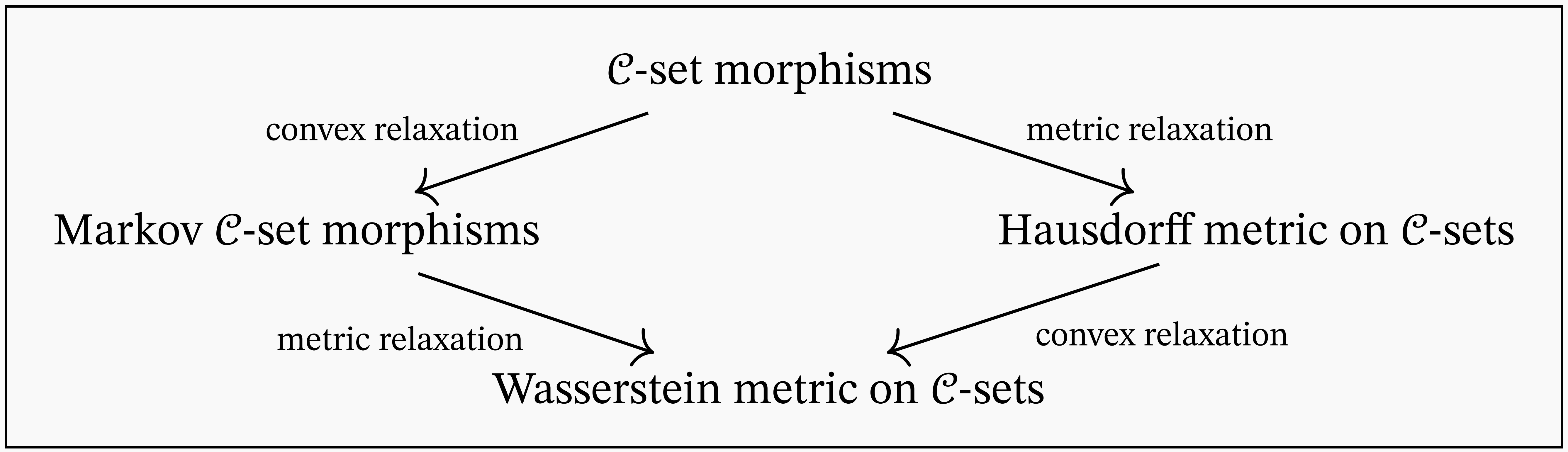
Optimal transport on graphs and other structured data



Evan Patterson
Stanford University, Statistics Department

Special Session on Applied Category
AMS Fall Western Sectional Meeting, November 9-10, 2019

Graph matching

Is there a correspondence between two graphs?

- Formalized in different ways:
    - Graph homomorphism
    - Graph isomorphism
    - Maximum common subgraph
    - Graph edit distance
    - ...
- Exact formulations are mostly NP-hard
- Exact and inexact matching heavily studied by computer scientists

Conte et al 2004; Emmert-Streib et al 2016; more...

Relaxation of graph matching

- Hardness due to combinatorics of matching
- Can we relax the matching problem into an easier one?
- A common method for relaxing matching problems is optimal transport
- Numerous efforts to match graphs using optimal transport

Aflalo et al 2015; Alvarez-Melis et al 2017; Vayer et al 2018

Optimal transport

Monge problem (1781): Given measures $\mu \in \mathrm{Prob}(X)$ and $\nu \in \mathrm{Prob}(Y)$ and cost function $c : X \times Y \to \mathbb{R}$,

$$\underset{\substack{T:X\to Y:\\ T\mu=\nu}}{\text{minimize}} \int_X c(x, T(x))\, \mu(dx)$$

