Bisimulation Metric for Continuous MDPs

COMP 767

Pascale Gourdeau

March 24th, 2017

Motivation

- Very large state space.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.
  - As we refine the space, we have an exponential blowup in the number of states.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.
  - As we refine the space, we have an exponential blowup in the number of states.
- Second thought: aggregate states with similar values.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.
  - As we refine the space, we have an exponential blowup in the number of states.
- Second thought: aggregate states with similar values.
  - May aggregate states that require totally different policies.

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.
  - As we refine the space, we have an exponential blowup in the number of states.
- Second thought: aggregate states with similar values.
  - May aggregate states that require totally different policies.
- Bisimulation!

Motivation

- Very large state space.
- CMDPs $\rightarrow$ *continuous state space*.
- First thought: discretization. Inconvenient because:
  - We may aggregate states that are "physically" close together, but that have very different state values.
  - As we refine the space, we have an exponential blowup in the number of states.
- Second thought: aggregate states with similar values.
  - May aggregate states that require totally different policies.
- Bisimulation!
  - Develop a notion of *behavioural distance* between states and aggregate states that are "close" to each other.

What is a Bisimulation Metric?

Definition

Let (S, Σ, A, P, r) be an MDP satisfying certain assumptions[1]. An equivalence relation R on S is a *bisimulation relation* if and only if it satisfies

$$sRs' \iff \begin{array}{l} \text{for every } a \in A,\ r_s^a = r_{s'}^a \text{ and} \\ \text{for every } X \in \Sigma(R),\ P_s^a(X) = P_{s'}^a(X). \end{array}$$

[1] Outlined in the appendix

What is a Bisimulation Metric?

Definition

Let (S, Σ, A, P, r) be an MDP satisfying certain assumptions[1]. An equivalence relation R on S is a *bisimulation relation* if and only if it satisfies

$$sRs' \iff \begin{array}{l} \text{for every } a \in A,\ r_s^a = r_{s'}^a \text{ and} \\ \text{for every } X \in \Sigma(R),\ P_s^a(X) = P_{s'}^a(X). \end{array}$$

Two states are *bisimilar* ($s \sim s'$) if and only if there exists a bisimulation relation R such that sRs'.

[1] Outlined in the appendix

What is a Bisimulation Metric?

Definition

Let (S, Σ, A, P, r) be an MDP satisfying certain assumptions[1]. An equivalence relation R on S is a *bisimulation relation* if and only if it satisfies

$$sRs' \iff \begin{array}{l}\text{for every } a \in A,\ r_s^a = r_{s'}^a \text{ and} \\ \text{for every } X \in \Sigma(R),\ P_s^a(X) = P_{s'}^a(X).\end{array}$$

Two states are *bisimilar* ($s \sim s'$) if and only if there exists a bisimulation relation R such that sRs'.

Definition

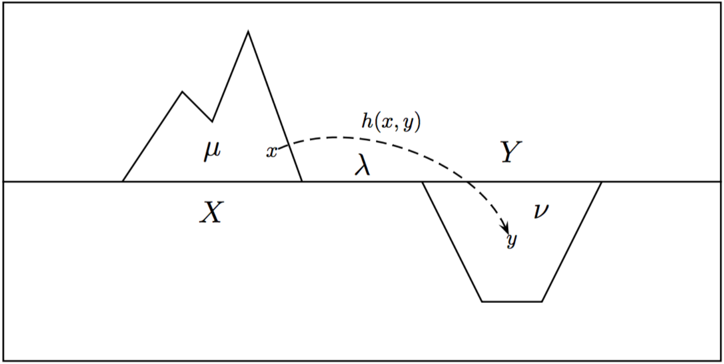
A pseudometric $\rho : S \times S \to [0, +\infty)$ on the states of an MDP is a *bisimulation metric* if it satisfies

$$\rho(s, s') = 0 \iff s \sim s'.$$

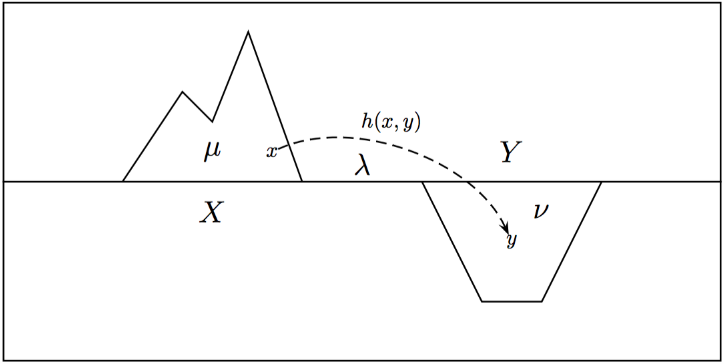
---

[1] Outlined in the appendix

Kantorovich Metric



Kantorovich Metric



Goal: determine a plan for transferring all the mass from X to Y while minimizing the cost.

