $\longrightarrow$ **Policy Space Identification (PSI)**
- Applications
  - Configurable MDPs (Metelli et al., 2018)
  - Imitation Learning (Osa et al., 2018)



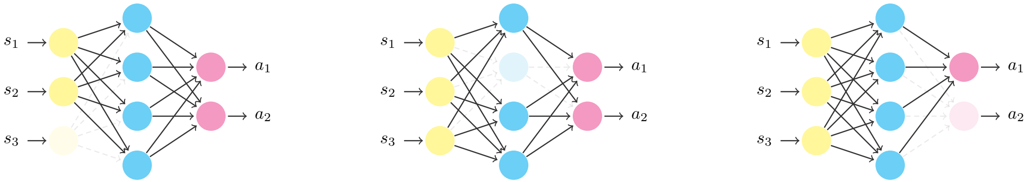
Motivations and Problem

- The **policy space** Π defines the **perception**, **actuation**, and **mapping** capabilities of an agent
- **Research Question**: How to identify the **policy space** of an agent by observing its behavior π^{Ag}? $\rightarrow$ **Policy Space Identification (PSI)**
- Applications
  - Configurable MDPs (Metelli et al., 2018)
  - Imitation Learning (Osa et al., 2018)



Motivations and Problem

- The **policy space** Π defines the **perception**, **actuation**, and **mapping** capabilities of an agent
- **Research Question**: How to identify the **policy space** of an agent by observing its behavior π^{Ag}? $\rightarrow$ **Policy Space Identification (PSI)**
- Applications
  - Configurable MDPs (Metelli et al., 2018)
  - Imitation Learning (Osa et al., 2018)



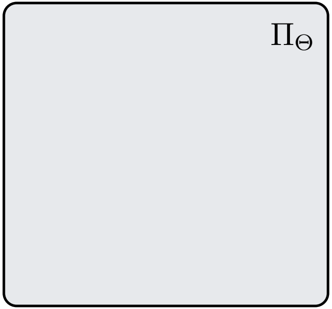
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\mathrm{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\mathrm{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \setminus I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\mathrm{Ag}} \in \Pi_{\Theta_{I^{\mathrm{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\mathrm{Ag}} : \pi^{\mathrm{Ag}} \notin \Pi_{\Theta_{I^{\mathrm{Ag}} \setminus \{i\}}}}_{\textbf{necessary}}$$



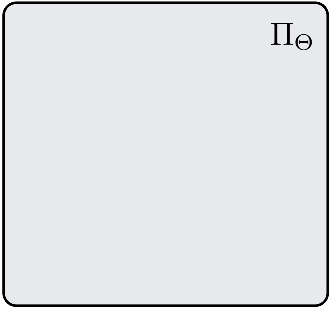
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\mathrm{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\mathrm{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \setminus I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\mathrm{Ag}} \in \Pi_{\Theta_{I^{\mathrm{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\mathrm{Ag}} : \pi^{\mathrm{Ag}} \notin \Pi_{\Theta_{I^{\mathrm{Ag}} \setminus \{i\}}}}_{\textbf{necessary}}$$



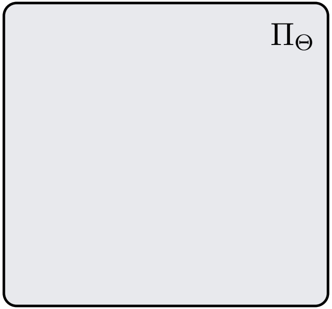
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\text{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\text{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \setminus I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\text{Ag}} \in \Pi_{\Theta_{I^{\text{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\text{Ag}} : \pi^{\text{Ag}} \notin \Pi_{\Theta_{I^{\text{Ag}} \setminus \{i\}}}}_{\textbf{necessary}}$$



