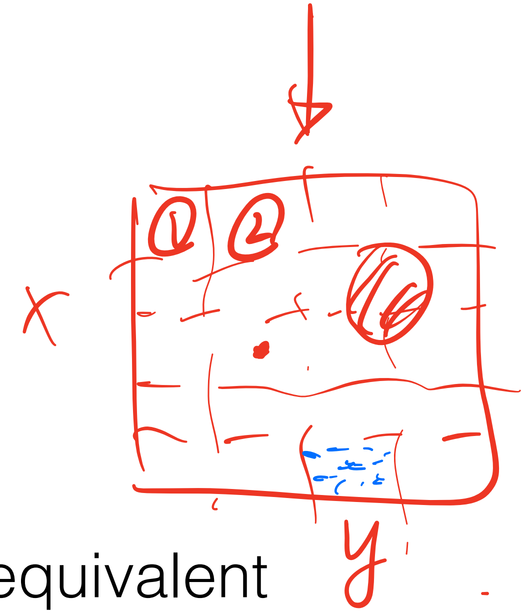
State Abstractions

Notations and Setup



- MDP $M = (S, A, P, R, \gamma)$
- Abstraction $\phi : S \rightarrow S_\phi$
- Surjection — aggregate states and treat as equivalent
- Are the aggregated states really equivalent?
- Do they have the same…
  - optimal action?
  - Q* values?
  - dynamics and rewards?

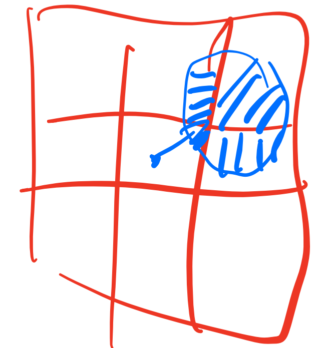
(x, y, z, w)

$S = (x, y, z, w) \xmapsto{\phi} (x, y)$

Outline of the lecture

1. Define various notions/criteria of abstractions
2. Study their relationships
3. Analyze how to use them (e.g., building an abstract model) in planning and learning
   - Abstract model will also appear in 1 & 2

Abstraction hierarchy



An abstraction ϕ is … if … $\forall\, s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$

- π^*-irrelevant: $\exists\ \pi_M{}^*$ s.t. $\pi_M{}^*(s^{(1)}) = \pi_M{}^*(s^{(2)})$
- Q^*-irrelevant: $\forall\, a$, $Q_M{}^*(s^{(1)}, a) = Q_M{}^*(s^{(2)}, a)$
- Model-irrelevant: $\forall\, a \in A$, $R(s^{(1)}, a) = R(s^{(2)}, a)$
  (bisimulation) $\forall\, a \in A,\ x' \in S_\phi$, $P(x' \mid s^{(1)}, a) = P(x' \mid s^{(2)}, a)$

$\sum_{s' \in \phi^{-1}(x')} P(s' \mid s^{(1)}, a)$

$\forall\, s' \in S.$

$P(s' \mid s^{(1)}, a) = P(s' \mid s^{(2)}, a)$

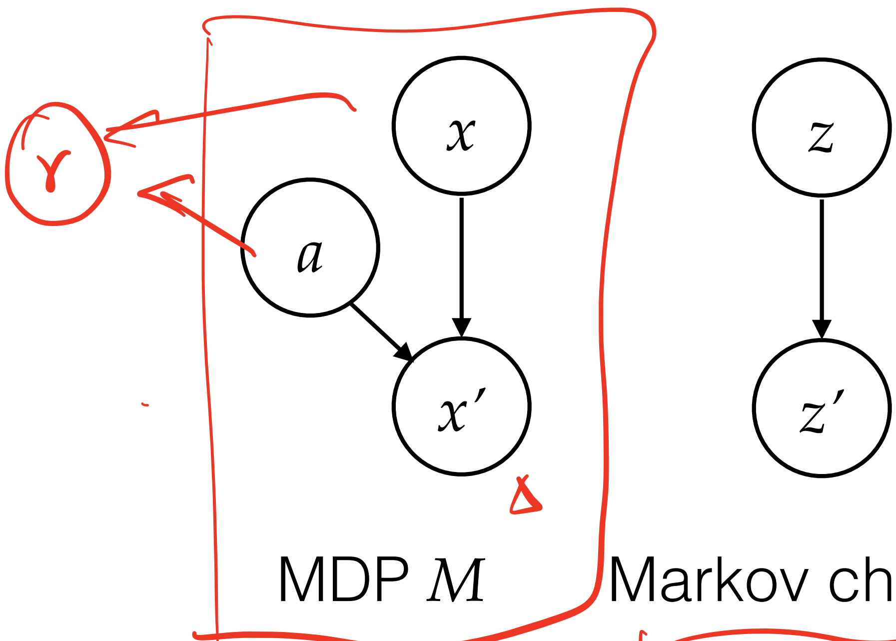
Theorem: Model-irrelevance $\Rightarrow$ Q^*-irrelevance $\Rightarrow$ π^*-irrelevance

Why not $P(s' \mid s^{(1)}, a) = P(s' \mid s^{(2)}, a)$?

violate.

