# State Abstractions

tabular $\quad Q(s,a) \quad |S| \times |A|$

$\downarrow$

function approximation

$Q(s,a) = Q_\theta(x)$

$\theta \in \mathbb{R}^d$

embedding

large state space $|S|$

$M = (S, A, P, R, \gamma)$

## State Abstractions / State Aggregation / State Compression

Abstraction: $\phi : S \to S_\phi$ $\langle$ $S$: original / primitive / raw space.

$S_\phi$: finite abstract space.

1) $\phi$ surjection

2) $\phi(s^{(1)}) = \phi(s^{(2)})$ $\Rightarrow$ treat $s^{(1)}, s^{(2)}$ as the same state

$\Downarrow$

equivalence class

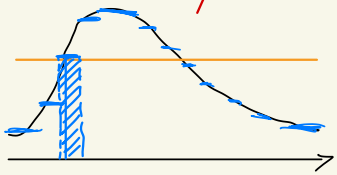
- reflexivity
- symmetry
- transitivity

3) Partition the state space into equivalence classes

4) Piecewise-Constant function approximation

5) Certainty Equivalence:

$(s, a, r, s') \to D_{s,a}, \quad |D_{s,a}|$

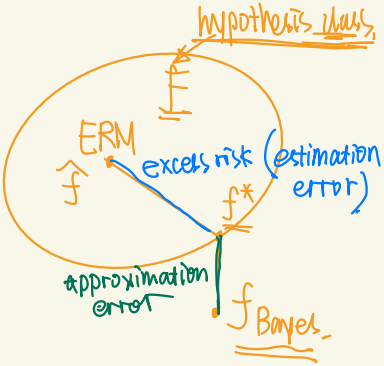


$\Rightarrow (\phi(s), a, r, \phi(s')) \to D_{x,a} = \bigcup_{s:\phi(s)=x} D_{s,a}$

$\Rightarrow$ effective sample size $\uparrow$ $\quad \Rightarrow$ estimation error $\downarrow$

6) Caveat: if we aggregate states that are very different from each other, then we do not converge to optimal policy even in the limit of infinite data.

$\Rightarrow$ approximation error




7) Trade-off: estimation - approximation
(variance) (bias)

8) How to build state abstraction $\phi$ ?

- optimal action ; $Q^*$ value ; dynamics and rewards.

## Abstraction Hierarchy

$M = (S, A, P, R, \gamma)$

An abstraction $\phi$ is ... if ... $\forall s^{(1)}, s^{(2)}$ s.t. $\phi(s^{(1)}) = \phi(s^{(2)})$

① $\pi^*$- irrelevant: $\exists \pi_M^*$ s.t. $\pi_M^*(s^{(1)}) = \pi_M^*(s^{(2)})$

② $Q^*$- irrelevant: $\forall a \in A$. $Q_M^*(s^{(1)}, a) = Q_M^*(s^{(2)}, a)$

③ Model-irrelevant: $\forall a \in A$. $R(s^{(1)}, a) = R(s^{(2)}, a)$.

(bisimulation) $\forall a \in A. \forall x' \in S_\phi$: $\sum_{s' \in \phi^{-1}(x')} P(s' | s^{(1)}, a) = \sum_{s' \in \phi^{-1}(x')} P(s' | s^{(2)}, a)$

1. Define various notions / criteria of state abstractions.

2. Study their relationships

3. Analyze how to use them (e.g. build an abstract model in planning & learning)

Bisimulation: $\forall a \in A.\ R(s^{(1)}, a) = R(s^{(2)}, a)$

$$\forall a \in A.\ \forall x' \in S_\phi:\ \sum_{s' \in \phi^{-1}(x')} P(s' \mid s^{(1)}, a) = \sum_{s' \in \phi^{-1}(x')} P(s' \mid s^{(2)}, a)$$

$$\Longleftrightarrow \boxed{P(x' \mid s^{(1)}, a) = P(x' \mid s^{(2)}, a)} \quad \triangle$$

$$\boxed{\Phi \in \mathbb{R}^{|S_\phi| \times |S|}}$$

operator

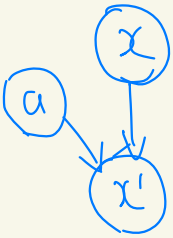
$$\boxed{\Phi(x, s) = \mathbb{1}\{\phi(s) = x\}}$$

$$\Rightarrow \Phi P(s, a) = P(\cdot \mid s, a)$$

in $S_\phi$

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

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




MDP M

Markov Chain C

State-dependent transition ?

