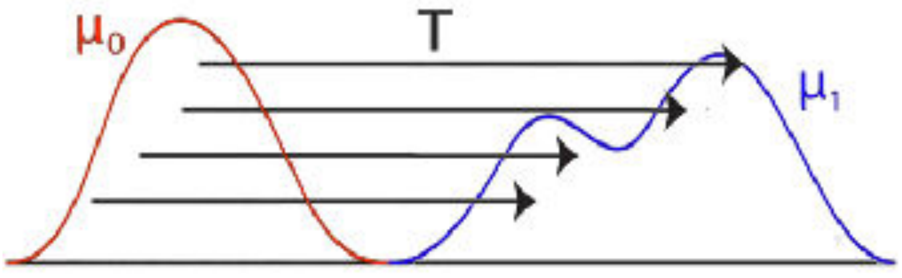


Learning to Match Distributions (L2MD)

Optimal Transport, Flow Matching & Applications

CSCI-GA 3033-148 · New York University · Fall 2026

Romain Lopez

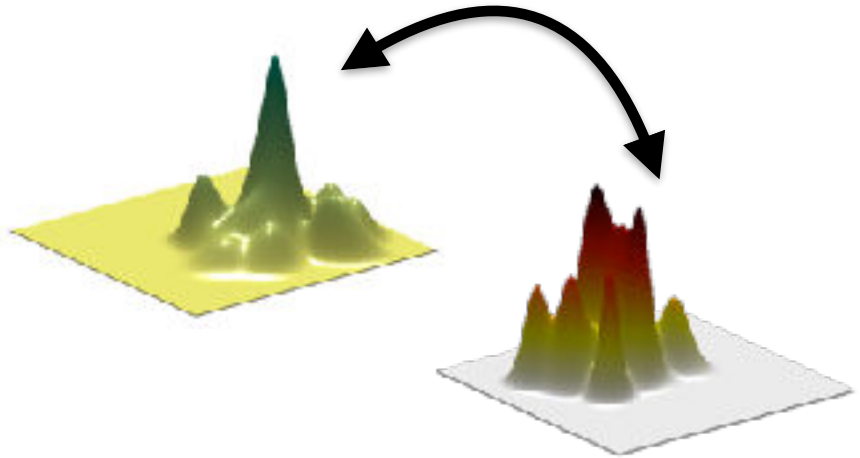
Monday 14 September

Today's schedule

1. Motivation & scope of the class
2. Logistics
3. Primer on optimal transport and flow matching
4. Frontier problems & topics for reading group
5. Next steps

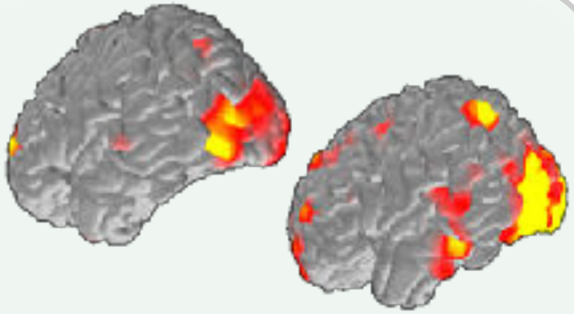
Motivation and scope of the class

How to compare two distributions?



Wasserstein distance
(optimal transport)

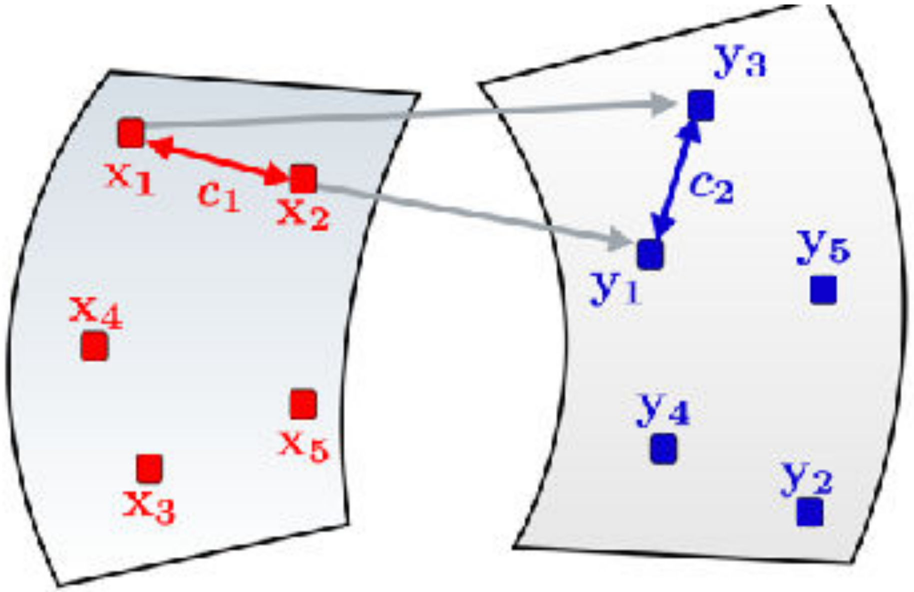
Bags of words



Brain Activation Maps

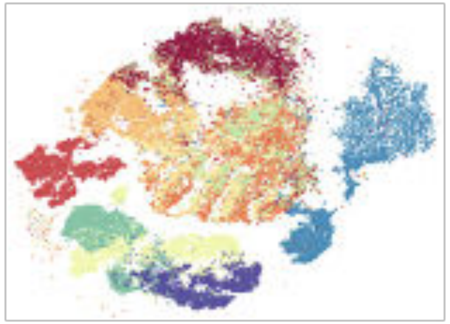
Cuturi & Salomon, NeurIPS 2017 tutorial

How to align two datasets with no shared coordinates?

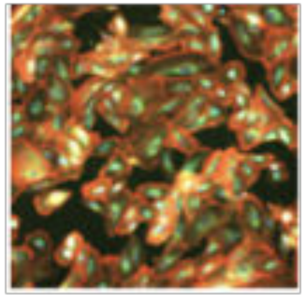


Gromov-Wasserstein Optimal Transport

Modality 1



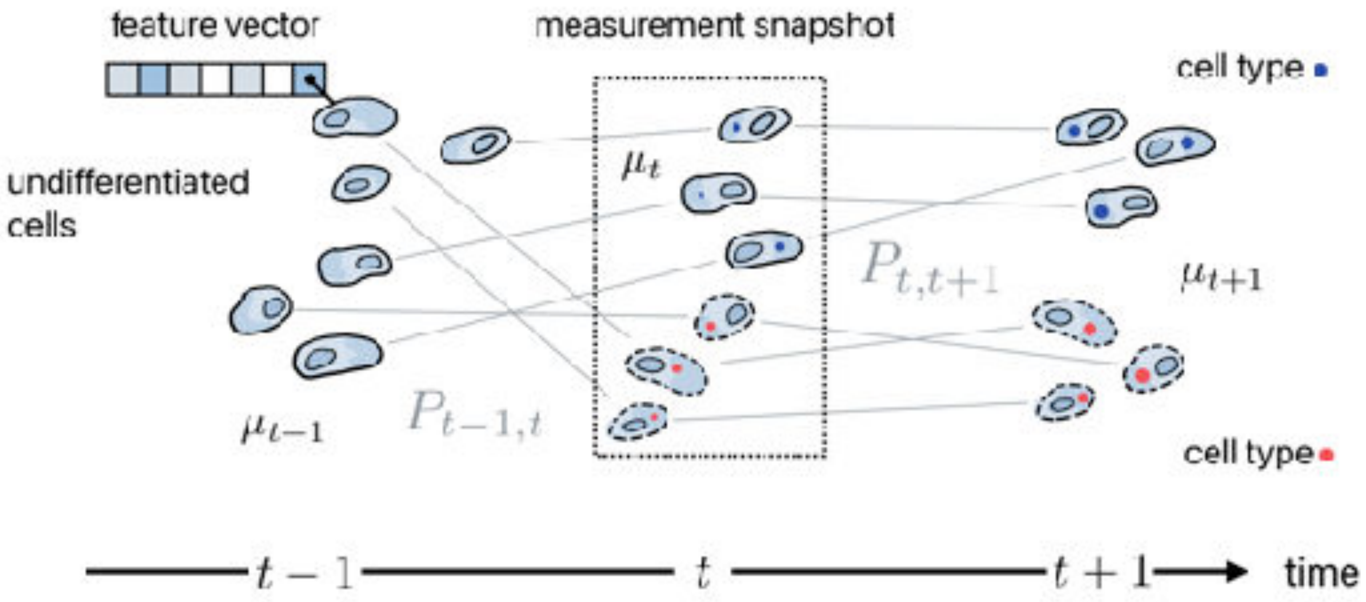
Modality 2



Multi-modal data

Bunne & Cuturi, ICML 2023 tutorial

How can we reconstruct dynamics based solely on population snapshots?



(Dynamic) Optimal Transport

Bunne & Cuturi, ICML 2023 tutorial

How to acquire new samples from a complex data distribution?



Flow Matching

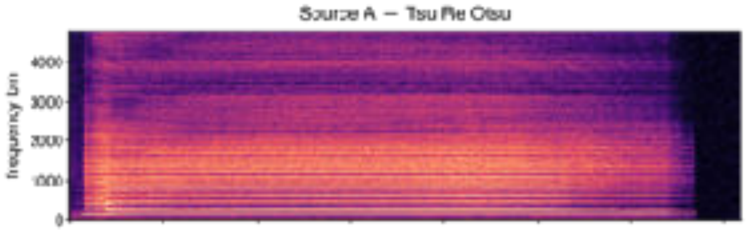
Stable Diffusion 3

How to morph one sound wave into another?

Glissando via Wasserstein Barycenters

Collaboration with Armand Bernardi

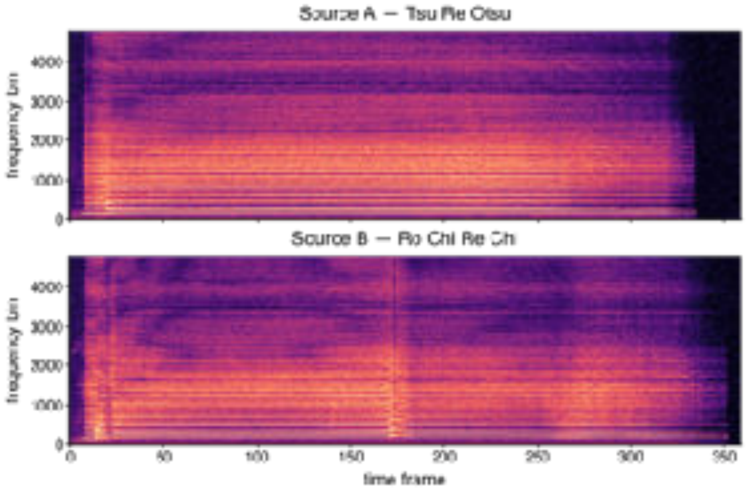
How to morph one sound wave into another?



