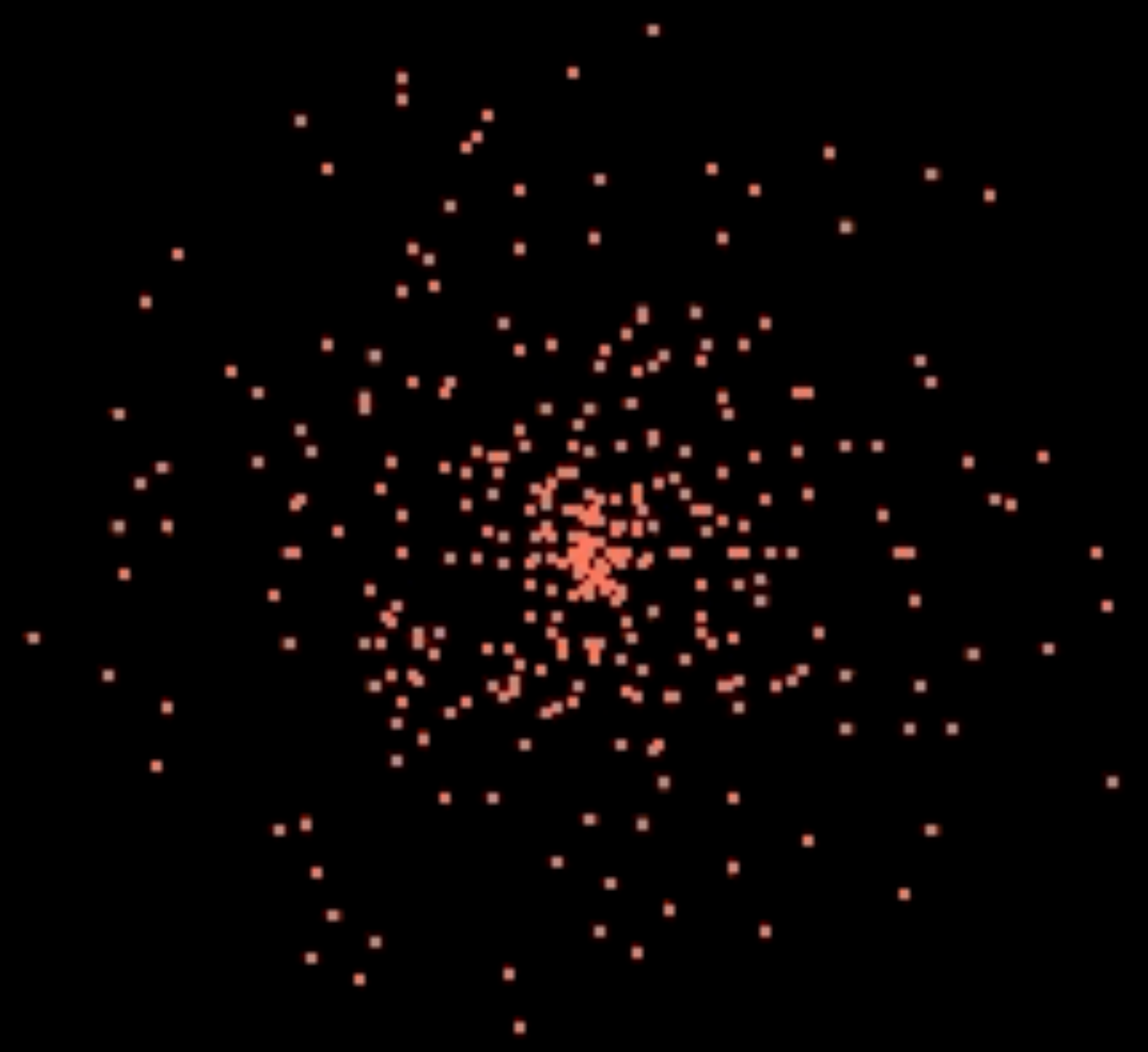
State-similarity metrics

Pablo Samuel Castro - Google Research, Brain Team

Most problems of practical interest are MDPs with very large (or continuous) state spaces.

Unstructured states

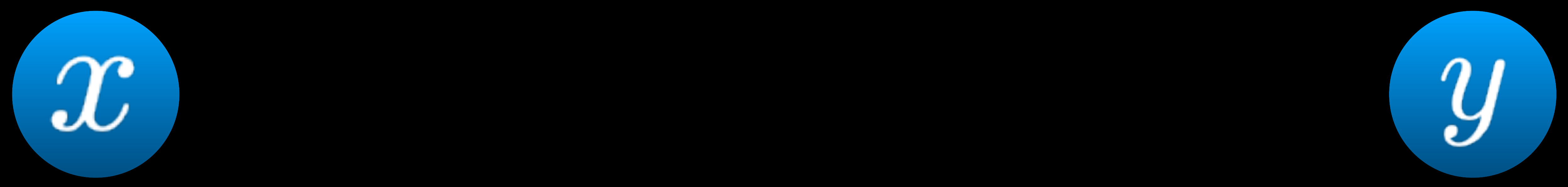
x



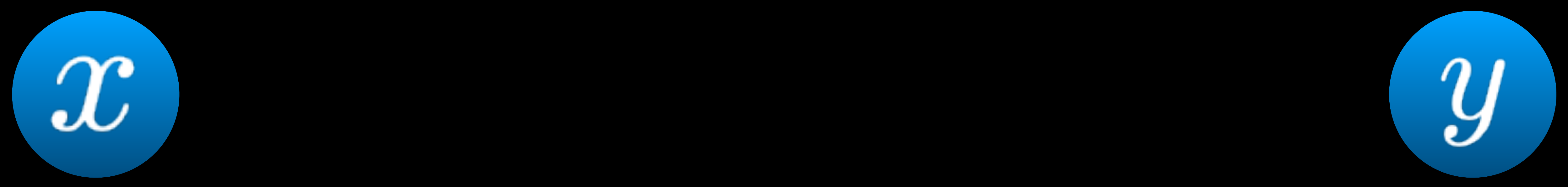
How to structure these states?

$$\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$$

$$\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$$



$$\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$$



$$x \stackrel{?}{=} y$$

$$\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$$

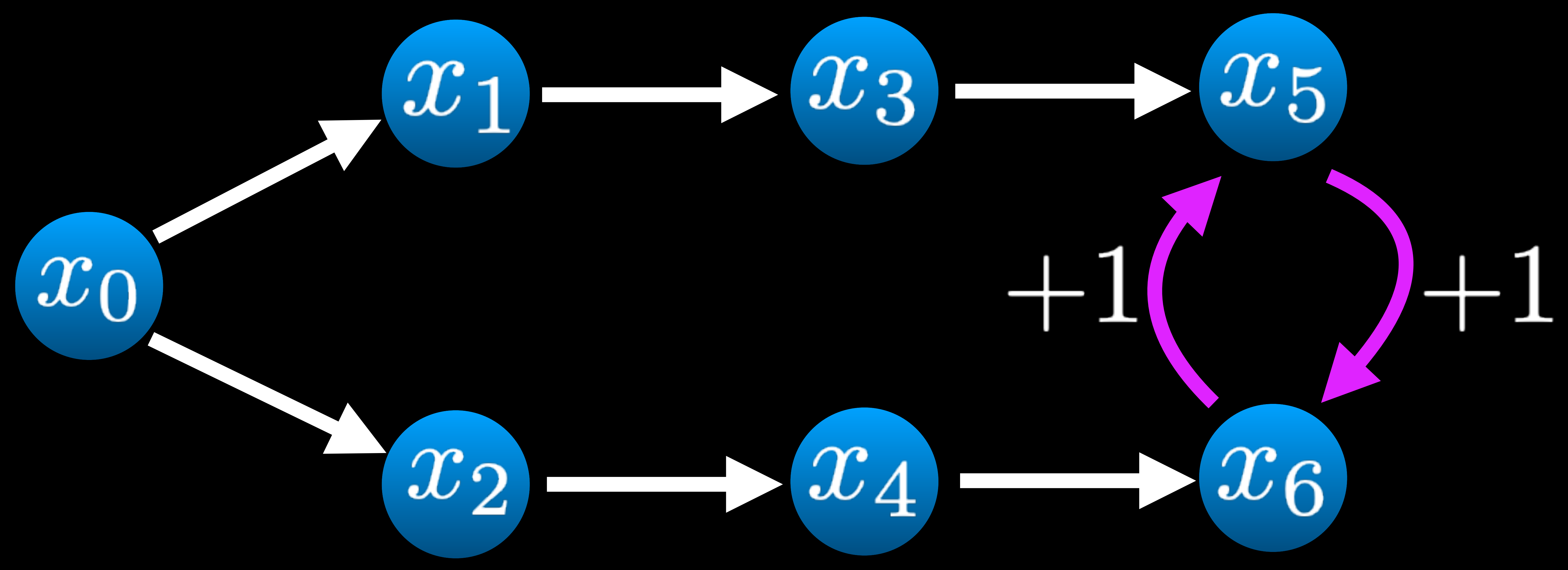
x

- Equal rewards
- Equal transitions

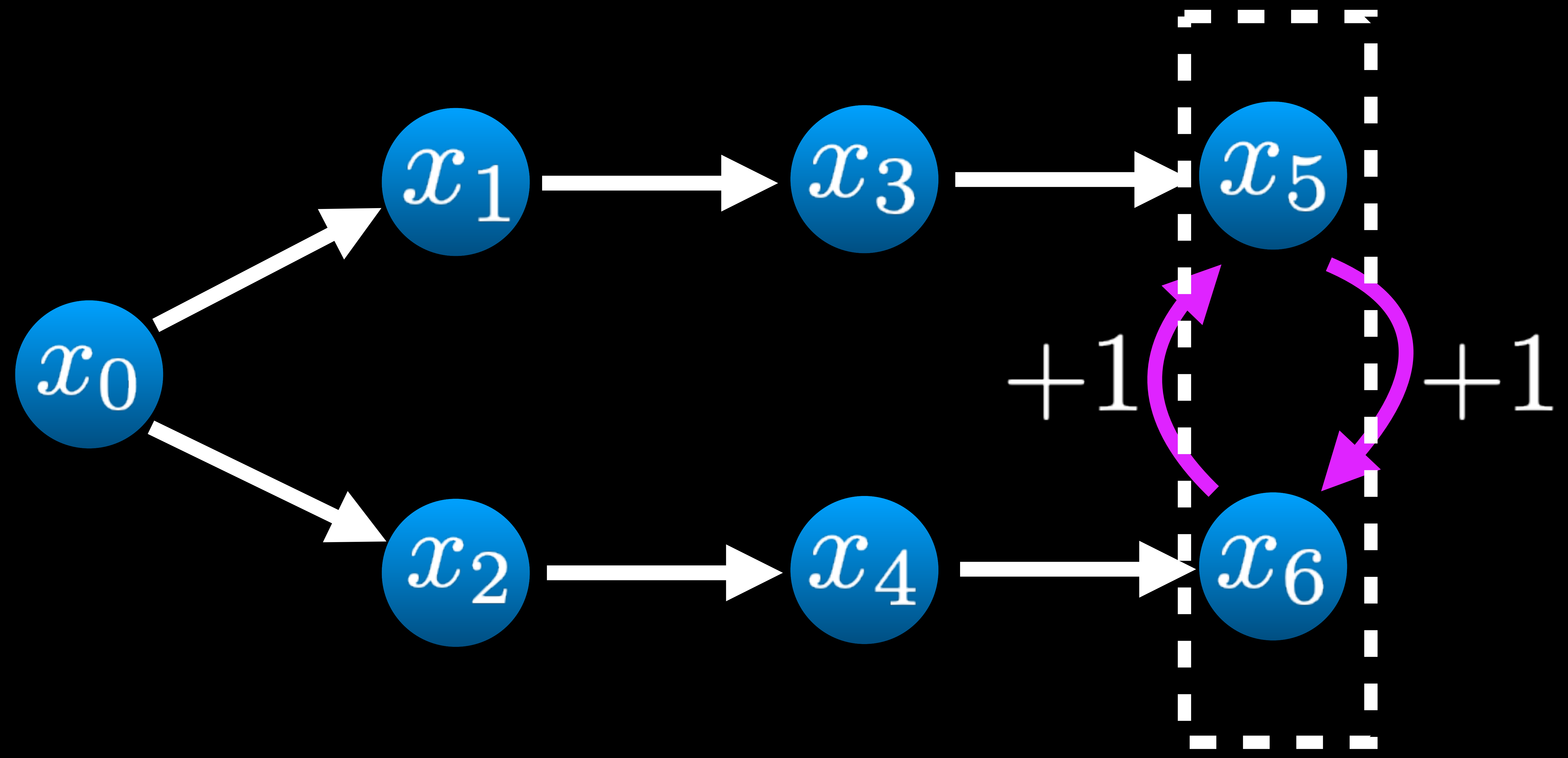
y

$$x \stackrel{?}{=} y$$

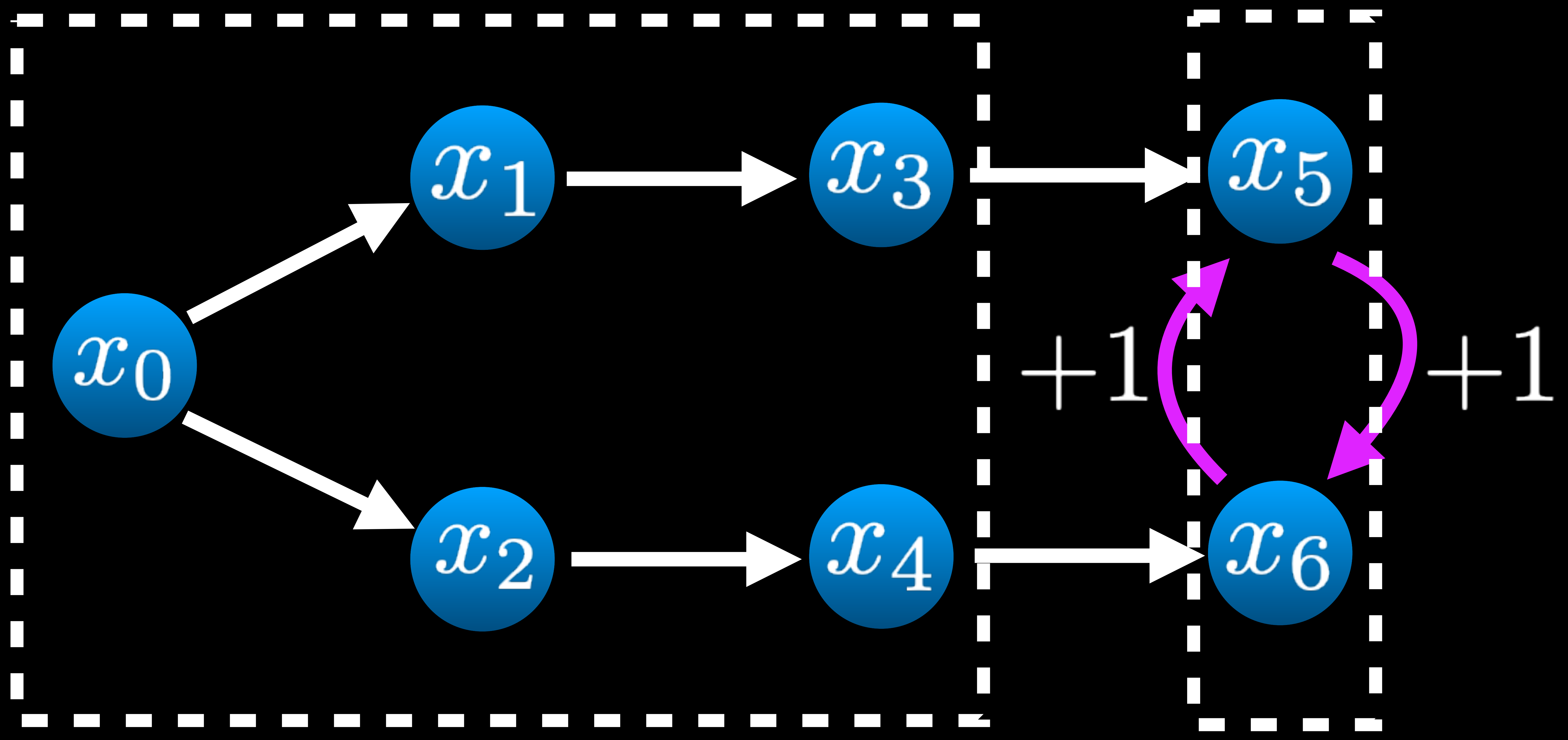
Which states are equivalent?