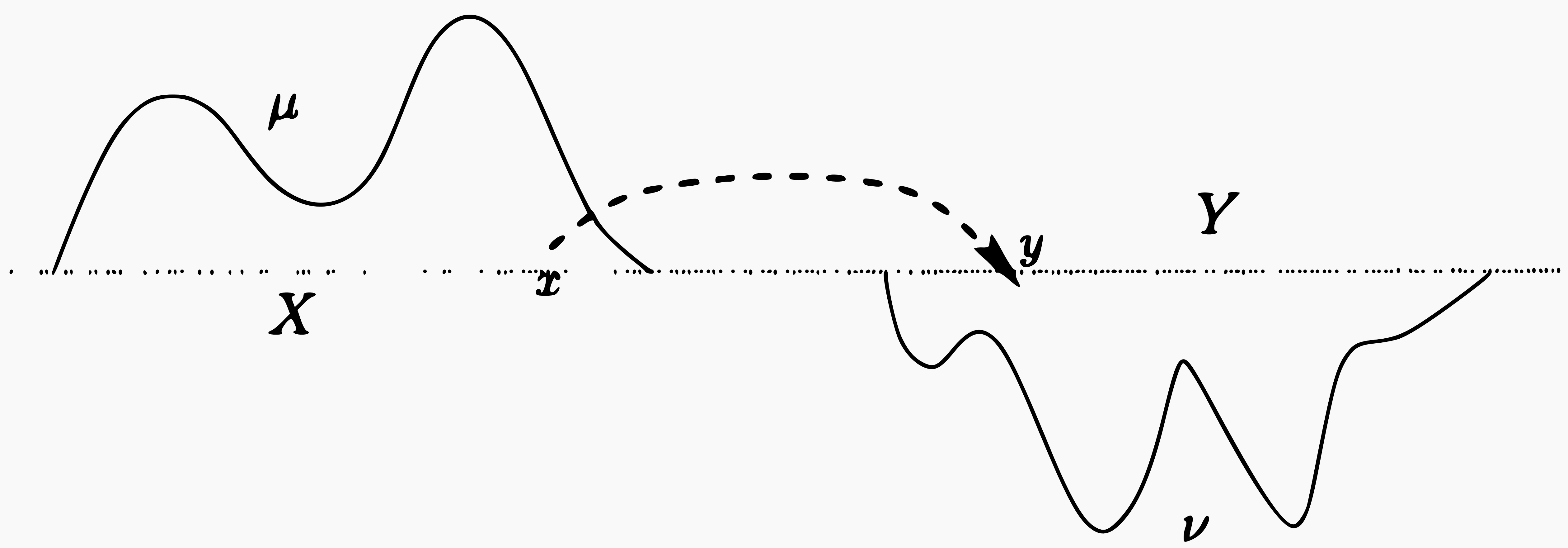


Image: Villani 2003

Optimal transport

Monge problem is combinatorial and nonconvex.

Kantorovich's relaxation (1942): Replace deterministic map with probabilistic coupling:

$$\underset{\pi \in \mathrm{Coup}(X,Y)}{\text{minimize}} \int_{X \times Y} c(x, y)\, \pi(dx, dy)$$

where

$$\mathrm{Coup}(X, Y) := \{\pi \in \mathrm{Prob}(X \times Y) : \mathrm{proj}_X\, \pi = \mu,\ \mathrm{proj}_Y\, \pi = \nu\}.$$

New problem is convex, in fact a linear program.

Villani 2003; Villani 2009; more...

Wasserstein metric in graph matching

When cost is a metric $d : X \times X \to \mathbb{R}$, we get the **Wasserstein metric** on $\mathrm{Prob}(X)$:

$$W_p(\mu, \nu) := \inf_{\pi \in \mathrm{Coup}(\mu,\nu)} \left(\int_{X \times X} d(x, x')^p \, \pi(dx, dx') \right)^{1/p}, \qquad 1 \leq p < \infty.$$