Glissando via Wasserstein Barycenters

Collaboration with Armand Bernardi

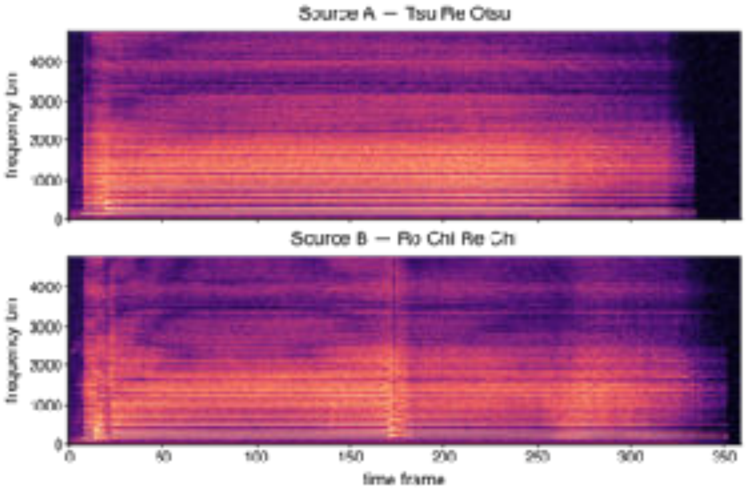
How to morph one sound wave into another?



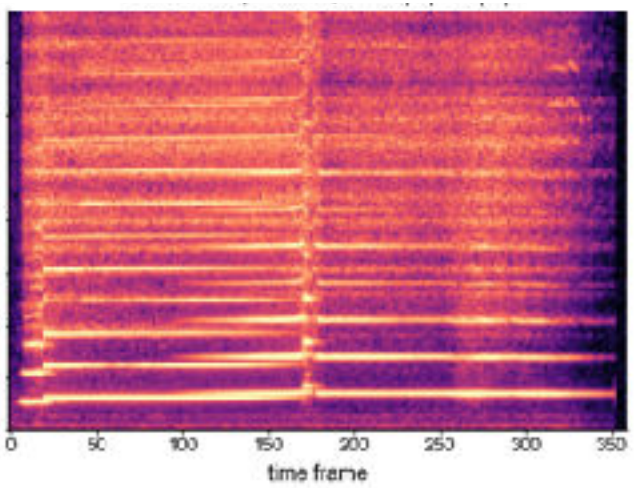
Glissando via Wasserstein Barycenters

Collaboration with Armand Bernardi

How to morph one sound wave into another?



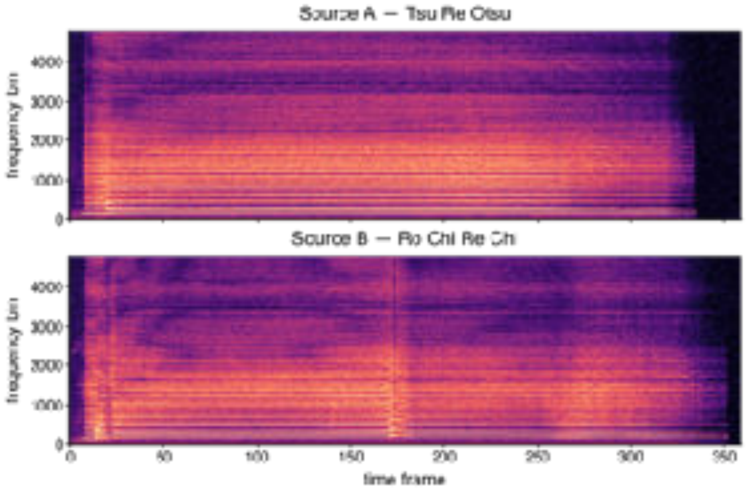
Euclidean interpolation



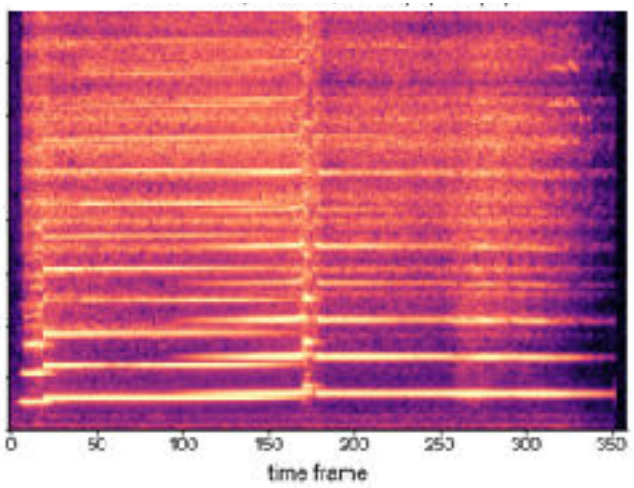
Glissando via Wasserstein Barycenters

Collaboration with Armand Bernardi

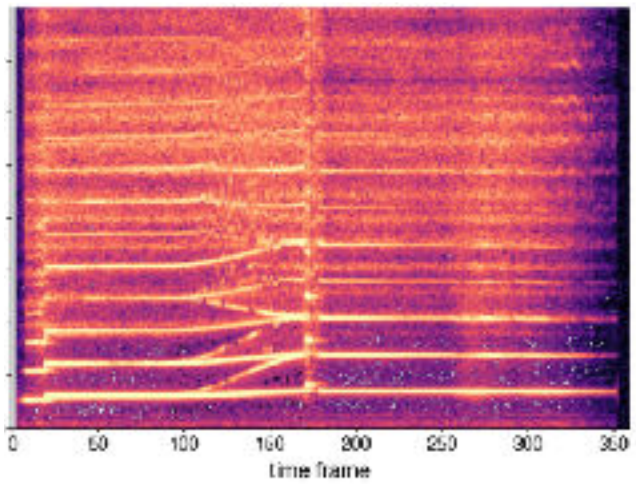
How to morph one sound wave into another?



Euclidean interpolation



Wasserstein interpolation



Glissando via Wasserstein Barycenters

Collaboration with Armand Bernardi

Overarching technical questions

- How far apart are two (or more) distributions?
- How do we map them onto each other?
- What lies on the path between them?

We will develop a basic machinery to answer those, and study applications for original research problem

What this course is

- A graduate seminar course. Three seeding lectures from the instructor. The rest will be paper presentations.
- All papers will be recent and related to machine learning and distribution modeling.
- Two papers a session. Four of you present each one, everyone reads both.
- A semester-long project of your own, in teams.

What this course is not

- **Not a math class.**

The course will focus on ML, and less on proofs. Convex analysis, asymptotic estimation rates, high-dimensional guarantees are out of scope. For this, see Prof. Guillen's MATH-GA.2610-001, Advanced Topics in Optimal Transport Theory and Applications course.

- **Not a diffusion class.**

Stochastic differential equations, Schrödinger bridges, diffusion models are out of scope.

Course Logistics
(subject to change, refer to website)

Find all this and more on the course website!
https://romain-lopez.github.io/learning-to-match-distributions/

Meet the instructor

Romain Lopez
say: « **Ra**w **Man**go »

OH: Mon 3:30-4:30
60FA#304

Meet the TA

- Second-year MS in Computer Science at Courant — personalized retrieval and long-context reasoning in multi-agent systems.
- Research intern at Siemens. Previously a software engineer at Adobe and Salesforce, with research visits to EPFL and Spain on scholarship.
- Come to Vaibhavi for help with scoping your project, and seminar logistics.
- Fun fact: deep into fountain pens, archival paper and niche stationery from Japan, France and Germany.

Vaibhavi Singh
say: « **Va**lid **Buh Vee** »

OH: Fri 12:30-1:30
60FA#502

Communication

- **Discord**

  Discussion. The invite is on Brightspace and does not expire.

- **Brightspace**

  Announcements, submissions, various private links (sign-up sheets, etc..)

- **Email to the team**

  Anything personal or urgent.

Grading

- 50% Class Project
- 30% Reading Group Presentation
- 10% Homework
- 10% Attendance

Policies

- **AI and collaboration**

  Both allowed. Disclose who you worked with and how you used any assistant.

- **Late work**

  Not accepted.

How the semester runs

- **Three instructor lectures from**

  Today with slides. Sep 21 and Sep 28 are whiteboard sessions.

- **Nine seminar sessions, two papers each**

  Eighteen papers between Oct 5 and Nov 30.

- **Two final sessions**

  Project presentations on Dec 7, a guest talk on Dec 14.

An example « role-play seminar » session

Paper 1

🔬 📊 👾 💡

Four presentations
(each from a different role)

Paper 2

🏺 👾 📊 💡

Four presentations
(each from a different role)

- Roughly eight minutes for each student, about ten slides.
- **Presentations materials are due by 11 AM on the day you present ([form upload]).**

The five roles (full description on website)

- 🔬 **Scientific Peer Reviewer** · review it as if it were under submission, and commit to accept or reject.
- 🏺 **Archaeologist** · papers it built on, newer papers it inspired.
- 📊 **Data Analyst** · datasets, curation, benchmarks, metrics.
- 👾 **Hacker** · how would you implement it? What exists, what runs, what it costs.
- 💡 **Academic Researcher** · propose a follow-up only possible because this paper worked.

Signing up

- The sheet goes live on Brightspace on **Sep 16th** and closes **Sep 28th** (add your NetID).
- Four sign-ups per paper, one per role.
- Students will present 3 to 4 times throughout the semester.
- Remaining slots (if any) will be randomly assigned.

The preparatory homework

Due Monday Sep 21, at the start of class.

Handwritten, scanned, submitted as a single PDF on Brightspace. Graded for genuine attempt.

This is the mathematical foundation for the whole course. Please spend time studying it so you follow the white boarding sessions next weeks (and the papers later on).

The preparatory homework

Due Monday Sep 21, at the start of class.

Handwritten, scanned, submitted as a single PDF on Brightspace. Graded for genuine attempt.