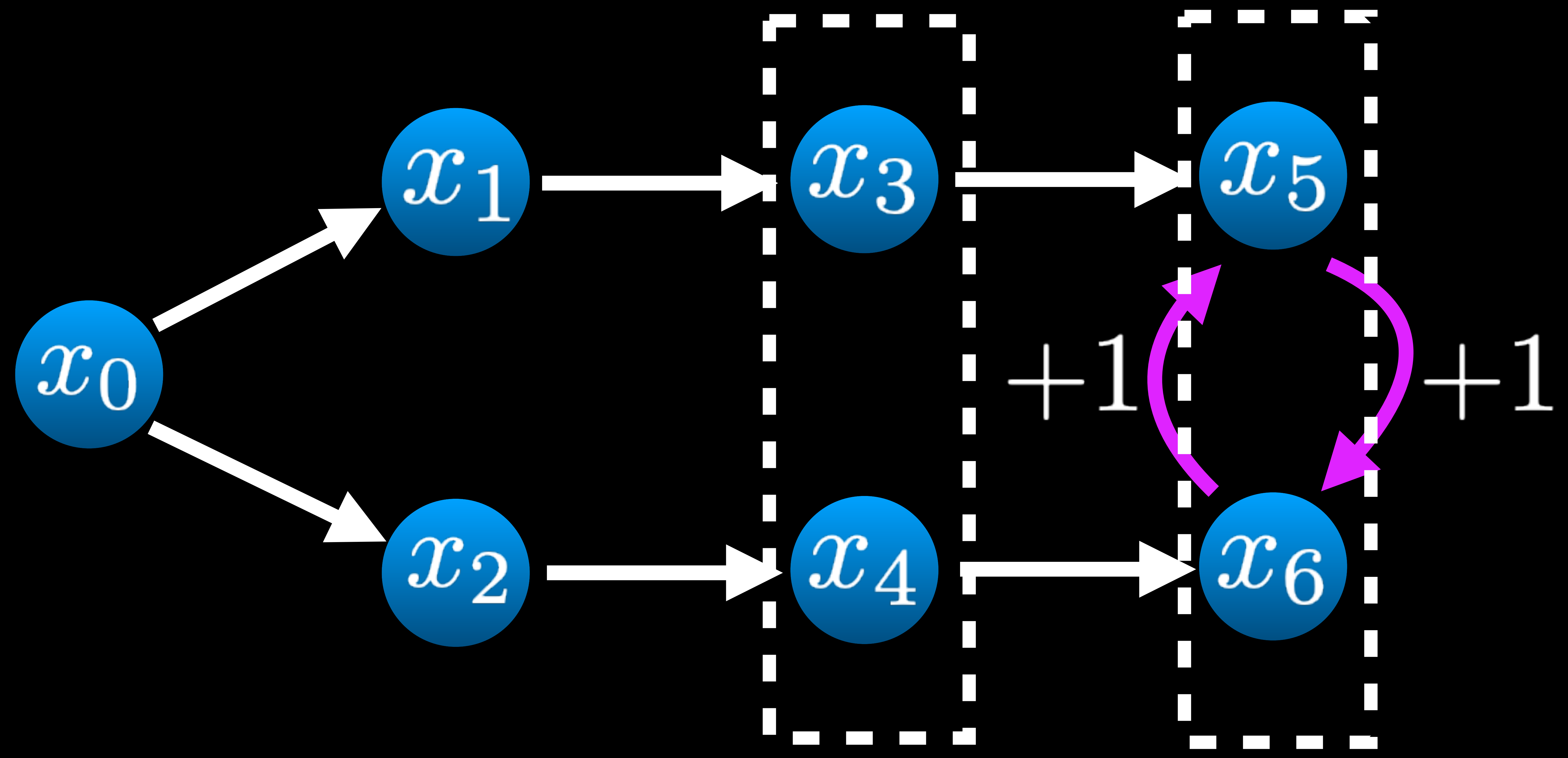
Which states are equivalent?



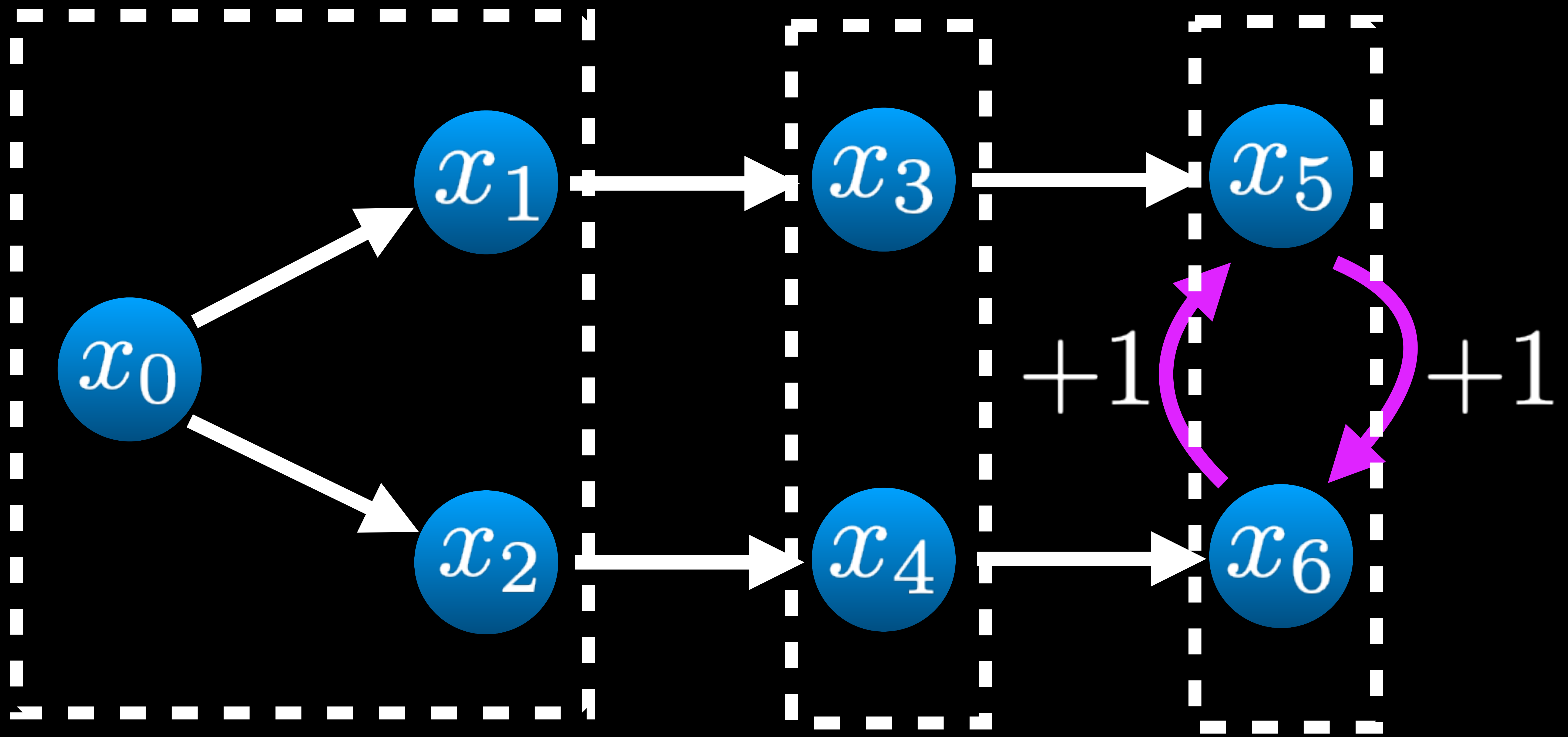
Which states are equivalent?



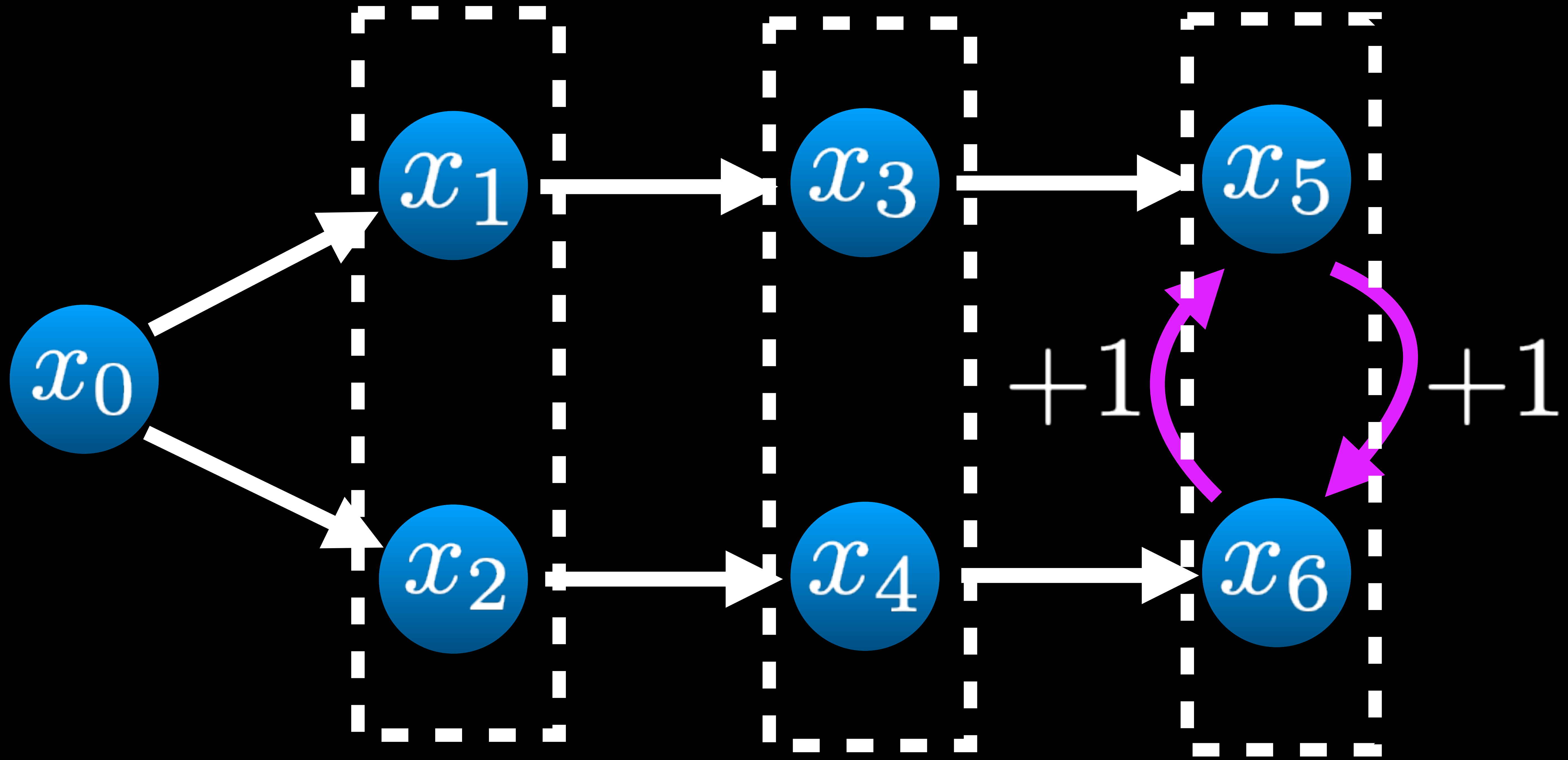
Which states are equivalent?



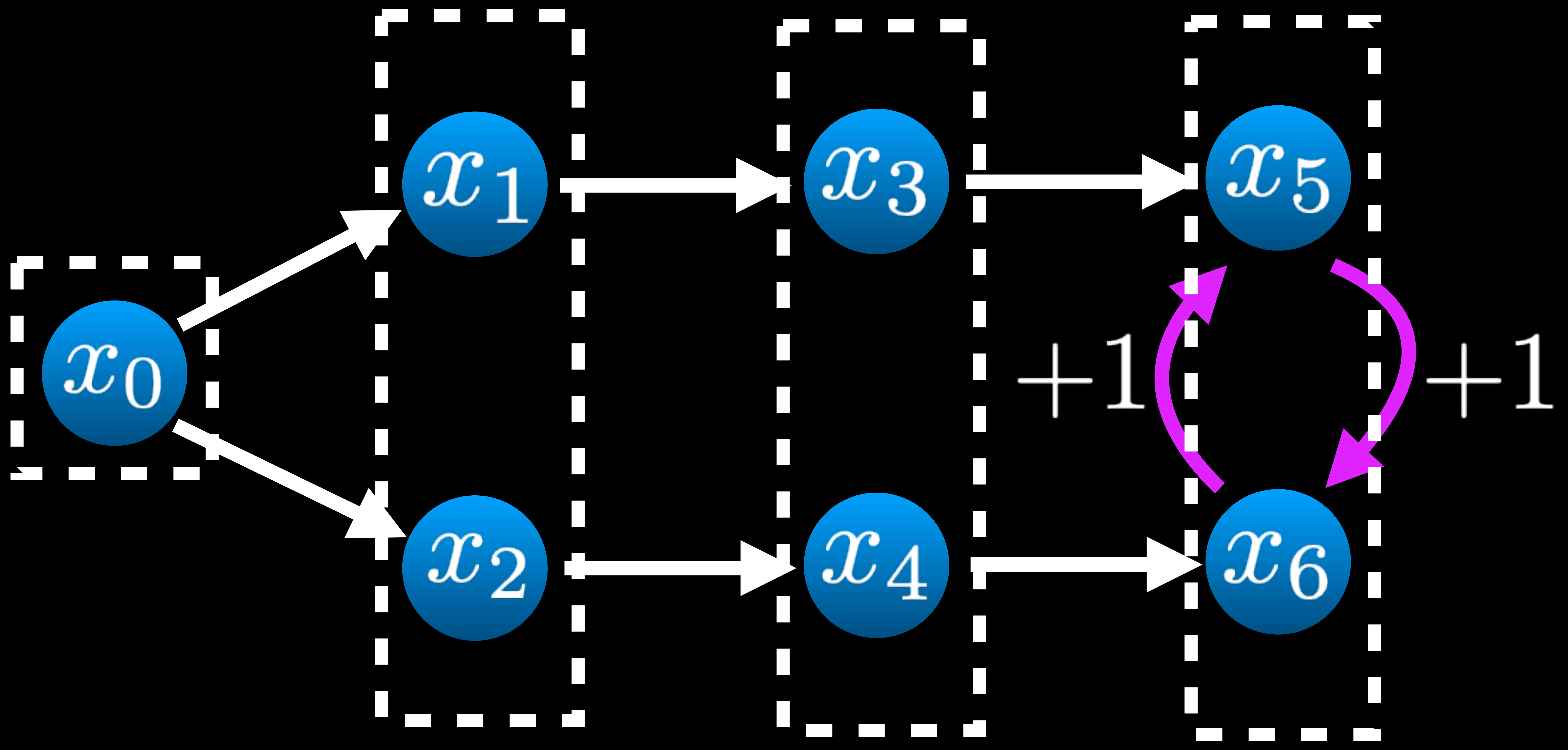
Which states are equivalent?



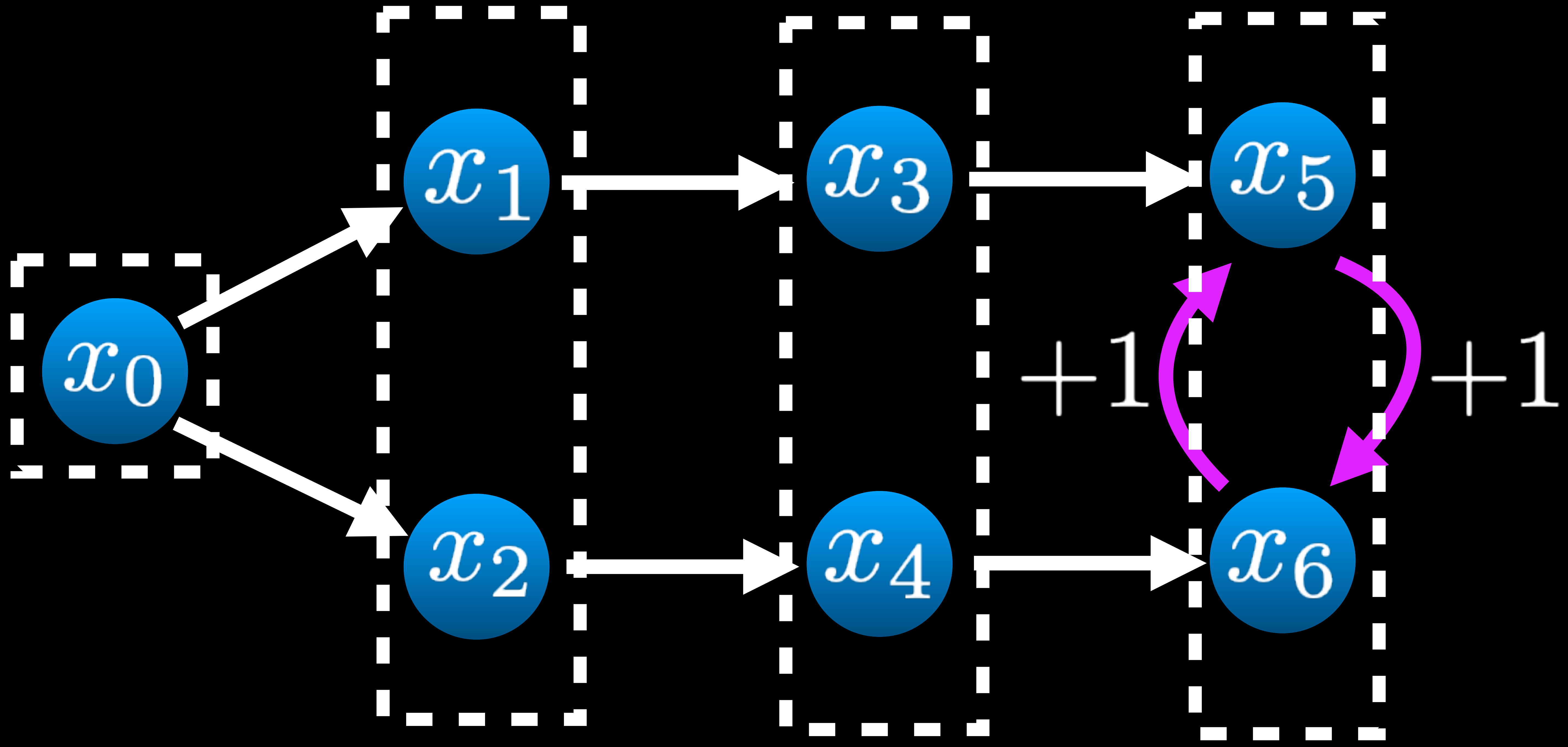
Which states are equivalent?



Which states are equivalent?



Which states are equivalent?



8 states => 4 states!