NOT i.i.d process

MDP $M' = M \times C$ → state space $\mathcal{X} \times \mathcal{Z}$. $(x, z)$

transition. $P_{M'}\left((x', z') \mid (x, z), a\right) = P_M\left(x' \mid x, z\right)$

$\phi : (x, z) \mapsto x$

$\Rightarrow P_{M'}\left((x', z') \mid (x, z), a\right) = P_M\left(x' \mid x, a\right) \cdot P_C\left(z' \mid z\right)$

same! $(x, z^{(1)}), (x, z^{(2)})$ ×

integrated out of bisimulation.

## Coarsest Bisimulation

**Claim**: a minimal bisimulation uniquely exists.

Proof: for any bisimulation $\phi_1, \phi_2$ of $M$.

their common coarsening is also a bisimulation. $\phi_{12}$

(two states are aggregated if they are aggregated under either $\phi_1, \phi_2$)

Assume $\phi_1(s^{(1)}) = \phi_2(s^{(1)})$ $\forall y' \in \phi_{12}(S)$

$$P(y' \mid s^{(1)}, a) = \sum_{s' \in \phi_{12}^{-1}(y')} P(s' \mid s^{(1)}, a)$$

$s' \in S$

$x' \in \phi_1(S)$

$y' \in \phi_{12}(S)$

$$= \sum_{x' \in \phi_1(\phi_{12}^{-1}(y'))} \sum_{s' \in \phi_1^{-1}(x') \cancel{\cap \phi_{12}^{-1}(y')}} P(s' \mid s^{(1)}, a)$$

$$= \sum_{x' \in \phi_1(\phi_2^{-1}(y'))} \underbrace{\sum_{s' \in \phi_1^{-1}(x')} P(s' \mid s^{(2)}, a)}_{\phi_1 : \text{bisimulation}}$$

$$= P(y' \mid s^{(2)}, a) \qquad \square_1$$

Note: this minimal bisimulation is NP-hard to find.

---

**Recall**: state abstractions $\phi : S \to S_\phi$ $\qquad (s, a) \mapsto (\phi(s), a)$

- piecewise-constant function approximation
  - $\mathcal{F}_\phi$ $\qquad \phi(s^{(1)}) = \phi(s^{(2)}) \Rightarrow s^{(1)}, s^{(2)}$ same embedding in RKHS $\mathcal{F}_\phi$.
- equivalence class
- approximation error / estimation error $\to$ trade-off.
- $\pi^*$-irrelevant: $\pi_M^*(s^{(1)}) = \pi_M^*(s^{(2)})$

  $Q^*$-irrelevant: $Q_M^*(s^{(1)}, a) = Q_M^*(s^{(2)}, a)$

  Model-irrelevant (bisimulation): $R(s^{(1)}, a) = R(s^{(2)}, a)$, $\Phi P(s^{(1)}, a) = \Phi P(s^{(2)}, a)$

lenient $\uparrow$ strict

This video: Model-irrelevant $\Rightarrow$ $Q^*$-irrelevant $\Rightarrow$ $\pi^*$-irrelevant.

$\pi_M^*(s) = \arg\max_{a \in A} Q_M^*(s, a)$

$Q^*$-irrelevant $\to$ realizability ($Q^* \in \mathcal{F}$)

Model-irrelevant $\to$ Bellman completeness ($\forall f \in \mathcal{F}.\ \mathcal{T} f \in \mathcal{F}$)

# Bisimulation

$M \leftrightarrow \hat{M}$

$M \leftrightarrow M_\phi$

$M = (S, A, P, R, \gamma)$

$\Downarrow$

abstract MDP $M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$

· for any $x \in S_\phi$, $a \in A$, $x' \in S_\phi$. arbitrary $P_x$.

$$R_\phi(x, a) = \mathbb{E}_{s \sim P_x}\left[R(s, a)\right] \text{ for all } s \in \phi^{-1}(x)$$

$$P_\phi(x' \mid x, a) = \mathbb{E}_{s \sim P_x}\left[P(x' \mid s, a)\right] \text{ for all } s \in \phi^{-1}(x)$$

e.g. learning

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

---

**def** (Lifting) $f$ : any function operates on $S_\phi$

$[f]_M$ : lifted version

$$\begin{cases} V: & [f]_M(s) = f(\phi(s)) \\ Q: & [f]_M(s, a) = f(\phi(s), a) \end{cases}$$

$[f]_M = \Phi^T f$

$\pi : S_\phi \to A$.