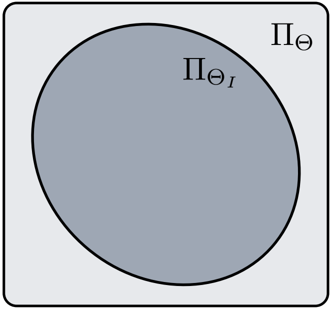
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\mathrm{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\mathrm{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \setminus I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\mathrm{Ag}} \in \Pi_{\Theta_{I^{\mathrm{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\mathrm{Ag}} : \pi^{\mathrm{Ag}} \notin \Pi_{\Theta_{I^{\mathrm{Ag}} \setminus \{i\}}}}_{\textbf{necessary}}$$



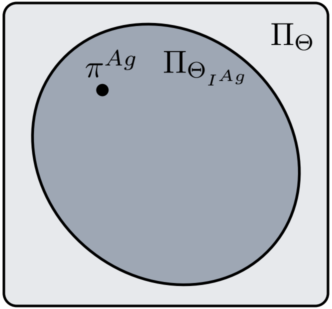
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\mathrm{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\mathrm{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \backslash I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\mathrm{Ag}} \in \Pi_{\Theta_{I^{\mathrm{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\mathrm{Ag}} : \pi^{\mathrm{Ag}} \notin \Pi_{\Theta_{I^{\mathrm{Ag}} \backslash \{i\}}}}_{\textbf{necessary}}$$



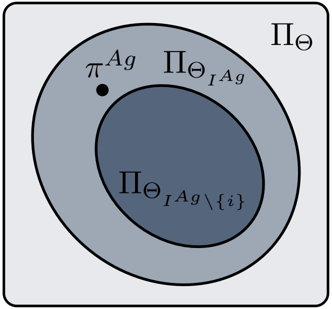
Policy Spaces and Correctness

- **Parametric** policy space Π_Θ, with $\Theta \subset \mathbb{R}^d$
- Agent's policy $\pi^{\text{Ag}} \in \Pi_\Theta$
- The agent **controls** (i.e., can change) $d^{\text{Ag}} < d$ parameters
- $I \subseteq \{1, \dots, d\}$ subset of indexes

$$\Theta_I = \{\boldsymbol{\theta} \in \Theta : \theta_i = 0, \forall i \in \{1, \dots, d\} \backslash I\}$$

- I^{Ag} is **correct** for the agent's policy π^{Ag} iff

$$\underbrace{\pi^{\text{Ag}} \in \Pi_{\Theta_{I^{\text{Ag}}}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in I^{\text{Ag}} : \pi^{\text{Ag}} \notin \Pi_{\Theta_{I^{\text{Ag}} \backslash \{i\}}}}_{\textbf{necessary}}$$



Hypothesis Tests

- **Idea**: perform **hypothesis test** for $I \subseteq \{1, ..., d\}$

$$\mathcal{H}_{0,I} : \pi^{\mathrm{Ag}} \in \Pi_{\Theta_I} \quad \text{vs} \quad \mathcal{H}_{1,I} : \pi^{\mathrm{Ag}} \in \Pi_{\Theta \setminus \Theta_I}$$

- Dataset of samples $\{(S_i, A_i)\}_{i=1}^{n}$ collected with the agent's policy π^{Ag}
- Likelihood of a parameter $\theta \in \Theta$

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{\boldsymbol{\theta}}(A_i|S_i) \qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\widehat{\mathcal{L}}(\boldsymbol{\theta})]$$

- Generalized **likelihood ratio** statistic (Casella and Berger, 2002)

$$\Lambda_I = \frac{\sup_{\boldsymbol{\theta} \in \Theta_I} \widehat{\mathcal{L}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \widehat{\mathcal{L}}(\boldsymbol{\theta})}$$

$\Lambda_I \simeq 0 \rightarrow$ reject $\mathcal{H}_{0,I}$

$\Lambda_I \simeq 1 \rightarrow$ do not reject $\mathcal{H}_{0,I}$

Hypothesis Tests

- **Idea**: perform **hypothesis test** for $I \subseteq \{1, ..., d\}$

$$\mathcal{H}_{0,I} : \pi^{\mathrm{Ag}} \in \Pi_{\Theta_I} \quad \text{vs} \quad \mathcal{H}_{1,I} : \pi^{\mathrm{Ag}} \in \Pi_{\Theta \setminus \Theta_I}$$