This is the mathematical foundation for the whole course. Please spend time studying it so you follow the white boarding sessions next weeks (and the papers later on).

Dealing with difficulty and calibrating fit for this course:

- Struggling with one or several exercises is fine, especially where the material is new to you. Convex optimization, for instance.
- Not being able to make progress on any of them is a signal: the papers we read will be harder to follow than they should be.

The project

- Teams of 2 to 4.
- Anything that builds on optimal transport or flow matching. Theoretical, methodological, or purely empirical and applicative work is welcome.
- Well-scoped negative results are welcome.

Mon Oct 5	Team formation
Mon Nov 2	Project check-in due
Mon Dec 7	Poster Presentation

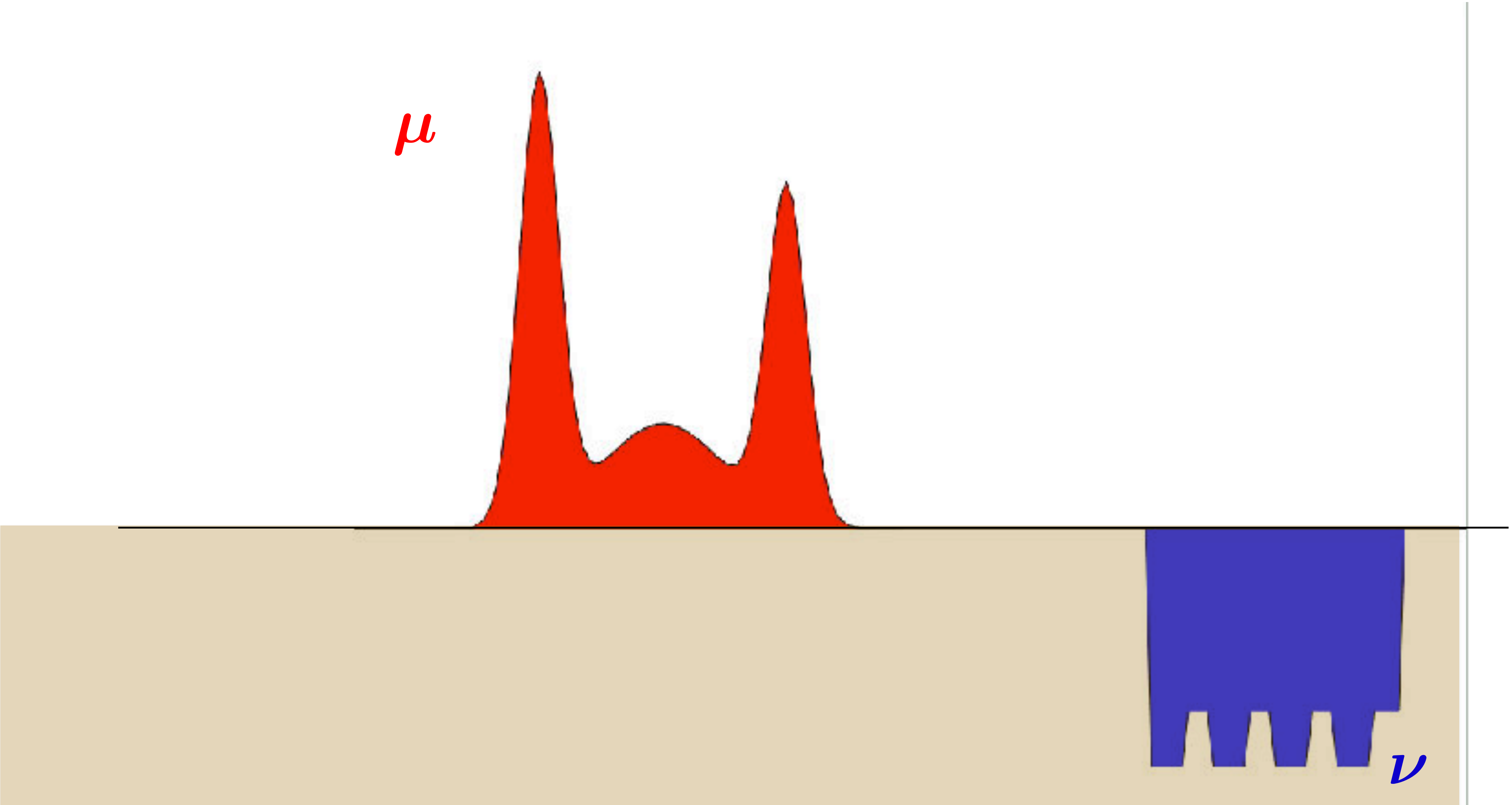
Foundations of OT and FM

Overview today. Derivations during the next two weeks.

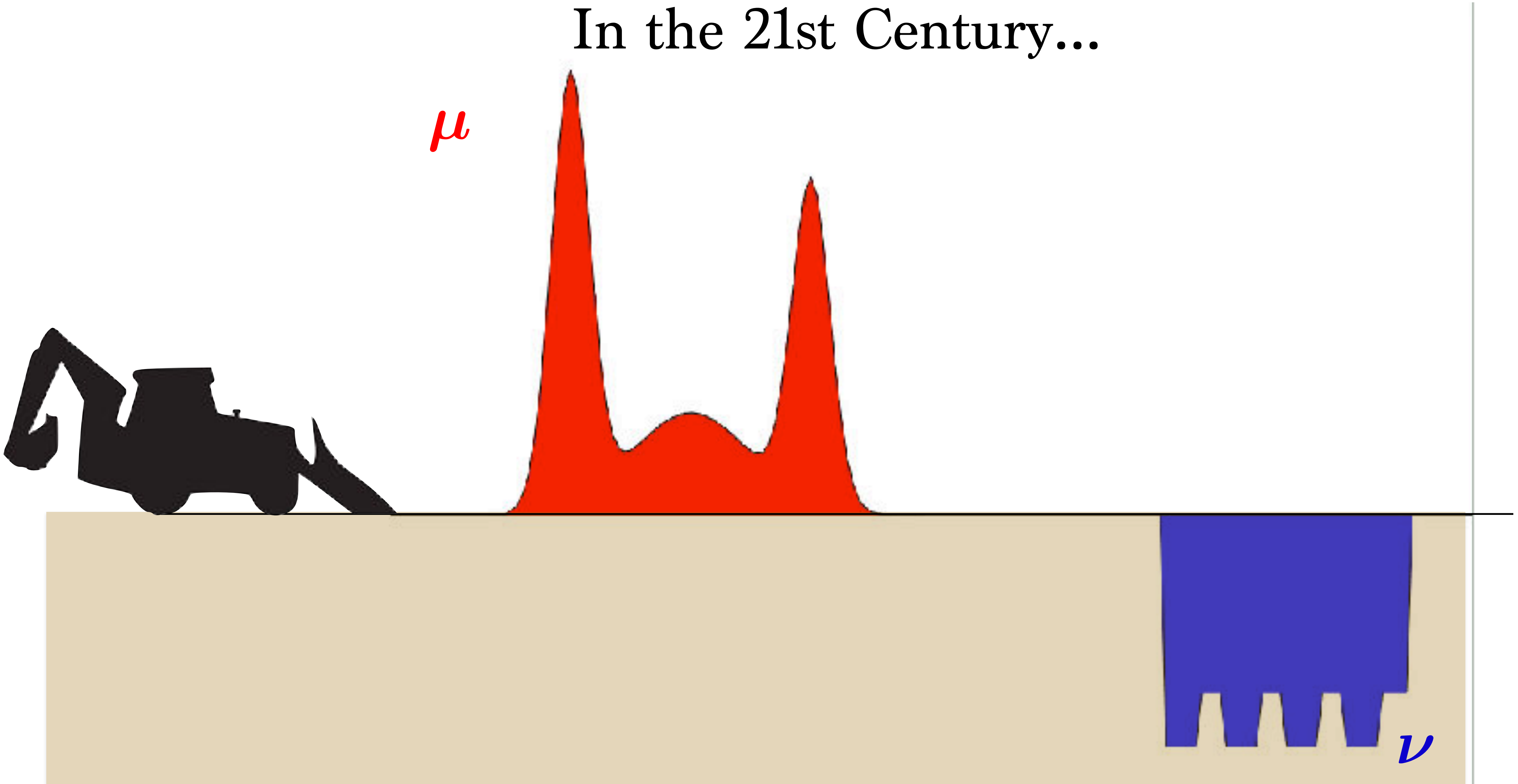
Cuturi & Salomon, NeurIPS 2017 tutorial
Bunne & Cuturi, ICML 2023 tutorial