$[\pi]_M : S \to A$ s.t. $[\pi]_M(s) = \pi(\phi(s))$

$$\| V_M^* - V_M^{[\pi_{M_\phi}^*]_M} \|_\infty, \quad \| Q_M^* - [Q_{M_\phi}^*]_M \|_\infty$$

# Approximate Abstractions

$M = (S, A, P, R, \gamma)$. $\phi$.

1) $\phi$ is an $\varepsilon_{\pi^*}$-approximate $\pi^*$-irrelevant abstraction

if $\exists\, \pi : S_\phi \to A$ s.t. $\| V_M^* - V_M^{[\pi]_M} \|_\infty \le \varepsilon_{\pi^*}$.

2) $\phi$ is an $\varepsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction.

if $\exists\, f : S_\phi \times A \to \mathbb{R}$ s.t. $\| [f]_M - Q_M^* \|_\infty \le \varepsilon_{Q^*}$.

3) $\phi$ is an $(\varepsilon_P, \varepsilon_Q)$-approximate model-irrelevant abstraction.

if $\forall\, s^{(1)}, s^{(2)} \in S$ s.t. $\phi(s^{(1)}) = \phi(s^{(2)})$. $\forall\, a \in A$:

$|R(s^{(1)}, a) - R(s^{(2)}, a)| \le \varepsilon_R$ $\quad \| \Phi P(s^{(1)}, a) - \Phi P(s^{(2)}, a) \|_1 \le \varepsilon_P$.

Hierarchy : Model-irrelevant $\Rightarrow$ $Q^*$-irrelevant $\Rightarrow$ $\pi^*$-irrelevant.

Theorem 1) If $\phi$ is an $\varepsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction,

then $\phi$ is also an $\varepsilon_{\pi^*}$-approximate $\pi^*$-irrelevant abstraction

with $\varepsilon_{\pi^*} = \dfrac{2\varepsilon_{Q^*}}{1-\gamma}$

$$\| V_M^{\pi_f} - V_M^* \|_\infty \le \frac{2}{1-\gamma} \cdot \| f - Q_M^* \|_\infty$$

where $f$ is well-defined by $Q^*$-irrelevance.

(2) If $\phi$ is an $(\varepsilon_P, \varepsilon_R)$-approximate bisimulation.

then $\phi$ is also an $\varepsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction

with $\varepsilon_{Q^*} = \frac{\varepsilon_R}{1-\gamma} + \gamma \cdot \frac{\varepsilon_P}{2(1-\gamma)} \cdot V_{max}$

Similar to bound of simulation lemma

$M \leftrightarrow \hat{M}$

$M \leftrightarrow M_\phi$

bisimulation $\phi \Rightarrow \boxed{M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)}$

$$P_\phi(x'|x,a) = \mathbb{E}_{s\sim p_x}[P(x'|s,a)]$$

$$R_\phi(x,a) = \mathbb{E}_{s\sim p_x}[R(s,a)]$$

$\forall$ probability distribution $p_x$ over $s \in \phi^{-1}(x)$

Lemma ("~~estimation~~ approximation error")

$$\begin{cases} |R_\phi(\phi(s),a) - R(s,a)| \le \varepsilon_R \\ \| P_\phi(\phi(s),a) - \Phi \cdot P(s,a) \|_1 \le \varepsilon_P \end{cases}$$

Proof: reward trivial

by $|R(s^{(1)},a) - R(s^{(2)},a)| \le \varepsilon_R$, any $\phi(s^{(1)}) = \phi(s^{(2)})$

Let $q_s \triangleq \Phi \cdot P(s,a)$ $\Rightarrow \|q_{s^{(1)}} - q_{s^{(2)}}\| \le \varepsilon_P$. any. $\phi(s^{(1)}) = \phi(s^{(2)})$

$$\Rightarrow \| P_\phi(\underbrace{\phi(s)}_{x}, a) - \Phi P(s,a) \|_1$$

$$= \| \sum_{\tilde{s} \in \phi^{-1}(x)} p_x(\tilde{s}) \cdot (q_{\tilde{s}} - q_s) \|_1$$