$V^* \equiv \hat{V}^*$

Bisimulation relations

Equivalence notions and model minimization in Markov decision processes

Robert Givan [a,*], Thomas Dean [b], Matthew Greig [a]

[a] *School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Computer Science, Brown University, Providence, RI 02912, USA*

Received 22 June 2001; received in revised form 17 April 2002

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \to \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. Same rewards
2. Same transitions

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. $\forall a \in \mathcal{A}, \quad \mathcal{R}(x, a) = \mathcal{R}(y, a)$
2. Same transitions

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. $\forall a \in \mathcal{A}, \quad \mathcal{R}(x, a) = \mathcal{R}(y, a)$
2. $\forall a \in \mathcal{A}, \forall c \in \mathcal{S}/_E, \quad \mathcal{P}(x, a)(c) = \mathcal{P}(y, a)(c)$

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. $\forall a \in \mathcal{A}, \quad \mathcal{R}(x, a) = \mathcal{R}(y, a)$
2. $\forall a \in \mathcal{A}, \forall c \in \mathcal{S}/_E, \quad \mathcal{P}(x, a)(c) = \mathcal{P}(y, a)(c)$

$$\left(\mathcal{P}(x, a)(c) = \sum_{s' \in c} \mathcal{P}(x, a)(x') \right)$$

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. $\forall a \in \mathcal{A}, \quad \mathcal{R}(x, a) = \mathcal{R}(y, a)$
2. $\forall a \in \mathcal{A}, \forall c \in \mathcal{S}/_E, \quad \mathcal{P}(x, a)(c) = \mathcal{P}(y, a)(c)$

$$\left(\mathcal{P}(x, a)(c) = \sum_{s' \in c} \mathcal{P}(x, a)(x')\right)$$

Two states x and y are **bisimilar** if there exists a bisimulation relation E such that xEy.

Bisimulation relations

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, an equivalence relation $E : \mathcal{S} \times \mathcal{S} \to \{0, 1\}$ is a **bisimulation relation** if whenever xEy we have:

1. $\forall a \in \mathcal{A}, \quad \mathcal{R}(x, a) = \mathcal{R}(y, a)$
2. $\forall a \in \mathcal{A}, \forall c \in \mathcal{S}/_E, \quad \mathcal{P}(x, a)(c) = \mathcal{P}(y, a)(c)$

$$\left(\mathcal{P}(x, a)(c) = \sum_{s' \in c} \mathcal{P}(x, a)(x')\right)$$

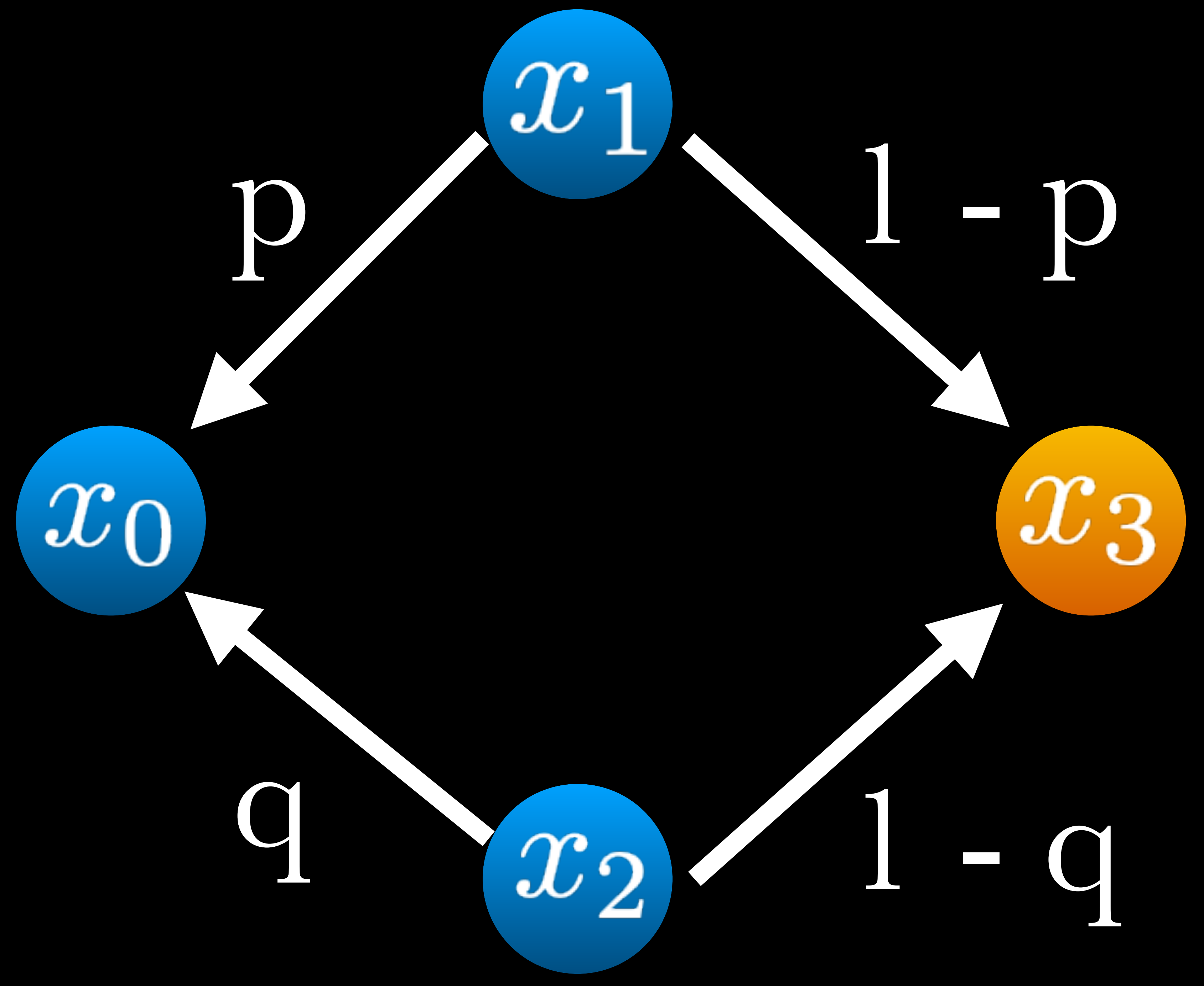
Two states x and y are **bisimilar** if there exists a bisimulation relation E such that xEy.

Let ~ be the maximal bisimulation relation.

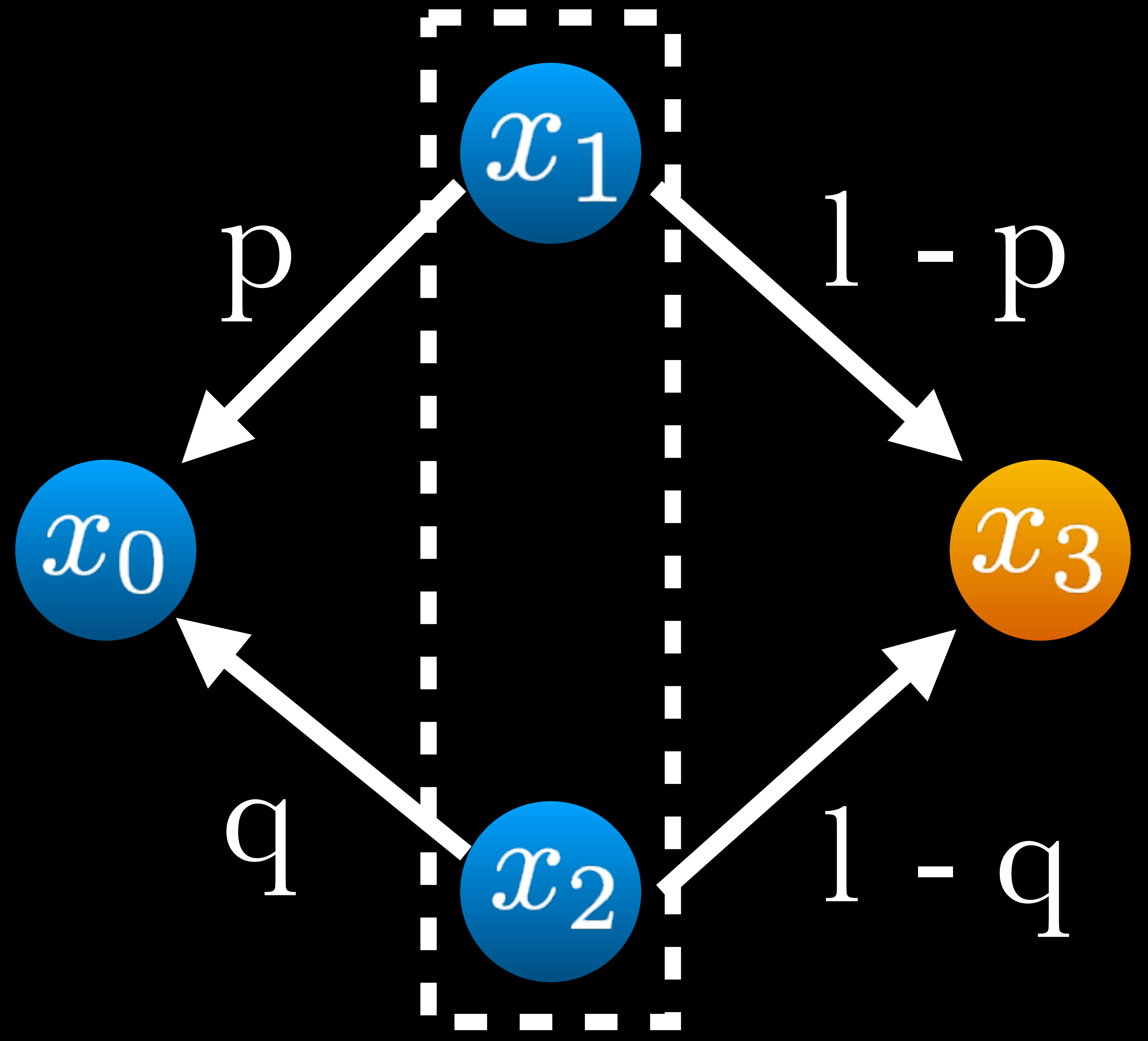
Bisimulation implies value equivalence

$$x \sim y \implies V^*(x) = V^*(y)$$

Are x1 and x2 bisimilar?

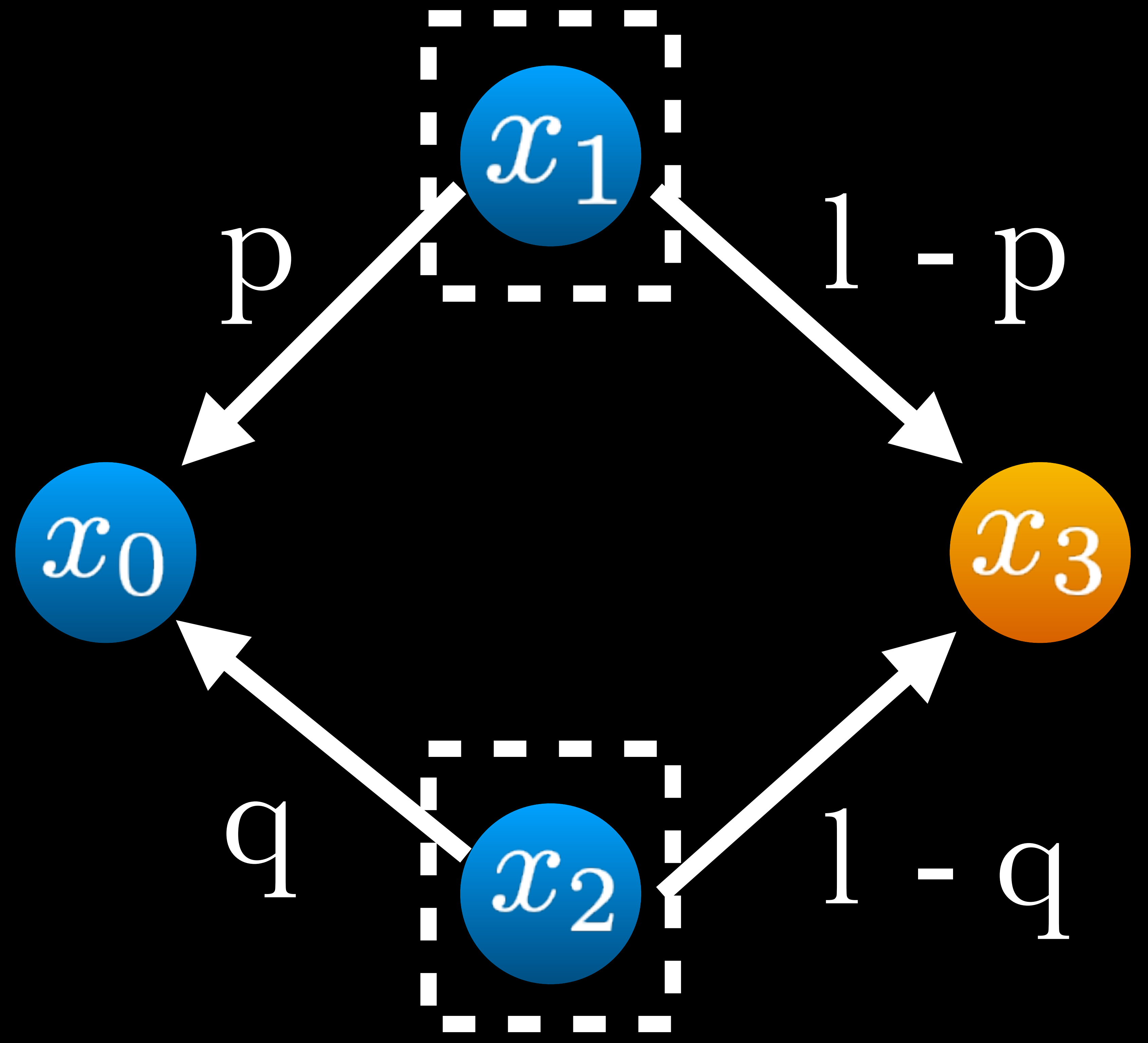


Are x1 and x2 bisimilar?



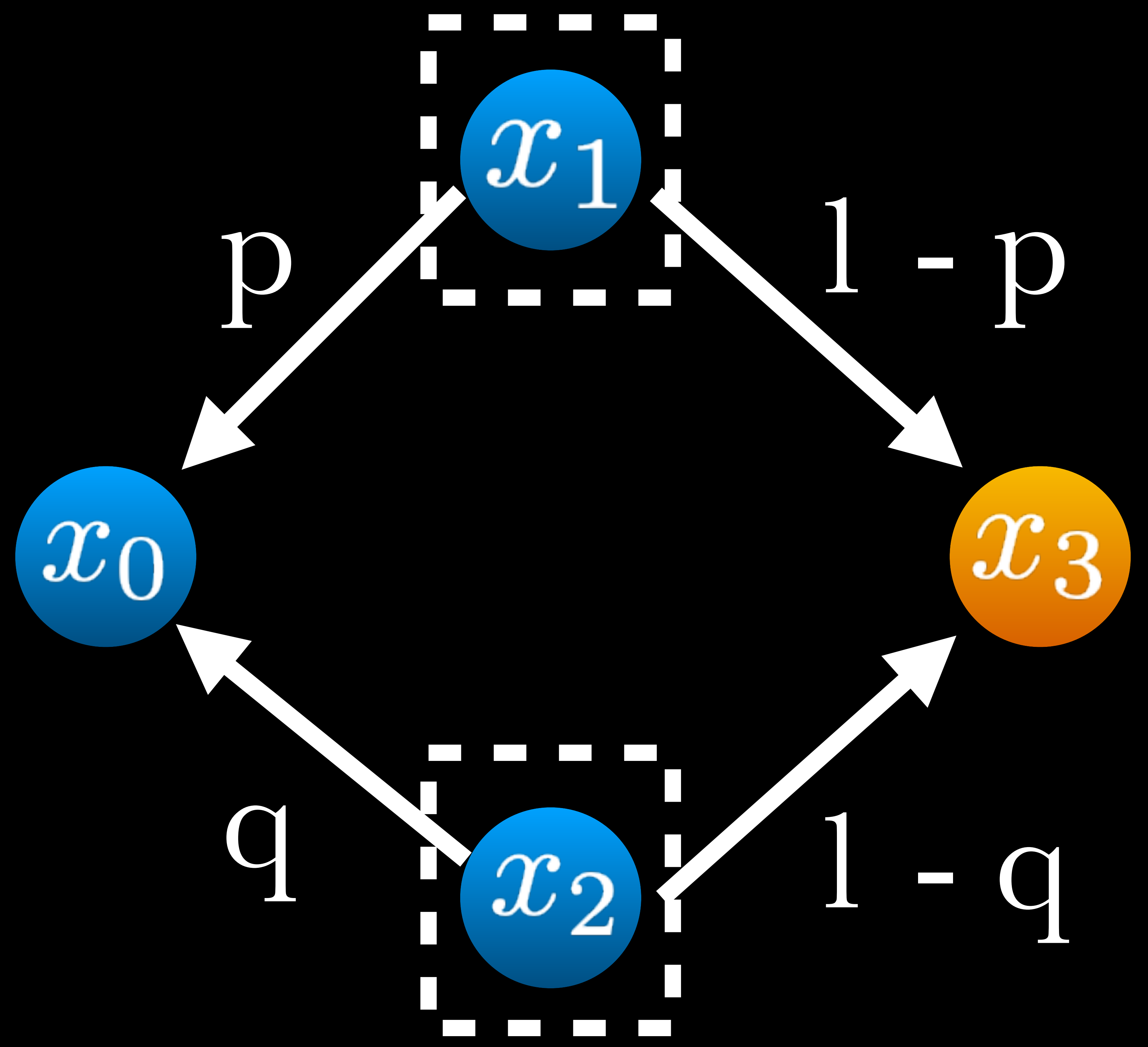
If p = q, then **yes**!