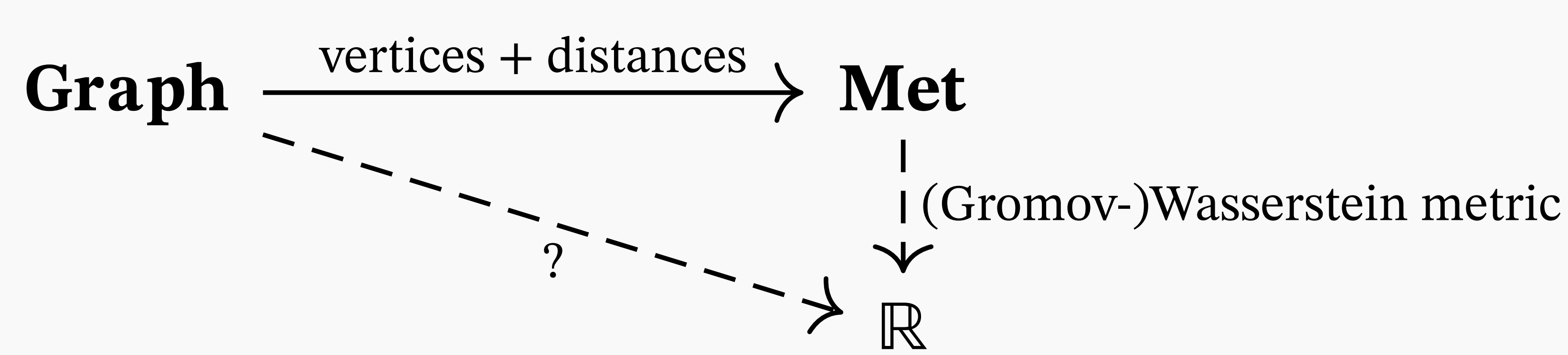
Applied to graph matching in two ways:

1. Featurize vertices of both graphs in common metric space, then compute Wasserstein distance.
2. Convert graphs into distinct metric spaces on vertices (via shortest path metric), then compute Gromov-Wasserstein distance.

Problem: Discards significant information about edges.

Wasserstein metric on graphs?

Goal: Construct a Wasserstein-style metric on graphs that respects both vertices and edges, in a sense to be defined.



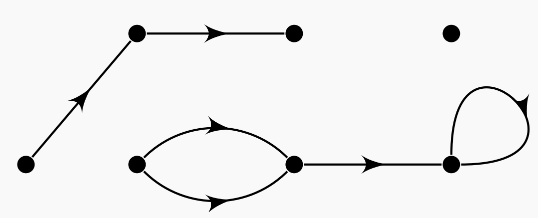
So, this informal diagram should not commute!

In fact, no reason to restrict to graphs; generalizing to $\mathcal{C}$-sets even points the way towards a solution.

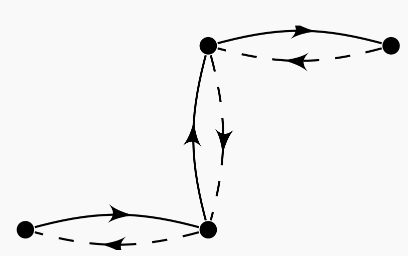
Graphs and other C-sets

Recall: For $\mathcal{C}$ a small category, a $\mathcal{C}$-set is a functor $X : \mathcal{C} \to \mathbf{Set}$.
The category of $\mathcal{C}$-sets is the functor category $[\mathcal{C}, \mathbf{Set}]$.

Example: When $\mathcal{C} = \left\{ E \overset{\text{src}}{\underset{\text{tgt}}{\rightrightarrows}} V \right\}$, a $\mathcal{C}$-set is a (directed) graph.



Example: When $\mathcal{C} = \left\{ \text{inv} \circlearrowright E \overset{\text{src}}{\underset{\text{tgt}}{\rightrightarrows}} V \;\middle|\; \begin{array}{r} \text{inv}^2 = 1_E \\ \text{inv} \cdot \text{src} = \text{tgt} \\ \text{inv} \cdot \text{tgt} = \text{src} \end{array} \right\}$, a $\mathcal{C}$-set is a symmetric graph.