$s = (x, z)$.

$\phi : (x, z) \mapsto \boxed{x}$.



$(x, z^{(1)})$ and $(x, z^{(2)})$ cannot be aggregated under the s'-based condition

$$P((x', z') \mid (x, z), a) = \boxed{P_M(x' \mid x, a)} \cdot P_C(z' \mid z)$$

$(x, z^{(1)})$

$(x, z^{(2)})$.

integrated out by bisimulation

Abstraction induces an **equivalence relation**

- Reflexivity, symmetry, transitivity
- Equivalence notion is a canonical representation of abstraction (i.e., what symbol you associate with each abstract state doesn't matter; what matters is which states are aggregated together)
- Partition the state space into **equivalence classes**
- Coarsest bisimulation is unique (see proof in notes)
    - sketch: if ϕ_1 and ϕ_2 are both bisimulations, their *common coarsening* is also a bisimulation (two states are aggregated if they are aggregated under *either* ϕ_1 or ϕ_2)

1, 2 3.   1, 2, 3

The abstract MDP implied by bisimulation

$\leftarrow S_\phi$

ϕ is bisimulation: $R(s^{(1)}, a) = R(s^{(2)}, a)$, $P(x' \mid s^{(1)}, a) = P(x' \mid s^{(2)}, a)$

- MDP $M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$
- For any $x \in S_\phi$, $a \in A$, $x' \in S_\phi$
  - $R_\phi(x, a) = R(s, a)$ for any $s \in \phi^{-1}(x)$
  - $P_\phi(x' \mid x, a) = P(x' \mid s, a)$ for any $s \in \phi^{-1}(x)$
- No way to distinguish between the two routes:

$$
\begin{tikzcd}
M \arrow[r, "\text{generate data}"] \arrow[d, "\text{compress w/ } \phi"'] \arrow[dr, bend left] & \{(s, a, r, s')\} \arrow[d, "\text{compress w/ } \phi"] \\
M_\phi \arrow[r, "\text{generate data}"'] & \{(\phi(s), a, r, \phi(s'))\}
\end{tikzcd}
$$

(s, a, r, s')

$\downarrow$

$(\phi(s), a, r, \phi(s'))$

$\boxed{\begin{matrix} s^{(1)} \\ s^{(2)} \end{matrix}}$ $\begin{matrix} , a \to r. \\ , a \to r. \end{matrix}$

$\Downarrow$

$x, a \to 2r.$

Implications of bisimulation

$\boxed{\pi}: S_\phi \to \Delta(A)$

- Q^* is preserved
- $Q_M{}^\pi$ is preserved *for any* π *lifted from an abstract policy*
  - the policy must take the same action (distribution) across aggregated states