- Dataset of samples $\{(S_i, A_i)\}_{i=1}^{n}$ collected with the agent's policy π^{Ag}
- Likelihood of a parameter $\theta \in \Theta$

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{\boldsymbol{\theta}}(A_i|S_i) \qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\widehat{\mathcal{L}}(\boldsymbol{\theta})]$$

- Generalized **likelihood ratio** statistic (Casella and Berger, 2002)

$$\Lambda_I = \frac{\sup_{\boldsymbol{\theta} \in \Theta_I} \widehat{\mathcal{L}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \widehat{\mathcal{L}}(\boldsymbol{\theta})}$$

$\Lambda_I \simeq 0 \rightarrow$ reject $\mathcal{H}_{0,I}$

$\Lambda_I \simeq 1 \rightarrow$ do not reject $\mathcal{H}_{0,I}$

Hypothesis Tests

- **Idea**: perform **hypothesis test** for $I \subseteq \{1, ..., d\}$

$$\mathcal{H}_{0,I} \;:\; \pi^{\text{Ag}} \in \Pi_{\Theta_I} \quad \text{vs} \quad \mathcal{H}_{1,I} \;:\; \pi^{\text{Ag}} \in \Pi_{\Theta \setminus \Theta_I}$$

- Dataset of samples $\{(S_i, A_i)\}_{i=1}^{n}$ collected with the agent's policy π^{Ag}
- Likelihood of a parameter $\boldsymbol{\theta} \in \Theta$

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{\boldsymbol{\theta}}(A_i | S_i) \qquad\qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\widehat{\mathcal{L}}(\boldsymbol{\theta})]$$

- Generalized **likelihood ratio** statistic (Casella and Berger, 2002)

$$\Lambda_I = \frac{\sup_{\boldsymbol{\theta} \in \Theta_I} \widehat{\mathcal{L}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \widehat{\mathcal{L}}(\boldsymbol{\theta})} \qquad \begin{array}{l} \Lambda_I \simeq 0 \;\rightarrow\; \text{reject } \mathcal{H}_{0,I} \\ \Lambda_I \simeq 1 \;\rightarrow\; \text{do not reject } \mathcal{H}_{0,I} \end{array}$$

Hypothesis Tests

- **Idea**: perform **hypothesis test** for $I \subseteq \{1, ..., d\}$

$$\mathcal{H}_{0,I} \, : \, \pi^{\mathrm{Ag}} \in \Pi_{\Theta_I} \quad \text{vs} \quad \mathcal{H}_{1,I} \, : \, \pi^{\mathrm{Ag}} \in \Pi_{\Theta \setminus \Theta_I}$$

- Dataset of samples $\{(S_i, A_i)\}_{i=1}^{n}$ collected with the agent's policy π^{Ag}
- Likelihood of a parameter $\boldsymbol{\theta} \in \Theta$

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{\boldsymbol{\theta}}(A_i | S_i) \qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\widehat{\mathcal{L}}(\boldsymbol{\theta})]$$

- Generalized **likelihood ratio** statistic (Casella and Berger, 2002)

$$\Lambda_I = \frac{\sup_{\boldsymbol{\theta} \in \Theta_I} \widehat{\mathcal{L}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \widehat{\mathcal{L}}(\boldsymbol{\theta})}$$

$\Lambda_I \simeq 0 \;\rightarrow\;$ reject $\mathcal{H}_{0,I}$
$\Lambda_I \simeq 1 \;\rightarrow\;$ do not reject $\mathcal{H}_{0,I}$

Hypothesis Tests

- **Idea**: perform **hypothesis test** for $I \subseteq \{1, \dots, d\}$

$$\mathcal{H}_{0,I} \; : \; \pi^{\mathrm{Ag}} \in \Pi_{\Theta_I} \quad \text{vs} \quad \mathcal{H}_{1,I} \; : \; \pi^{\mathrm{Ag}} \in \Pi_{\Theta \setminus \Theta_I}$$

