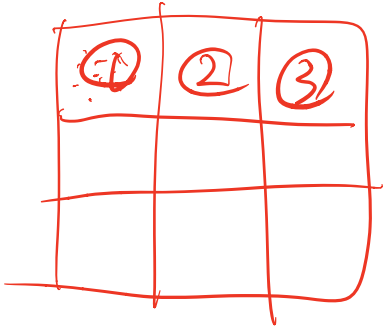
State Abstractions

Notations and Setup



- MDP $M = (S, A, P, R, \gamma)$
- Abstraction $\phi : S \rightarrow \phi(S)$ S_ϕ
- 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?

$\phi : S \rightarrow \{1, 2, 3, \ldots, 9\}$

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 (bisimulation): $\forall\, a \in A$, $R(s^{(1)}, a) = R(s^{(2)}, a)$
  $\forall\, a \in A,\ x' \in \phi(S)$, $P(x' \mid s^{(1)}, a) = P(x' \mid s^{(2)}, a)$

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

stronger.

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

Theorem: Model-irrelevance $\overset{?}{\Rightarrow}$ Q^*-irrelevance $\overset{?}{\Rightarrow}$ π^*-irrelevance

$S = X \times Z$.

"naive condition"

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

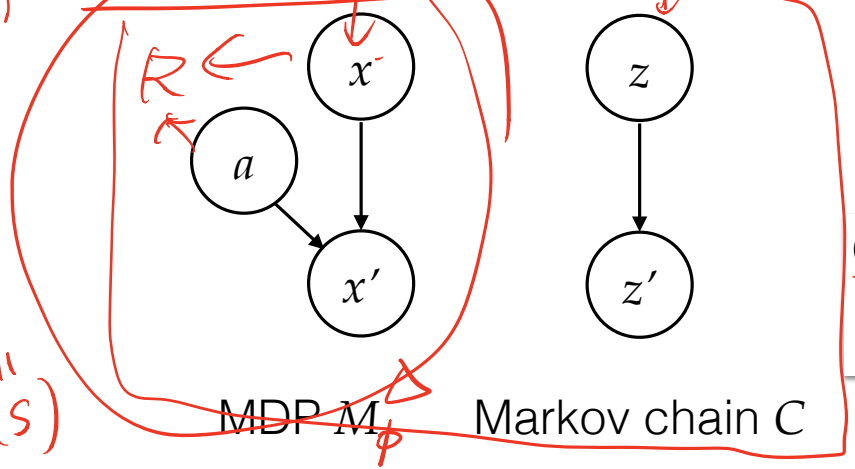
S

$\parallel$

$(X, Z) \xrightarrow{\phi} X$

$X = \{x^{(1)}, x^{(2)}, \ldots\}$

$Z = \{z^{(1)}, \ldots z^{(n)}, \ldots\}$



S_ϕ

$\parallel$

$\phi(s)$

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

(x', z')

bisimulation

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

I can aggregate $(x, z^{(1)})$ & $(x, z^{(2)})$

$P(x' \mid (x, z^{(1)}), a) = P(x' \mid (x, z^{(2)}), a)$.

integrated out by bisimulation

The abstract MDP implied by bisimulation

whenever $\phi(s^{(1)}) = \phi(s^{(2)})$

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

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

$\forall x' \in S_\phi$.

$\forall s' \in S$.

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

$\phi(s)$ $\phi(s')$.

$\phi(S) = S_\phi$.

(x, a, r, x').

Implications of bisimulation

$f: S_\phi \times A \to \mathbb{R}$.

Proof outline:

- Q^*-irrelevance
  - Plan in $\boxed{M_\phi}$ and get $\boxed{Q^\star_{M_\phi}}$ (dimension: $\boxed{\phi(S)\times A}$)
  - *Lift* $Q^\star_{M_\phi}$ from $\phi(S)$ to S (populate aggregated states with the same value)
  - Useful notation: Φ is a $|\phi(\mathcal{S})| \times |\mathcal{S}|$ 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: $\Phi\, P(s, a)$
  - Claim: $\left[Q^\star_{M_\phi}\right]_M = Q^\star_M$ (proof on board)

$S_\phi \times A$.