$|a+b| \le |a| + |b|$

$$\le \sum_{\tilde{s}} \| p_x(\tilde{s}) \cdot (q_{\tilde{s}} - q_s) \|_1$$

$$\leq \sum_{\tilde{s}} P_x(\tilde{s}) \cdot \varepsilon_P = \varepsilon_P. \qquad \square.$$

Proof of $\varepsilon_{Q^*} \leq \frac{\varepsilon_R}{1-\gamma} + \frac{\gamma}{2(1-\gamma)} \cdot \varepsilon_P \cdot V_{max}$

$$\exists\, f : S_\phi \times A \to \mathbb{R} \text{ s.t. } \| [f]_M - Q^*_M \|_\infty \leq \varepsilon_{Q^*}.$$

Let $f = Q^*_{M_\phi} \Rightarrow \| [Q^*_{M_\phi}]_M - Q^*_M \|_\infty$

$$\leq \frac{1}{1-\gamma} \cdot \| [Q^*_{M_\phi}]_M - \mathcal{T}_M [Q^*_{M_\phi}]_M \|_\infty$$

Bellman error

$$\boxed{\|f - Q^*\|_\infty \leq \frac{1}{1-\gamma} \cdot \|f - \mathcal{T} f\|_\infty}$$

$$= \frac{1}{1-\gamma} \cdot \| [\mathcal{T}_{M_\phi} Q^*_{M_\phi}]_M - \mathcal{T}_M [Q^*_{M_\phi}]_M \|_\infty$$

$\forall (s,a)$:

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

$$= \left| (\mathcal{T}_{M_\phi} Q^*_{M_\phi})(\phi(s), a) - (\mathcal{T}_M [Q^*_{M_\phi}]_M)(s,a) \right|$$

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

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

$$\Phi^T V^*_{M_\phi}$$

$$= \varepsilon_R + \gamma \cdot \left| \langle P_\phi(\phi(s), a), V^*_{M_\phi} \rangle - \langle \Phi P(s,a), V^*_{M_\phi} \rangle \right|$$

$$\leq \varepsilon_R + \gamma \cdot \underbrace{\| P_\phi(\phi(s), a) - \Phi \cdot P(s,a) \|_1}_{} \cdot \underbrace{\| V^*_{M_\phi} \|_\infty}_{} \qquad V^*_{M_\phi} - V_{max} \cdot \vec{1}$$

(by Hölder)

$$\leq \varepsilon_R + \gamma \cdot \varepsilon_P \cdot V_{max}$$

$$\Rightarrow \| [Q^*_{M_\phi}]_M - Q^*_M \|_\infty \leq \frac{1}{1-\gamma} \cdot (\varepsilon_R + \gamma \cdot \varepsilon_P \cdot V_{max}) \qquad \square$$

---

<u>Recall</u>: state abstractions $\phi$

1. Define various notions / creteria of state abstractions.
2. Study their relationships

---

3. Analyze how to use them (e.g. build an abstract model) in planning & learning

bisimulation $\Rightarrow$ $Q^*$-irrelevant $\Rightarrow$ $\pi^*$-irrelevant.

① $(\varepsilon_P, \varepsilon_R) \Rightarrow \varepsilon_{Q^*} \leq \frac{\varepsilon_R}{1-\gamma} + \frac{\gamma}{1-\gamma} \varepsilon_P \cdot V_{max}$

② $\varepsilon_{Q^*} \Rightarrow \varepsilon_{\pi^*} \leq \frac{2}{1-\gamma} \cdot \varepsilon_{Q^*}$

approximation error

$\|V^* - V^{\pi_\phi}\|_\infty \leq \frac{2\|Q^* - \phi\|_\infty}{1-\gamma}$

$\Phi$ (representation)

realizability : $Q^* \in \mathcal{F}$. $\rightarrow$ $\mathcal{F}$ richer ✓

completeness : $f \in \mathcal{F}$. $\mathcal{T} f \in \mathcal{F}$. $\rightarrow$ $\mathcal{F}$ richer ✗

---

# Bounding the loss of abstract model

## $\phi$ is approximate bisimulation

$$M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$$

$P_x$

$$\| V_M^{[\pi^*_{M_\phi}]_M} - V_M^* \|_\infty \le \frac{2}{1-\gamma}\left( \frac{\varepsilon_R}{1-\gamma} + \frac{\gamma}{1-\gamma} \cdot \varepsilon_p \cdot V_{max} \right)$$

$\frac{1}{1-\gamma}$ : cubic dependency.