Building a Bisimulation Metric for CMDPs

MDP M $\mathrm{img}(r) \subseteq [0,1]$

and discount factor $\gamma \in (0,1)$

Kantorovich Metric

$\mathcal{K} : \mathfrak{lsc}_m \times \mathcal{D}(S) \times \mathcal{D}(S) \to \mathbb{R}^+$

Building a Bisimulation Metric for CMDPs

$$
\begin{tikzcd}
\boxed{\begin{array}{l}\text{MDP } M \quad \mathrm{img}(r) \subseteq [0,1] \\ \text{and discount factor } \gamma \in (0,1)\end{array}} \arrow[dr] & & \boxed{\begin{array}{c}\text{Kantorovich Metric} \\ \mathcal{K} : \mathfrak{lsc}_m \times \mathcal{D}(S) \times \mathcal{D}(S) \to \mathbb{R}^+\end{array}} \arrow[dl] \\
& \boxed{\text{Map } F_c : \mathfrak{lsc}_m \to \mathfrak{lsc}_m} &
\end{tikzcd}
$$

$$F(h)(s, s') = \max_{a \in A} \left[(1-c)|r_s^a - r_{s'}^a| + c T_K(h)(P_s^a, P_{s'}^a) \right]$$

Building a Bisimulation Metric for CMDPs

MDP M $\mathrm{img}(r) \subseteq [0,1]$
and discount factor $\gamma \in (0,1)$

Kantorovich Metric
$\mathcal{K} : \mathfrak{lsc}_m \times \mathcal{D}(S) \times \mathcal{D}(S) \to \mathbb{R}^+$

Map $F_c : \mathfrak{lsc}_m \to \mathfrak{lsc}_m$

BFPT

Bisimulation metric
$\rho^* = F_c(\rho^*)$

$$F(h)(s,s') = \max_{a \in A} \left[(1-c)|r_s^a - r_{s'}^a| + cT_K(h)(P_s^a, P_{s'}^a) \right]$$

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').
- For $\lceil \frac{\ln \delta}{\ln c} \rceil$ iterations:

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').
- For $\lceil \frac{\ln \delta}{\ln c} \rceil$ iterations:
  - For each tuple (s, s', a)

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').
- For $\lceil \frac{\ln \delta}{\ln c} \rceil$ iterations:
  - For each tuple (s, s', a)
    - $TK_a(s, s') = Hungarian\_alg(\rho, P_s^a, P_{s'}^a)$

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').
- For $\lceil \frac{\ln \delta}{\ln c} \rceil$ iterations:
  - For each tuple (s, s', a)
    - $TK_a(s, s') = Hungarian\_alg(\rho, P_s^a, P_{s'}^a)$
- For each state pair (s, s')

Approximation Algorithm

For a finite MDP, we can use Banach's fixed point theorem to derive an approximation algorithm:
Given tolerance δ,

- Initialize $\rho(s, s') = 0$ for all state pairs (s, s').
- For $\lceil \frac{\ln \delta}{\ln c} \rceil$ iterations:
  - For each tuple (s, s', a)
    - $TK_a(s, s') = Hungarian\_alg(\rho, P_s^a, P_{s'}^a)$
- For each state pair (s, s')
  - $\rho(s, s') = \max_a[(1-c)|r_s^a - r_{s'}^a| + cTK_a(s, s')]$

Approximation Algorithm

- If we have a continuous MDP, we need a finite set of states to approximate the result.

[2] all points are at distance ϵ apart and $U \subseteq \bigcup_{x \in X} B(x, \epsilon)$.

Approximation Algorithm

- If we have a continuous MDP, we need a finite set of states to approximate the result.
- Assume we are provided with $U \subseteq S$, where U is a countable and dense in S, and a metric d on $U \times U$.

[2] all points are at distance ϵ apart and $U \subseteq \bigcup_{x \in X} B(x, \epsilon)$.

Approximation Algorithm

- If we have a continuous MDP, we need a finite set of states to approximate the result.
- Assume we are provided with $U \subseteq S$, where U is a countable and dense in S, and a metric d on $U \times U$.
- We construct a subset X of U such that X is an ϵ-net[2]

[2] all points are at distance ϵ apart and $U \subseteq \bigcup_{x \in X} B(x, \epsilon)$.