$[f]_M : S \times A \to \mathbb{R}$

$\forall s, a, [f]_M(s,a) = f(\phi(s), a)$

$\phi(S)$

$\mathbb{R}^{S_\phi}$

$P(s,a) \in \mathbb{R}^S$.

$\hookrightarrow$ implies Q^*_M is "piecewise const" under ϕ. $\Rightarrow Q^*$-irrelevance

$\Phi = [\mathbb{I}[\phi(s) = x]]_{x,s} \in \mathbb{R}^{S_\phi \times S}$ $\quad \boxed{S_\phi = \phi(S).}$

bisimulation: $\forall\, s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$.

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

w.t.s. $[Q^*_{M_\phi}]_M = Q^*_M$. $\Leftarrow$ $\| \cdot - Q^*_M \|$

$S \times A \to \mathbb{R}$. where $[Q^*_{M_\phi}]_M(s,a) = Q^*_{M_\phi}(\phi(s), a)$.

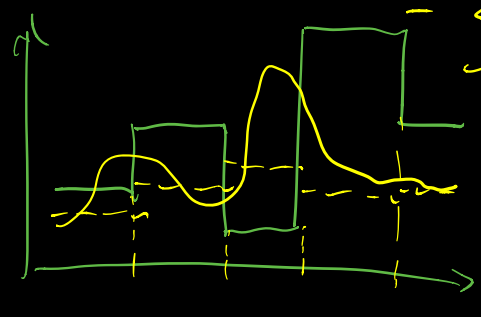
it suffices to show: $\mathcal{T}_M [Q^*_{M_\phi}]_M = [Q^*_{M_\phi}]_M$.

$$\begin{cases} \text{LHS}(s,a) = R(s,a) + \gamma \langle \underset{|S|,|S|}{P(\cdot \mid s,a)}, [V^*_{M_\phi}]_M \rangle = [\mathcal{T}_{M_\phi} Q^*_{M_\phi}]_M. \\ \text{RHS}(s,a) = R_\phi(\phi(s), a) + \gamma \langle \underset{|S_\phi|,|S_\phi|}{P_\phi(\cdot \mid \phi(s), a)}, V^*_{M_\phi} \rangle \end{cases}$$

$R(s,a) \Leftrightarrow R_\phi(\phi(s), a)$

$\langle P(\cdot \mid s,a), [V^*_{M_\phi}]_M \rangle = \langle P(\cdot \mid s,a), \Phi^T V^*_{M_\phi} \rangle$

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



$\Phi =$

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

Implications of bisimulation

- Q^*-irrelevance
- $Q_M{}^{\pi}$ is preserved *for any π lifted from an abstract policy*
- Given any lifted π, distribution over reward seq. is preserved (assuming reward is deterministic function of s, a) (Is this sufficient?)
  - Can be extended to features of state to define a notion of saliency (think: what happens when the reward criterion is missing?)
  - For deeper thoughts along these lines, read Erik Talvitie's thesis

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 (proof)