Lifting abstract $\pi$ or $f$ :

$[\pi]_M(s) = \pi(\phi(s))$

$[f]_M(s) = f(\phi(s))$

Thm (tighter bound)

$$\| V_M^* - V_M^{[\pi^*_{M_\phi}]_M} \|_\infty \le \frac{2}{1-\gamma} \left( \varepsilon_R + \gamma \cdot \varepsilon_p \cdot V_{max} \right)$$

V-gap $\xrightarrow{\frac{1}{1-\gamma}}$ Q-gap $\xrightarrow{\frac{1}{1-\gamma}}$ Approx. $\left(\frac{1}{1-\gamma}\right)$

Proof: for any abstract policy $\pi$ : $S_\phi \to A$

Claim: $\max_\pi \| [V^\pi_{M_\phi}]_M - V_M^{[\pi]_M} \|_\infty \le \frac{1}{1-\gamma} \left( \varepsilon_R + \gamma \cdot \varepsilon_p \cdot V_{max} \right)$

$$V_M^*(s) - V_M^{[\pi^*_{M_\phi}]_M}(s) = V_M^*(s) - V^*_{M_\phi}(\phi(s)) + V^*_{M_\phi}(\phi(s)) - V_M^{[\pi^*_{M_\phi}]_M}(s)$$

$$\leq \underbrace{\| Q_M^* - [Q_{M\phi}^*]_M \|_\infty}_{Q^*\text{-irrelevance}} + \underbrace{\| [V_{M\phi}^{\pi_{M\phi}^*}]_M - V_M^{[\pi_{M\phi}^*]_M} \|_\infty}_{\text{claim.}}$$

$$\leq \frac{2}{1-\gamma} \cdot \left( \varepsilon_R + \gamma \cdot \varepsilon_p \cdot V_{max} \right).$$

Proof of Claim: $\| [V_{M\phi}^{\pi}]_M - V_M^{[\pi]_M} \|_\infty$

$$\leq \frac{1}{1-\gamma} \| [V_{M\phi}^{\pi}]_M - \mathcal{T}_M^{[\pi]_M} [V_{M\phi}^{\pi}]_M \|_\infty$$

$$= \frac{1}{1-\gamma} \| [\mathcal{T}_{M\phi}^{\pi} V_{M\phi}^{\pi}]_M - \mathcal{T}_M^{[\pi]_M} [V_{M\phi}^{\pi}]_M \|_\infty$$

for any $s \in S$: $\left| [\mathcal{T}_{M\phi}^{\pi} V_{M\phi}^{\pi}]_M (s) - (\mathcal{T}_M^{[\pi]_M} [V_{M\phi}^{\pi}]_M)(s) \right|$

$$= \left| (\mathcal{T}_{M\phi}^{\pi} V_{M\phi}^{\pi})(\phi(s)) - (\mathcal{T}_M^{[\pi]_M} [V_{M\phi}^{\pi}]_M)(s) \right|$$

$$= \left| R_\phi(\phi(s), \pi) + \gamma \cdot \langle P_\phi(\phi(s), \pi), V_{M\phi}^{\pi} \rangle - R(s, [\pi]_M) - \gamma \cdot \langle P(s, [\pi]_M), [V_{M\phi}^{\pi}]_M \rangle \right|$$

$$\leq \varepsilon_R + \gamma \cdot \left| \langle P_\phi(\phi(s), \pi), V_{M\phi}^{\pi} \rangle - \langle \Phi \cdot P(s, [\pi]_M), V_{M\phi}^{\pi} \rangle \right|$$

$$= \varepsilon_R + \gamma \cdot \left| \langle P_\phi(\phi(s), \pi) - \Phi \cdot P(s, [\pi]_M), V_{M\phi}^{\pi} \rangle \right|$$

$\leq \varepsilon_R + \gamma \cdot \varepsilon_P \cdot V_{max}$ $\square$

---

$\phi$ is approximate $Q^*$-irrelevant.

$$Q^*_{M_\phi} \approx Q^*_M \qquad M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma) \qquad M$$

$P_x$

near-optimal policy

$\cancel{Q^*_{M_\phi}}$ $\cancel{\varepsilon_P}$ $\cancel{\varepsilon_R}$

Thm $\varepsilon_{Q^*}$: $\exists f : S_\phi \times A \to \mathbb{R}$ s.t. $\| [f]_M - Q^*_M \|_\infty \leq \varepsilon_{Q^*}$

$$\Rightarrow \| [Q^*_{M_\phi}]_M - Q^*_M \|_\infty \leq \frac{2\varepsilon_{Q^*}}{1-\gamma}.$$