Nearly the same as an undirected graph.

Graphs and other C-sets

Other examples

- Reflexive and symmetric reflexive graphs
- Bipartite graphs
- Hypergraphs
- Higher-dimensional (semi-)simplicial sets

Reyes et al 2004; Spivak 2009; more...

For applications: Attributes can be modeled in C, to get vertex-attributed graphs, edge-attributed graphs, and so on.

Functorial semantics of C-sets

A $\mathcal{C}$-set in a category $\mathcal{S}$ is a functor $X : \mathcal{C} \to \mathcal{S}$.

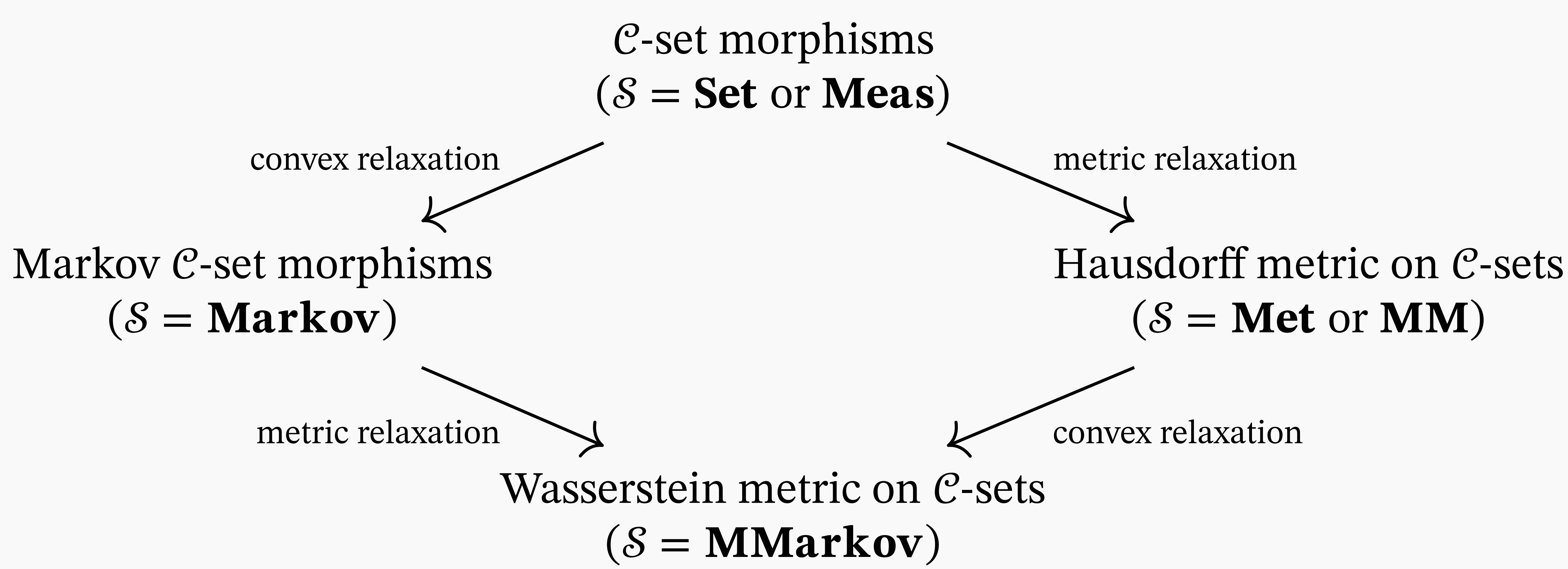
For us, useful categories $\mathcal{S}$ include:

- $\mathbf{Set}$, the category of sets and functions
- $\mathbf{Meas}$, the category of measurable spaces and measurable functions
- $\mathbf{Meas}_*$, the category of measure spaces and measurable functions
- $\mathbf{Met}$, the category of metric spaces and functions
- $\mathbf{MM}$, the category of metric measure spaces (mm spaces) and measurable functions
- $\mathbf{Markov}$, the category of measurable spaces and Markov kernels

Leads to $\mathcal{C}$-sets, measurable $\mathcal{C}$-spaces, measure $\mathcal{C}$-spaces, metric $\mathcal{C}$-spaces, and so on.