$s \mapsto \pi(\phi(s))$

$\to R(s^{(1)}, a) = R(s^{(2)}, a).$

$\to P(x' | s^{(1)}, a) = P(x' | s^{(2)}, a)$

Extension to handle action aggregation

$(S, \underline{A}, P, R, \gamma)$

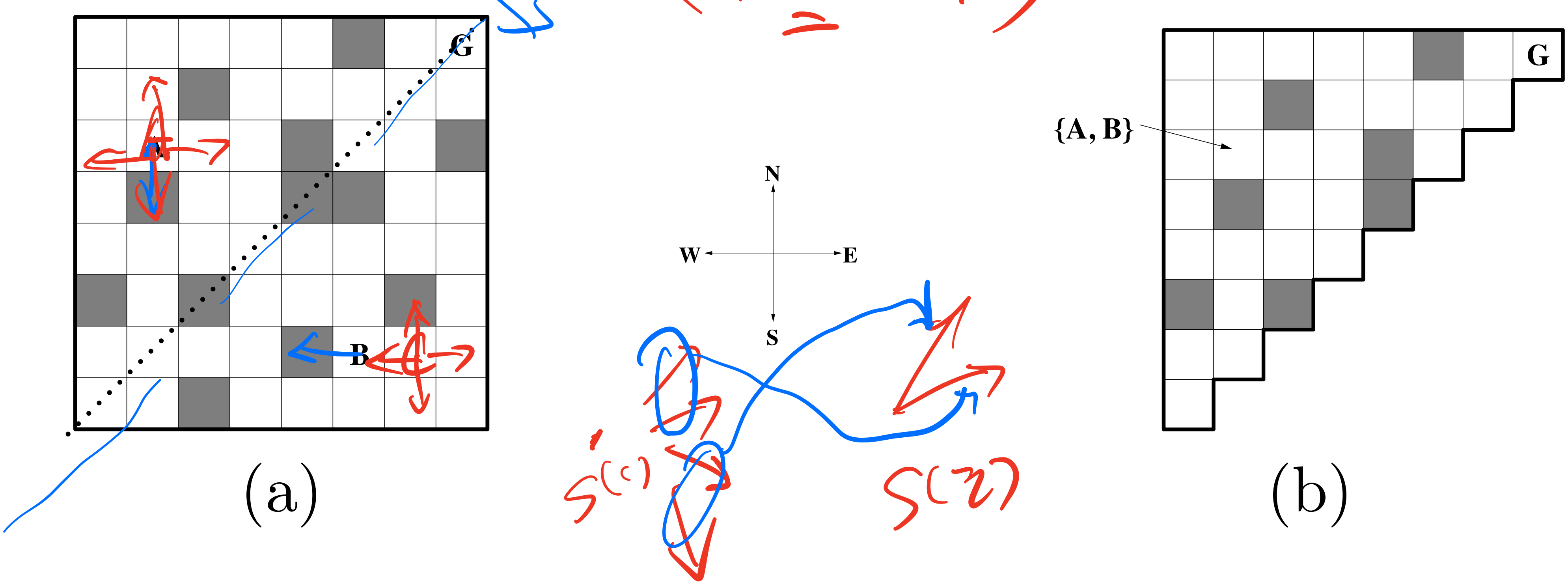


Figure from: Ravindran & Barto. Approximate Homomorphisms: A framework for non-exact minimization in Markov Decision Processes. 2004.

$[\pi]_M$

Definition 3 (Approximate abstractions). Given MDP $M = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$ and state abstraction ϕ that operates on $\mathcal{S}$, define the following types of abstractions:

δ

1. ϕ is an $\epsilon_{\pi^\star}$-approximate $\pi^\star$-irrelevant abstraction, if there exists an abstract policy $\pi : \mathcal{S}_\phi \to \mathcal{A}$, such that $\|V_M^\star - V_M^{[\pi]_M}\|_\infty \leq \epsilon_{\pi^\star}$.
2. ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction if there exists an abstract Q-value function $f : \mathcal{S}_\phi \times \mathcal{A} \to \mathbb{R}$, such that $\|[f]_M - Q_M^\star\|_\infty \leq \epsilon_{Q^\star}$.
3. ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction if for any $s^{(1)}$ and $s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)}), \forall a \in \mathcal{A}$,

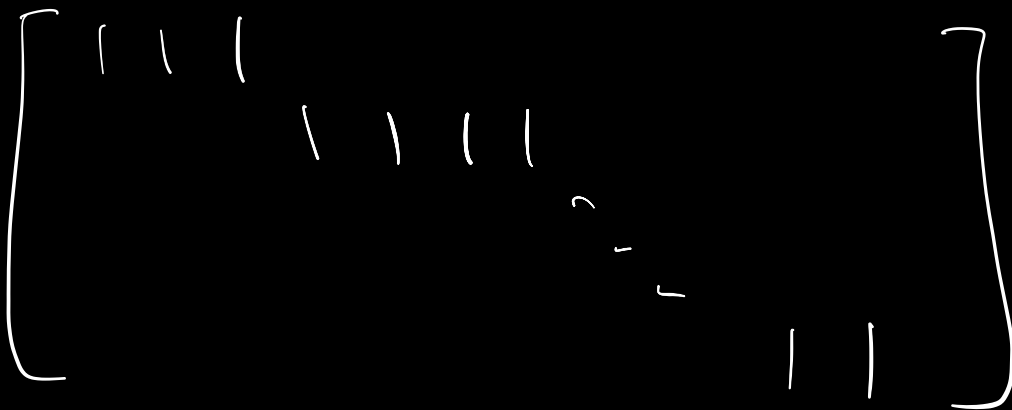
$P(x' | s^{(1)}, a)$

$$|R(s^{(1)}, a) - R(s^{(2)}, a)| \leq \epsilon_R, \quad \left\|\Phi P(s^{(1)}, a) - \Phi P(s^{(2)}, a)\right\|_1 \leq \epsilon_P. \tag{3}$$

Useful notation: Φ is a $|\mathcal{S}_\phi| \times |\mathcal{S}|$ matrix, with

$$\Phi(x, s) = \mathbb{I}[\phi(s) = x]$$

- lifting a state-value function: $[V^\star_{M_\phi}]_M = \Phi^\top V^\star_{M_\phi}$
- collapsing the transition distribution: $\Phi\, P(s, a)$

$\Phi =$ 

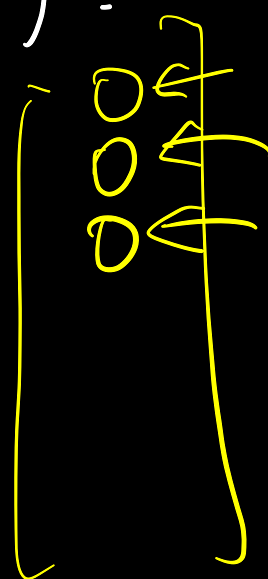
(1) $p \in \Delta(S) \subseteq \mathbb{R}^{|S|}$.