Are x1 and x2 bisimilar?



If p ≠ q, then **no**!

Are x1 and x2 bisimilar?



Bisimulation relations can be brittle!

Equivalence relations

Equivalence relations

Metrics

Equivalence relations

1. Reflexivity
2. Symmetry
3. Transitivity

Metrics

Equivalence relations

Metrics

1. Reflexivity

$$x \sim x$$

2. Symmetry

3. Transitivity

Equivalence relations

1. Reflexivity

$$x \sim x$$

2. Symmetry

$$x \sim y \iff y \sim x$$

3. Transitivity

Metrics

Equivalence relations

Metrics

1. Reflexivity

$$x \sim x$$

2. Symmetry

$$x \sim y \iff y \sim x$$

3. Transitivity

$$x \sim y \text{ and } y \sim z \implies x \sim z$$

Equivalence relations	Metrics
1. Reflexivity $x \sim x$	Identity of indescernibles
2. Symmetry $x \sim y \iff y \sim x$	Symmetry
3. Transitivity $x \sim y \text{ and } y \sim z \implies x \sim z$	Triangle inequality

Equivalence relations

1. Reflexivity

$$x \sim x$$

2. Symmetry

$$x \sim y \iff y \sim x$$

3. Transitivity

$$x \sim y \text{ and } y \sim z \implies x \sim z$$

Metrics

Identity of indescernibles

$$d(x, y) = 0 \iff x = y$$

Symmetry

Triangle inequality

Equivalence relations

1. Reflexivity

$$x \sim x$$

2. Symmetry

$$x \sim y \iff y \sim x$$

3. Transitivity

$$x \sim y \text{ and } y \sim z \implies x \sim z$$

Metrics

Identity of indescernibles

$$d(x, y) = 0 \iff x = y$$

Symmetry

$$d(x, y) = d(y, x)$$

Triangle inequality

Equivalence relations

1. Reflexivity

$$x \sim x$$

2. Symmetry

$$x \sim y \iff y \sim x$$

3. Transitivity

$$x \sim y \text{ and } y \sim z \implies x \sim z$$

Metrics

Identity of indescernibles

$$d(x, y) = 0 \iff x = y$$

Symmetry

$$d(x, y) = d(y, x)$$

Triangle inequality

$$d(x, z) \leq d(x, y) + d(y, z)$$

Metrics

1. Identity of indescernibles

$$d(x, y) = 0 \iff x = y$$

2. Symmetry

$$d(x, y) = d(y, x)$$

3. Triangle inequality

$$d(x, z) \leq d(x, y) + d(y, z)$$

Metrics

1. Identity of indescernibles

$$d(x,y) = 0 \iff x = y$$

2. Symmetry

$$d(x,y) = d(y,x)$$

3. Triangle inequality

$$d(x,z) \leq d(x,y) + d(y,z)$$

Pseudo-metrics

$$d(x,x) = 0$$
$$d(x,y) \geq 0$$

The Kantorovich metric

The Kantorovich metric

(also known as Wasserstein metric)

The Kantorovich metric

(also known as Wasserstein metric)

(also known as Optimal Transport)

The Kantorovich metric

(also known as Wasserstein metric)

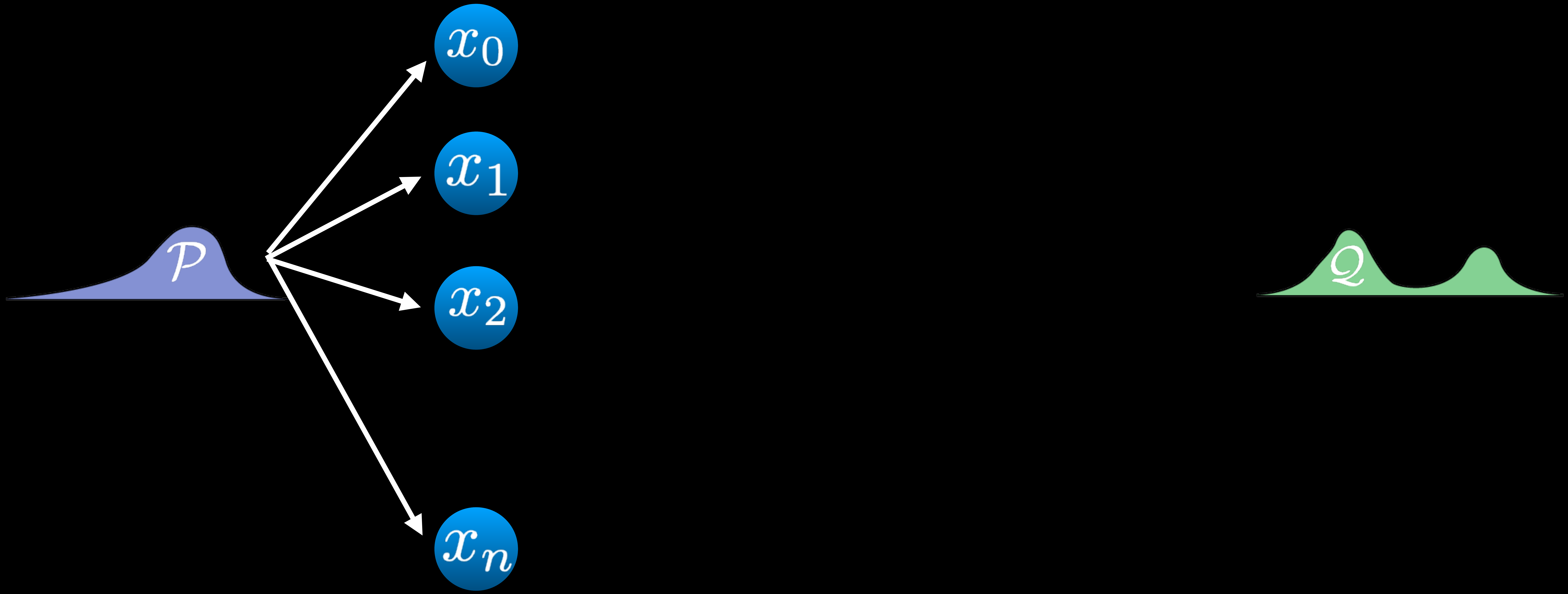
(also known as Optimal Transport)

(also known as Earth Movers Distance)

