State Abstractions

What are abstractions and why study them?

- When we use more sophisticated function approximation, we are always generalizing the knowledge learned from one state to other similar states.
  - When is such generalization valid? **What states can be considered as “similar”?**
- To answer these questions, it is worth studying the simplest form of generalization: abstractions
- State abstractions ≈ aggregate equivalent (or similar) states and run tabular algorithms

Examples of state abstractions

- Multiple ways of expressing an abstraction
  - Mapping ϕ from original (or *raw*) states to abstract states
  - Partition over the state space
  - An equivalence notion over raw states
- Example 1: discretize a continuous state space
  - Mapping from continuous state to the grid
  - Partition is obvious
  - Two original states are equivalent if they fall in the same grid
- Example 2: Suppose the original state is described by some state variables $s = (x, y)$. $\phi(s) = x$ is an abstraction
  - mapping $\phi : (x, y) \mapsto x$
  - Partition over $\{(x, y)\}$
  - $s_1 = (x_1, y_1)$ is equiv to $s_2 = (x_2, y_2)$ iff $x_1 = x_2$ (i.e., $\phi(s_1) = \phi(s_2)$)

Notations and Formal 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?

Abstraction hierarchy

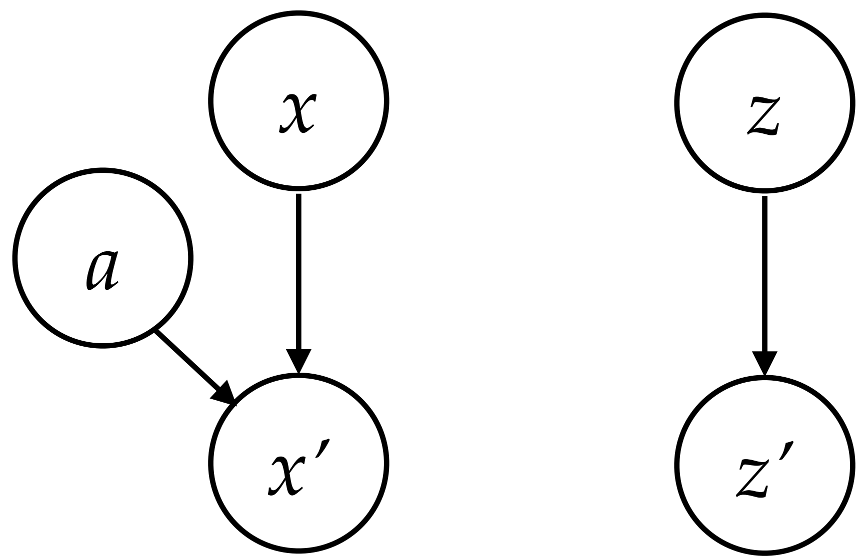
An abstraction ϕ is … if … $\forall\ s^{(1)}, s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$

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

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

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

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



MDP M Markov chain C

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

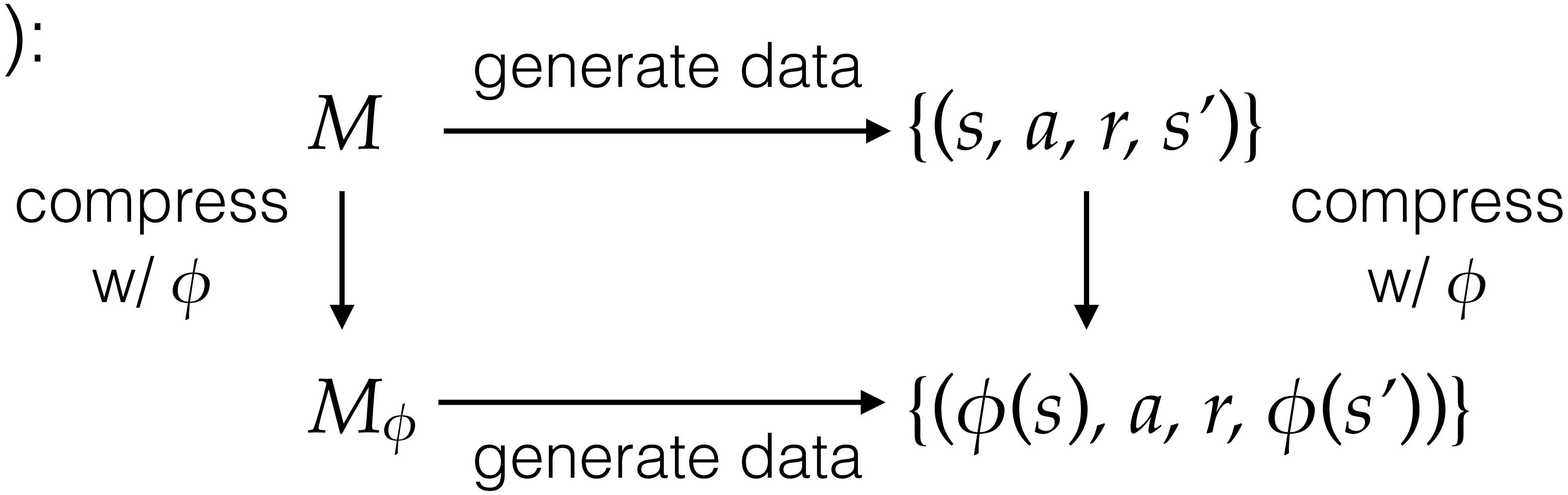
$$P((x', z') \mid (x, z), a) = P_M(x' \mid x, a) \cdot \underline{P_C(z' \mid z)}$$

integrated out by bisimulation

The abstract MDP implied by bisimulation

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 (if a only depends on $\phi(s)$):