I. Introduction to OT

Origins: Monge Problem

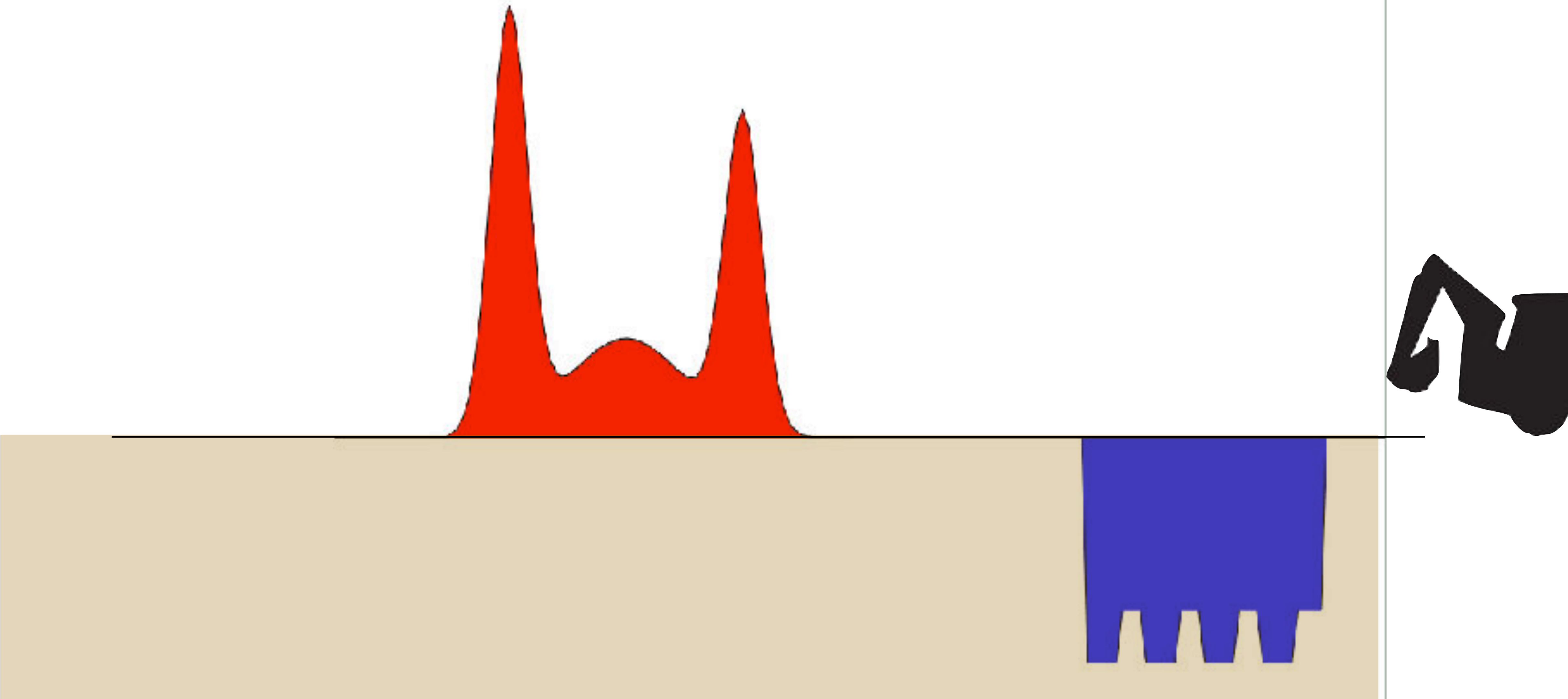


Origins: Monge Problem



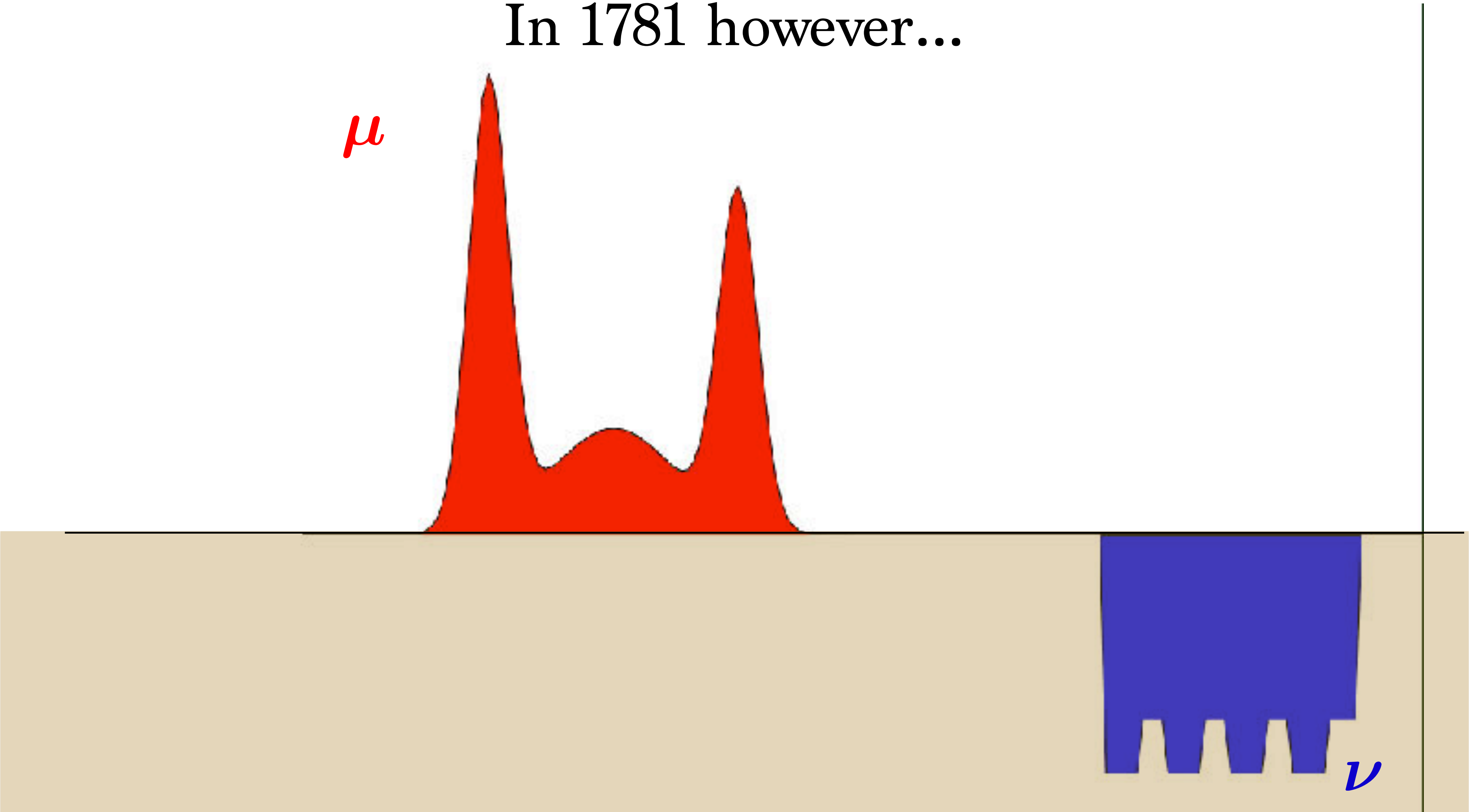
Origins: Monge Problem

In the 21st Century...



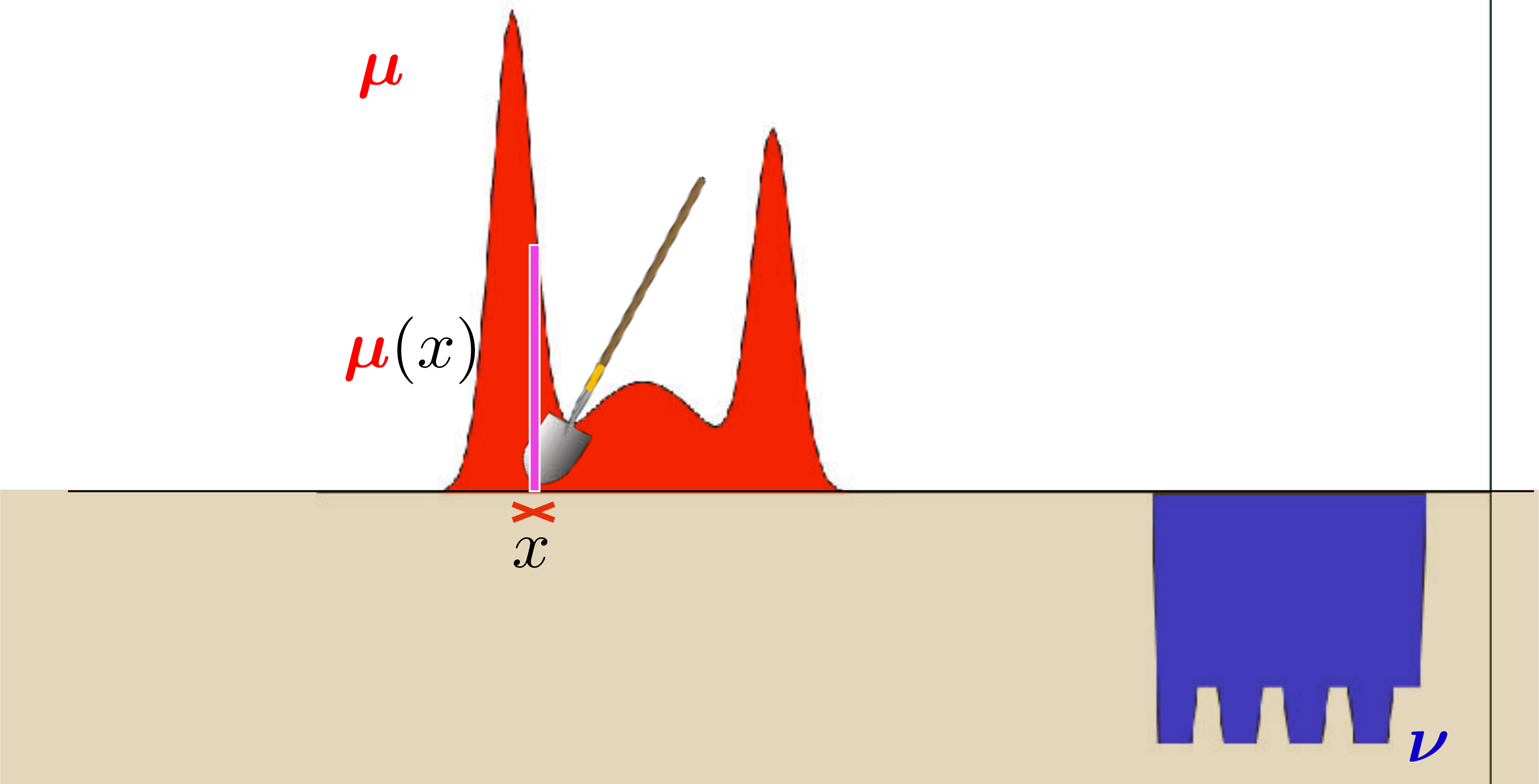
Origins: Monge's Problem

In 1781 however...