Project overview

Explore relaxations of the notion of homomorphism (natural transformation):



In this talk, I give a sketch. A systematic development is in the paper.

The category of Markov kernels

A Markov kernel $M : X \to Y$ is a measurable assignment of each point $x \in X$ to a probability measure $M(x) \in \mathrm{Prob}(Y)$.

Other names for Markov kernels:

- Probability kernels
- Stochastic kernels
- Stochastic relations

There is a category **Markov** of measurable spaces and Markov kernels.

Čencov 1982; Panangaden 1998; Fritz 2019; more...

Markov kernels and couplings

Let $\mu \in \mathrm{Prob}(X)$ and $\nu \in \mathrm{Prob}(Y)$.

For any coupling $\pi \in \mathrm{Coup}(\mu, \nu)$, the disintegration (conditional probability distribution) $M : X \to Y$ satisfies

$$\mu \cdot M = \nu.$$

Conversely, for any Markov kernel $M : X \to Y$ with $\mu \cdot M = \nu$, there is a product

$$\mu \otimes M \in \mathrm{Coup}(\mu, \nu).$$

Markov kernels and optimal transport

In fact, this correspondence is functorial.

Proposition (folkore?): Under regularity conditions, there is an isomorphism between

- the category of probability spaces and couplings, with composition defined by the gluing lemma, and
- the category of probability spaces and measure-preserving Markov kernels (defined up to almost-everywhere equality).

Interpretation: Markov kernels allow a "directed" version of optimal transport.

Markov morphisms of C-sets

$\mathbf{Meas}$ embeds in $\mathbf{Markov}$ as the deterministic Markov kernels:

$$\mathcal{M} : \mathbf{Meas} \hookrightarrow \mathbf{Markov}.$$

Induces a relaxation functor by post-composition:

$$\mathcal{M}_* : [\mathcal{C}, \mathbf{Meas}] \to [\mathcal{C}, \mathbf{Markov}].$$

Definition. A Markov morphism $X \to Y$ of measurable $\mathcal{C}$-spaces X and Y is a morphism $\mathcal{M}_*(X) \to \mathcal{M}_*(Y)$.

Markov morphisms of graphs

So, a Markov morphism $\Phi : X \to Y$ of graphs X and Y consists of Markov kernels $\Phi_V : X(V) \to Y(V)$ and $\Phi_E : X(E) \to Y(E)$ such that

$$\begin{array}{ccc} X(E) & \xrightarrow{\text{src}} & X(V) \\ {\scriptstyle \Phi_E}\downarrow & & \downarrow{\scriptstyle \Phi_V} \\ Y(E) & \xrightarrow{\text{src}} & Y(V) \end{array} \qquad \begin{array}{ccc} X(E) & \xrightarrow{\text{tgt}} & X(V) \\ {\scriptstyle \Phi_E}\downarrow & & \downarrow{\scriptstyle \Phi_V} \\ Y(E) & \xrightarrow{\text{tgt}} & Y(V) \end{array}$$

Important: Graph homomorphism is NP-hard, but Markov graph morphism is a linear feasibility problem.