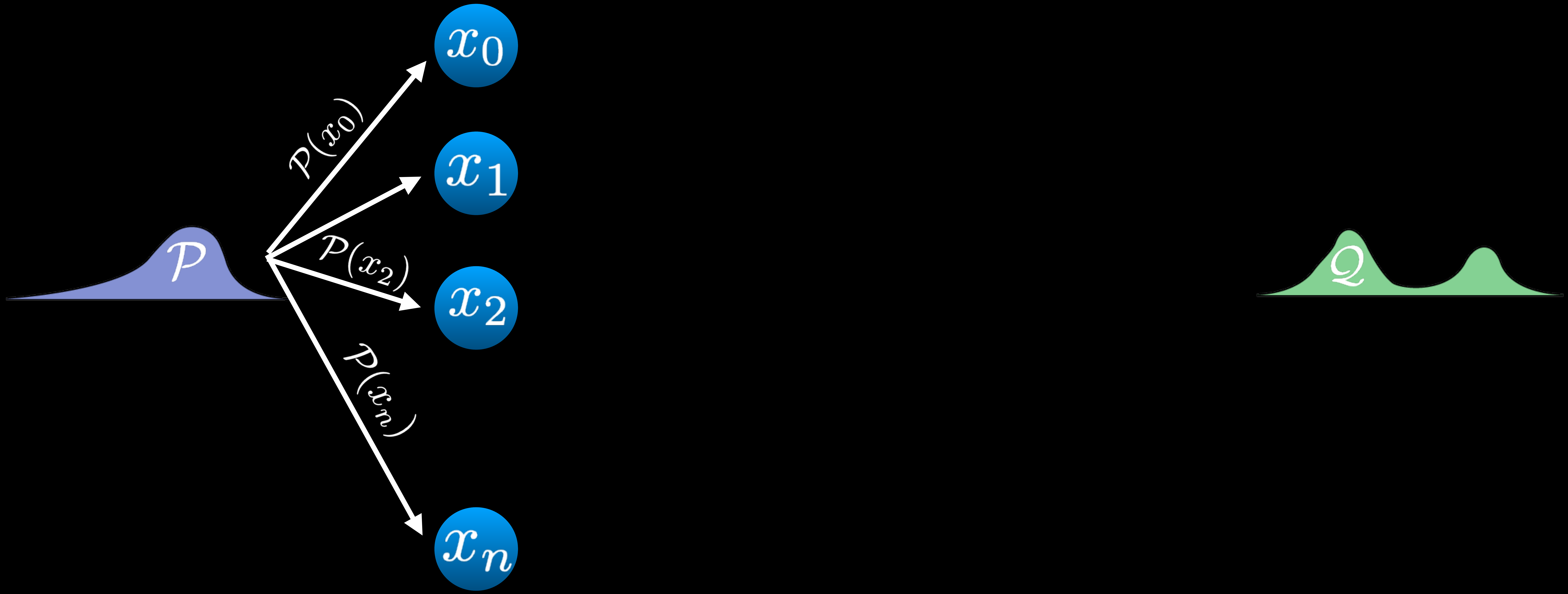
The Kantorovich metric

The Kantorovich metric

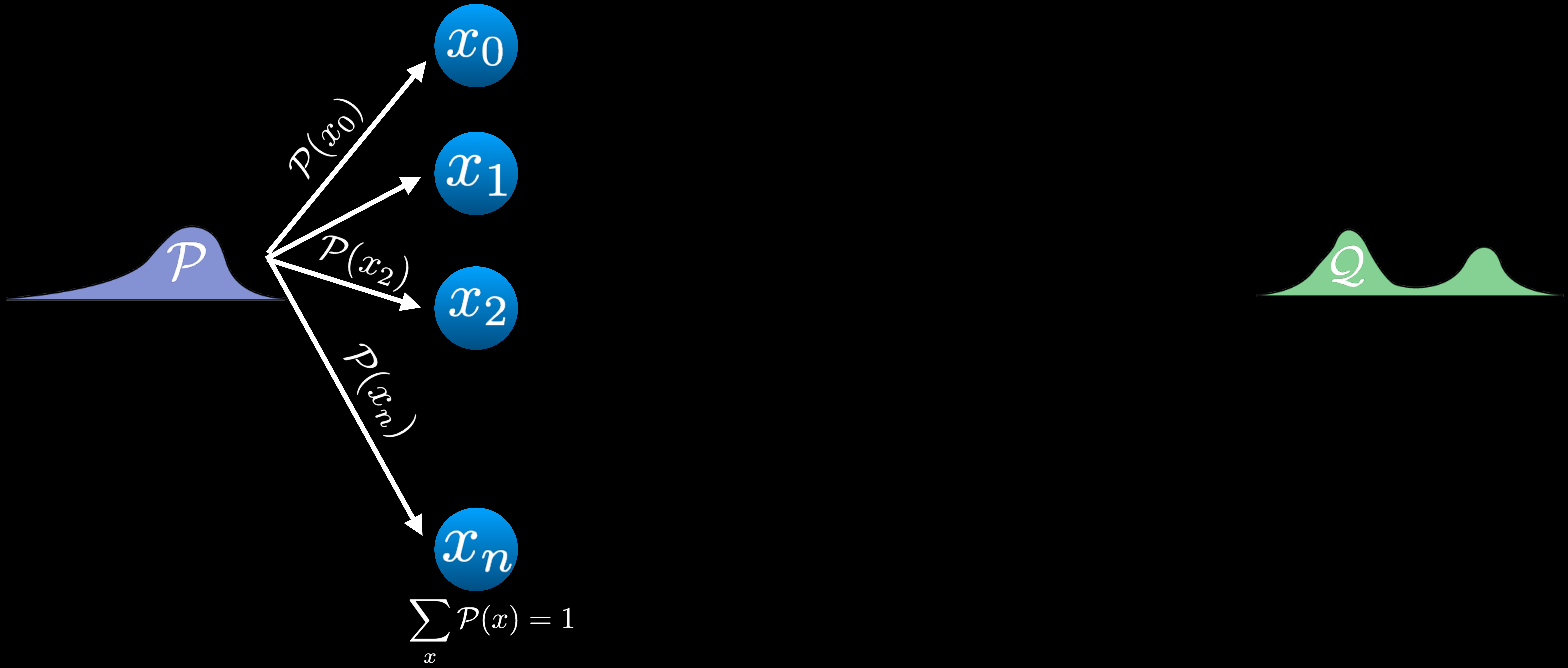
The Kantorovich metric



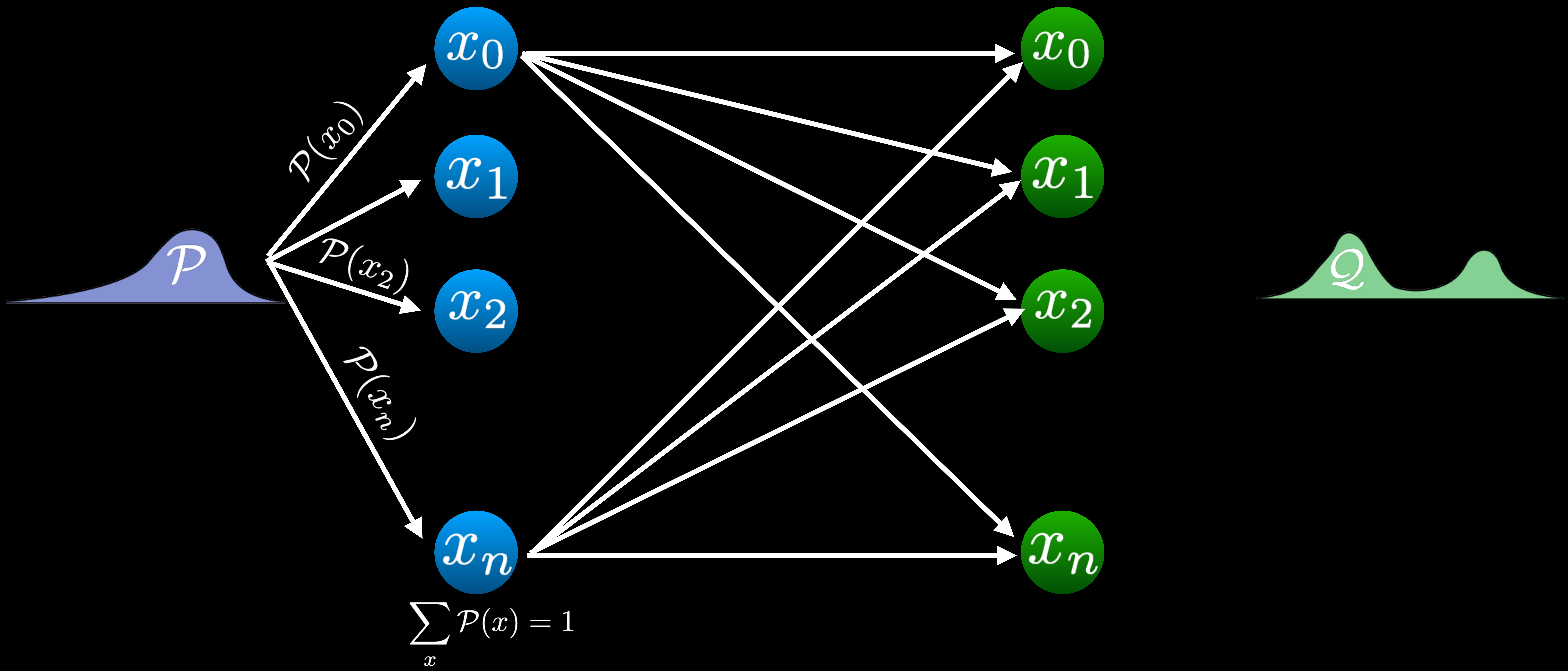
The Kantorovich metric



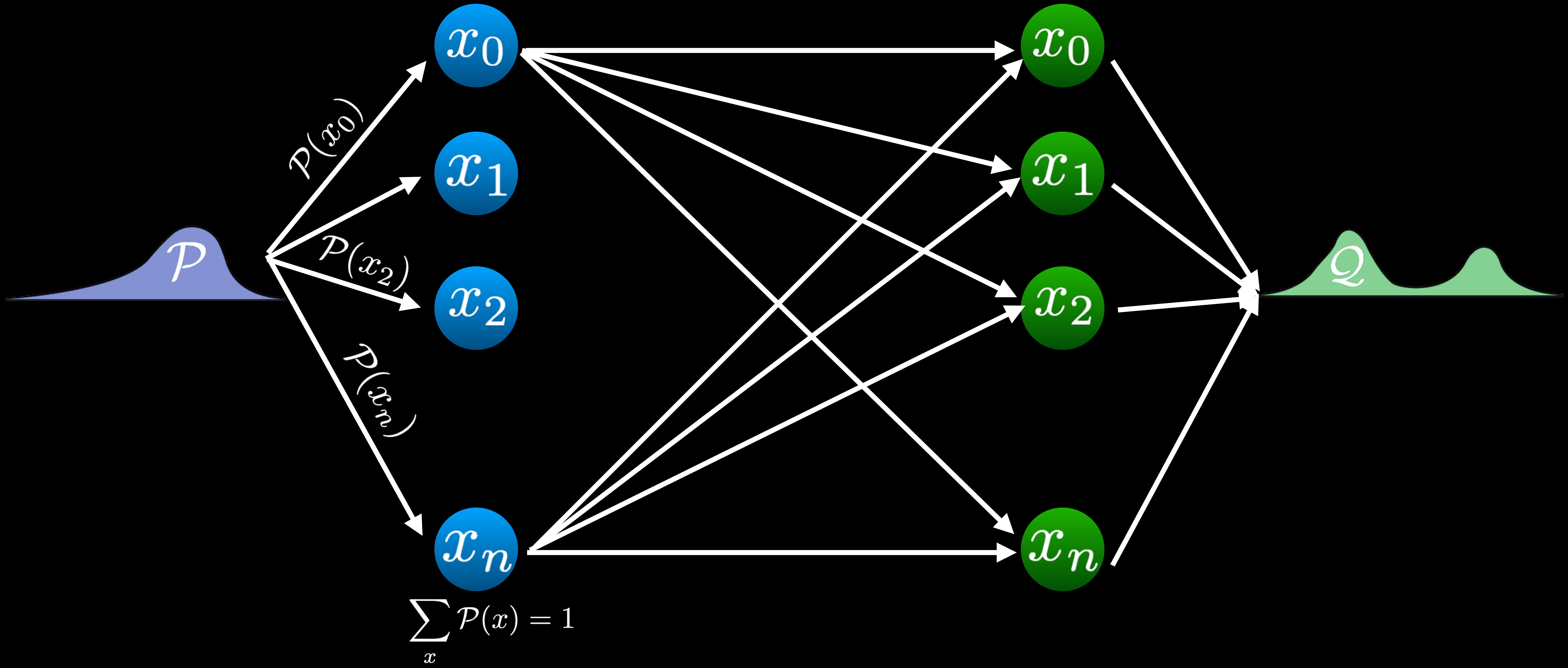
The Kantorovich metric



The Kantorovich metric

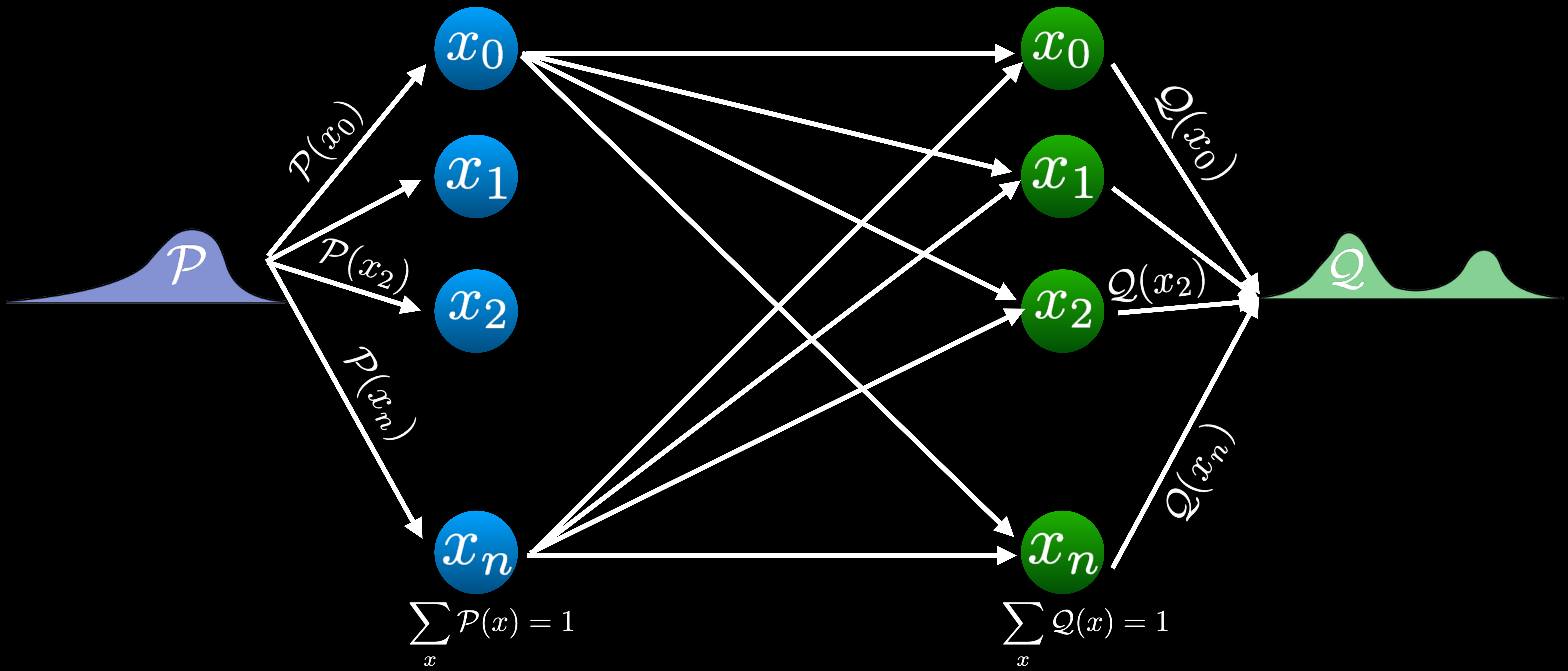


The Kantorovich metric

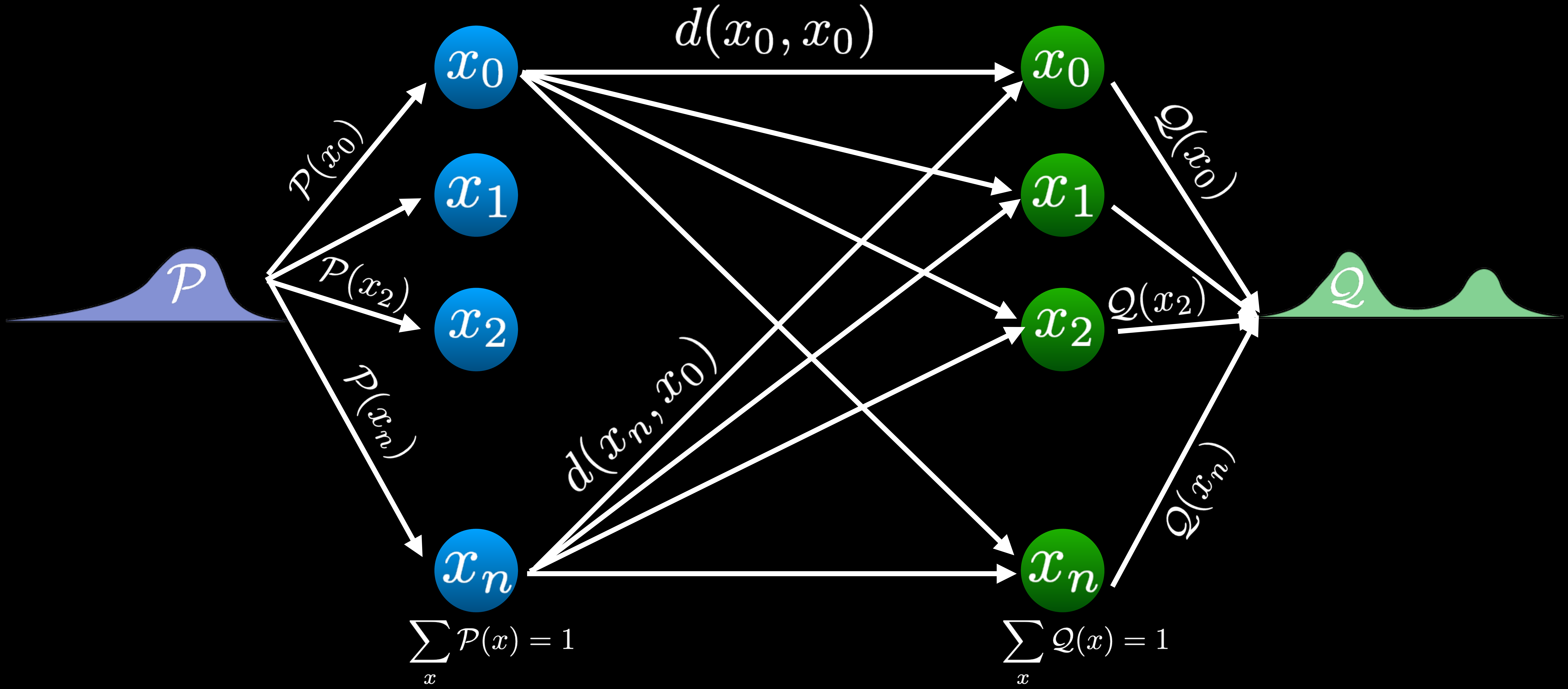


The Kantorovich metric

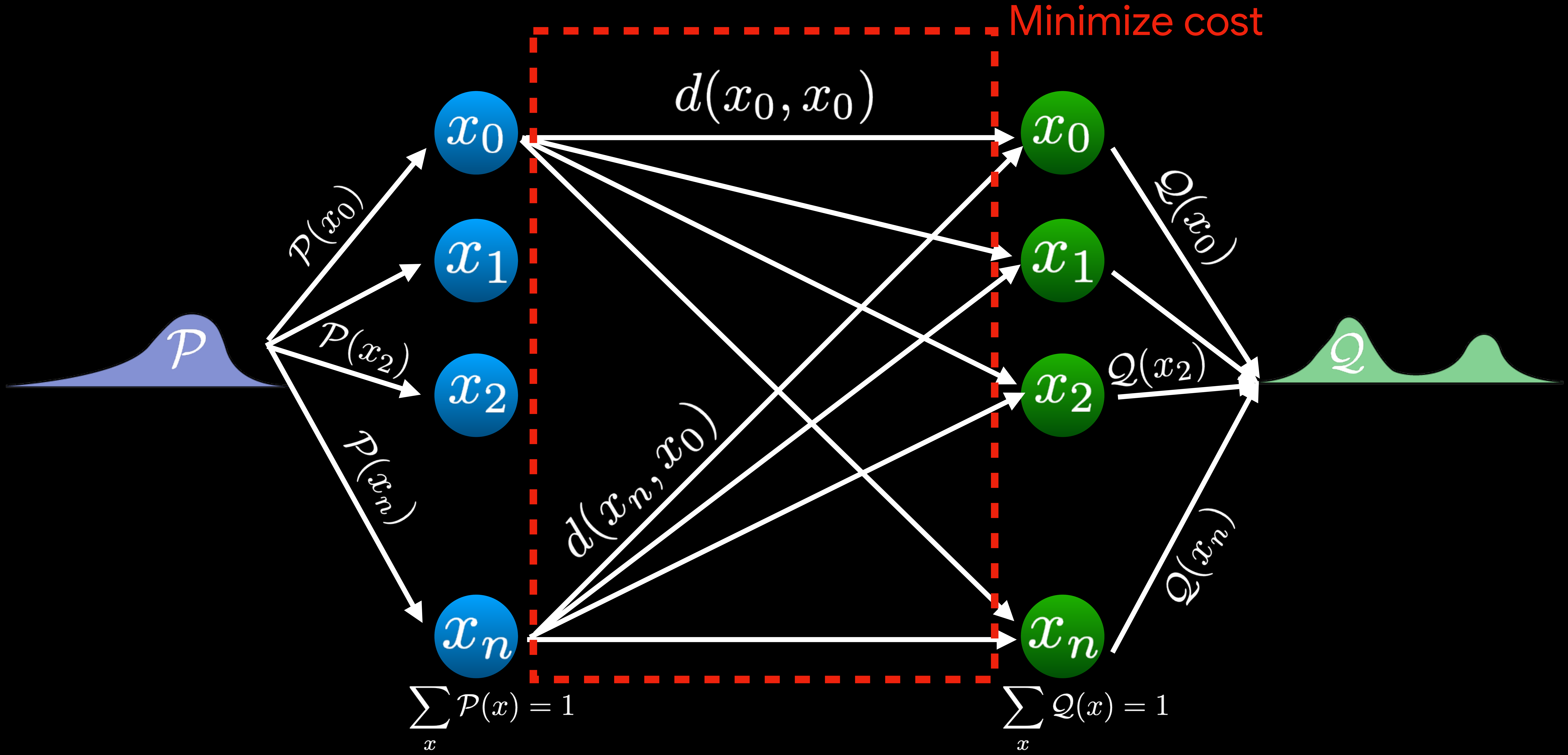
The Kantorovich metric



The Kantorovich metric



The Kantorovich metric



The Kantorovich metric

$$\max_{\mu} \sum_{x \in \mathcal{S}} (\mathcal{P}(x) - \mathcal{Q}(x))\mu_x$$

$$\text{subject to}$$

$$\mu_x - \mu_y \leq d(x, y) \quad \forall x, y \in \mathcal{S}$$

$$\mu_x \geq 0 \quad \forall x \in \mathcal{S}$$

The Kantorovich metric

Primal

$$\max_{\mu} \sum_{x \in \mathcal{S}} (\mathcal{P}(x) - \mathcal{Q}(x))\mu_x$$

$$\text{subject to}$$

$$\mu_x - \mu_y \leq d(x, y) \quad \forall x, y \in \mathcal{S}$$

$$\mu_x \geq 0 \quad \forall x \in \mathcal{S}$$

The Kantorovich metric

Primal

$$\max_{\mu} \sum_{x \in \mathcal{S}} (\mathcal{P}(x) - \mathcal{Q}(x))\mu_x$$

$$\text{subject to}$$

$$\mu_x - \mu_y \leq d(x, y) \quad \forall x, y \in \mathcal{S}$$

$$\mu_x \geq 0 \quad \forall x \in \mathcal{S}$$

Dual

$$\min_{\lambda} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \lambda_{x,y} d(x, y)$$

$$\text{subject to}$$

$$\sum_{y \in \mathcal{S}} \lambda_{x,y} = \mathcal{P}(x) \quad \forall x \in \mathcal{S}$$

$$\sum_{x \in \mathcal{S}} \lambda_{x,y} = \mathcal{Q}(y) \quad \forall y \in \mathcal{S}$$

$$\lambda_{x,y} \geq 0 \quad \forall x, y \in \mathcal{S}$$

The Kantorovich metric

Primal

$$\max_{\mu} \sum_{x\in\mathcal{S}} (\mathcal{P}(\ldots$$

$$\mu_x - \mu_y \leq d(x, \ldots$$

$$\mu_x \ldots$$

$$T_K(d)(\mathcal{P}, \mathcal{Q})$$

Dual

$$\ldots \sum_{\ldots\in\mathcal{S}} \lambda_{x,y} d(x,y)$$

subject to

$$\ldots(x) \quad \forall x \in \mathcal{S}$$

$$\ldots\mathcal{Q}(y) \quad \forall y \in \mathcal{S}$$

$$\ldots_{x\in\mathcal{S}}$$

$$\lambda_{x,y} \geq 0 \quad \forall x, y \in \mathcal{S}$$

Bisimulation metrics

Metrics for Finite Markov Decision Processes

Norm Ferns
School of Computer Science
McGill University
Montréal, Canada, H3A 2A7
nferns@cs.mcgill.ca

Prakash Panangaden
School of Computer Science
McGill University
Montréal, Canada, H3A 2A7
prakash@cs.mcgill.ca

Doina Precup
School of Computer Science
McGill University
Montréal, Canada, H3A 2A7
dprecup@cs.mcgill.ca

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x, y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x, y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

1. Compute bisimulation equivalence relation ~
2. Assign distances as:

$$d(x, y) = 0 \text{ if } x \sim y, \quad d(x, y) = \infty \text{ otherwise.}$$

3. Profit!

Bisimulation metrics

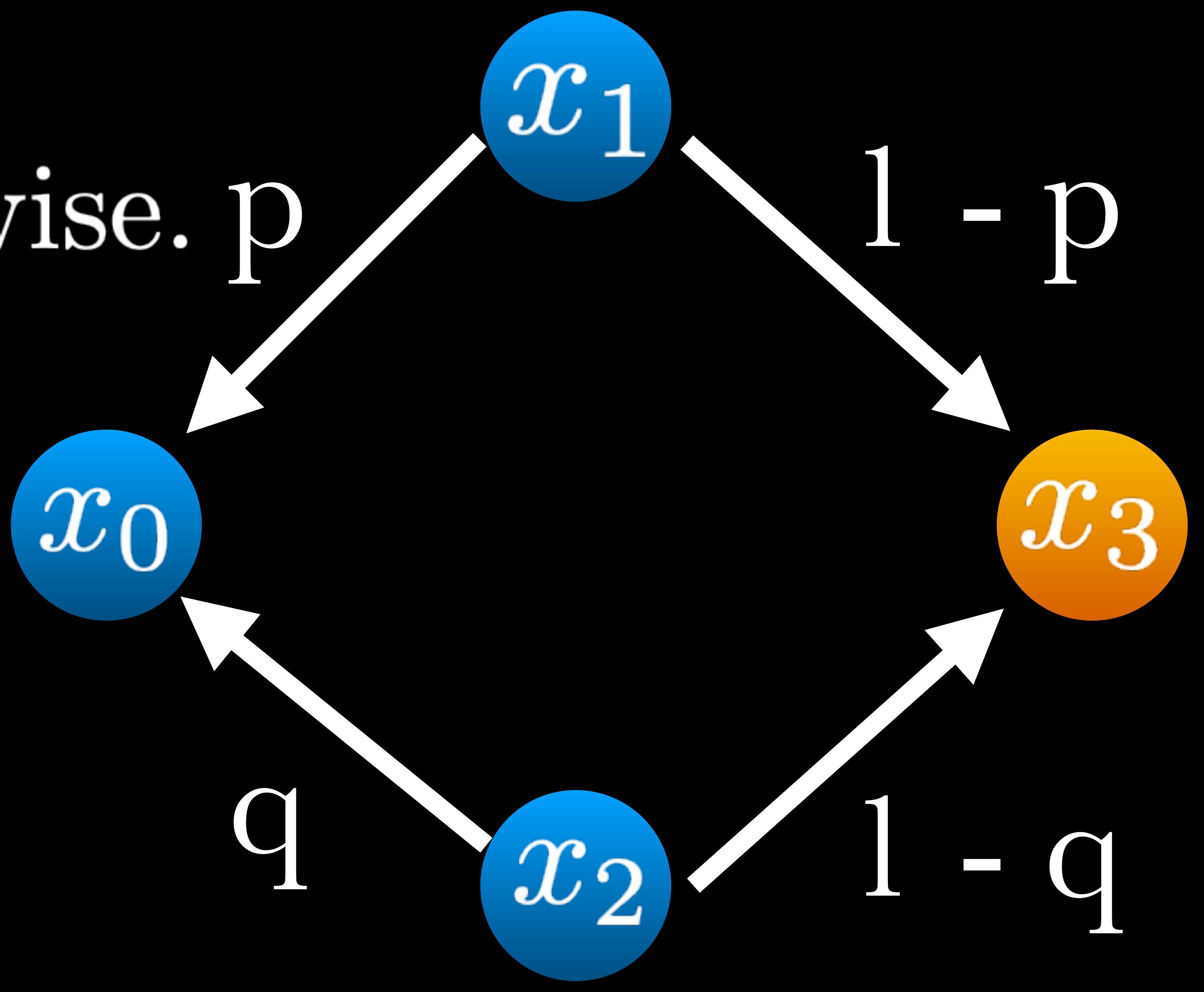
Definition: A metric d is a bisimulation metric if