Extension to handle action aggregation

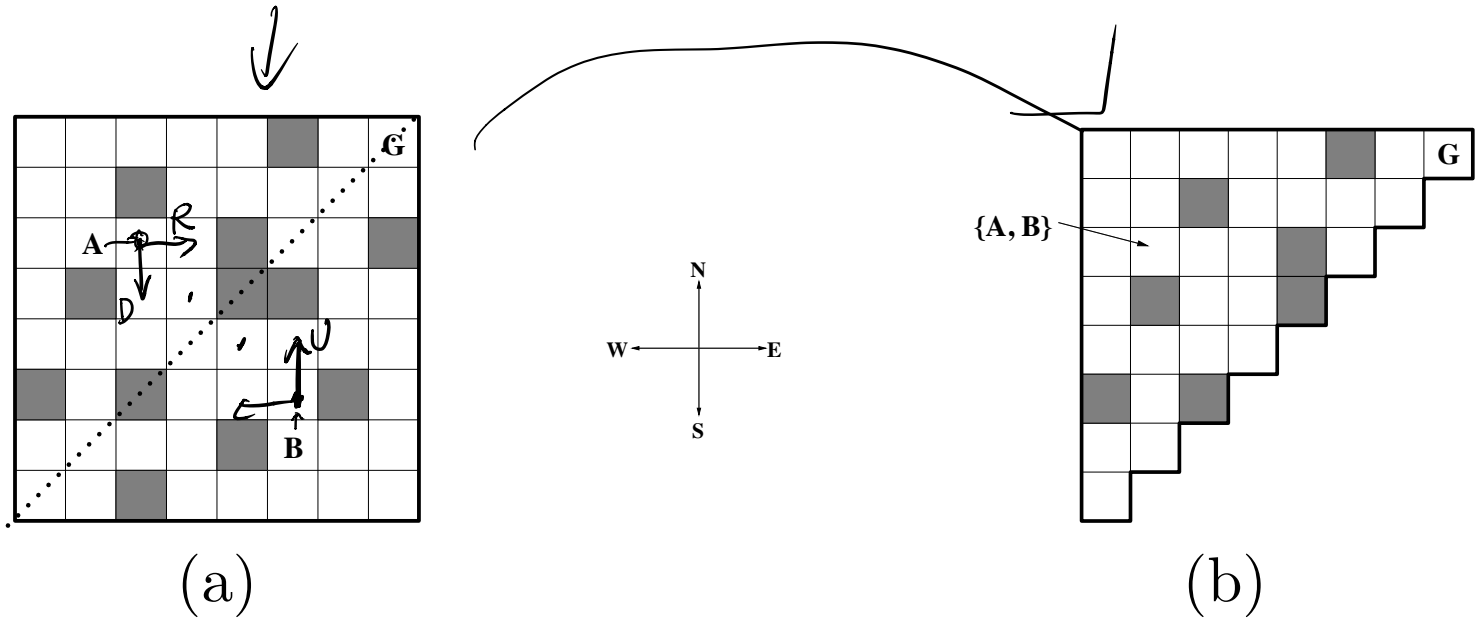


Figure from: Ravindran & Barto. Approximate Homomorphisms: A framework for non-exact minimization in Markov Decision Processes. 2004.

Approximate Abstractions

Hierarchy of approximate abstractions

1. ϕ is an $\epsilon_{\pi^\star}$-approximate $\pi^\star$-irrelevant abstraction, if there exists an abstract policy $\pi : \phi(\mathcal{S}) \to \mathcal{A}$, 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 $f : \phi(\mathcal{S}) \times \mathcal{A} \to \mathbb{R}$, 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 $\phi(s^{(1)}) = \phi(s^{(2)})$, $\forall a \in \mathcal{A}$,

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

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 $\epsilon_{\pi^\star} = 2\epsilon_{Q^\star}/(1-\gamma)$.

- (2) has been proved; remains to prove (1)
- 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 $f : \phi(\mathcal{S}) \times \mathcal{A} \to \mathbb{R}$, such that $\|[f]_M - Q^\star_M\|_\infty \leq \epsilon_{Q^\star}$.

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

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

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

$|R_\phi(\phi(s), a) - R(s, a)| \le \varepsilon_R, \quad |P_\phi(\phi(s), a) - \Phi\, P(s, a)| \le \varepsilon_P.$

$\varepsilon_R = 0$

$\varepsilon_P = 0.$

$$\left\|[Q^\star_{M_\phi}]_M - Q^\star_M\right\|_\infty \le \frac{1}{1-\gamma}\left\|[Q^\star_{M_\phi}]_M - \mathcal{T}[Q^\star_{M_\phi}]_M\right\|_\infty = \frac{1}{1-\gamma}\left\|[\mathcal{T}_{M_\phi} Q^\star_{M_\phi}]_M - \mathcal{T}[Q^\star_{M_\phi}]_M\right\|_\infty.$$

For any (s, a),

$$\begin{aligned}
&|([\mathcal{T}_{M_\phi} Q^\star_{M_\phi}]_M)(s,a) - (\mathcal{T}[Q^\star_{M_\phi}]_M)(s,a)| \\
&= |(\mathcal{T}_{M_\phi} Q^\star_{M_\phi})(\phi(s),a) - (\mathcal{T}[Q^\star_{M_\phi}]_M)(s,a)| \\
&= |R_\phi(\phi(s),a) + \gamma\langle P_\phi(\phi(s),a), V^\star_{M_\phi}\rangle - R(s,a) - \gamma\langle P(s,a), [V^\star_{M_\phi}]_M\rangle| \\
&\le \epsilon_R + \gamma\left|\langle P_\phi(\phi(s),a), V^\star_{M_\phi}\rangle - \langle P(s,a), \Phi^\top V^\star_{M_\phi}\rangle\right| \\
&= \epsilon_R + \gamma\left|\langle P_\phi(\phi(s),a), V^\star_{M_\phi}\rangle - \langle \Phi P(s,a), V^\star_{M_\phi}\rangle\right| \\
&\le \epsilon_R + \gamma\epsilon_P \|V^\star_{M_\phi} - \tfrac{R_{\max}}{2(1-\gamma)}\mathbf{1}\|_\infty \\
&\le \epsilon_R + \gamma\epsilon_P R_{\max}/(2(1-\gamma)).
\end{aligned}$$