$\Phi \times p =$ 

$\Phi P(\cdot \mid s^{(1)}, a) = \Phi P(\cdot \mid s^{(2)}, a)$.

(2) $f : S_\phi \to \mathbb{R}$. $\Phi^T f$.

 $= [f]_M$.



Theorem 2. *(1) If ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction, then ϕ is also an approximate $Q^\star$-irrelevant abstraction with approximation error $\epsilon_{Q^\star} = \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{2(1-\gamma)^2}$.*
(2) If ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction, then ϕ is also an approximate $\pi^\star$-irrelevant abstraction with approximation error $\epsilon_{\pi^\star} = 2\epsilon_{Q^\star}/(1-\gamma)$.

$\|V^* - V^{\pi_f}\|_\infty \le \frac{2\|f - Q^*\|_\infty}{1-\gamma}$.

- (2) follows directly from a known result; can you see?

- Construct the f in the definition of approx. Q*-irrelevance:

ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction if there exists an abstract Q-value function $f : \mathcal{S}_\phi \times \mathcal{A} \to \mathbb{R}$, such that $\|[f]_M - Q^\star_M\|_\infty \le \epsilon_{Q^\star}$.

  - Define $M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$ w/ any weighting distributions $\{p_x: x \in S_\phi\}$, where each p_x is supported on $\phi^{-1}(x)$

$$R_\phi(x, a) = \Sigma_{s\in\phi^{-1}(x)}\, p_x(s)\, R(s, a), \quad P_\phi(x, a) = \Sigma_{s\in\phi^{-1}(x)}\, p_x(s)\, \Phi\, P(s, a).$$

  - $|R_\phi(\phi(s), a) - R(s, a)| \le \varepsilon_R$, $\|P_\phi(\phi(s), a) - \Phi\, P(s, a)\| \le \varepsilon_P$.
  - Set $f := Q^\star_{M_\phi}$, bound $\|[f]_M - Q^\star_M\|_\infty$

$$\| \boxed{[Q^*_{M_\phi}]_M}^{g} - Q^*_M \|_\infty$$

$\|g - Q^*\|_\infty$
$= \|g - Tg + Tg - TQ^*\|_\infty$
$\le \|g - Tg\|_\infty + \gamma \|g - Q^*\|_\infty$

$$= \frac{1}{1-\gamma} \| \boxed{[Q^*_{M_\phi}]_M}^{g} - T_M \boxed{[Q^*_{M_\phi}]_M}_{g} \|_\infty$$

$$= \frac{1}{1-\gamma} \| \underbrace{[T_{M_\phi} Q^*_{M_\phi}]_M - T_M [Q^*_{M_\phi}]_M}_{\triangle} \|_\infty$$

$\forall (s,a)$.

$$\left| (T_{M_\phi} Q^*_{M_\phi})(\phi(s),a) - \underline{T_M [Q^*_{M_\phi}]_M (s,a)} \right|$$

$$= \left| R_\phi(\phi(s),a) + \underline{\gamma \langle P_\phi(\phi(s),a), V^*_{M_\phi} \rangle} - R(s,a) - \gamma \langle P(s,a), [V^*_{M_\phi}]_M \rangle \right|$$

$R(s,a) + \gamma \mathbb{E}_{s \sim P(\cdot|s,a)} [\max_{a'} Q^*(s',a')]$

$s' \to \max_{a'} Q(s',a')$.