Approximation Algorithm

- If we have a continuous MDP, we need a finite set of states to approximate the result.
- Assume we are provided with $U \subseteq S$, where U is a countable and dense in S, and a metric d on $U \times U$.
- We construct a subset X of U such that X is an ϵ-net[2]
  - Start with $X = \{s\}$ for some $s \in U$, then pick $s' \in U \setminus X$ such that maximizes $\min\{d(x, s) : x \in X\}$.

[2] all points are at distance ϵ apart and $U \subseteq \bigcup_{x \in X} B(x, \epsilon)$.

Approximation Algorithm

- If we have a continuous MDP, we need a finite set of states to approximate the result.
- Assume we are provided with $U \subseteq S$, where U is a countable and dense in S, and a metric d on $U \times U$.
- We construct a subset X of U such that X is an ϵ-net[2]
  - Start with $X = \{s\}$ for some $s \in U$, then pick $s' \in U \setminus X$ such that maximizes $\min\{d(x, s) : x \in X\}$.
  - In practice, we use sampling to find such an s' (if U is countably infinite).

[2] all points are at distance ϵ apart and $U \subseteq \bigcup_{x \in X} B(x, \epsilon)$.

Bisimulation Metrics are Optimal Value Functions

$$|V^*(s) - V^*(s')| \leq \frac{1}{1-c}\rho_c^*(s, s')$$

Bisimulation Metrics are Optimal Value Functions

$$|V^*(s) - V^*(s')| \leq \frac{1}{1-c}\rho_c^*(s, s')$$

- The closer the distance (relative to bisimilarity) the more likely they will share optimal value functions (and hence policies).

Bisimulation Metrics are Optimal Value Functions

$$|V^*(s) - V^*(s')| \leq \frac{1}{1-c}\rho_c^*(s, s')$$

- The closer the distance (relative to bisimilarity) the more likely they will share optimal value functions (and hence policies).
- Aggregating states that are close in behaviour (w.r.t. bisimilarity) implies aggregating states with similar value functions.

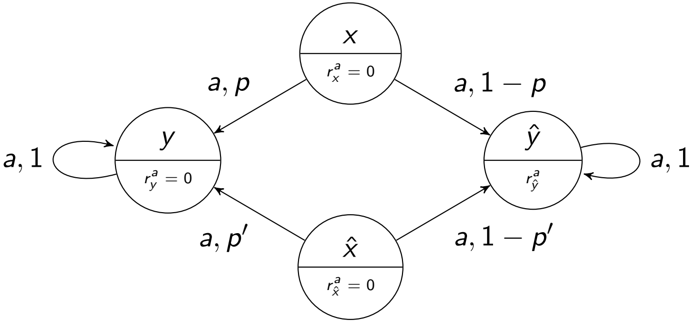
Bisimulation Metrics are Optimal Value Functions

$$|V^*(s) - V^*(s')| \leq \frac{1}{1-c}\rho_c^*(s, s')$$

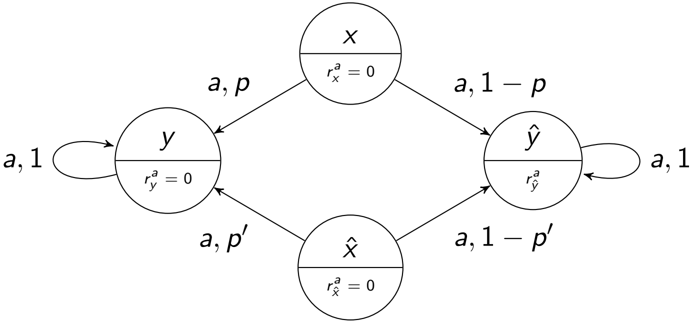
- The closer the distance (relative to bisimilarity) the more likely they will share optimal value functions (and hence policies).
- Aggregating states that are close in behaviour (w.r.t. bisimilarity) implies aggregating states with similar value functions.
- In fact, for a given MDP M, there exists a coupling K^* of M with itself, such that $\rho_c^* = V_c^*(K^*)$.

Is there time for an example?