Using Abstract Models

Outline of the remaining material

- Consider planning, e.g., want to plan in the abstract model instead of the original model to reduce computation cost
- Approach: compress the model (M_ϕ), and plan in M_ϕ
- Want to know: if ϕ is not an exact bisimulation, how lossy is the resulting policy as a function of (ε_R, ε_P) ?
- 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 $\left[\pi^\star_{M_\phi}\right]_M$: approx. bisimulation

- Recall: M_ϕ defined using any weighting distributions $\{p_x\}$ satisfies $|R_\phi(\phi(s), a) - R(s, a)| \leq \varepsilon_R$, $\|P_\phi(\phi(s), a) - \Phi\, P(s, a)\|_1 \leq \varepsilon_P$.
- Apply both claims of the Thm: $\left\|V^\star_M - V_M^{[\pi^\star_{M_\phi}]_M}\right\|_\infty \leq \frac{2\epsilon_R}{(1-\gamma)^2} + \frac{\gamma\epsilon_P R_{\max}}{(1-\gamma)^3}$
- Can improve: $\left\|V^\star_M - V_M^{[\pi^\star_{M_\phi}]_M}\right\|_\infty \leq \frac{2\epsilon_R}{1-\gamma} + \frac{\gamma\epsilon_P R_{\max}}{(1-\gamma)^2}$
- Idea: for any $\pi : \phi(S) \rightarrow A$, $\left\|[V^\pi_{M_\phi}]_M - V_M^{[\pi]_M}\right\|_\infty \leq \frac{\epsilon_R}{1-\gamma} + \frac{\gamma\epsilon_P R_{\max}}{2(1-\gamma)^2}$
- Finally,

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

- 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

Loss of $\left[\pi^\star_{M_\phi}\right]_M$: approx. Q*-irrelevance

M_ϕ $\phi(s^{(1)}) = \phi(s^{(2)})$ a.l.a. $Q^*(s^{(1)}, a) = Q^*(s^{(2)}, a)$ $\forall a$,

- 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 (RHS=0): ∀ $s^{(1)}$, $s^{(2)}$ 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:
$$\begin{aligned}
(\mathcal{T}_{M_\phi}[Q^\star_M]_\phi)(x, a) &= R_\phi(x, a) + \gamma\langle P_\phi(x, a), [V^\star_M]_\phi\rangle \\
&= \sum_{s\in\phi^{-1}(x)} \boxed{p_x(s)}\left(R(s, a) + \gamma\langle \Phi P(s, a), [V^\star_M]_\phi\rangle\right) \\
&= \sum_{s\in\phi^{-1}(x)} p_x(s)\left(R(s, a) + \gamma\langle P(s, a), \boxed{V^\star_M}\rangle\right) \\
&= \sum_{s\in\phi^{-1}(x)} p_x(s)\,[Q^\star_M]_\phi(x, a) = [Q^\star_M]_\phi(x, a).
\end{aligned}$$

$S_\phi \times A$.

$[Q^*_M]_\phi(x, a) = Q^*_M(s, a)$ $\forall s \in \phi^{-1}(x)$.

in general $(\mathcal{T}_M f)(s^{(1)}, a) \neq (\mathcal{T}_M f)(s^{(2)}, a)$.

but, "=" if $f = Q^*_M$.