$$d(x, y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

1. Compute bisimulation equivalence relation ~
2. Assign distances as:

$$d(x, y) = 0 \text{ if } x \sim y, \quad d(x, y) = \infty \text{ otherwise.}$$

3. Profit!



Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x,y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F} : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}(d)(x,y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a), \mathcal{P}(y,a)) \}$$

has a unique fixed point $d_\sim$ and $d_\sim$ is a bisimulation metric

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x,y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F} : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}(d)(x,y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a), \mathcal{P}(y,a)) \}$$

has a unique fixed point $d_\sim$ and $d_\sim$ is a bisimulation metric

Difference in rewards

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x, y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F} : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}(d)(x, y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x, a) - \mathcal{R}(y, a)| + \gamma T_K(d)(\mathcal{P}(x, a), \mathcal{P}(y, a)) \}$$

has a unique fixed point $d_\sim$ and $d_\sim$ is a bisimulation metric

Difference in transitions

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x,y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F} : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}(d)(x,y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a), \mathcal{P}(y,a)) \}$$

has a unique fixed point $d_\sim$ and $d_\sim$ is a bisimulation metric

Over all actions

Bisimulation metrics

Definition: A metric d is a bisimulation metric if

$$d(x,y) = 0 \iff x \sim y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F} : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}(d)(x,y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a), \mathcal{P}(y,a)) \}$$

has a unique fixed point $d_\sim$ and $d_\sim$ is a bisimulation metric

Theorem: $|V^*(x) - V^*(y)| \leq d_\sim(x,y) \quad \forall x, y \in \mathcal{S}$

A brief overview of some (tabular) extensions

Lax bisimulation metrics

Bounding Performance Loss in Approximate MDP Homomorphisms

Jonathan J. Taylor
Dept. of Computer Science
University of Toronto
Toronto, Canada, M5S 3G4
jonathan.taylor@utoronto.ca

Doina Precup
School of Computer Science
McGill University
Montreal, Canada, H3A 2A7
dprecup@cs.mcgill.ca

Prakash Panangaden
School of Computer Science
McGill University
Montreal, Canada, H3A 2A7
prakash@cs.mcgill.ca

Lax bisimulation metrics

Definition 5. Given a finite 1-bounded metric space $(\mathcal{M}, d)$, let $\mathcal{P}(\mathcal{M})$ be the set of compact spaces (e.g. closed and bounded in $\mathbb{R}$). The *Hausdorff metric* $H(d) : \mathcal{P}(\mathcal{M}) \times \mathcal{P}(\mathcal{M}) \to [0,1]$ is defined as:

$$H(d)(X,Y) = \max(\sup_{x \in X} \inf_{y \in Y} d(x,y), \sup_{y \in Y} \inf_{x \in X} d(x,y))$$

Definition 6. Denote $X_s = \{(s,a) | a \in A\}$. Let $\mathcal{M}$ be the set of all semimetrics on S. We define the operator $F : \mathcal{M} \to \mathcal{M}$ as $F(d)(s,u) = H(\delta(d))(X_s, X_u)$

Theorem 8. *Let e_{fix} be the metric defined in (Ferns et al., 2004). Then we have:*

$$c_r |V^*(s) - V^*(u)| \leq d_{fix}(s,u) \leq e_{fix}(s,u)$$

Bisimulation metrics for options

On planning, prediction and knowledge transfer in Fully and Partially Observable Markov Decision Processes

by

Pablo Samuel Castro

Bisimulation metrics for options

Definition 4.16. A relation $E \subseteq S \times S$ is said to be an option-bisimulation relation if whenever sEt:

1. $\forall o, R(s,o) = R(t,o)$
2. $\forall o, \forall C \in S/E. \sum_{s' \in C} Pr(s'|s,o) = \sum_{s' \in C} Pr(s'|t,o)$

Theorem 4.17. *The functional $F : \mathcal{M} \to \mathcal{M}$ defined as*

$$F(d)(s,t) = \max_{o \in OPT} (|\mathfrak{R}(s,o) - \mathfrak{R}(t,o)| + \gamma T_K(d)(Pr(\cdot|s,o), Pr(\cdot|t,o))$$

has a greatest fixed-point, $d_\sim$, and $d_\sim$ is an option-bisimulation metric.

Theorem 4.18. *If $s \sim_O t$, then $W^*(s) = W^*(t)$.*

Bisimulation metrics for policy transfer

Using Bisimulation for Policy Transfer in MDPs

Pablo Samuel Castro and **Doina Precup**
School of Computer Science, McGill University, Montreal, QC, Canada
pcastr@cs.mcgill.ca and dprecup@cs.mcgill.ca

Bisimulation metrics for policy transfer

$$M_1 = \{\mathcal{S}_1, \mathcal{A}, \mathcal{P}_1, \mathcal{R}_1, \gamma\} \longrightarrow M_2 = \{\mathcal{S}_2, \mathcal{A}, \mathcal{P}_2, \mathcal{R}_2, \gamma\}$$

Bisimulation metrics for policy transfer

$$M_1 = \{\mathcal{S}_1, \mathcal{A}, \mathcal{P}_1, \mathcal{R}_1, \gamma\} \longrightarrow M_2 = \{\mathcal{S}_2, \mathcal{A}, \mathcal{P}_2, \mathcal{R}_2, \gamma\}$$

$$\pi_d(y) = \pi^* \left(\arg\min_{x \in \mathcal{S}_1} d_{\sim}(x, y) \right)$$

Bisimulation metrics for policy transfer

$$M_1 = \{\mathcal{S}_1, \mathcal{A}, \mathcal{P}_1, \mathcal{R}_1, \gamma\} \longrightarrow M_2 = \{\mathcal{S}_2, \mathcal{A}, \mathcal{P}_2, \mathcal{R}_2, \gamma\}$$

$$\pi_d(y) = \pi^*\left(\arg\min_{x\in\mathcal{S}_1} d_\sim(x, y)\right)$$

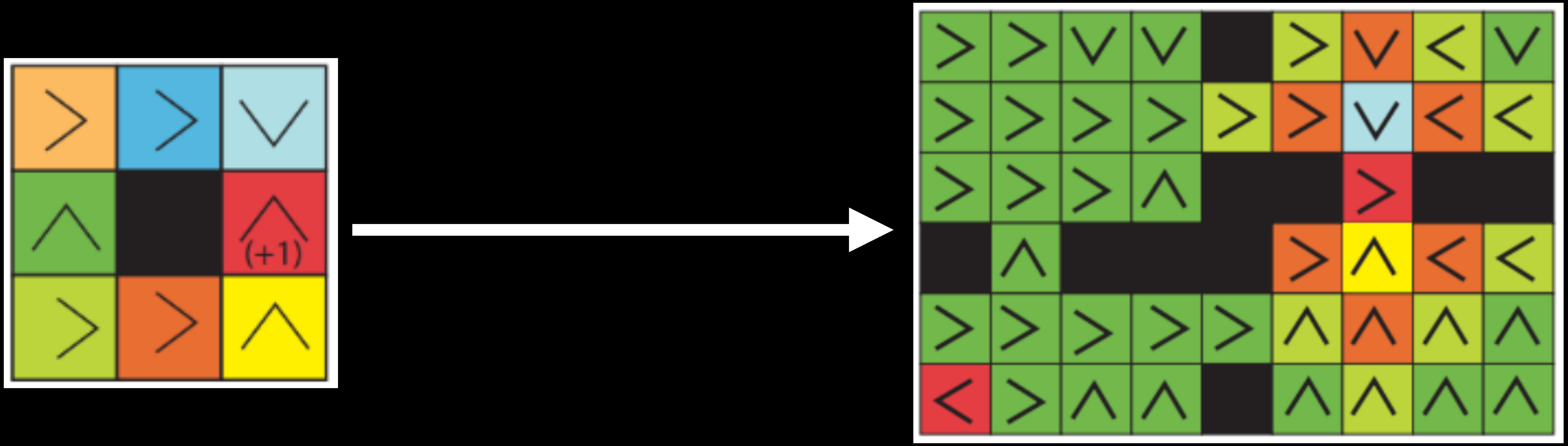
Theorem: $|Q_2^*(y, \pi_d(y)) - V_2^*(y)| \leq 2 \min_{x\in\mathcal{S}_1} d_\sim(x, y)$

Bisimulation metrics for policy transfer

$$M_1 = \{\mathcal{S}_1, \mathcal{A}, \mathcal{P}_1, \mathcal{R}_1, \gamma\} \longrightarrow M_2 = \{\mathcal{S}_2, \mathcal{A}, \mathcal{P}_2, \mathcal{R}_2, \gamma\}$$

$$\pi_d(y) = \pi^* \left(\arg\min_{x \in \mathcal{S}_1} d_\sim(x, y) \right)$$

Theorem: $|Q_2^*(y, \pi_d(y)) - V_2^*(y)| \leq 2 \min_{x \in \mathcal{S}_1} d_\sim(x, y)$



Break!

Bisimulation metrics are great

Bisimulation metrics are great but...

Bisimulation metrics are great but...

1. They're inherently pessimistic and only for π*

$$\mathcal{F}(d)(x,y) = \max_{a\in\mathcal{A}}\{|\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a),\mathcal{P}(y,a))\}$$

Bisimulation metrics are great but...

1. They're inherently pessimistic and only for π*

$$\mathcal{F}(d)(x,y) = \max_{a\in\mathcal{A}}\{|\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a),\mathcal{P}(y,a))\}$$

2. They're expensive to compute

$$\tilde{O}\left(\frac{|\mathcal{S}|^5|\mathcal{A}|\log(\epsilon)}{\log(\gamma)}\right)$$

Bisimulation metrics are great but...

1. They're inherently pessimistic and only for π*

$$\mathcal{F}(d)(x,y) = \max_{a \in \mathcal{A}} \{ |\mathcal{R}(x,a) - \mathcal{R}(y,a)| + \gamma T_K(d)(\mathcal{P}(x,a), \mathcal{P}(y,a)) \}$$

2. They're expensive to compute

$$\tilde{O}\left(\frac{|\mathcal{S}|^5 |\mathcal{A}| \log(\epsilon)}{\log(\gamma)}\right)$$

3. They require a full model and full state enumerability

$$T_K(\mathcal{P}(x,a), \mathcal{P}(y,a))$$

Scalable Methods for Computing State Similarity in Deterministic Markov Decision Processes

Pablo Samuel Castro
Google Brain
psc@google.com

1. They’re inherently pessimistic

1. They're inherently pessimistic

Solution: π-bisimulation!

1. They're inherently pessimistic

Solution: π-bisimulation!

Given an MDP $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$ and policy π, an equiv. relation $E : \mathcal{S} \times \mathcal{S} \to \{0, 1\}$ is a π-**bisimulation relation** if whenever xEt we have:

1. $\mathcal{R}^\pi_x = \mathcal{R}^\pi_y$
2. $\forall c \in \mathcal{S}/_E, \quad \mathcal{P}^\pi_x(c) = \mathcal{P}^\pi_y(c)$

Two states x and y are π-**bisimilar** if there exists a bisimulation relation E such that xEy.

Let $\sim_\pi$ be the maximal bisimulation relation.

1. They're inherently pessimistic

Solution: π-bisimulation!

Definition: A metric d is a π-bisimulation metric if

$$d(x,y) = 0 \iff x \sim_\pi y \qquad \forall x, y \in \mathcal{S}$$

Theorem: The functional $\mathcal{F}^\pi : \mathcal{M} \mapsto \mathcal{M}$ defined as

$$\mathcal{F}^\pi(d)(x,y) = |\mathcal{R}_x^\pi - \mathcal{R}_y^\pi| + \gamma T_K(d)(\mathcal{P}_x^\pi, \mathcal{P}_y^\pi)$$

has a unique fixed point $d_{\sim_\pi}$ and $d_{\sim_\pi}$ is a π-bisimulation metric

Theorem: $|V^\pi(x) - V^\pi(y)| \leq d_{\sim_\pi}(x,y) \quad \forall x, y \in \mathcal{S}$

2. They’re expensive to compute

2. They're expensive to compute

Solution: Sampling!

2. They're expensive to compute

Solution: Sampling!

$$d_n(s,t) = d_{n-1}(s,t), \qquad \forall s \neq s_n, t \neq t_n$$

$$d_n(s_n,t_n) = \max\left[\begin{array}{c} d_{n-1}(s_n,t_n), \\ |\mathcal{R}(s_n,a_n) - \mathcal{R}(t_n,a_n)| + \\ \gamma d_{n-1}(\mathcal{N}(s_n,a_n), \mathcal{N}(t_n,a_n)) \end{array}\right]$$

Theorem: If d_n is updated as above and $d_0 \equiv 0$, then $\lim_{n\to\infty} d_n = d_{\sim_\pi}$ almost surely.

2. They're expensive to compute

Solution: Sampling!

$$d_n(s,t) = d_{n-1}(s,t), \qquad \forall s \neq s_n, t \neq t_n$$

$$d_n(s_n,t_n) = \max\left[\begin{array}{c} d_{n-1}(s_n,t_n), \\ |\mathcal{R}(s_n,a_n) - \mathcal{R}(t_n,a_n)| + \\ \gamma d_{n-1}(\mathcal{N}(s_n,a_n), \mathcal{N}(t_n,a_n)) \end{array}\right]$$

Theorem: If d_n is updated as above and $d_0 \equiv 0$, then $\lim_{n\to\infty} d_n = d_{\sim_\pi}$ almost surely.

Caveat: Only holds for deterministic MDPs.

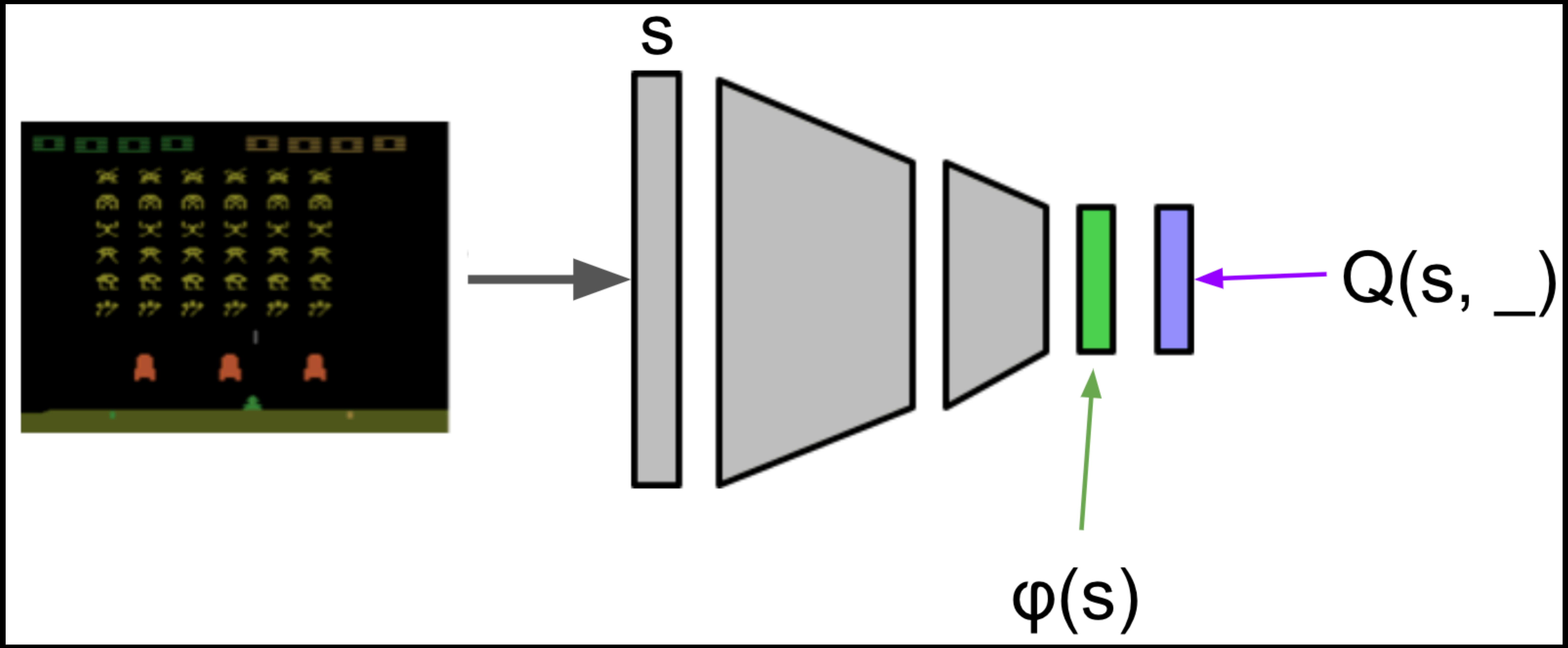
3. They require full state enumerability

3. They require full state enumerability

Solution: Use neural nets!