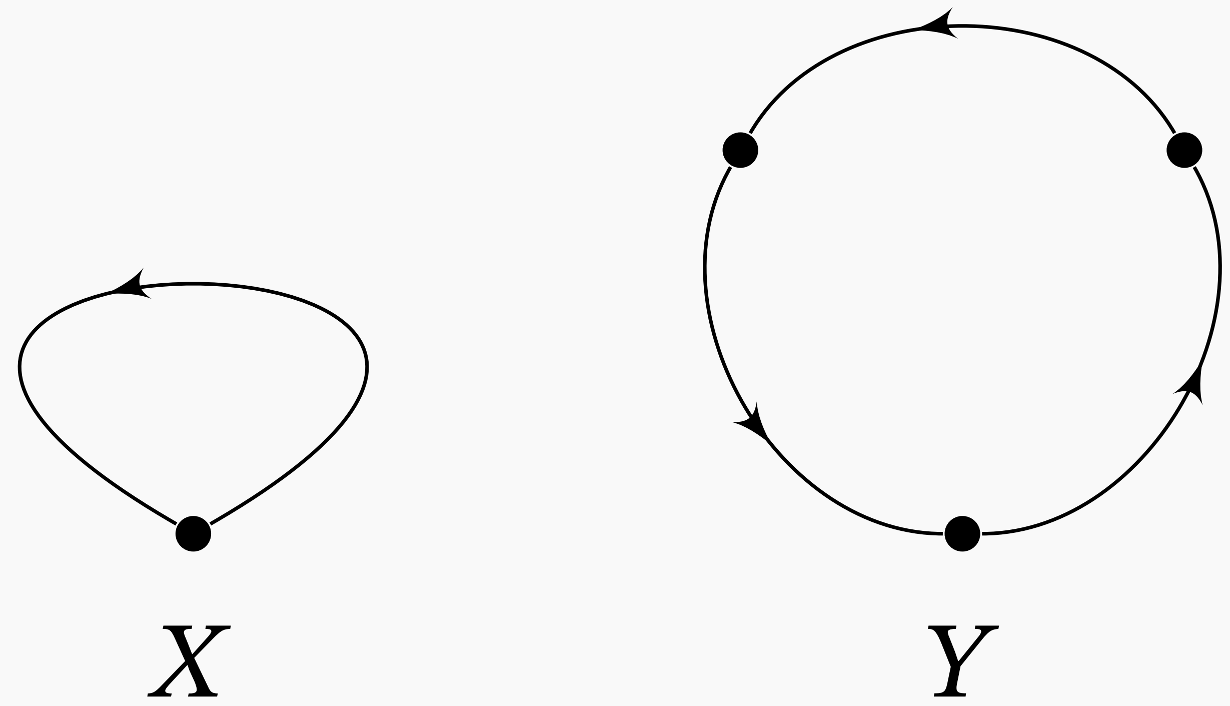
Examples of Markov morphisms:

- any graph homomorphism
- any probabilistic mixture of graph homomorphisms

Markov morphisms of graphs

But more exotic things can happen because mass can be "split".

Example: X = self loop, Y = directed cycle.



No graph homomorphisms $X \to Y$, but there is a unique Markov morphism $\Phi : X \to Y$:

$$\Phi_V(*) \sim \mathrm{Unif}(Y(V)), \qquad \Phi_E(*) \sim \mathrm{Unif}(Y(E)).$$

Metric categories

Now, the metric side of matching $\mathcal{C}$-sets.

Let $\mathbf{Met}$ be the category of Lawvere metric spaces and maps. (Note choice of morphisms.)

Definition. A metric category is a category $\mathcal{S}$ enriched in $\mathbf{Met}$, i.e., the hom-sets $\mathcal{S}(X, Y)$ are Lawvere metric spaces.

Definition. A morphism $f : X \to Y$ in $\mathcal{S}$ is short if for all morphisms $g, g' : Y \to Z$ and $h, h' : W \to X$,

$$d(fg, fg') \leq d(g, g') \quad \text{and} \quad d(hf, h'f) \leq d(h, h').$$

Short morphisms of $\mathcal{S}$ form a subcategory $\mathrm{Short}(\mathcal{S})$.