Loss of $\left[\pi^\star_{M_\phi}\right]_M$: 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 $\left[\pi^\star_{M_\phi}\right]_M$: 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 $M_\phi = (\phi(S), A, P_\phi, R_\phi, \gamma)$

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

- 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

$$\left\| V_M^\star - V_M^{[\pi^\star_{M_\phi}]_M} \right\|_\infty \leq \frac{2\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{(1-\gamma)^2}$$

  - If approx. Q*-irrelevance, use "$\|Q_M^\star - Q_{\widehat{M}}^\star\|$" type analysis

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

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 $\forall f : \phi(\mathcal{S}) \to \mathbb{R},\ \mathcal{T}_M[f]_M = [\mathcal{T}_{M_\phi} f]_M$
- 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
  - 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

$Q_M^* = \mathcal{T}_M Q_M^\star = [\mathcal{T}_{M_\phi}[Q_M^\star]_\phi]_M$

Gordon '95.

The Learning Setting and Finite Sample Analysis

The learning setting

$D_{s,a} = \{(r, s') : r = R(s,a),\ s' \sim P(\cdot|s,a)\}$.

- Given: $D = \{D_{s,a}\}_{(s,a)\in\mathcal{S}\times\mathcal{A}}$ and $\boxed{\phi}$

(s, a, r, s')

$\downarrow$

$\underbrace{(\phi(s), a, r, \phi(s'))}_{S_\phi}$.

- 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?

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

  - At the mercy of data to be exploratory (as always for batch RL)

The learning setting

- Analysis varies according to whether ϕ is (approx.) bisimulation or Q*-irrelevant and the style ($\max_\pi \|V^\pi_M - V^\pi_{\widehat{M}}\|_\infty$ vs $\|Q^\star_M - Q^\star_{\widehat{M}}\|$)
- Will show analysis of Q*-irrelevance (can only use "$\|Q^\star_M - Q^\star_{\widehat{M}}\|$")
- Let $\boxed{\widehat{M}_\phi}$ be the estimated model over S_ϕ.
- Let M_ϕ be an abstract model w/ weighting distributions $\boxed{p_x(s) \propto |D_{s,a}|}$
- M_ϕ is the "expected model" of $\widehat{M}_\phi$

alg output.

$\lim_{|D|\to\infty} \widehat{M}_\phi = M_\phi$

- $$\left\|Q^\star_M - [Q^\star_{\widehat{M}_\phi}]_M\right\|_\infty \le \underbrace{\left\|Q^\star_M - [Q^\star_{M_\phi}]_M\right\|_\infty}_{\text{Approximation error}} + \underbrace{\left\|\boxed{[Q^\star_{M_\phi}]_M} - [Q^\star_{\widehat{M}_\phi}]_M\right\|_\infty}_{\text{Estimation error}}$$

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

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

- 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

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

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

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

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

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

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

$\mathcal{F}_\phi \subseteq [0, V_{max}]^{S \times A}$ is the space of func. piecewise const under ϕ.

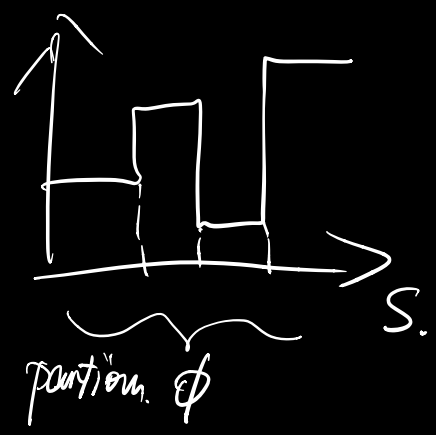
$\hookrightarrow \frac{R_{max}}{1-\gamma}$.

$\forall f \in \mathcal{F}_\phi$.

$(\mathcal{T}f)(s^{(1)}, a) = R(s^{(1)}, a) + \gamma \langle \boxed{P(\cdot \mid s^{(1)}, a)}, V_f \rangle$

$\| \qquad \downarrow$

$(\mathcal{T}f)(s^{(2)}, a) = \cdots$



$V_f := \arg\max_a f(\cdot, a)$.

$\mathcal{T}f$ is also piecewise const.

$(\mathcal{T}f) \in \mathcal{F}_\phi$.

(Thm): Let ϕ be abstraction and $\mathcal{F}_\phi \cdots$

Claim: (1) ϕ is bimulation $\Rightarrow$ $\boxed{\forall f \in \mathcal{F}_\phi,\ \mathcal{T}f \in \mathcal{F}_\phi}$.

(2) $\Leftarrow ?$