- Dataset of samples $\{(S_i, A_i)\}_{i=1}^{n}$ collected with the agent's policy π^{Ag}
- Likelihood of a parameter $\boldsymbol{\theta} \in \Theta$

$$\widehat{\mathcal{L}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{\boldsymbol{\theta}}(A_i | S_i) \qquad\qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\widehat{\mathcal{L}}(\boldsymbol{\theta})]$$

- Generalized **likelihood ratio** statistic (Casella and Berger, 2002)

$$\Lambda_I = \frac{\sup_{\boldsymbol{\theta} \in \Theta_I} \widehat{\mathcal{L}}(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} \widehat{\mathcal{L}}(\boldsymbol{\theta})} \qquad\qquad \begin{array}{l} \Lambda_I \simeq 0 \;\rightarrow\; \text{reject } \mathcal{H}_{0,I} \\ \Lambda_I \simeq 1 \;\rightarrow\; \text{do not reject } \mathcal{H}_{0,I} \end{array}$$

Combinatorial Identification Rule

- **Combinatorial Identification Rule**: retain all the **approximately correct** $\widehat{I} \subseteq \{1, \dots, d\}$:

$$\underbrace{\text{do not reject } \mathcal{H}_{0,\widehat{I}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in \widehat{I} : \text{reject } \mathcal{H}_{0,\widehat{I}\setminus\{i\}}}_{\textbf{necessary}}$$

- Requires $O(2^d)$ tests! $\longrightarrow$ combinatorial
- Works under multiple representations of π^{Ag}

What if there exists a unique representation of π^{Ag}?

Combinatorial Identification Rule

- **Combinatorial Identification Rule**: retain all the **approximately correct** $\widehat{I} \subseteq \{1, \dots, d\}$:

$$\underbrace{\text{do not reject } \mathcal{H}_{0,\widehat{I}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in \widehat{I} : \text{reject } \mathcal{H}_{0,\widehat{I}\setminus\{i\}}}_{\textbf{necessary}}$$

- Requires $O(2^d)$ tests! $\longrightarrow$ combinatorial
- Works under multiple representations of π^{Ag}

What if there exists a unique representation of π^{Ag}?

Combinatorial Identification Rule

- **Combinatorial Identification Rule**: retain all the **approximately correct** $\widehat{I} \subseteq \{1, \dots, d\}$:

$$\underbrace{\text{do not reject } \mathcal{H}_{0,\widehat{I}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in \widehat{I} : \text{reject } \mathcal{H}_{0,\widehat{I}\setminus\{i\}}}_{\textbf{necessary}}$$

- Requires $O(2^d)$ tests! $\longrightarrow$ combinatorial
- Works under multiple representations of π^{Ag}

What if there exists a unique representation of π^{Ag}?

Combinatorial Identification Rule

- **Combinatorial Identification Rule**: retain all the **approximately correct** $\widehat{I} \subseteq \{1, \dots, d\}$:

$$\underbrace{\text{do not reject } \mathcal{H}_{0,\widehat{I}}}_{\textbf{sufficient}} \quad \wedge \quad \underbrace{\forall i \in \widehat{I} : \text{reject } \mathcal{H}_{0,\widehat{I}\setminus\{i\}}}_{\textbf{necessary}}$$

- Requires $O(2^d)$ tests! $\longrightarrow$ combinatorial
- Works under multiple representations of π^{Ag}

What if there exists a unique representation of π^{Ag}?