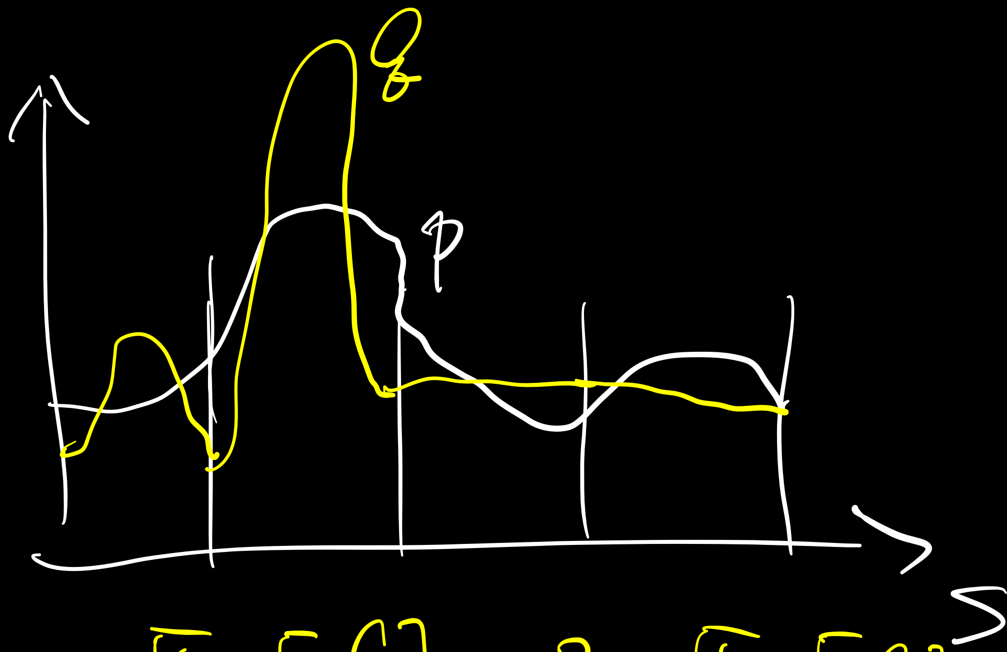
$$\le \varepsilon_R + \gamma \times \left(\langle \boxed{P_\phi(\phi(s),a)}, \boxed{V^*_{M_\phi}} \rangle - \langle \boxed{P(s,a)} \Phi^T V^*_{M_\phi} \rangle \right).$$

$\wedge$ $\varepsilon_P \cdot \frac{V_{max}}{2}$.

$= \langle \Phi P(s,a), V^*_{M_\phi} \rangle$

$\langle u, Av \rangle = (u^T A) v$
$= \langle A^T u, v \rangle$.

$$\Phi\, P(s^{(1)}, a) = \Phi\, P(s^{(2)}, a).$$



$$\mathbb{E}_p[f] \overset{?}{=} \mathbb{E}_q[f].$$

Outline of the lecture

1. Define various notions/criteria of abstractions
2. Study their relationships
3. Analyze how to use them (e.g., building an abstract model) in planning and learning

- e.g., plan in M_ϕ to reduce computational cost
- If ϕ is not exact bisimulation, what's sub-optimality as a function of $(\varepsilon_R, \varepsilon_P)$? (Partially answered; will take a closer look)
- What if ϕ is only approximately Q*-irrelevant? Is the abstract model still useful? Can we still bound loss as a function of ε_{Q^*}?
- Learning setting?

Loss of $\pi^\star_{M_{\phi M}}$: approx. bisimulation

$\frac{2\|\Phi - Q^\star\|_\infty}{1-\gamma} \; \epsilon$

- Recall: M_ϕ defined using any weighting distributions $\{p_x\}$ satisfies $|R_\phi(\phi(s), a) - R(s, a)| \le \varepsilon_R$, $\|P_\phi(\phi(s), a) - \Phi\, P(s, a)\|_1 \le \varepsilon_P$.

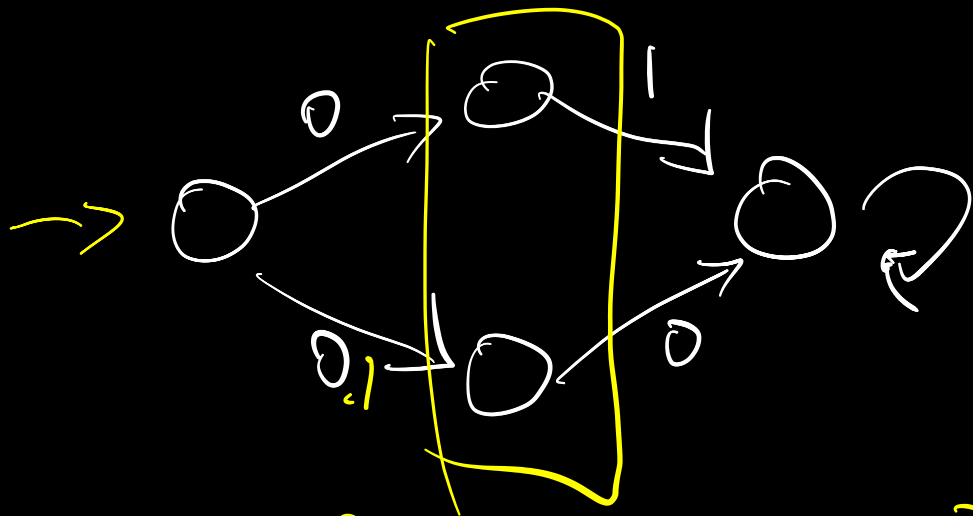
$[Q^\star_{M_\phi}]_M$.

