State Abstractions
Notations and Setup

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

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

- $\pi^*$-irrelevant: $\exists \pi^*_M \text{ s.t. } \pi^*_M(s^{(1)}) = \pi^*_M(s^{(2)})$
- $Q^*$-irrelevant: $\forall a, Q^*_M(s^{(1)}, a) = Q^*_M(s^{(2)}, a)$
- Model-irrelevant: $\forall a \in A, \forall x' \in \phi(S), R(s^{(1)}, a) = R(s^{(2)}, a), P(x' | s^{(1)}, a) = P(x' | s^{(2)}, a)$

Theorem: Model-irrelevance $\Rightarrow Q^*$-irrelevance $\Rightarrow \pi^*$-irrelevance
Why not \( P(s' \mid s^{(1)}, a) = P(s' \mid s^{(2)}, a) \)?

\[ S = X \times Z. \]

MDP \( M \)  

Markov chain \( C \)

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

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

\[ P(x' \mid (x, z^{(i)}), a) = P(x' \mid (x, z^{(i)}), a). \]

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

integrated out by bisimulation
The abstract MDP implied by bisimulation

\[ \phi(s^{(1)}) = \phi(s^{(2)}) \]

\[ \phi \text{ is bisimulation: } R(s^{(1)}, a) = R(s^{(2)}, a), \quad P(x' \mid s^{(1)}, a) = P(x' \mid s^{(2)}, a) \]

- 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 \( s \in \phi^{-1}(x) \)
  - \( P_\phi(x' \mid x, a) = P(x' \mid s, a) \) for any \( s \in \phi^{-1}(x) \)
- No way to distinguish between the two routes:

\[ \{ (s, a, r, s') \} \]

\[ \{ (\phi(s), a, r, \phi(s')) \} \]
Implications of bisimulation

Proof outline:

• **Q*'-irrelevance**

  - Plan in $\mathcal{M}_{\phi}$ and get $Q^*_{\mathcal{M}_{\phi}}$ (dimension: $\phi(S) \times A$)

  - Lift $Q^*_{\mathcal{M}_{\phi}}$ from $\phi(S)$ to $S$ (populate aggregated states with the same value)

• Useful notation: $\Phi$ is a $|\phi(S)| \times |S|$ matrix, with

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

• Lifting a state-value function: $[V^*_{\mathcal{M}_{\phi}}]_M = \Phi^T V^*_{\mathcal{M}_{\phi}}$

• Collapsing the transition distribution: $\Phi P(s, a)$

• Claim: $[Q^*_{\mathcal{M}_{\phi}}]_M = Q^*_M$ (proof on board)

$\Rightarrow$ implies $Q^*_M$ is "piecewise const" under $\phi$. $\Rightarrow$ Q*'-irrelevance
\[\Phi = \begin{bmatrix} \Phi(x) = x \end{bmatrix} \in \mathbb{R}^{S \times S}, \quad S_\Phi = \Phi(s).\]

Bisimulation: \( \forall \, s^{(1)}, s^{(2)} \text{ where } \Phi(s^{(1)}) = \Phi(s^{(2)}). \)

\( R(s^{(1)}, a) = R(s^{(2)}, a) \quad \& \quad P(1|s^{(1)}, a) = P(1|s^{(2)}, a) \)

w.r.t.

\[ [Q_{\Phi}^*]_M = Q_{\Phi}^* \leq \| \cdot - Q_{\Phi}^* \| \]

\( S \times A \rightarrow \mathbb{R}, \text{ where } [Q_{\Phi}^*]_M(s, a) = Q_{\Phi}^*(\Phi(s), a). \)

It suffices to show:

\[ J_\Phi [Q_{\Phi}^*]_M = [Q_{\Phi}^*]_M. \]

\( \text{LHS } (s, a) = R(s, a) + \gamma < P(1|s, a), V_{\Phi}^* > \)