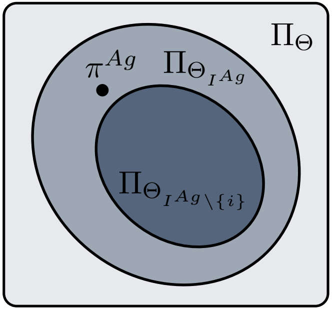
Identifiability and Correctness

- Π_Θ is **identifiable** if

$$\forall \boldsymbol{\theta}', \boldsymbol{\theta} \in \Theta \qquad \pi_{\boldsymbol{\theta}'}(\cdot|s) = \pi_{\boldsymbol{\theta}}(\cdot|s) \text{ a.s.} \implies \boldsymbol{\theta}' = \boldsymbol{\theta}$$

- We can reason on the **parameters** only!
- The **only** correct I^{Ag} for the agent's policy $\pi_{\boldsymbol{\theta}^{\mathrm{Ag}}}$ is

$$I^{\mathrm{Ag}} = \left\{ i \in \{1, \dots, d\} : \theta_i^{\mathrm{Ag}} \neq 0 \right\}$$



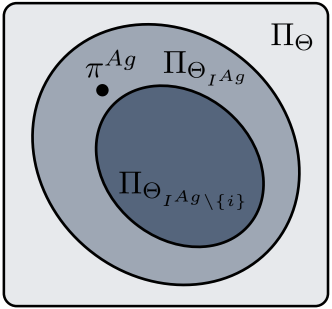
Identifiability and Correctness

- Π_Θ is **identifiable** if

$$\forall \boldsymbol{\theta}', \boldsymbol{\theta} \in \Theta \qquad \pi_{\boldsymbol{\theta}'}(\cdot|s) = \pi_{\boldsymbol{\theta}}(\cdot|s) \text{ a.s.} \implies \boldsymbol{\theta}' = \boldsymbol{\theta}$$

- We can reason on the **parameters** only!
- The **only** correct I^{Ag} for the agent's policy $\pi_{\boldsymbol{\theta}^{\text{Ag}}}$ is

$$I^{\text{Ag}} = \left\{ i \in \{1, \dots, d\} : \theta_i^{\text{Ag}} \neq 0 \right\}$$



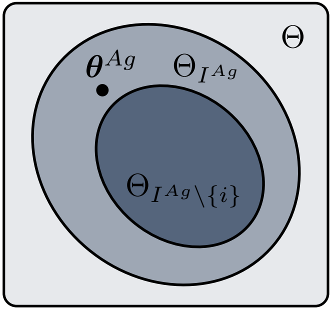
Identifiability and Correctness

- Π_Θ is **identifiable** if

$$\forall \boldsymbol{\theta}', \boldsymbol{\theta} \in \Theta \qquad \pi_{\boldsymbol{\theta}'}(\cdot|s) = \pi_{\boldsymbol{\theta}}(\cdot|s) \text{ a.s.} \implies \boldsymbol{\theta}' = \boldsymbol{\theta}$$

- We can reason on the **parameters** only!
- The **only** correct I^{Ag} for the agent's policy $\pi_{\boldsymbol{\theta}^{\text{Ag}}}$ is

$$I^{\text{Ag}} = \left\{ i \in \{1, \dots, d\} : \theta_i^{\text{Ag}} \neq 0 \right\}$$



Simplified Identification Rule

- **Idea**: perform **hypothesis test** for $i \in \{1, ..., d\}$

$$\mathcal{H}_{0,i} : \theta_i^{\text{Ag}} = 0 \quad \text{vs} \quad \mathcal{H}_{1,i} : \theta_i^{\text{Ag}} \neq 0$$

- **Simplified Identification Rule**: $\widehat{I}$ is the union of all $i \in \{1, ..., d\}$ such that:

$$\text{reject } \mathcal{H}_{0,i}$$

- Requires $O(d)$ tests! $\rightarrow$ linear
- Works under **identifiability** only!

Simplified Identification Rule

- **Idea**: perform **hypothesis test** for $i \in \{1, \dots, d\}$

$$\mathcal{H}_{0,i} : \theta_i^{\text{Ag}} = 0 \quad \text{vs} \quad \mathcal{H}_{1,i} : \theta_i^{\text{Ag}} \neq 0$$

- **Simplified Identification Rule**: $\widehat{I}$ is the union of all $i \in \{1, \dots, d\}$ such that:

$$\text{reject } \mathcal{H}_{0,i}$$

- Requires $O(d)$ tests! $\rightarrow$ linear
- Works under **identifiability** only!