Bisimulation => Q^*-irrelevance

- Consider the Q* in M_ϕ , $Q^\star_{M_\phi}$ (dimension: $|S_\phi \times A|$)
- *Lift* $Q^\star_{M_\phi}$ from S_ϕ to S (populate aggregated states with the same value)
- Useful notation: Φ is a $|S_\phi| \times |S|$ 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)

Useful/fun facts about bisimulation

- ${Q_M}^\pi$ is preserved *for any* π *lifted from an abstract policy*
- Given any lifted π, distribution over reward sequence is preserved (assuming reward is deterministic function of s, a)
- Coarsest bisimulation always exists: in any MDP, the common coarsening of two bisimulations is always a bisimulation
  - e.g., ϕ_1 tells you to ignore some state variables, ϕ_2 tells you to ignore some others => can ignore both sets of variables!
  - Intuitive but nontrivial; needs proof (see notes)

Useful/fun facts about bisimulation

- Recall that bisimulation is defined by a reward condition and a transition condition
- Guess what's the coarsest bisimulation if we drop the reward condition and only require the transition condition?
  - Aggregate all states together!
  - reward function defines a notion of (short-term) saliency
  - can extend the definition by replacing reward function with other functions (even not real-valued ones) whose codomain is equipped with an equivalence notion

The abstract model

- 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 (and lift the policy back to M)
- We already showed: if ϕ is bisimulation, this approach produces an optimal policy of M
- What if ϕ is Q^*-irrelevant? or π^*-irrelevant?
- π^*-irrelevant: learned policy can be suboptimal (see refs in Li et al'06)
- Q^*-irrelevant: surprisingly, optimality is preserved; for details and further reading, see ref notes.

Extension to handle action aggregation/permutation: Homomorphisms

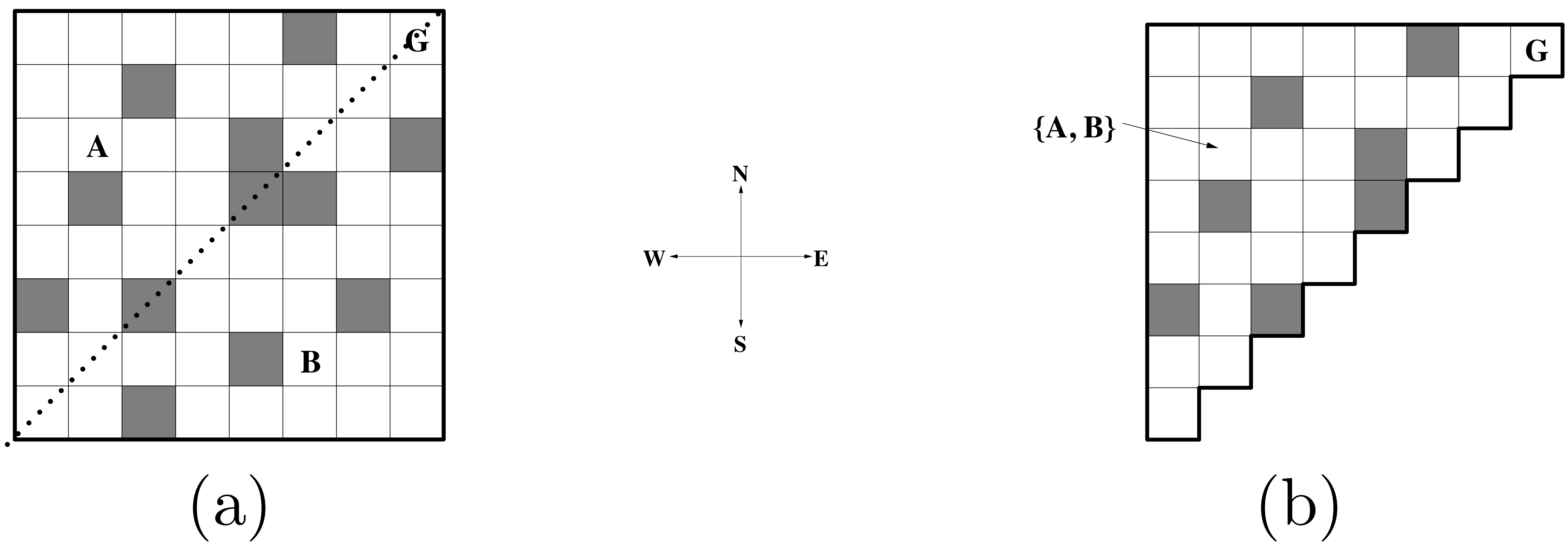


Figure from: Ravindran & Barto. Approximate Homomorphisms: A framework for non-exact minimization in Markov Decision Processes. 2004.

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 $\|V_M^\star - V_M^{[\pi]_M}\|_\infty \le \epsilon_{\pi^\star}$.
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 $\|[f]_M - Q_M^\star\|_\infty \le \epsilon_{Q^\star}$.
3. ϕ is an (ϵ_R, ϵ_P)-approximate model-irrelevant abstraction if for any $s^{(1)}$ and $s^{(2)}$ where $\phi(s^{(1)}) = \phi(s^{(2)})$, $\forall a \in \mathcal{A}$,

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

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