\( \text{RHS } (s, a) = R_\Phi(s, a) + \gamma < P_\Phi(1|\Phi(s), a), V_{\Phi}^* > \)

\[ < P(1|s, a), [V_{\Phi}^*]_M > = < P(1|s, a), \Phi^T V_{\Phi}^* > \]

\[ = < \Phi P(1|s, a), V_{\Phi}^* > \]

\[ [V_{\Phi}^*]_M \leq P(1|s, a). \]

\( \Phi(x, s) = \Pi(\Phi(s) = x). \)
Implications of bisimulation

• $Q^*$-irrelevance

• $Q_M^\pi$ is preserved for any $\pi$ lifted from an abstract policy

• Given any lifted $\pi$, 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. $\phi$ is an $\epsilon_{\pi^*}$-approximate $\pi^*$-irrelevant abstraction, if there exists an abstract policy $\pi : \phi(S) \rightarrow A$, such that $\|V^*_M - V^{[\pi]}_M\|_{\infty} \leq \epsilon_{\pi^*}$.

2. $\phi$ is an $\epsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction if there exists an abstract $Q$-value function $f : \phi(S) \times A \rightarrow \mathbb{R}$, such that $\|[f]_M - Q^*_M\|_{\infty} \leq \epsilon_{Q^*}$.

3. $\phi$ is an $(\epsilon_R, \epsilon_P)$-approximate model-irrelevant abstraction if for any $s^{(1)}$ and $s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)}), \forall a \in A$,

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

(3)
Theorem 2. (1) If $\phi$ is an $(\epsilon_R, \epsilon_P)$-approximate model-irrelevant abstraction, then $\phi$ is also an approximate $Q^*$-irrelevant abstraction with approximation error $\epsilon_{Q^*} = \frac{\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\text{max}}}{2(1-\gamma)^2}$.
(2) If $\phi$ is an $\epsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction, then $\phi$ is also an approximate $\pi^*$-irrelevant abstraction with approximation error $\epsilon_{\pi^*} = 2\epsilon_{Q^*}/(1-\gamma)$.

• (2) has been proved; remains to prove (1)

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

$\phi$ is an $\epsilon_{Q^*}$-approximate $Q^*$-irrelevant abstraction if there exists an abstract $Q$-value function $f : \phi(S) \times A \rightarrow \mathbb{R}$, such that $\|[f]_M - Q^*_M\|_\infty \leq \epsilon_{Q^*}$.

• 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 $\phi^{-1}(x)$

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

• $|R_\phi(\phi(s), a) - R(s, a)| \leq \epsilon_R, \quad |P_\phi(\phi(s), a) - \Phi P(s, a)| \leq \epsilon_P$.

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

\left\| [Q_{M_\phi}^*]_M - Q_M^* \right\|_\infty \leq \frac{1}{1 - \gamma} \left\| [Q_{M_\phi}^*]_M - \mathcal{T}[Q_{M_\phi}^*]_M \right\|_\infty = \frac{1}{1 - \gamma} \left\| [\mathcal{T}_{M_\phi} Q_{M_\phi}^*]_M - \mathcal{T}[Q_{M_\phi}^*]_M \right\|_\infty.

For any (s, a),

\begin{align*}
&\left| ([\mathcal{T}_{M_\phi} Q_{M_\phi}^*]_M)(s, a) - (\mathcal{T}[Q_{M_\phi}^*]_M)(s, a) \right|
= |(\mathcal{T}_{M_\phi} Q_{M_\phi}^*)(\phi(s), a) - (\mathcal{T}[Q_{M_\phi}^*]_M)(s, a)|
= |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|
\leq \varepsilon_R + \gamma |\langle P_\phi(\phi(s), a), V_{M_\phi}^* \rangle - \langle P(s, a), \Phi^T V_{M_\phi}^* \rangle|
= \varepsilon_R + \gamma |\langle P_\phi(\phi(s), a), V_{M_\phi}^* \rangle - \langle \Phi P(s, a), V_{M_\phi}^* \rangle|
\leq \varepsilon_R + \gamma \varepsilon_P \left\| V_{M_\phi}^* - \frac{R_{\max}}{2(1 - \gamma)} \mathbf{1} \right\|_\infty
\leq \varepsilon_R + \gamma \varepsilon_P R_{\max} / (2(1 - \gamma)).
\end{align*}
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 ($M_\phi$), and plan in $M_\phi$

• Want to know: if $\phi$ is not an exact bisimulation, how lossy is the resulting policy as a function of ($\epsilon_R$, $\epsilon_P$)?