① Exact $Q^*$-irrelevance $\Rightarrow$ $[Q^*_{M_\phi}]_M = Q^*_M$

despite the dynamics and rewards in $M_\phi$ do not make sense

$\phi(s^{(1)}) = \phi(s^{(2)})$ : $\boxed{Q^*_M(s^{(1)}, a) = Q^*_M(s^{(2)}, a)}$

$\Rightarrow R(s^{(1)}, a) + \gamma \cdot \langle P(s^{(1)}, a), V^*_M \rangle = R(s^{(2)}, a) + \gamma \cdot \langle P(s^{(2)}, a), V^*_M \rangle$

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

(inverse of lifting)

Note: $[\cdot]_\phi$ can only applied to piecewise constant functions under $\phi$.

$\Rightarrow$ $\boxed{[Q_M^*]_\phi \text{ is a fixed point of } \mathcal{T}_{M\phi}}$ (to be proved)

i.e. $\mathcal{T}_{M\phi}[Q_M^*]_\phi = [Q_M^*]_\phi$

$\Rightarrow$ $\underline{[Q_M^*]_\phi} = Q_{M\phi}^*$ is optimal in $M\phi$

$\xRightarrow{\text{Lifting}}$ $Q_M^* = [Q_{M\phi}^*]_M$ is optimal in $M$. ✓

$$\left(\mathcal{T}_{M\phi}[Q_M^*]_\phi\right)(x,a) = R_\phi(x,a) + \gamma \cdot \langle P_\phi(x,a), [V_M^*]_\phi \rangle$$

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

$$= \sum_{s \in \phi^{-1}(x)} P_x(s) \cdot \left( R(s,a) + \gamma \cdot \langle P(s,a), V_M^* \rangle \right)$$

$$= \sum_{s \in \phi^{-1}(x)} P_x(s) \cdot Q_M^*(s,a) = \cancel{\sum_{s \in \phi^{-1}(x)} P_x(s)} \cdot [Q_M^*]_\phi(x,a)$$

normalized

② Approximate $Q^*$-irrelevant: $Q_M^*$ is not piecewise constant under $\phi$.

$\rightarrow$ we cannot apply $\mathcal{T}_{M\phi}$.

$\star$ $M'_\phi = (S, A, P'_\phi, R'_\phi, \gamma)$. original state space!

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

$\hookrightarrow p_x$.

$$M \xleftrightarrow{M'_\phi} M_\phi$$

Key: 1) Optimal Q-function of $M'_\phi$ coincides with that of $M_\phi$ (under lifting)

2) We can apply $\mathcal{T}_M$ to $Q^*_{M'_\phi}$.

$\Leftrightarrow$ We can apply $\mathcal{T}_{M'_\phi}$ to $Q^*_M$ $\star$

Claim: $Q^*_{M'_\phi} = [Q^*_{M_\phi}]_M$.

Proof: $(\mathcal{T}_{M'_\phi}[Q^*_{M_\phi}]_M)(s,a)$

$$= R'_\phi(s,a) + \gamma \cdot \langle P'_\phi(s,a), [V^*_{M_\phi}]_M \rangle.$$

$$= \sum_{\tilde{s}:\phi(\tilde{s})=\phi(s)} p_x(\tilde{s}) \cdot \left( R(\tilde{s},a) + \gamma \cdot \langle P(\tilde{s},a), [V^*_{M_\phi}]_M \rangle \right)$$

$$= \sum_{\tilde{s}:\phi(\tilde{s})=\phi(s)} p_x(\tilde{s}) \cdot \left( R(\tilde{s},a) + \gamma \cdot \langle \Phi P(\tilde{s},a), V^*_{M_\phi} \rangle \right)$$

$$= \underline{R_\phi(\phi(s), a)} + \gamma \cdot \langle P_\phi(\phi(s), a), V^*_{M_\phi} \rangle$$

$$= Q^*_{M_\phi}(\phi(s), a) = [Q^*_{M_\phi}]_M(s, a) \qquad \square$$

Goal: $\| \underline{[Q^*_{M_\phi}]_M} - Q^*_M \|_\infty = \| \underline{Q^*_{M'_\phi}} - \underline{Q^*_M} \|_\infty$