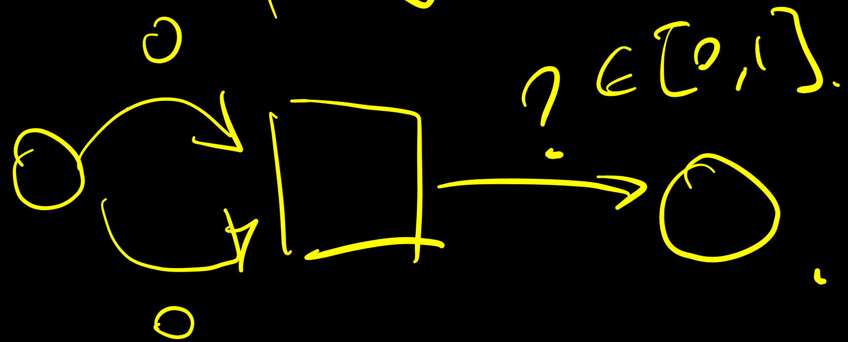
- Apply earlier Theorem: $\left\| V^\star_M - V^{[\pi^\star_{M_\phi}]_M}_M \right\|_\infty \le \frac{2\epsilon_R}{(1-\gamma)^2} + \frac{\gamma \epsilon_P R_{\max}}{(1-\gamma)^3}$
- Can improve: $\left\| V^\star_M - V^{[\pi^\star_{M_\phi}]_M}_M \right\|_\infty \le \frac{2\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{(1-\gamma)^2}$
- Idea: for any $\pi : S_\phi \to A$, $\left\| [V^\pi_{M_\phi}]_M - V^{[\pi]_M}_M \right\|_\infty \le \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{2(1-\gamma)^2}$
- Finally,

$$
\begin{aligned}
V^\star_M(s) - V^{[\pi^\star_{M_\phi}]_M}_M(s) &= V^\star_M(s) - V^\star_{M_\phi}(\phi(s)) + V^\star_{M_\phi}(\phi(s)) - V^{[\pi^\star_{M_\phi}]_M}_M(s) \\
&\le \left\| Q^\star_M - [Q^\star_{M_\phi}]_M \right\|_\infty + \left\| [V^{\pi^\star_{M_\phi}}_{M_\phi}]_M - V^{[\pi^\star_{M_\phi}]_M}_M \right\|_\infty
\end{aligned}
$$

- Lesson: w/ approx. bisimulation, take the $\max_\pi \|V^\pi_M - V^\pi_{\widehat{M}}\|_\infty$ route instead of the $\|Q^\star_M - Q^\star_{\widehat{M}}\|$ route to save dependence on horizon

$[\pi_{M_\phi^*}]_M$ can be bad. even if ϕ is π^*-irrelevant.





Loss of $\pi^\star_{M_{\phi_M}}$: approx. Q*-irrelevance

- M_ϕ well defined, but transitions/rewards don't make sense
- Can still show: $\|[Q^\star_{M_\phi}]_M - Q^\star_M\|_\infty \leq 2\epsilon_{Q^\star}/(1-\gamma)$
- Exact case ($\epsilon_{Q^\star} = 0$): $\forall\, s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$

$$R(s^{(1)}, a) + \gamma\langle P(s^{(1)}, a), V^\star_M\rangle = Q^\star(s^{(1)}, a) = Q^\star(s^{(2)}, a) = R(s^{(2)}, a) + \gamma\langle P(s^{(2)}, a), V^\star_M\rangle$$

"inverse" of lifting (can only be applied to piece-wise constant functions)

So:
$$\begin{aligned}
(\mathcal{T}_{M_\phi}[Q^\star_M]_\phi)(x,a) &= R_\phi(x,a) + \gamma\langle P_\phi(x,a), [V^\star_M]_\phi\rangle \\
&= \sum_{s\in\phi^{-1}(x)} p_x(s)\left(R(s,a) + \gamma\langle \Phi P(s,a), [V^\star_M]_\phi\rangle\right) \\
&= \sum_{s\in\phi^{-1}(x)} p_x(s)\left(R(s,a) + \gamma\langle P(s,a), V^\star_M\rangle\right) \\
&= \sum_{s\in\phi^{-1}(x)} p_x(s)\,[Q^\star_M]_\phi(x,a) = [Q^\star_M]_\phi(x,a).
\end{aligned}$$

$$[Q^*_{M_\phi}]_M = Q^*_M \in \mathbb{R}^{S\times A}$$ fixed point of T_M.

$$\Longleftrightarrow \quad Q^*_M \text{ (compressed) } =: [Q_{M^*}]_\phi$$

is fixed point of T_{M_ϕ}.

$$(T_{M_\phi}[Q^*_M]_\phi)(x,a) \overset{?}{=} \boxed{[Q^*_M]_\phi(x,a).}$$

$$= R_\phi(x,a) + \gamma \langle P_\phi(x,a), [V^*_M]_\phi \rangle$$

$$= \sum_{s\in\phi^{-1}(x)} p_x(s)\left(R(s,a) + \gamma \langle \Phi P(s,a), [V^*_M]_\phi \rangle \right)$$

$$= \underbrace{\sum_{s\in\phi^{-1}(x)} p_x(s)\Big(\underbrace{R(s,a) + \gamma \langle P(s,a), V^*_M \rangle}_{Q^*_M(s,a).} \Big).}$$