• What if $\phi$ is only approximately $Q^*$-irrelevant? Is the abstract model still useful? Can we still bound loss as a function of $\epsilon_{Q^*}$?

• Learning setting?
Loss of $\left[\pi_{M,\phi}^*\right]_M^*$: approx. bisimulation

- Recall: $M_\phi$ 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: $\left\|V^*_M - V_{M,\phi}^{[\pi_{M,\phi}^*]_M}\right\|_\infty \leq \frac{2\varepsilon_R}{(1-\gamma)^2} + \frac{\gamma\varepsilon_P R_{\max}}{(1-\gamma)^3}$
- Can improve: $\left\|V^*_M - V_{M,\phi}^{[\pi_{M,\phi}^*]_M}\right\|_\infty \leq \frac{2\varepsilon_R}{1-\gamma} + \frac{\gamma\varepsilon_P R_{\max}}{(1-\gamma)^2}$
- Idea: for any $\pi : \phi(S) \rightarrow A$, $\left\|V_{M,\phi}^{\pi} - V_{M}^{[\pi_{M,\phi}^*]_M}\right\|_\infty \leq \frac{\varepsilon_R}{1-\gamma} + \frac{\gamma\varepsilon_P R_{\max}}{2(1-\gamma)^2}$
- Finally,

$$V^*_M(s) - V_{M,\phi}^{[\pi_{M,\phi}^*]_M}(s) = V^*_M(s) - V_{M,\phi}^*(\phi(s)) + V_{M,\phi}^*(\phi(s)) - V_{M,\phi}^{[\pi_{M,\phi}^*]_M}(s)$$

$$\leq \left\|Q^*_M - [Q_{M,\phi}^*]_M\right\|_\infty + \left\|V_{M,\phi}^{\pi_{M,\phi}^*} - [V_{M,\phi}^{[\pi_{M,\phi}^*]_M}]_M\right\|_\infty$$

- Lesson: w/ approx. bisimulation, take the $\max_\pi \|V_{M,\phi}^{\pi} - V_{M,\phi}^{[\pi_{M,\phi}^*]_M}\|_\infty$ route instead of the $\|Q^*_M - Q_{M,\phi}^*\|_\infty$ route to save dependence on horizon
Loss of \([\pi_{M\phi}^*]_M\) : approx. Q*-irrelevance

- \(M\phi\) well defined, but transitions/rewards don’t make sense
- Can still show: \(\|[Q_{M\phi}^*]_M - Q_M^*\|_\infty \leq 2\epsilon_{Q^*}/(1 - \gamma)\)
- Exact case (RHS=0): \(\forall s^{(1)}, s^{(2)}\) where \(\phi(s^{(1)}) = \phi(s^{(2)})\)

\[
R(s^{(1)}, a) + \gamma\langle P(s^{(1)}, a), V_M^*\rangle = Q^*(s^{(1)}, a) = Q^*(s^{(2)}, a) = R(s^{(2)}, a) + \gamma\langle P(s^{(2)}, a), V_M^*\rangle
\]

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

So:

\[
(T_{M\phi}[Q_M^*]_\phi)(x, a) = R_\phi(x, a) + \gamma\langle P_\phi(x, a), [V_M^*]_\phi\rangle
= \sum_{s \in \phi^{-1}(x)} p_x(s) (R(s, a) + \gamma\langle \Phi P(s, a), [V_M^*]_\phi\rangle)
= \sum_{s \in \phi^{-1}(x)} p_x(s) (R(s, a) + \gamma\langle P(s, a), V_M^*\rangle)
= \sum_{s \in \phi^{-1}(x)} p_x(s) [Q_M^*]_\phi(x, a) = [Q_M^*]_\phi(x, a).
\]