Simplified Identification Rule

- **Idea**: perform **hypothesis test** for $i \in \{1, \ldots, d\}$

$$\mathcal{H}_{0,i} : \theta_i^{\mathrm{Ag}} = 0 \quad \text{vs} \quad \mathcal{H}_{1,i} : \theta_i^{\mathrm{Ag}} \neq 0$$

- **Simplified Identification Rule**: $\widehat{I}$ is the union of all $i \in \{1, \ldots, d\}$ such that:

$$\text{reject } \mathcal{H}_{0,i}$$

- Requires $O(d)$ tests! $\rightarrow$ linear
- Works under **identifiability** only!

Simplified Identification Rule

- **Idea**: perform **hypothesis test** for $i \in \{1, \ldots, d\}$

$$\mathcal{H}_{0,i} : \theta_i^{\mathrm{Ag}} = 0 \quad \text{vs} \quad \mathcal{H}_{1,i} : \theta_i^{\mathrm{Ag}} \neq 0$$

- **Simplified Identification Rule**: $\widehat{I}$ is the union of all $i \in \{1, \ldots, d\}$ such that:

$$\text{reject } \mathcal{H}_{0,i}$$

- Requires $O(d)$ tests! $\rightarrow$ linear
- Works under **identifiability** only!

Ambiguous Identification

What can we conclude when $\theta_i = 0$?

1. The agent **does not control** θ_i or...
2. ...the agent has **consciously** chosen to set $\theta_i = 0$

- **Problem**: How to distinguish between these two scenarios?
- **Idea**: change the environment to make the parameter "maximally important" $\rightarrow$ **Configurable Environment** (Metelli et al., 2018)

Ambiguous Identification

What can we conclude when $\theta_i = 0$?

1. The agent **does not control** θ_i or...
2. ...the agent has **consciously** chosen to set $\theta_i = 0$

- **Problem**: How to distinguish between these two scenarios?
- **Idea**: change the environment to make the parameter "maximally important" $\rightarrow$ **Configurable Environment** (Metelli et al., 2018)

Ambiguous Identification

What can we conclude when $\theta_i = 0$?

1. The agent **does not control** θ_i or...
2. ...the agent has **consciously** chosen to set $\theta_i = 0$

- **Problem**: How to distinguish between these two scenarios?
- **Idea**: change the environment to make the parameter "maximally important" $\longrightarrow$ **Configurable Environment** (Metelli et al., 2018)

Ambiguous Identification

What can we conclude when $\theta_i = 0$?

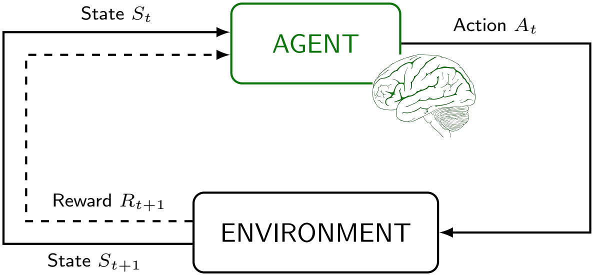
1. The agent **does not control** θ_i or...
2. ...the agent has **consciously** chosen to set $\theta_i = 0$

- **Problem**: How to distinguish between these two scenarios?
- **Idea**: change the environment to make the parameter "maximally important" $\rightarrow$ **Configurable Environment** (Metelli et al., 2018)