Loss of $\pi^\star_{M_{\phi_M}}$: approx. Q*-irrelevance

- Approximate case: proof breaks as $Q_M{}^*$ not piece-wise constant
- Workaround: define a new model $M_\phi{}'$ over S

$$R'_\phi(s,a) = \mathbb{E}_{\tilde{s}\sim p_{\phi(s)}}[R(\tilde{s},a)], \qquad P'_\phi(s'|s,a) = \mathbb{E}_{\tilde{s}\sim p_{\phi(s)}}[P(s'|\tilde{s},a)].$$

- Can show: M_ϕ and $M_\phi{}'$ share the same Q^* (up to lifting)
- $\left\| [Q^\star_{M_\phi}]_M - Q^\star_M \right\|_\infty = \left\| Q^\star_{M'_\phi} - Q^\star_M \right\|_\infty \leq \frac{1}{1-\gamma} \left\| \mathcal{T}_{M'_\phi} Q^\star_M - Q^\star_M \right\|_\infty$

$$\begin{aligned}
&|(\mathcal{T}_{M'_\phi} Q^\star_M)(s,a) - Q^\star_M(s,a)| \\
&= |R'_\phi(s,a) + \gamma\langle P'_\phi(s,a), V^\star_M\rangle - Q^\star_M(s,a)| \\
&= \left| \left(\sum_{\tilde{s}:\phi(\tilde{s})=\phi(s)} p_x(\tilde{s})\left(R(\tilde{s},a) + \gamma\langle P(\tilde{s},a), V^\star_M\rangle\right)\right) - Q^\star_M(s,a) \right| \\
&= \left| \sum_{\tilde{s}:\phi(\tilde{s})=\phi(s)} p_x(\tilde{s})\left(Q^\star_M(\tilde{s},a) - Q^\star_M(s,a)\right) \right| \leq \left| \sum_{\tilde{s}:\phi(\tilde{s})=\phi(s)} p_x(\tilde{s})(2\epsilon_{Q^\star}) \right| = 2\epsilon_{Q^\star}.
\end{aligned}$$

Loss of $\pi^\star_{M_{\phi_M}}$: approx. Q*-irrelevance

- Lesson: with Q*-irrelevance, the $\max_\pi \|V^\pi_M - V^\pi_{\widehat{M}}\|_\infty$ approach is not available; $\|Q^\star_M - Q^\star_{\widehat{M}}\|$ is the only choice
- If ϕ does not respect transition/reward, our analysis does not have to either!

Recap

- **Theorem 2.** *(1) If ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction, then ϕ is also an approximate $Q^\star$-irrelevant abstraction with approximation error $\epsilon_{Q^\star} = \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{2(1-\gamma)^2}$.*
  (2) If ϕ is an $\epsilon_{Q^\star}$-approximate $Q^\star$-irrelevant abstraction, then ϕ is also an approximate $\pi^\star$-irrelevant abstraction with approximation error $\epsilon_{\pi^\star} = 2\epsilon_{Q^\star}/(1-\gamma)$.
- Given weighting distributions $\{p_x\}$, define $M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$

  $R_\phi(x, a) = \Sigma_{s \in \phi^{-1}(x)}\ p_x(s)\, R(s, a)$, $\quad P_\phi(x, a) = \Sigma_{s \in \phi^{-1}(x)}\ p_x(s)\, \Phi\, P(s, a)$.
- How lossy is it to plan in M_ϕ and lift back to M?
  - If approx. bisimulation, use "$\max_\pi \|V_M^\pi - V_{\widehat{M}}^\pi\|_\infty$" type analysis

    $$\left\| V_M^\star - V_M^{[\pi^\star_{M_\phi}]_M} \right\|_\infty \le \frac{2\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{(1-\gamma)^2}$$
  - If approx. Q*-irrelevance, use "$\|Q_M^\star - Q_{\widehat{M}}^\star\|$" type analysis

    $$\left\| V_M^\star - V_M^{[\pi^\star_{M_\phi}]_M} \right\|_\infty \le \frac{2\epsilon_{Q^\star}}{(1-\gamma)^2}$$