\[
\phi(s^{(1)}) = \phi(s^{(2)}) \Rightarrow (T_f)(s^{(1)}, a) \neq (T_f)(s^{(2)}, a)
\]

but, “=” if \(f = Q_M^*\)
Loss of $[\pi^*_M]_M$ : approx. Q*-irrelevance

- Approximate case: proof breaks as $Q^*_M$ not piece-wise constant
- Workaround: define a new model $M'_\phi$ over $S$
  $$R'_\phi(s, a) = \mathbb{E}_{\tilde{s} \sim p_{\phi(s)}}[R(\tilde{s}, a)], \quad P'_\phi(s'|s, a) = \mathbb{E}_{\tilde{s} \sim p_{\phi(s)}}[P(s'|\tilde{s}, a)].$$
- Can show: $M_\phi$ and $M'_\phi$ share the same $Q^*$ (up to lifting)
  $$\| [Q^*_{M_\phi}]_M - Q^*_M \|_\infty = \| Q^*_{M'_\phi} - Q^*_M \|_\infty \leq \frac{1}{1 - \gamma} \| T_{M'_\phi} Q^*_M - Q^*_M \|_\infty$$

$$\left| (T_{M'_\phi} Q^*_M)(s, a) - Q^*_M(s, a) \right|$$

$$\leq \left| Q^*_M(s, a) \right|$$

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

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

$$= \left| \sum_{\tilde{s} : \phi(\tilde{s}) = \phi(s)} p_x(\tilde{s}) (Q^*_M(\tilde{s}, a) - Q^*_M(s, a)) \right| \leq \left| \sum_{\tilde{s} : \phi(\tilde{s}) = \phi(s)} p_x(\tilde{s}) (2\epsilon_{Q^*}) \right| = 2\epsilon_{Q^*}.$$
Loss of $[\pi_{M\phi}^*]_M$: approx. $Q^*$-irrelevance

- Lesson: with $Q^*$-irrelevance, the $\max_{\pi} \|V_{M}^\pi - V_{M}^{\pi}\|_\infty$ approach is not available; $\|Q_{M}^* - Q_{M}^{*}\|$ is the only choice

- If $\phi$ does not respect transition/reward, our analysis does not have to either!
Recap

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

- Given weighting distributions \( \{p_x\} \), define \( M_\phi = (\phi(S), A, P_\phi, R_\phi, \gamma) \)
  \( R_\phi(x, a) = \sum_{s \in \phi^{-1}(x)} p_x(s) R(s, a), \quad P_\phi(x, a) = \sum_{s \in \phi^{-1}(x)} p_x(s) \Phi P(s, a). \)

- How lossy is it to plan in \( M_\phi \) and lift back to \( M \)?
  - If approx. bisimulation, use “\( \max_{\pi} \|V^\pi_M - V^\pi_{\bar{M}}\|_\infty \)” type analysis
    \[
    \left\| V^*_M - V^{[\pi_{\bar{M}_\phi}]_M}_M \right\|_\infty \leq \frac{2\epsilon_R}{1-\gamma} + \frac{\gamma \epsilon_P R_{\max}}{(1-\gamma)^2} \]

  - If approx. \( Q^* \)-irrelevance, use “\( \|Q^*_M - Q^*_{\bar{M}}\| \)” type analysis
    \[
    \left\| V^*_M - V^{[\pi_{\bar{M}_\phi}]_M}_M \right\|_\infty \leq \frac{2\epsilon_{Q^*}}{(1-\gamma)^2} \]
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

\[ \forall f : \phi(S) \rightarrow \mathbb{R}, \quad \mathcal{T}_M[f]_M = [\mathcal{T}_{M\phi} f]_M \]

• Q*-irrelevance: abstract VI initially behaves crazily. It only starts to resemble true VI when the function is close to \( Q^*_M \)

\[ Q^*_M = [\mathcal{T} Q^*_M]_M = \mathcal{T}_{M\phi}[Q^*_M]\]

• 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

Gordon ’95
The Learning Setting and Finite Sample Analysis
The learning setting

\[ D_{s,a} = \{ (r, s') : r = R(s, a), s' \sim P(s, a) \} \]