Example of bisimulation metric on MDPs



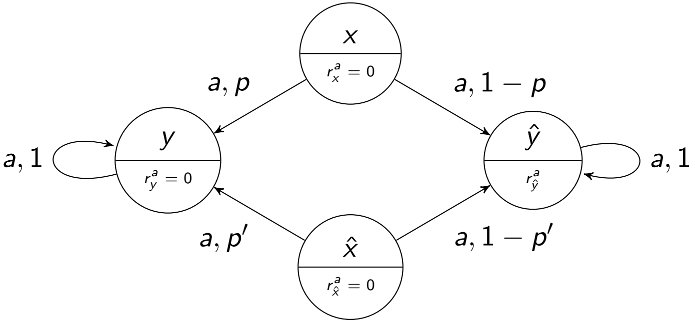
Example of bisimulation metric on MDPs



Note that:

1. There is only one action,

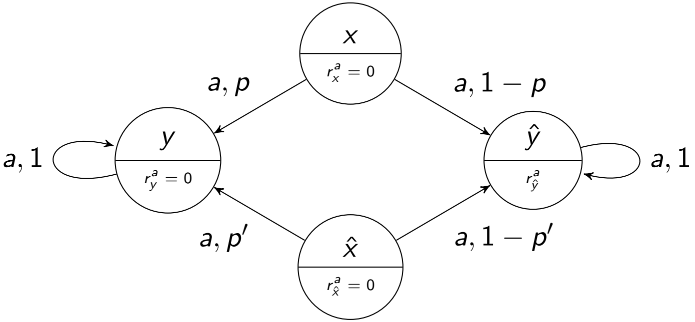
Example of bisimulation metric on MDPs



Note that:

1. There is only one action,
2. $T_K(\rho^*)(\delta_x, \delta_y) = \rho^*(x, y)$,

Example of bisimulation metric on MDPs



Note that:

1. There is only one action,
2. $T_K(\rho^*)(\delta_x, \delta_y) = \rho^*(x, y)$,
3. $F(\rho^*)(s, s') = (\rho^*)(s, s')$ and ρ^* is unique.

Questions?

Appendix

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite set of actions,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite set of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite set of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite set of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,

Definition of Continuous MDPs

Definition

A continuous MDP is a tuple (S, Σ, A, P, r) where

- (S, Σ) is a measurable space,
- A is a finite set of actions,
- $r : S \times A \to \mathbb{R}$ is a measurable reward function,
- $P : S \times A \times \Sigma \to [0, 1]$ is a labelled stochastic transition kernel:
  - $\forall a \in A$, $\forall s \in S$, $P(s, a, \cdot) : \Sigma \to [0, 1]$ is a probability measure,
  - $\forall a \in A$, $\forall X \in \Sigma$, $P(\cdot, a, X) : S \to [0, 1]$ is a measurable function.

Bisimulation

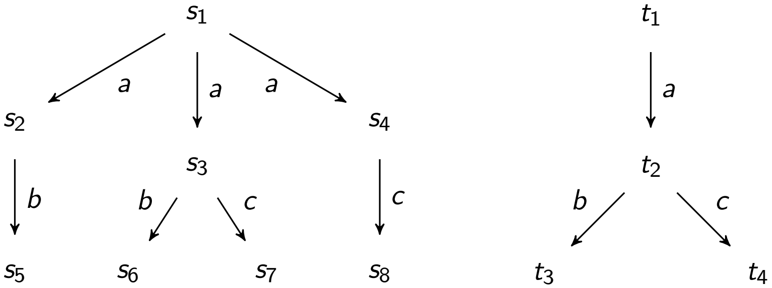
- Originally due to Park (1981) and extended to probabilistic systems by Larsen and Skou (1991).
- Abstract notion of *behavioural equivalence* between processes.
- If I have two bisimilar systems, I can replace one by the other and no test (sequence of experiments) can distinguish them.

Bisimulation as a game

- Two-way simulation
- Bisimulation

Bisimulation as a game

- Two-way simulation
- Bisimulation



Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.

Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.
- $\mathrm{img}(r) \subseteq [0, 1]$

Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.
- $\text{img}(r) \subseteq [0, 1]$
- For each $a \in A$, $r(\cdot, a)$ is continuous on S.

Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.
- $\mathrm{img}(r) \subseteq [0, 1]$
- For each $a \in A$, $r(\cdot, a)$ is continuous on S.
- For each $a \in A$, P_s^a is continuous as a function of s.

Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.
- $\text{img}(r) \subseteq [0, 1]$
- For each $a \in A$, $r(\cdot, a)$ is continuous on S.
- For each $a \in A$, P_s^a is continuous as a function of s.