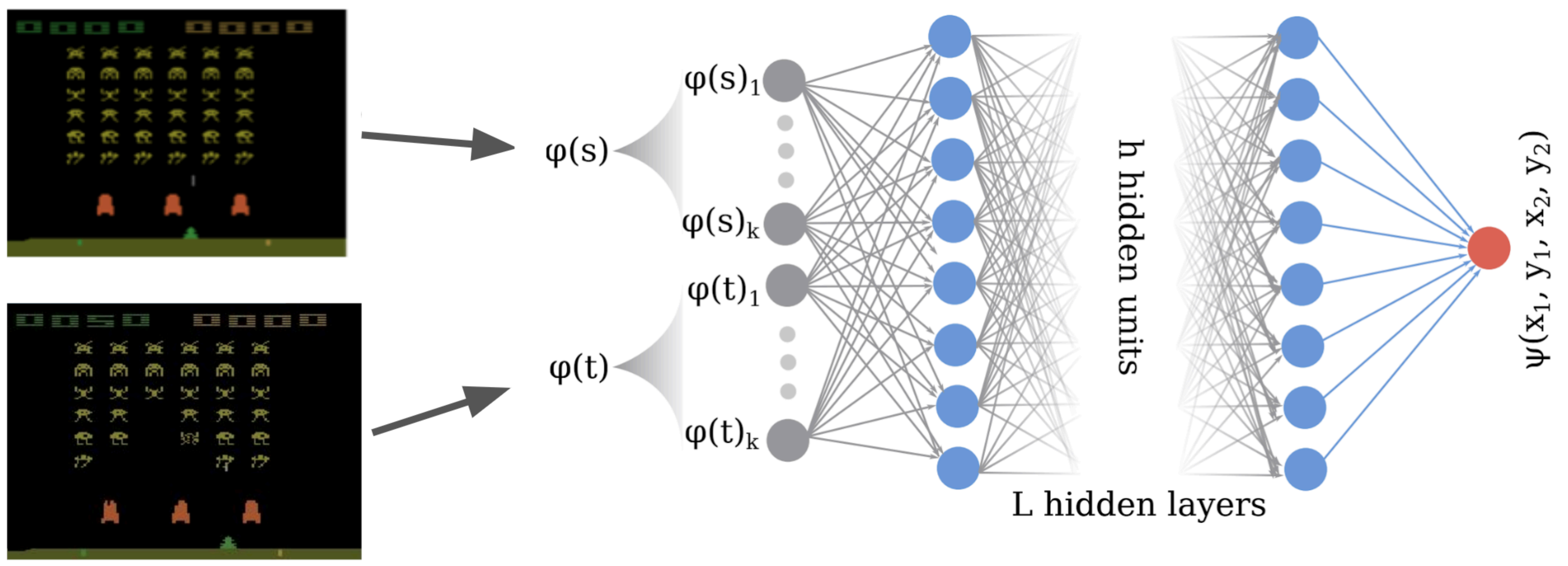
3. They require full state enumerability

Solution: Use neural nets!



3. They require full state enumerability

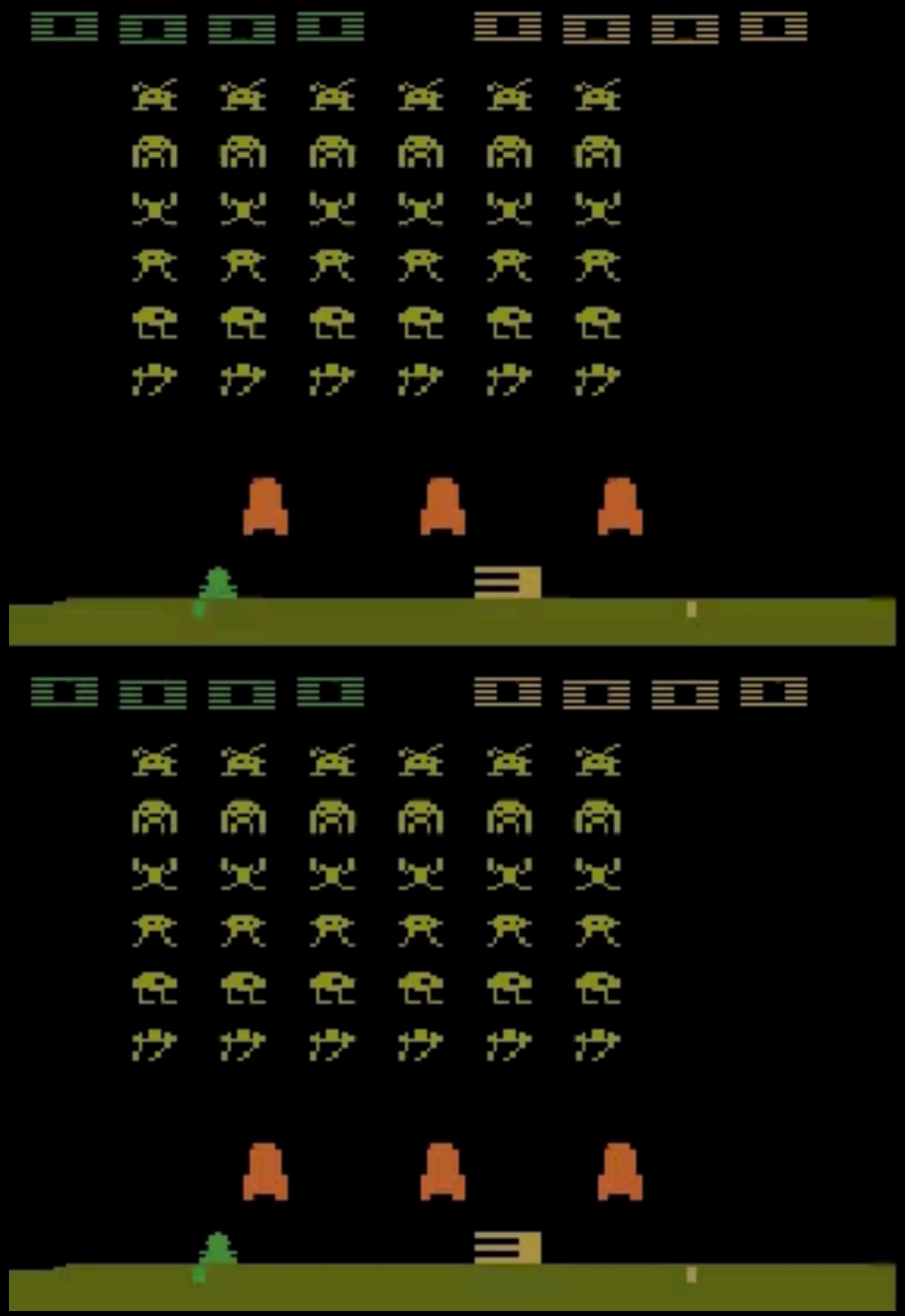
Solution: Use neural nets!



$$\mathbf{T}^{\pi}_{\theta_i^-}(s,t) = |\mathcal{R}(s,\pi(s)) - \mathcal{R}(t,\pi(t))| + \gamma \psi^{\pi}_{\theta_i^-}([\phi(\mathcal{N}(s,\pi(s))), \phi(\mathcal{N}(t,\pi(t)))])$$

$$\mathcal{L}^{(\pi)}_{s,t,a} = \mathbb{E}_{\mathcal{D}}\left(\mathbf{T}^{(\pi)}_{\theta_i^-}(s,t,a) - \psi^{(\pi)}_{\theta_i}([\phi(s),\phi(t)])\right)^2$$

Does it work?



π-bisimulation metrics are great

π-bisimulation metrics are great but...

π-bisimulation metrics are great but…

1. They require a pre-trained agent

2. They assume determinism

Learning Invariant Representations for Reinforcement Learning without Reconstruction

Amy Zhang[*12] **Rowan McAllister**[*3] **Roberto Calandra**[2] **Yarin Gal**[4] **Sergey Levine**[3]

[1]McGill University
[2]Facebook AI Research
[3]University of California, Berkeley
[4]OATML group, University of Oxford

Deep Bisimulation for Control (DBC)

$$J(\phi) = \Big(||\mathbf{z}_i - \mathbf{z}_j||_1 - |r_i - r_j| - \gamma W_2\big(\hat{\mathcal{P}}(\cdot|\bar{\mathbf{z}}_i, \mathbf{a}_i), \hat{\mathcal{P}}(\cdot|\bar{\mathbf{z}}_j, \mathbf{a}_j)\big)\Big)^2,$$

$$W_2(\mathcal{N}(\mu_i, \Sigma_i), \mathcal{N}(\mu_j, \Sigma_j))^2 = ||\mu_i - \mu_j||_2^2 + ||\Sigma_i^{1/2} - \Sigma_j^{1/2}||_{\mathcal{F}}^2$$

Deep Bisimulation for Control (DBC)

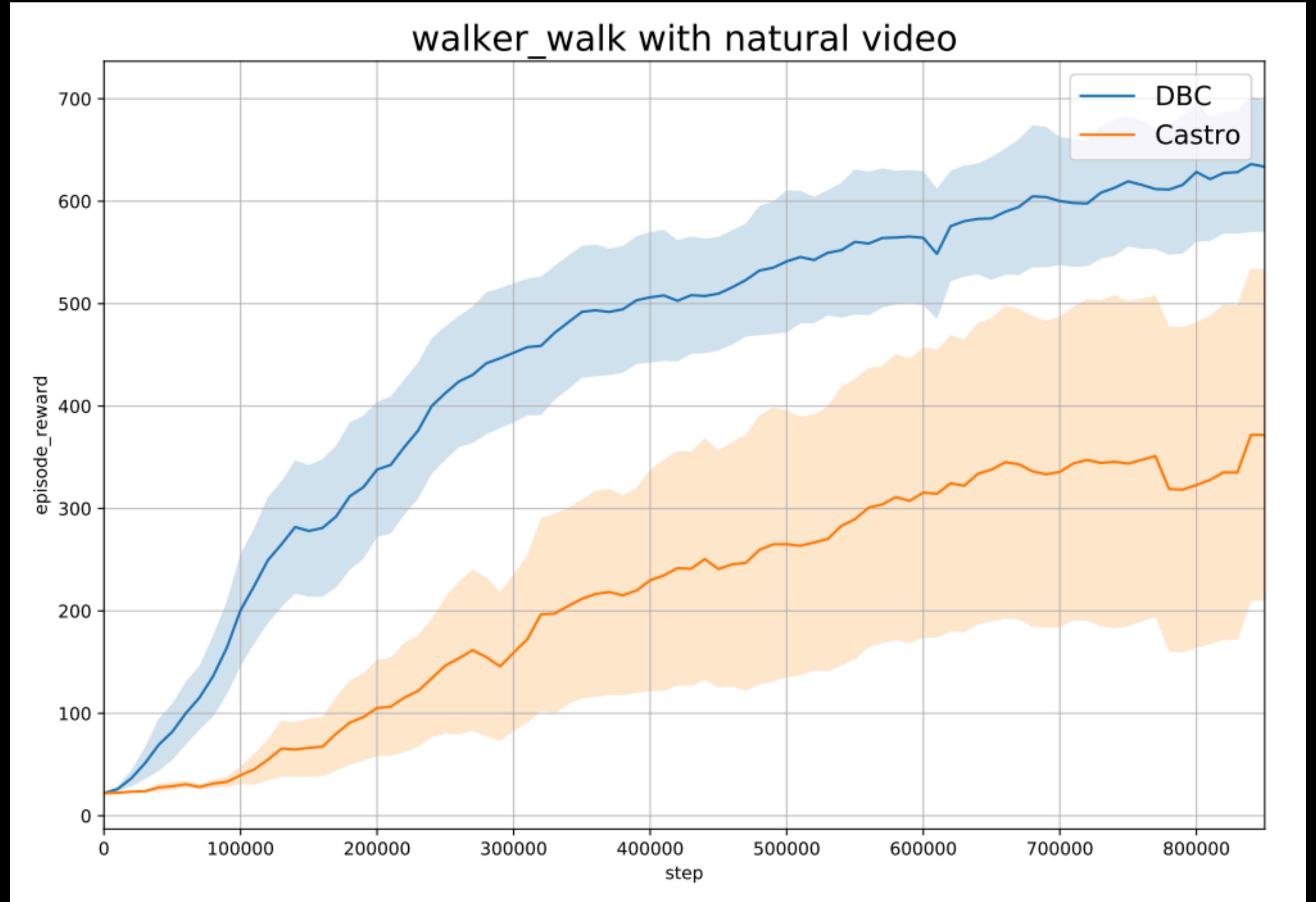
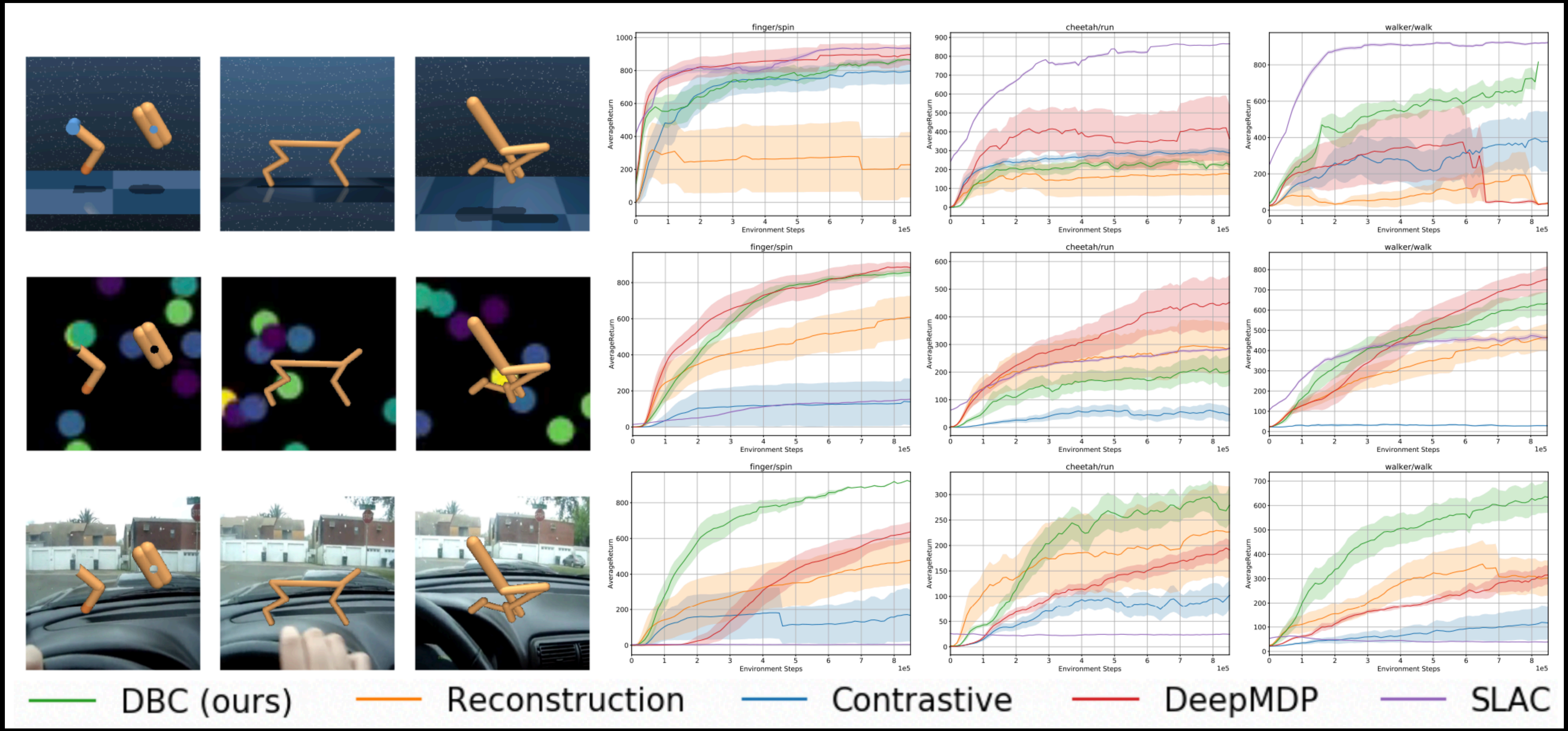


Figure 6: Bisim. results. Blue is DBC and orange is Castro (2020).