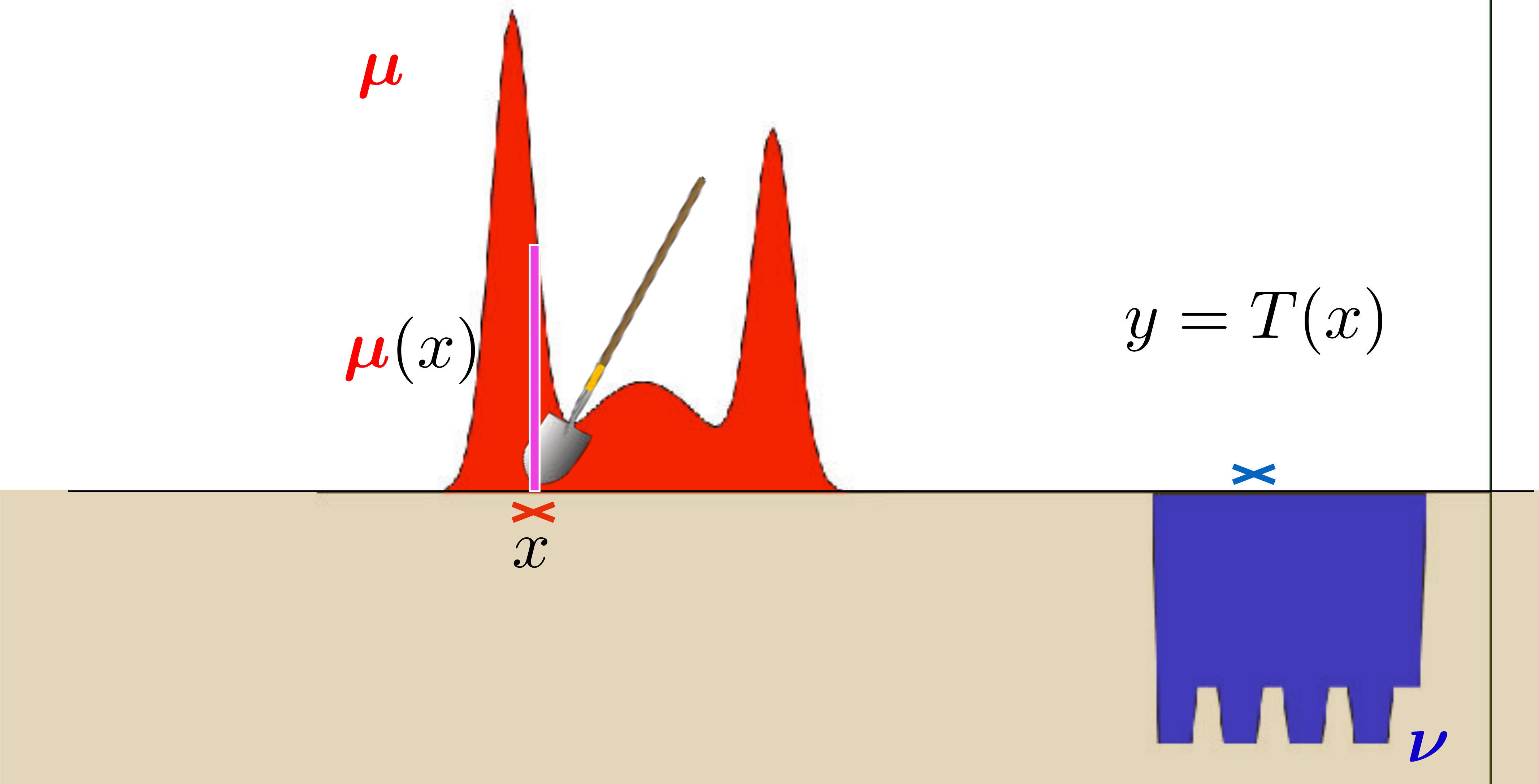
Origins: Monge's Problem

In 1781 however...



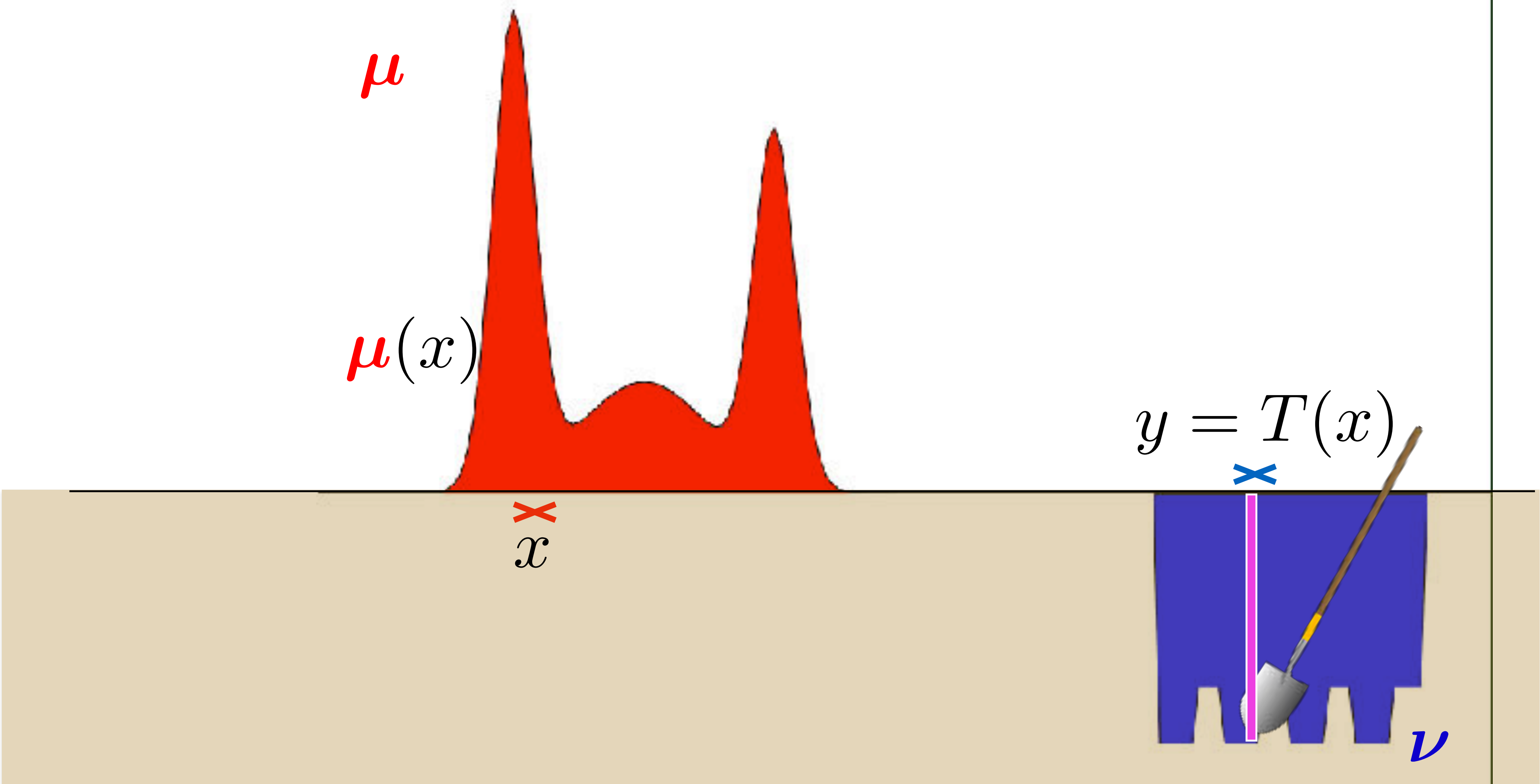
Origins: Monge's Problem

In 1781 however...



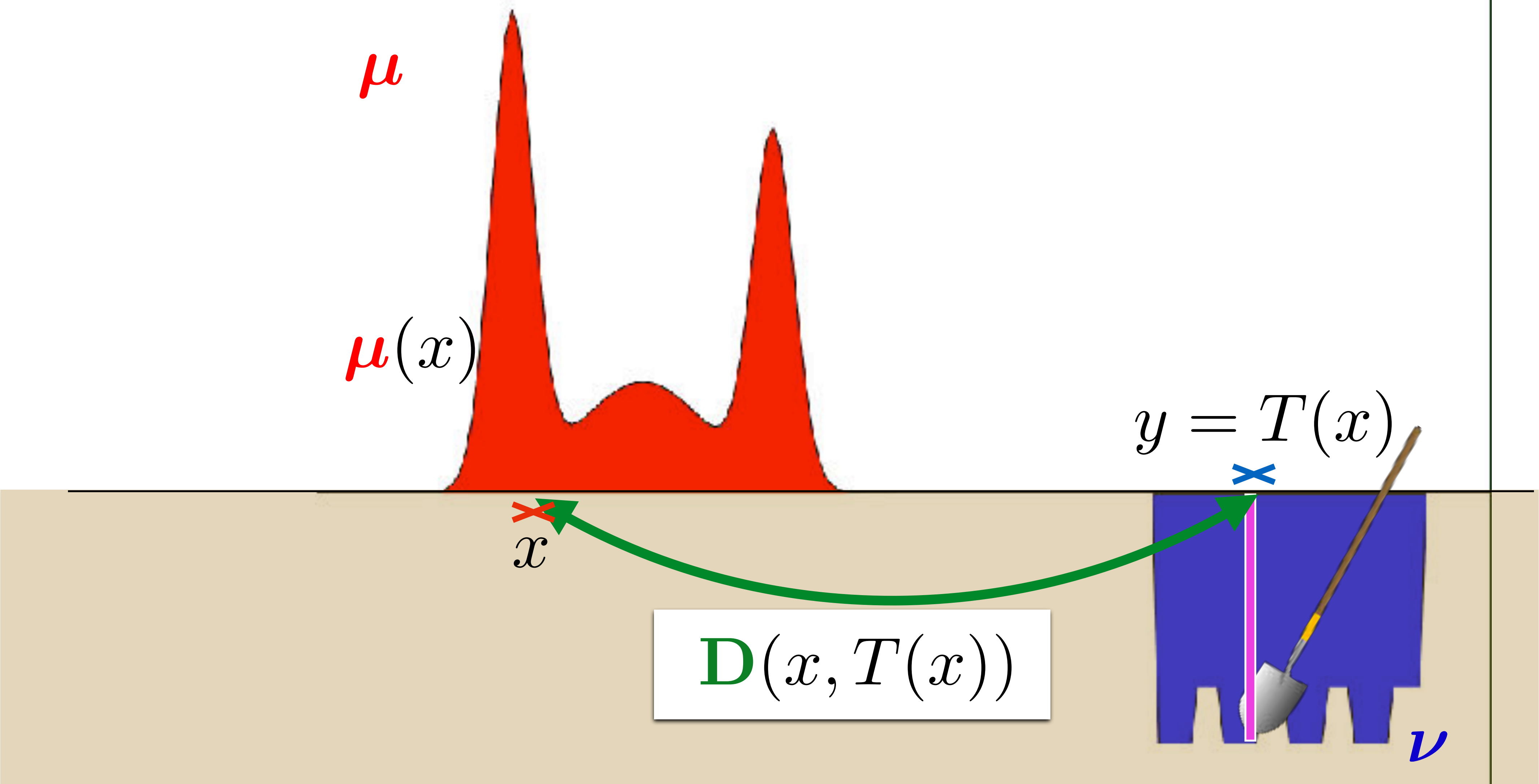
Origins: Monge's Problem

In 1781 however...