$$\leq \frac{1}{1-\gamma} \cdot \| \underline{\mathcal{T}_{M'_\phi} Q^*_M - Q^*_M} \|_\infty$$

$$\left| (\mathcal{T}_{M'_\phi} Q^*_M)(s,a) - Q^*_M(s,a) \right|$$

$$= \left| \underline{R'_\phi(s,a) + \gamma \cdot \langle P'_\phi(s,a), V^*_M \rangle} - Q^*_M(s,a) \right|$$

$$= \left| \sum_{\tilde{s} : \phi(\tilde{s}) = \phi(s)} P_x(\tilde{s}) \left( \underline{R(\tilde{s}, a) + \gamma \cdot \langle P(\tilde{s}, a), V^*_M \rangle} \right) - Q^*_M(s,a) \right|$$

$$= \left| \sum_{\tilde{s} : \phi(\tilde{s}) = \phi(s)} P_x(\tilde{s}) \cdot \left( \underline{Q^*_M(\tilde{s}, a) - Q^*_M(s,a)} \right) \right|$$

$$\leq \left| \sum_{\tilde{s} : \phi(\tilde{s}) = \phi(s)} P_x(\tilde{s}) \cdot 2 \varepsilon_{Q^*} \right|$$

$$= 2 \varepsilon_{Q^*}$$

$\exists f : \| [f]_M - Q^*_M \|_\infty \leq \varepsilon_{Q^*}$

$$\left| Q^*_M(\underline{s^{(1)}, a}) - Q^*_M(\underline{s^{(2)}, a}) \right|$$

$$\leq \| Q^*_M - [f]_M \|_\infty + \| f - Q^*_M \|_\infty$$

$$= 2 \varepsilon_{Q^*}$$

(add-and-subtract)

---

With $Q^*$-irrelevance,

$$\max_\pi \underline{\| V^\pi_M - V^\pi_{\hat{M}} \|_\infty} \times \quad \| Q^\pi_M - Q^\pi_{\hat{M}} \|_\infty \checkmark$$

# Planning with Model Abstraction

Given weighting distributions $\{P_x\}$: $M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma)$

- If approx. bisimulation: $\max_\pi \| V_M^\pi - V_{\hat{M}}^\pi \|_\infty$ type analysis

$$\Rightarrow \| V_M^* - V_M^{[\pi^*_{M_\phi}]_M} \|_\infty \leq \frac{2\varepsilon_R}{1-\gamma} + \frac{\gamma \cdot \varepsilon_P}{1-\gamma} \cdot V_{max}$$

- If approx. $Q^*$-irrelevance: $\| Q_M^* - Q_{\hat{M}}^\pi \|_\infty$ type analysis

$$\Rightarrow \| V_M^* - V_M^{[\pi^*_{M_\phi}]_M} \|_\infty \leq \frac{2\varepsilon_{Q^*}}{(1-\gamma)^2}.$$

Exact Case

---

$M$ $\quad$ $M_\phi$

Consider VI in true model VS abstract model:

- bisimulation: every step of abstract VI resembles that step in true VI, throughout all iterations. $\forall f: S_\phi \to \mathbb{R}$, $\mathcal{T}[f]_M = [\mathcal{T}_{M_\phi} f]_M$
- $Q^*$-irrelevance: abstract VI initially behaves crazily. It only start to resemble true VI when the function is close to $Q^*$.

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

Stability

Abstract Bellman Update.

$\rightarrow$ Projected Bellman Update.

$\hookrightarrow$ can break stability

---

## Finite Sample Analysis. (learning setting)

① pre-processing data via state abstraction $\phi$

② certainty-equivalence RL.

$$D = \{D_{s,a}\}_{(s,a)\in S\times A} \rightarrow D_{x,a} = \bigcup_{s\in\phi^{-1}(x)} D_{s,a}$$

effective sample size $n_\phi(D) = \min_{\substack{x\in S_\phi \\ a\in A}} |D_{x,a}|$

Intuition:

loss bound $\rightarrow$ $n_\phi(D)$

$\rightarrow$ approximation error of $\phi$

$\rightarrow$ ~~$|S|$~~

Analysis: CE $\rightarrow$ $\max_\pi \|V_M^\pi - V_{\hat{M}}^\pi\|_\infty \rightarrow$ bisimulation