Example 1 of metric category: metric spaces

Category $\mathbf{Met}$ with supremum metric

$$d_\infty(f, g) := \sup_{x \in X} d_Y(f(x), g(x)), \qquad f, g \in \mathbf{Met}(X, Y).$$

Short morphisms are short maps:

$$d_Y(f(x), f(x')) \leq d_X(x, x'), \quad \forall x, x' \in X.$$

Proposition: For any metric category $\mathcal{S}$, $\mathrm{Short}(\mathcal{S})$ is enriched in $\mathrm{Short}(\mathbf{Met})$.

Example 2 of metric category: metric measure spaces

Category $\mathbf{MM}$ of mm spaces and measurable maps, with L^p metric, $1 \leq p < \infty$:

$$d_p(f,g) := \left(\int_X d_Y(f(x), g(x))^p \, \mu_X(dx) \right)^{1/p}, \qquad f, g \in \mathbf{MM}(X, Y).$$

Proposition: A map $f : X \to Y$ is short iff

$$\mu_X f := \mu_X \circ f^{-1} \leq \mu_Y,$$

and

$$d_Y(f(x), f(x')) \leq d_X(x, x'), \quad \forall x, x' \in X.$$

Metrics on C-sets in metric categories

Let $\mathcal{C}$ be a finitely presented category and $\mathcal{S}$ a metric category.

Idea: For $X, Y \in [\mathcal{C}, \mathcal{S}]$, consider distance from naturality of transformation $\phi : X \to Y$ at $c \in \mathcal{C}$:

$$d(Xf \cdot \phi_{c'}, \phi_c \cdot Yf) \quad \text{“=”} \quad \begin{array}{ccc} X(c) & \xrightarrow{Xf} & X(c') \\ {\scriptstyle\phi_c}\downarrow & \swarrow & \downarrow{\scriptstyle\phi_{c'}} \\ Y(c) & \xrightarrow[Yf]{} & Y(c') \end{array}$$

Metrics on C-sets in metric categories

Theorem: For any $1 \leq p \leq \infty$, a Lawvere metric on $[\mathcal{C}, \mathcal{S}]$ is defined by

$$d(X, Y) := \inf_{\phi : X \to Y} \sum_{f : c \to c'}^{p} d(Xf \cdot \phi_{c'}, \phi_c \cdot Yf),$$

where

- infimum is over (unnatural) transformations with components in $\mathrm{Short}(\mathcal{S})$
- ℓ^p norm/sum is over a fixed, finite generating set of morphisms in $\mathcal{C}$.