Compare abstract model w/ bisimulation vs w/ Q*-irrelevance

Both guarantee optimality (exact case), but in different ways

- Consider value iteration (VI) in true model vs abstract model
- Bisimulation: every step of abstract VI resembles that step in true VI, throughout all iterations, b/c $\forall f : \phi(\mathcal{S}) \to \mathbb{R},\ \mathcal{T}[f]_M = [\mathcal{T}_{M_\phi} f]_M$
- Q*-irrelevance: abstract VI initially behaves crazily. It only starts to resemble true VI when the function is close to $Q_M{}^*$
  - This is a circular argument $\mathcal{T} Q_M^\star = [\mathcal{T}_{M_\phi} [Q_M^\star]_\phi]_M$
  - Secret is stability—contraction of abstract Bellman update
  - Abstract Bellman update is a special case of projected Bellman update, and in general stability is not guaranteed. In that case, "Q*-irrelevance" alone is not enough to guarantee optimality

The learning setting

- Given: $D = \{D_{s,a}\}_{(s,a)\in\mathcal{S}\times\mathcal{A}}$ and ϕ
- Algorithm: CE after processing data w/ ϕ
- Shouldn't assume $|D_{s,a}|$ is the same for all (s, a)
    - … as we want to handle $|D| << |S|$
    - What should appear in the bound to describe sample size?

$$n_\phi(D) := \min_{x\in\mathcal{S}_\phi, a\in\mathcal{A}} |D_{x,a}|, \quad \textbf{where} \quad D_{x,a} := \bigcup_{s\in\phi^{-1}(x)} D_{s,a}.$$

    - At the mercy of data to be exploratory

The learning setting

- Analysis varies according to whether ϕ is (approx.) bisimulation or Q*-irrelevant and the style ($\max_\pi \|V_M^\pi - V_{\widehat{M}}^\pi\|_\infty$ vs $\|Q_M^\star - Q_{\widehat{M}}^\star\|$)
- Will show analysis of Q*-irrelevance (can only use "$\|Q_M^\star - Q_{\widehat{M}}^\star\|$")
- Let $\widehat{M}_\phi$ be the estimated model
- Let M_ϕ be an abstract model w/ weighting distributions $p_x(s) \propto |D_{s,a}|$
- M_ϕ is the "expected model" of $\widehat{M}_\phi$
- $$\left\|Q_M^\star - [Q_{\widehat{M}_\phi}^\star]_M\right\|_\infty \le \underbrace{\left\|Q_M^\star - [Q_{M_\phi}^\star]_M\right\|_\infty}_{\text{Approximation error}} + \underbrace{\left\|[Q_{M_\phi}^\star]_M - [Q_{\widehat{M}_\phi}^\star]_M\right\|_\infty}_{\text{Estimation error}}$$

Approximation error
- "Bias", informally
- Doesn't vanish with more data
- Smaller with a **finer** ϕ

  (not w/ bisimulation; we will see why…)

Estimation error
- "Variance", informally
- Goes to 0 w/ infinite data
- Smaller with a **coarser** ϕ

$$\left\| Q^\star_M - [Q^\star_{\widehat{M}_\phi}]_M \right\|_\infty \le \Big\| \underbrace{Q^\star_M - [Q^\star_{M_\phi}]_M}_{\text{already handled}} \Big\|_\infty + \Big\| \underbrace{[Q^\star_{M_\phi}]_M - [Q^\star_{\widehat{M}_\phi}]_M}_{\text{to be analyzed}} \Big\|_\infty$$

- Reusing the analysis for $\|Q^\star_M - Q^\star_{\widehat{M}}\|$
- Challenge: data is not generated from M_ϕ
- Leverage the fact that Hoeffding can be applied to r.v.'s with non-identical distributions

$$\left\| [Q^\star_{M_\phi}]_M - [Q^\star_{\widehat{M}_\phi}]_M \right\|_\infty = \left\| Q^\star_{M_\phi} - Q^\star_{\widehat{M}_\phi} \right\|_\infty$$

$$\le \frac{1}{1-\gamma} \left\| Q^\star_{M_\phi} - \mathcal{T}_{\widehat{M}_\phi} Q^\star_{M_\phi} \right\|_\infty = \frac{1}{1-\gamma} \left\| \mathcal{T}_{\widehat{M}_\phi} Q^\star_{M_\phi} - \mathcal{T}_{M_\phi} Q^\star_{M_\phi} \right\|_\infty$$

$$|(\mathcal{T}_{\widehat{M}_\phi} Q^\star_{M_\phi})(x,a) - (\mathcal{T}_{M_\phi} Q^\star_{M_\phi})(x,a)|$$

$$= |\widehat{R}_\phi(x,a) + \gamma \langle \widehat{P}_\phi(x,a), V^\star_{M_\phi} \rangle - R_\phi(x,a) - \gamma \langle P_\phi(x,a), V^\star_{M_\phi} \rangle|$$

$$= \left| \frac{1}{|D_{x,a}|} \sum_{s \in \phi^{-1}(x)} \sum_{(r,s') \in D_{s,a}} \Big(r + \gamma V^\star_{M_\phi}(\phi(s')) - R(s,a) - \gamma \langle P(s,a), [V^\star_{M_\phi}]_M \rangle \Big) \right|$$