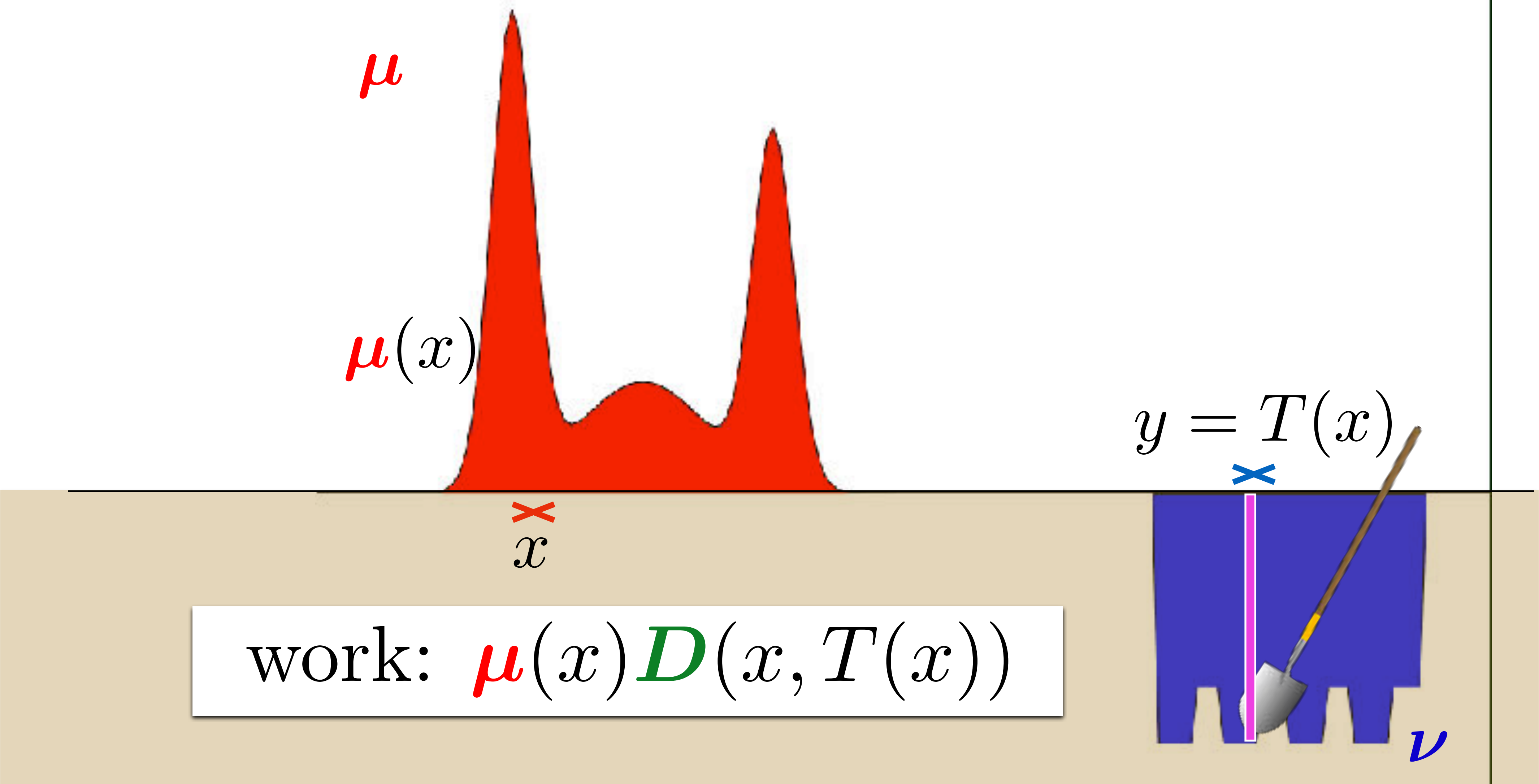
Origins: Monge's Problem

In 1781 however...



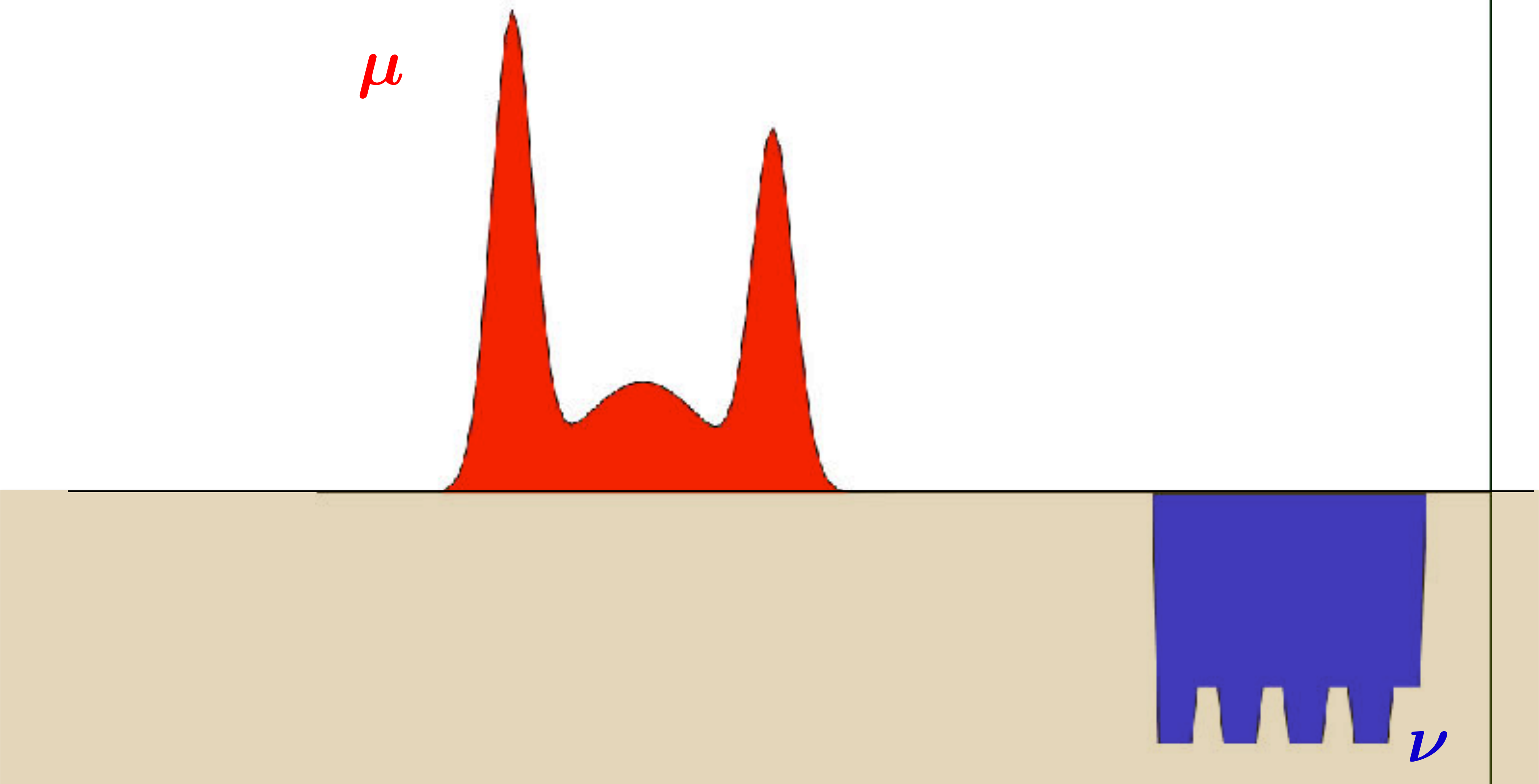
Origins: Monge's Problem

In 1781 however...



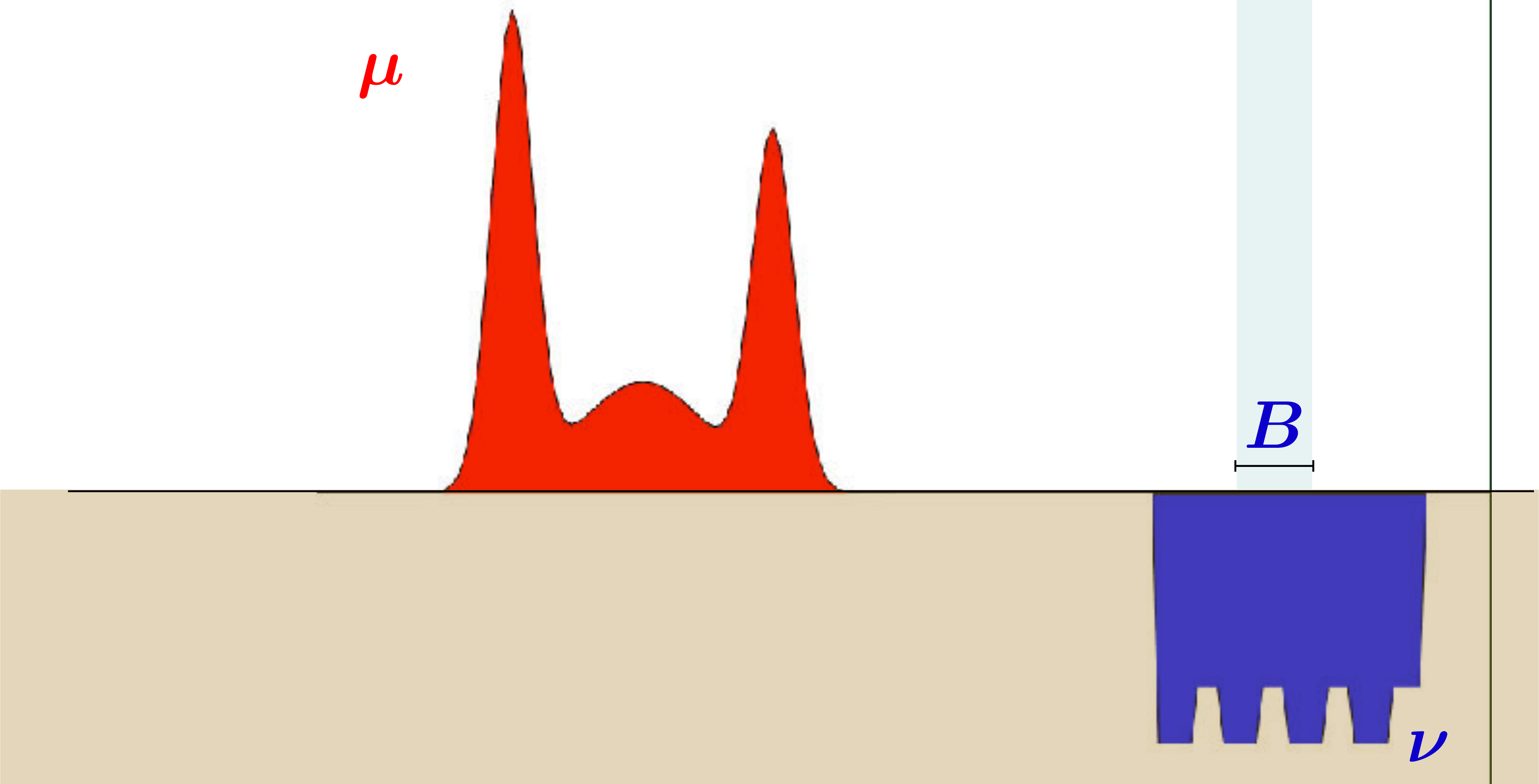
Origins: Monge's Problem

T must map red to blue.