Bisimulation on CMDPs

MDP assumptions:

- S is a Polish space with its Borel σ-algebra Σ.
- $\mathrm{img}(r) \subseteq [0, 1]$
- For each $a \in A$, $r(\cdot, a)$ is continuous on S.
- For each $a \in A$, P_s^a is continuous as a function of s.

Definition

Let (S, Σ, A, P, r) be an MDP satisfying the above assumptions. An equivalence relation R on S is a *bisimulation relation* if and only if it satisfies

$$sRs' \iff \begin{array}{l} \text{for every } a \in A,\ r_s^a = r_{s'}^a \text{ and} \\ \text{for every } X \in \Sigma(R),\ P_s^a(X) = P_{s'}^a(X). \end{array}$$

Appendix – Metric Space Definitions

Definition

A *metric* on a set X is a map $d : X \times X \to [0, \infty)$ such that for all $x, y, z \in X$:

1. $x = y \iff d(x, y) = 0$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Definition

We say that the tuple (X,d) where X is a set with a metric $d : X \times X \to [0, \infty)$ is a *metric space*.

Appendix – Properties of Metric Space

Definition

A metric space (X, d) is said to be *separable* if it has some countable dense subset.

Definition

A metric space (X, d) is said to be *complete* if every Cauchy sequence converges.

Definition

A metric space (X, d) is said to be *Polish* if it is both separable and complete.

Kantorovich Metric Definition

Definition

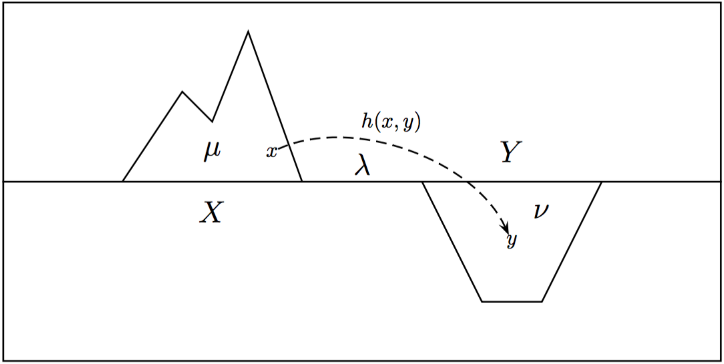
Let (S, d) be a Polish metric space, h a bounded pseudo-metric on S that is lower semi-continuous on $S \times S$ and $Lip(h)$ the set of all bounded functions $f : S \to \mathbb{R}$ that are measurable w.r.t. $\mathcal{B}(S)$ and satisfy the Lipschitz condition $f(x) - f(y) \leq h(x, y)$ for every $x, y \in S$. Given two probability measures P and Q, the *Kantorovich distance* $T_K(h)$ is defined by

$$T_K(h)(P, Q) = \sup_{f \in Lip(h)} (P(f) - Q(f)) = \sup_{f \in Lip(h)} \left(\int f dP - \int f dQ \right)$$

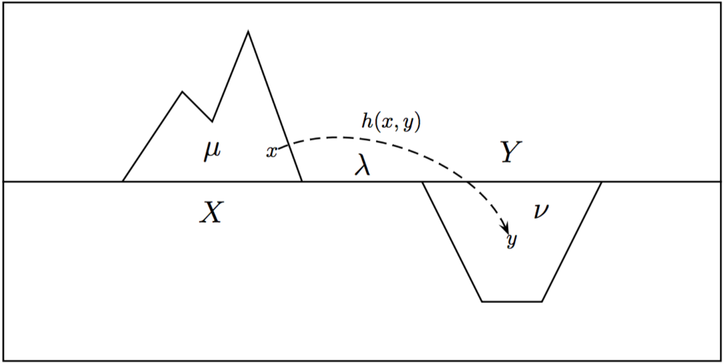
Theorem (Kantorovich-Rubinstein Duality Theorem)

$$T_K(h)(P, Q) = \sup_{f \in Lip(h)} (P(f) - Q(f)) = \inf_{\lambda \in \Lambda(P,Q)} h(\lambda)$$

Kantorovich Metric



Kantorovich Metric



Goal: determine a plan for transferring all the mass from X to Y while minimizing the cost.

Kantorovich Metric

Lemma

Let $\mathfrak{lsc}_{\mathfrak{m}}$ be the set of bounded pseudometrics on S which are lower semi-continuous on $S \times S$, $h \in \mathfrak{lsc}_{\mathfrak{m}}$ and $Rel(h)$ be the kernel of h. Then

$$T_K(h)(P, Q) = 0 \iff P(X) = Q(X)\ \forall X \in \Sigma(Rel(h)) \ .$$