Non-Configurable Environments

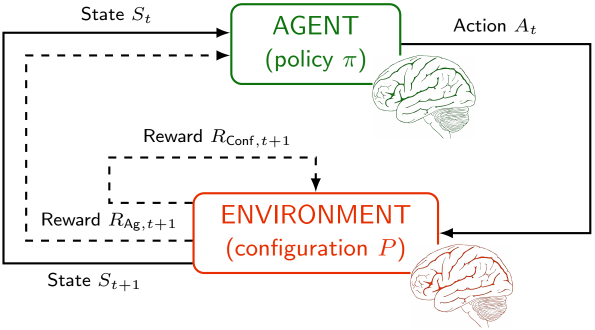
MDP (Puterman, 2014)

the environment is fixed and out of control



Configurable Environments

Conf-MDP (Metelli et al., 2018)
the environment can be configured



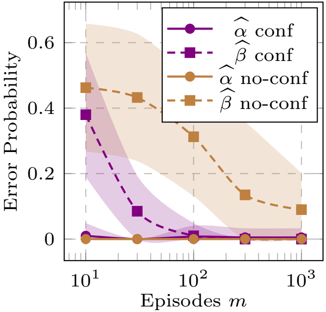
Performance of the Identification Rules

Discrete Grid world

- $\pi_{\boldsymbol{\theta}}$ linear in RBF features ϕ
- Agent observes a limited subset of ϕ
- Configure the initial state distribution

$$\alpha = \Pr\left(\exists i \notin I^{\text{Ag}} \,:\, i \in \widehat{I}\right)$$

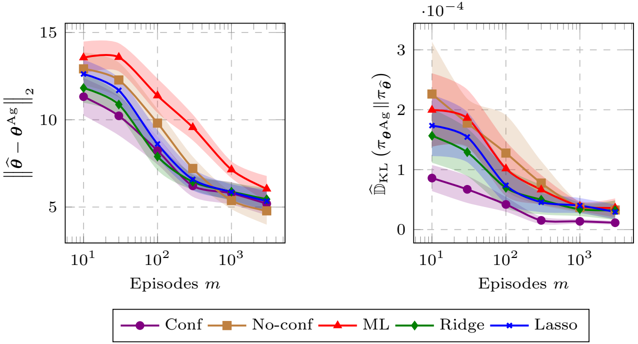
$$\beta = \Pr\left(\exists i \in I^{\text{Ag}} \,:\, i \notin \widehat{I}\right)$$



Application to Imitation Learning
Discrete Grid world

- Maximum likelihood policy estimation

$$-\sum_{i=1}^{n} \log \pi_{\boldsymbol{\theta}}(a_i|s_i) + \mathrm{Reg}(\boldsymbol{\theta})$$

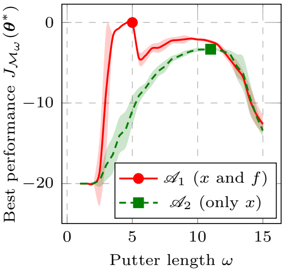


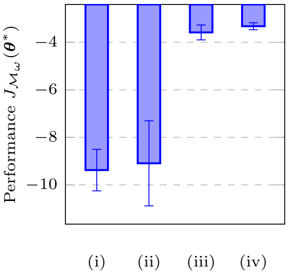
Application to Conf-MDP

Minigolf

- $\mathscr{A}_1$ observes *position* x and *friction* f
- $\mathscr{A}_2$ observes *position* x only

(i) random choice
(ii) optimal for $\mathscr{A}_1$
(iii) using identification rule
(iv) optimal for $\mathscr{A}_2$





Discussion and Conclusions

Contributions

- Identification rules
- Environment configurability to improve identification
- Applications of policy space identification

Future Works

- Bayesian statistical tests
- Multi-agent systems

Discussion and Conclusions

Contributions

- Identification rules
- Environment configurability to improve identification
- Applications of policy space identification

Future Works

- Bayesian statistical tests
- Multi-agent systems

Thank You for Your Attention!

Code: `github.com/albertometelli/policy-space-identification`

Contact: `albertomaria.metelli@polimi.it`

References I

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Alberto Maria Metelli, Mirco Mutti, and Marcello Restelli. Configurable markov decision processes. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 3488–3497. PMLR, 2018.

Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J. Andrew Bagnell, Pieter Abbeel, and Jan Peters. An algorithmic perspective on imitation learning. *Foundations and Trends in Robotics*, 7(1-2):1–179, 2018. doi: 10.1561/2300000053.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.