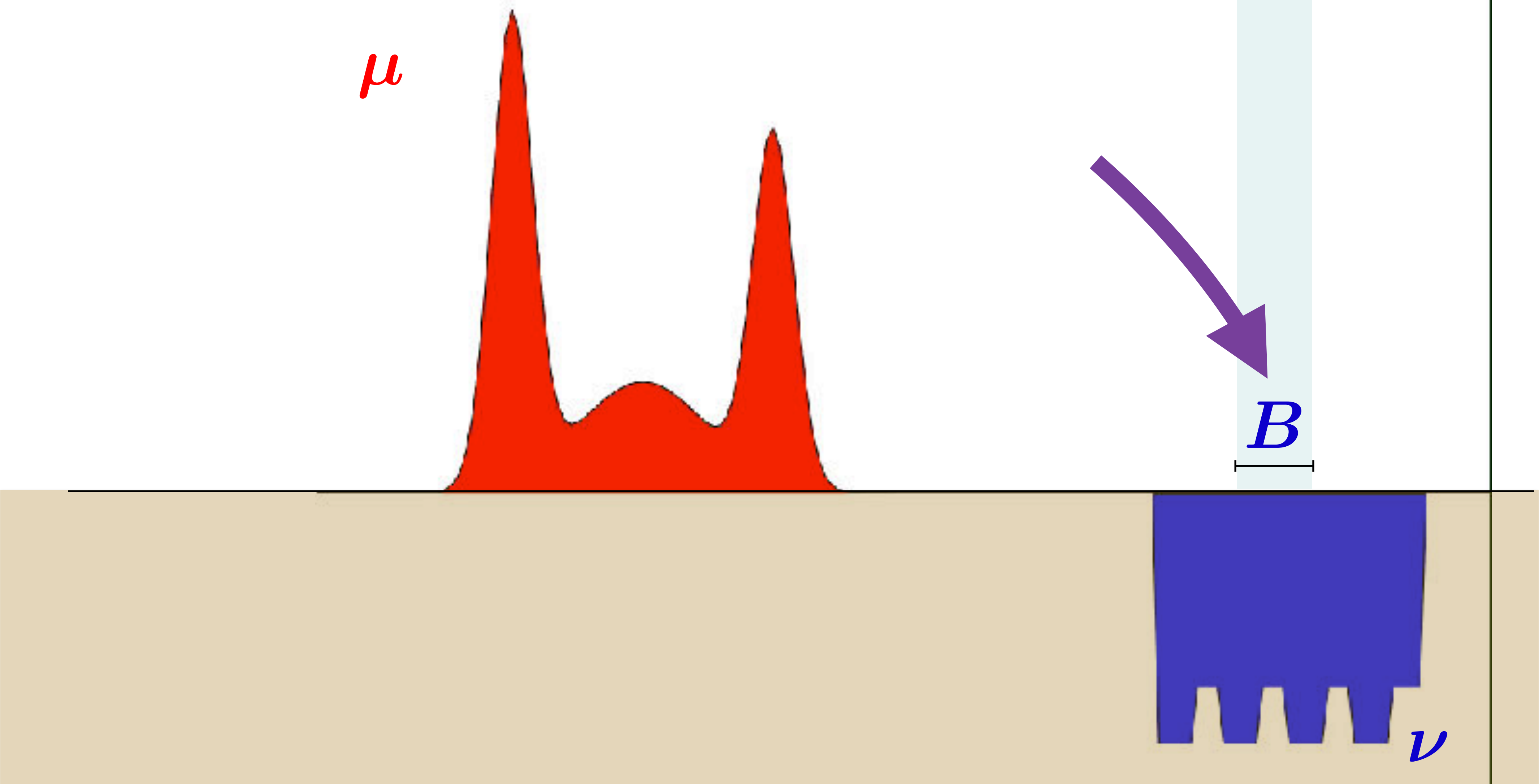
Origins: Monge's Problem

T must map red to blue.



Origins: Monge's Problem

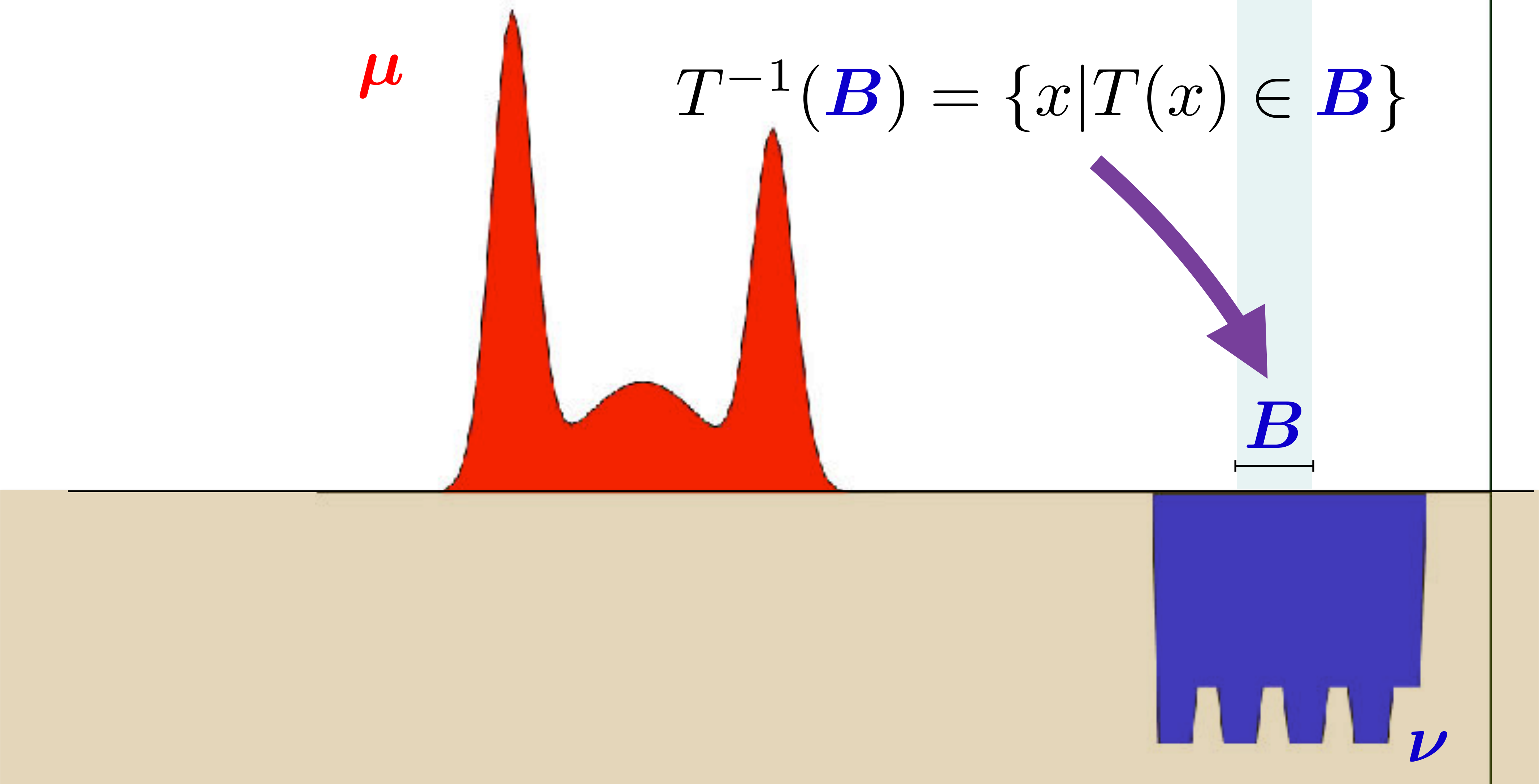
T must map red to blue.



Origins: Monge's Problem

T must map red to blue.