$\rightarrow$ $\|Q_M^* - Q_{\hat{M}}^*\|_\infty \rightarrow Q^*$-irrelevance

$$\hat{M}_\phi = (S_\phi, A, \hat{P}_\phi, \hat{R}_\phi, \gamma) \quad : \text{estimated model}$$

$$M_\phi = (S_\phi, A, P_\phi, R_\phi, \gamma) \quad \text{with} \quad P_x(s) = \frac{|D_{s,a}|}{|D_{\phi(s),a}|}$$

$$R_\phi(x,a) = \frac{\sum_{\tilde{s} \in \phi^{-1}(x)} |D_{\tilde{s},a}| \cdot R(\tilde{s},a)}{|D_{x,a}|}$$

$$P_\phi(x'|x,a) = \frac{\sum_{\tilde{s} \in \phi^{-1}(x)} |D_{\tilde{s},a}| \cdot P(x'|\tilde{s},a)}{|D_{x,a}|}$$

$M_\phi$ is the expectation of $\hat{M}_\phi$ w.r.t. the randomness in data.

$$\lim_{n_\phi(D) \to \infty} \hat{M}_\phi = M_\phi$$

bias-variance trade-off

Goal: $\| Q^*_M - [Q^*_{\hat{M}_\phi}]_M \|_\infty$

$$\leq \underbrace{\| Q^*_M - [Q^*_{M_\phi}]_M \|_\infty}_{\text{approximation error (bias)}} + \underbrace{\| [Q^*_{M_\phi}]_M - [Q^*_{\hat{M}_\phi}]_M \|_\infty}_{\text{estimation error (variance)}}$$

Approximation error $\downarrow$ $\Leftarrow$ finer $\phi$

Estimation error $\downarrow$ $\Leftarrow$ coarser $\phi$, more data.

excess risk $\begin{cases} \text{sample size} \\ \text{model complexity} \end{cases}$

$\Rightarrow$ Trade-off

$\rightarrow$ best balance depends on sample size.

$$\| Q^*_M - [Q^*_{M_\phi}]_M \|_\infty \le \frac{2}{1-\gamma} \cdot \varepsilon_{Q^*}.$$

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

$\star$ $D_{x,a} \leftarrow P_\phi(x,a)$

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

independent

not identical !!

We can apply Hoeffding's ineq. !

$$= \| Q^*_{M_\phi} - Q^*_{\hat{M}_\phi} \|_\infty \quad (\text{lifting})$$

$$\le \frac{1}{1-\gamma} \| Q^*_{M_\phi} - \mathcal{T}_{\hat{M}_\phi} Q^*_{M_\phi} \|_\infty$$

$$= \frac{1}{1-\gamma} \| \mathcal{T}_{M_\phi} Q^*_{M_\phi} - \mathcal{T}_{\hat{M}_\phi} Q^*_{M_\phi} \|_\infty$$

$$\left| (\mathcal{T}_{\hat{M}_\phi} Q^*_{M_\phi})(x,a) - (\mathcal{T}_{M_\phi} Q^*_{M_\phi})(x,a) \right|$$

$$= \Big| \hat{R}_\phi(x,a) + \gamma \cdot \langle \hat{P}_\phi(x,a), V^*_{M_\phi} \rangle$$

$$- R_\phi(x,a) - \gamma \cdot \langle P_\phi(x,a), V^*_{M_\phi} \rangle \Big|$$

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

$|D_{x,a}| \geq n_\phi(D)$

is the empirical estimate of

$\forall (x,a) \in S_\phi \times A.$

$\Rightarrow$ w.p. $\geq 1-\delta$ :

$$\| \mathcal{T}_{\hat{M}_\phi} Q^*_{M_\phi} - \mathcal{T}_{M_\phi} Q^*_{M_\phi} \|_\infty \leq V_{max} \cdot \sqrt{\frac{1}{2 n_\phi(D)} \log \frac{2 \cdot |S_\phi \times A|}{\delta}}$$

## Review of State Abstraction Theory :

1) piecewise-constant function approximation.

2) hierarchy : bisimulation $\Rightarrow$ $Q^*$-irrelevance

$\Rightarrow$ $\pi^*$-irrelevance

3). planning & learning

$\rightarrow$ $\max_\pi \| V^\pi_M - V^\pi_{\hat{M}} \|_\infty$

$\rightarrow$ Bellman Completeness

$\rightarrow$ $\| Q^*_M - Q^*_{\hat{M}} \|_\infty$

$\rightarrow$ realizability

4). error decomposition

approximation error

estimation error