- Given: \( D = \{ D_{s,a} \}_{(s,a) \in S \times A} \) and \( \phi \)
- Algorithm: CE after processing data w/ \( \phi \)
- 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?
  \[ n_{\phi}(D) := \min_{x \in \phi(S), a \in A} |D_{x,a}|, \quad \text{where} \quad |D_{x,a}| := \sum_{s \in \phi^{-1}(x)} |\Delta_{s,a}|. \]
- At the mercy of data to be exploratory (as always for batch RL)
The learning setting

- Analysis varies according to whether $\phi$ is (approx.) bisimulation or $Q^*$-irrelevant and the style ($\max_\pi \|V^\pi_M - V^\pi_M\|_\infty$ vs $\|Q^*_M - Q^*_M\|_\infty$) \leftarrow
- Will show analysis of $Q^*$-irrelevance (can only use $\|Q^*_M - Q^*_M\|$)
- Let $\hat{M}_\phi$ be the estimated model over $S_\phi$.
- Let $M_\phi$ be an abstract model w/ weighting distributions $p_x(s) \propto |D_{s,a}|$
- $M_\phi$ is the "expected model" of $\hat{M}_\phi$
- $\|Q^*_M - [Q^*_{\hat{M}_\phi}]_M\|_\infty \leq \|Q^*_M - [Q^*_{\hat{M}_\phi}]_M\|_\infty + \|[Q^*_{\hat{M}_\phi}]_M - [Q^*_{\hat{M}_\phi}]_M\|_\infty$

Approximation error
- "Bias", informally
- Doesn’t vanish with more data
- Smaller with a finer $\phi$
  (not w/ bisimulation; we will see why...)

Estimation error
- "Variance", informally
- Goes to 0 w/ infinite data
- Smaller with a coarser $\phi$
\[ \| Q^*_M - [Q^*_{M\phi}]_M \|_\infty \leq \| Q^*_M - [Q^*_{M\phi}]_M \|_\infty + \| [Q^*_{M\phi}]_M - [Q^*_{M\phi}]_M \|_\infty \]

already handled

to be analyzed

- Reusing the analysis for \( \| Q^*_M - Q^*_M \| \)
- Challenge: data is not generated from \( M_\phi \)
- Leverage the fact that Hoeffding can be applied to r.v.’s with non-identical distributions

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

\[
| (\mathcal{T}_{M\phi} Q^*_{M\phi})(x, a) - (\mathcal{T}_{M\phi} Q^*_{M\phi})(x, a) | \\
= | \widehat{R}_{\phi}(x, a) + \gamma \langle \widehat{P}_{\phi}(x, a), V^*_{M\phi} \rangle - R_{\phi}(x, a) - \gamma \langle P_{\phi}(x, a), V^*_{M\phi} \rangle | \\
= \left| \frac{1}{|D_{x,a}|} \sum_{s \in \phi^{-1}(x)} \sum_{(r,s') \in D_{s,a}} \left( r + \gamma V^*_{M\phi}(\phi(s')) - R(s, a) - \gamma \langle P(s, a), [V^*_{M\phi}]_M \rangle \right) \right|
\]
\( \forall s^{(1)}, s^{(2)} \in S, R(s^{(1)}, a) = R(s^{(2)}, a), \Phi P(s'' | s^{(1)}, a) = \Phi P(s'' | s^{(2)}, a) \)

\[ F_\phi \subseteq [0, V_{\text{max}}]^{s \times A} \text{ is the space of func. piecewise const under } \phi. \]

\[ \forall f \in F_\phi. \]

\[ (Tf)(s^{(1)}, a) = R(s^{(1)}, a) + \delta \left\{ P(s^{(1)}, a) \right\}_{V_f} \]

\[ (Tf)(s^{(2)}, a) = \]

\( Tf \) is also piecewise const.

\[ (Tf) \in F_\phi. \]

(Thm): Let \( \phi \) be abstraction and \( F_\phi \)

Claim (1) \( \Phi \) is bimulation \( \Rightarrow \forall f \in F_\phi, Tf \in F_\phi. \)