MICo: Improved representations via sampling-based state similarity for Markov decision processes

Pablo Samuel Castro* Google Research, Brain Team

Tyler Kastner* McGill University

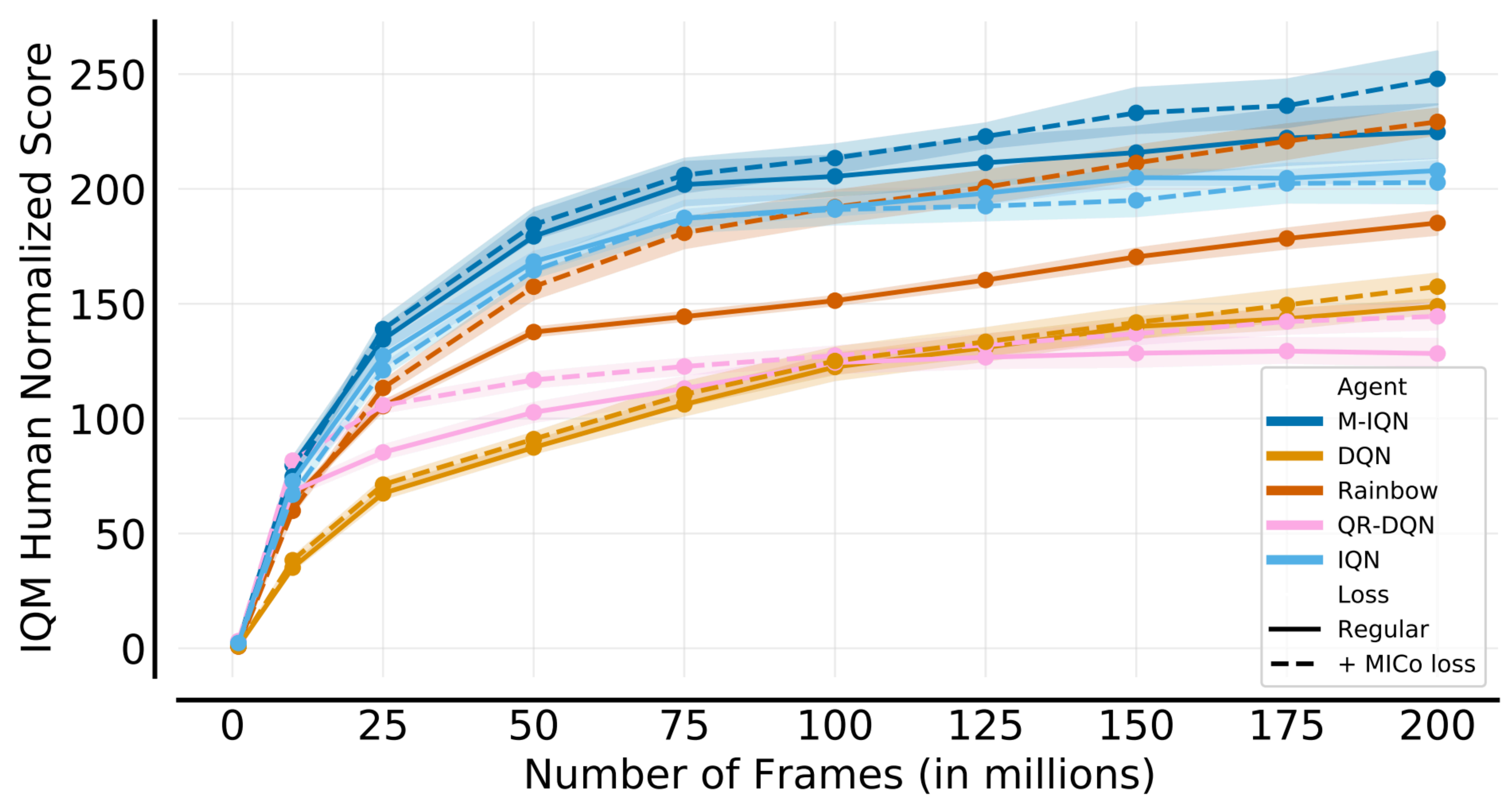
Prakash Panangaden McGill University

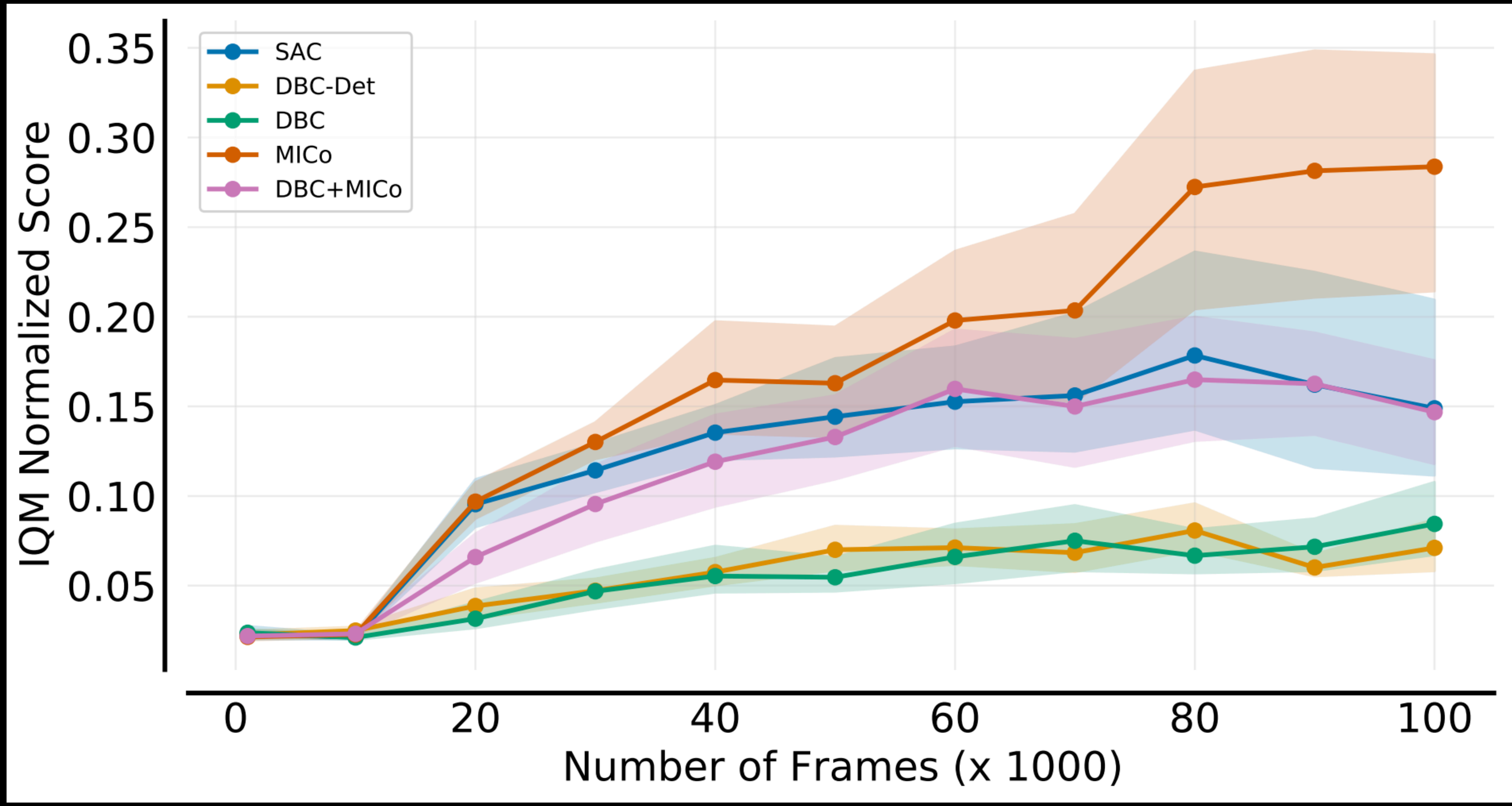
Mark Rowland DeepMind

What is a good distance?

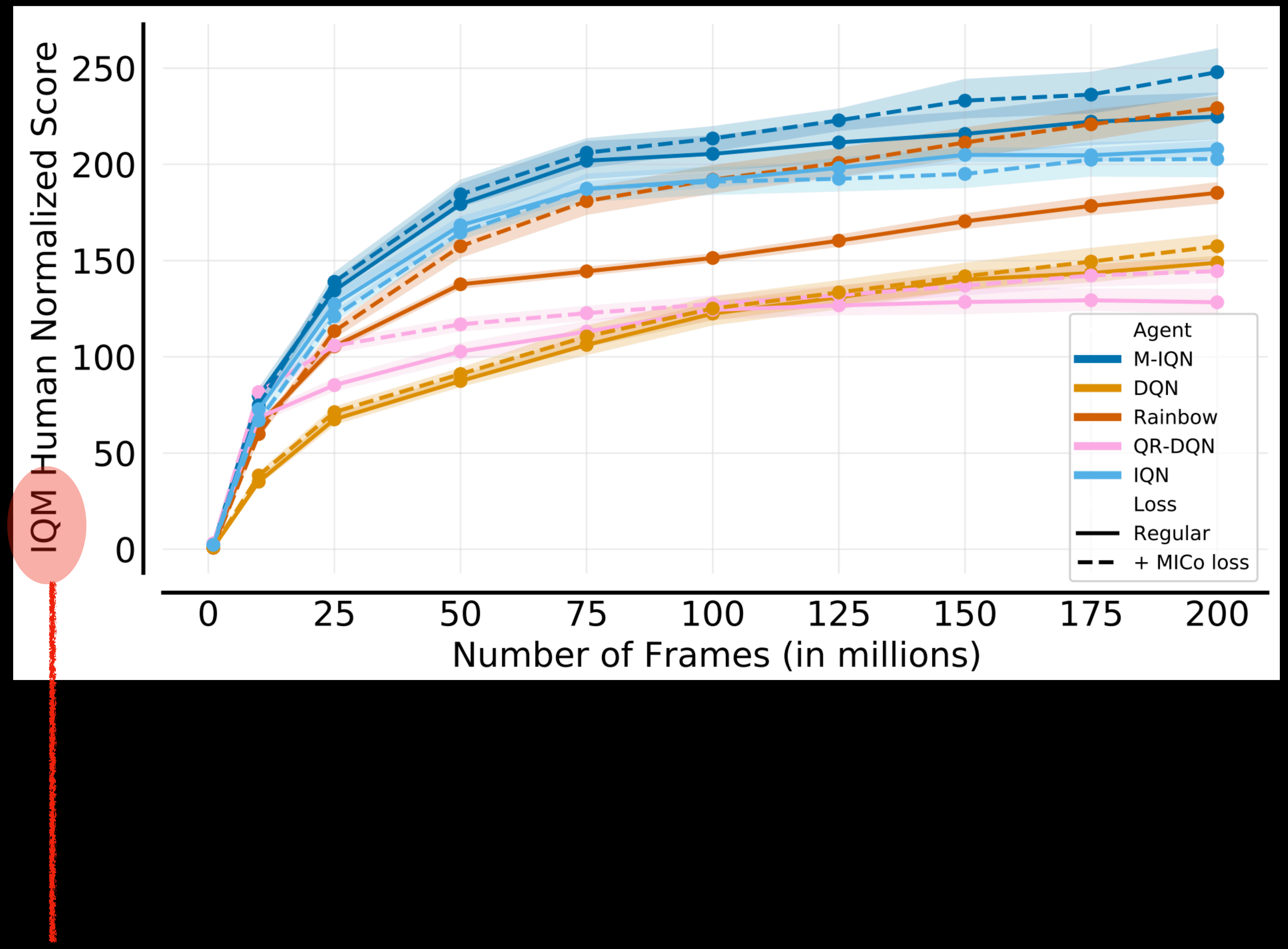
From https://psc-g.github.io/posts/research/rl/mico/

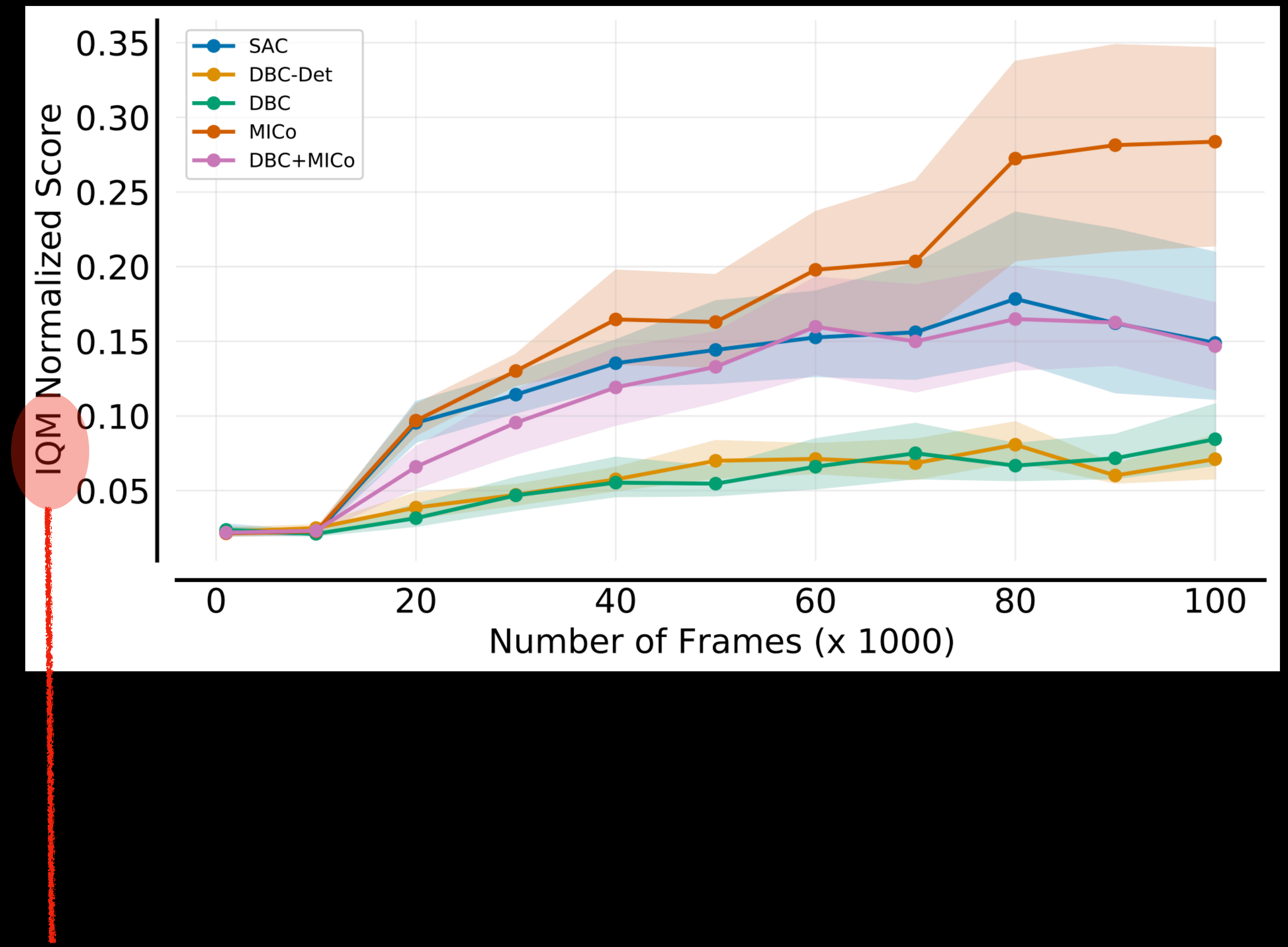
Experimental results





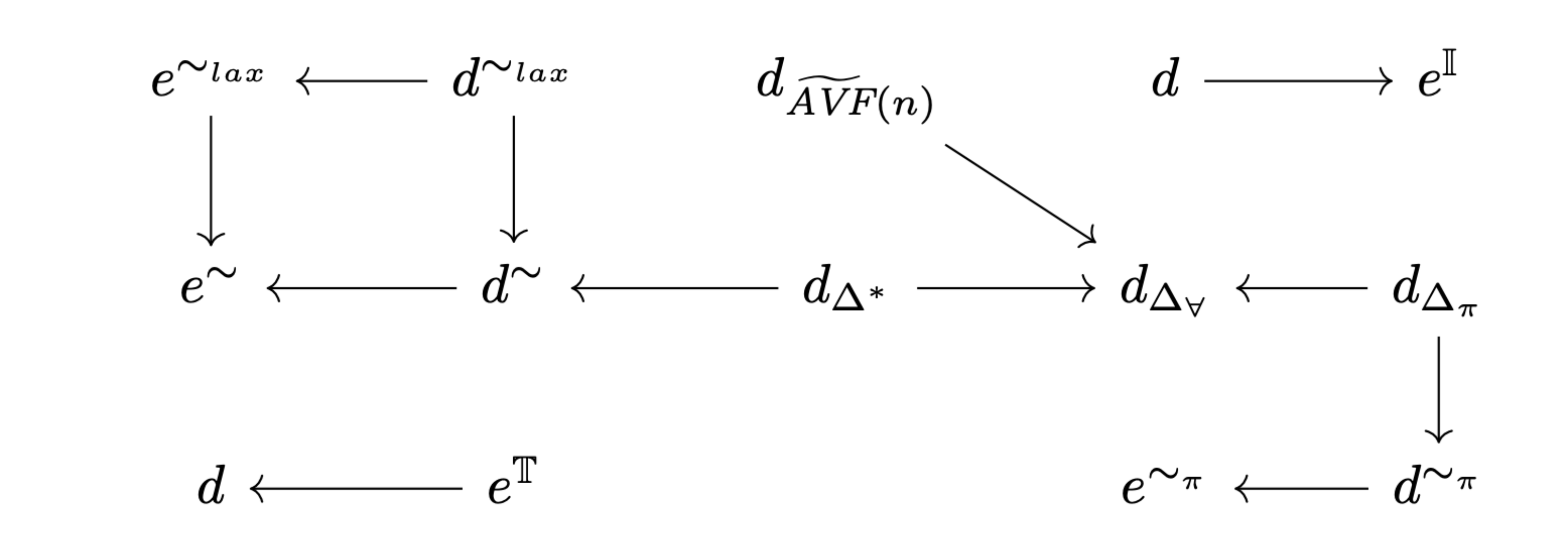
Experimental results





Agarwal, Schwarzer, Castro, Courville, & Bellemare, NeurIPS, 2021

Thanks! Some other recent work:



Metrics and continuity in reinforcement learning

LeLan, Bellemare, & Castro; AAAI 2021

$$d^*(x,y) = \underbrace{\textsc{Dist}\big(\pi^*(x), \pi^*(y)\big)}_{\textbf{(A)}} + \underbrace{\gamma \mathcal{W}_1(d^*)\big(P^{\pi^*}(\cdot\,|\,x), P^{\pi^*}(\cdot\,|\,y)\big)}_{\textbf{(B)}} . \quad (3)$$

Contrastive Behavioural Similarity Embeddings for Generalization in Reinforcement Learning

Agarwal, Machado, Castro, & Bellemare;
ICLR 2021