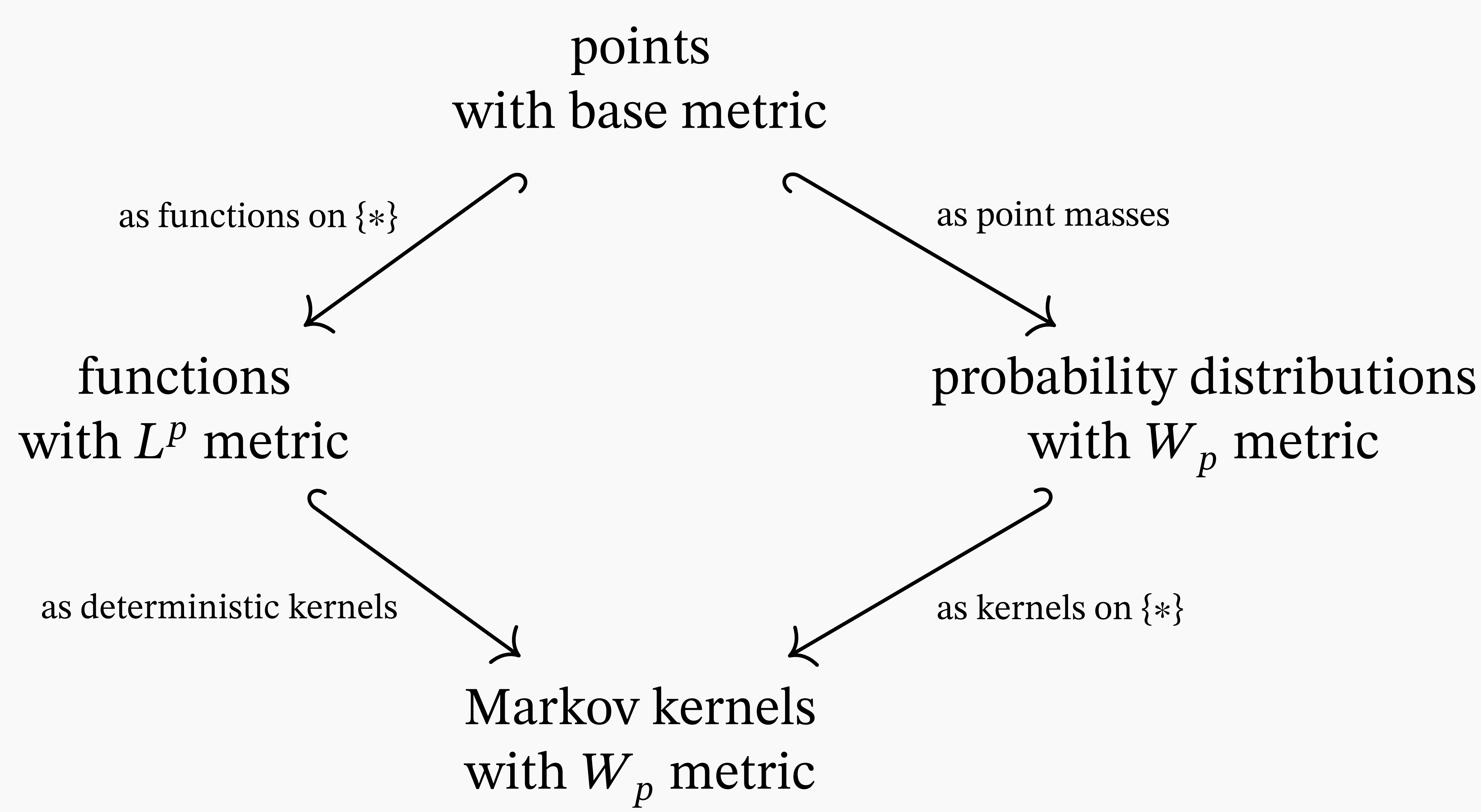
Note: Condition that each $\phi_c \in \mathrm{Short}(\mathcal{S})$ is needed for triangle inequality.

Example 3 of metric category: Markov kernels on mm spaces

Category **MMarkov** of mm spaces and Markov kernels, with Wasserstein metric:

$$W_p(M, N) := \inf_{\Pi \in \mathrm{Coup}(M,N)} \left(\int_{X \times Y \times Y} d_Y(y, y')^p \, \Pi(dy, dy' \mid x) \, \mu_X(dx) \right)^{1/p}$$

Generalizes both classical L^p and Wasserstein metrics:



Example 3 of metric category: Markov kernels on mm spaces

Proposition: Under regularity conditions, a Markov kernel $M : X \to Y$ is short iff

$$\mu_X M \leq \mu_Y$$

and there exists $\Pi \in \mathrm{Prod}(X, Y)$ such that

$$\int_{Y\times Y} d_Y(y, y')^p \, \Pi(dy, dy \mid x, x') \leq d_X(x, x')^p, \qquad \forall x, x' \in X.$$

Consequence: Via the theorem, a Wasserstein-style metric on metric measure $\mathcal{C}$-spaces, computable by solving a linear program.

Future work

- Beyond $\mathcal{C}$-sets
  - Sums (coproducts) and units (terminal objects) are easy
  - Products are less immediate
- Faster algorithms
  - Needed for practical use on graphs of even moderate size
  - Entropic regularization of both theoretical and algorithmic interest

Thanks!

Paper: Hausdorff and Wasserstein metrics on graphs and other structured data, 2019. arXiv:1907.00257.