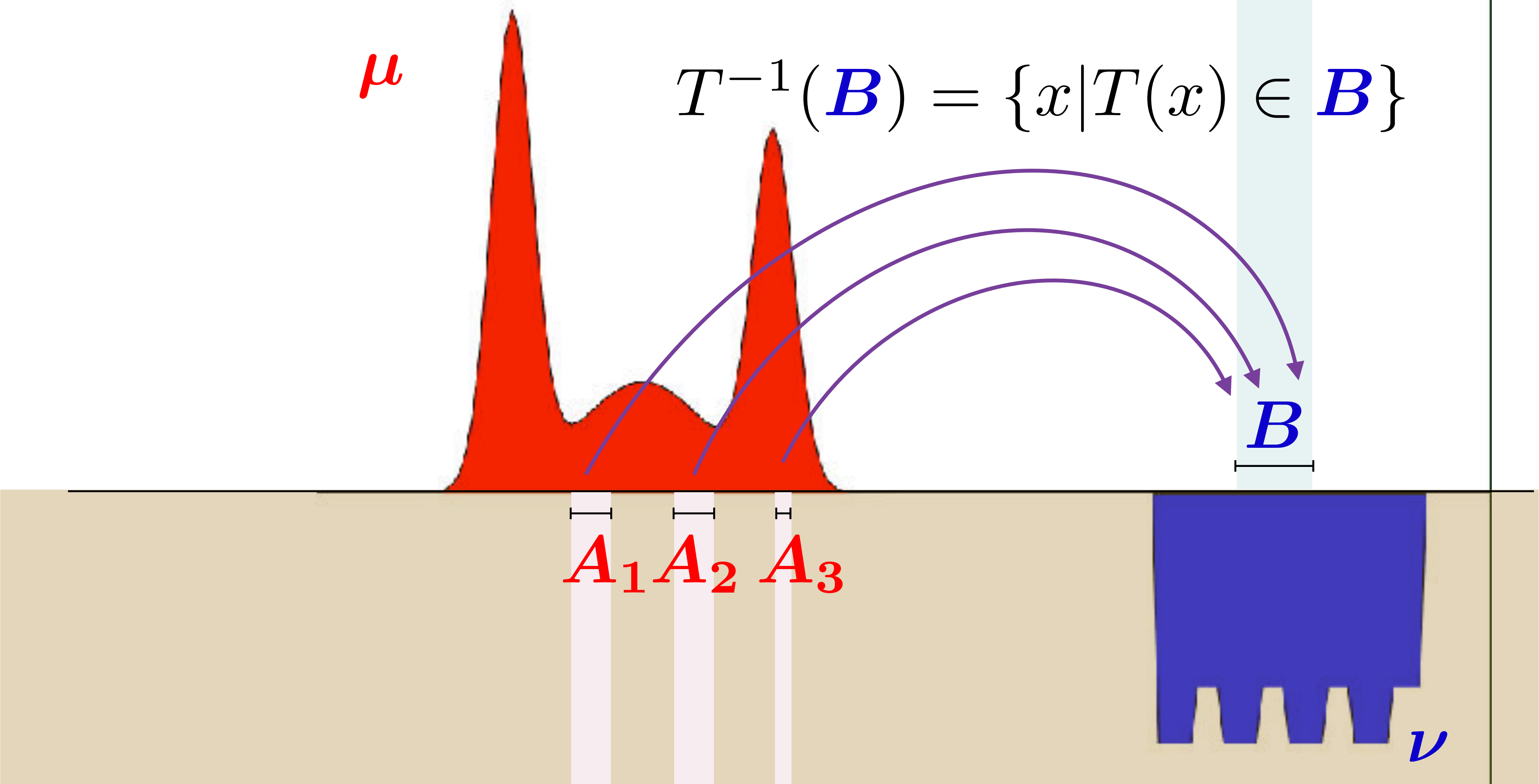
$$T^{-1}(B) = \{x | T(x) \in B\}$$



Origins: Monge's Problem

T must map red to blue.

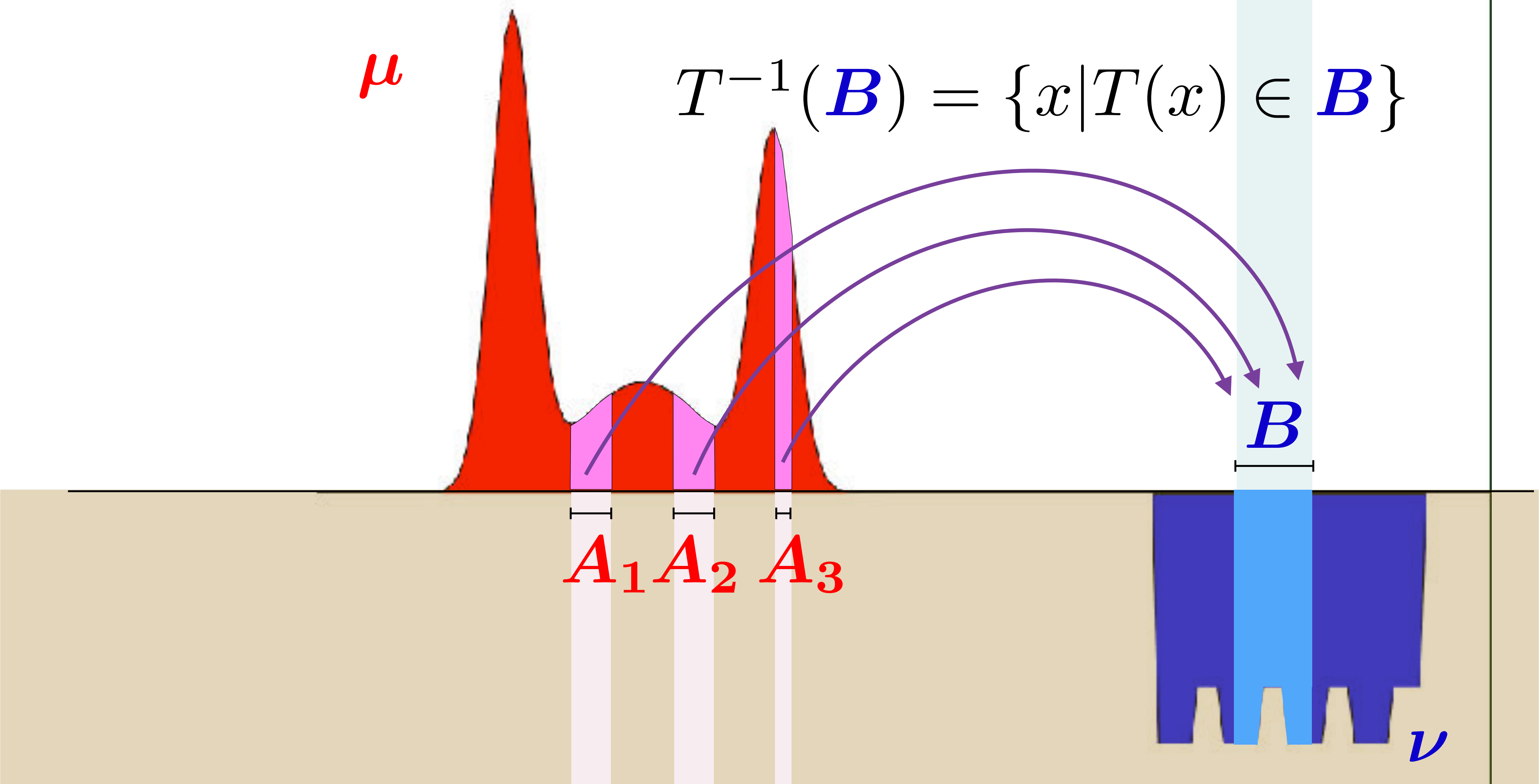
$$T^{-1}(\color{blue}{B}) = \{x | T(x) \in \color{blue}{B}\}$$



Origins: Monge's Problem

T must map red to blue.

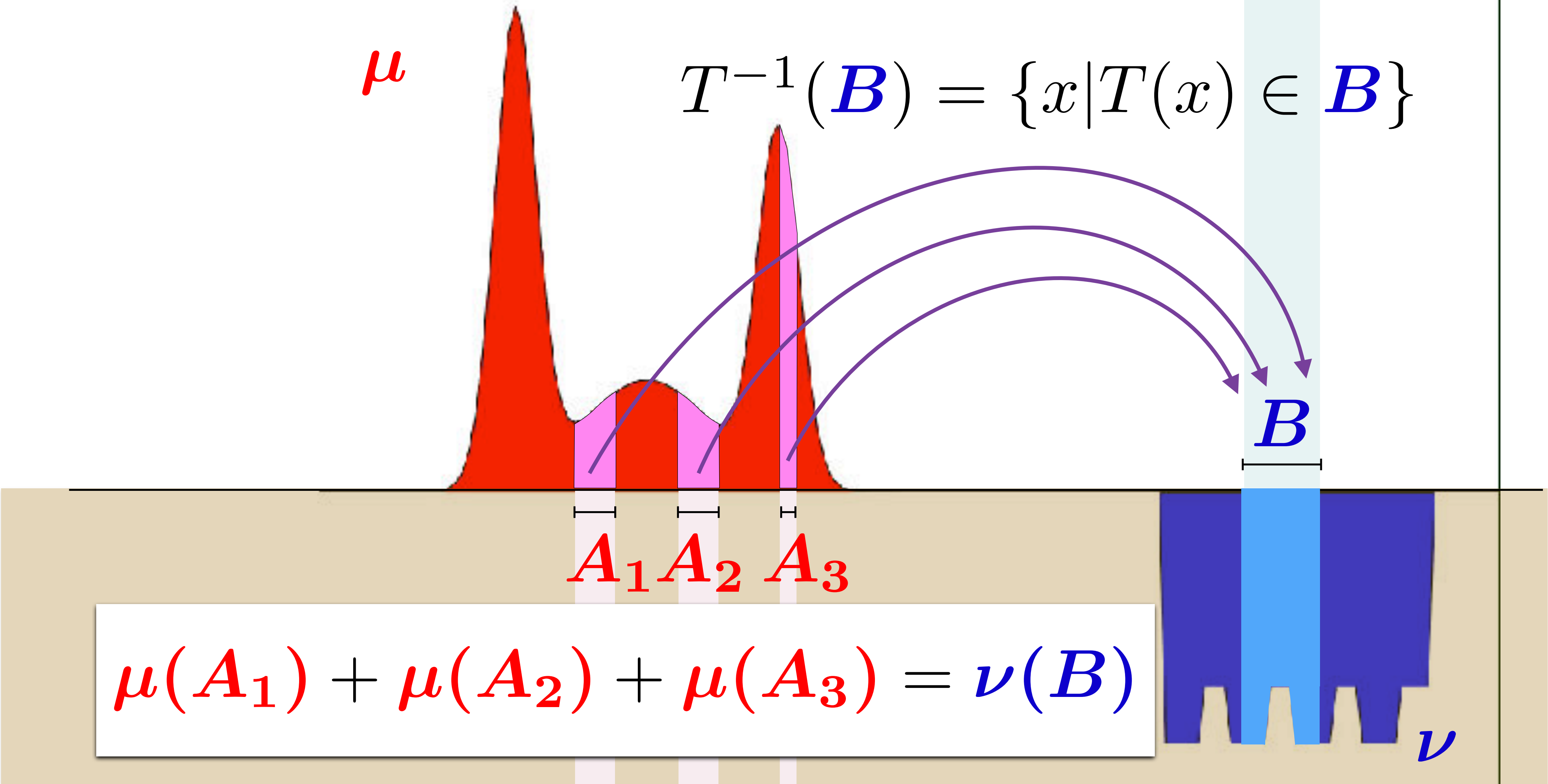
$$T^{-1}(B) = \{x \mid T(x) \in B\}$$



Origins: Monge's Problem

T must map red to blue.

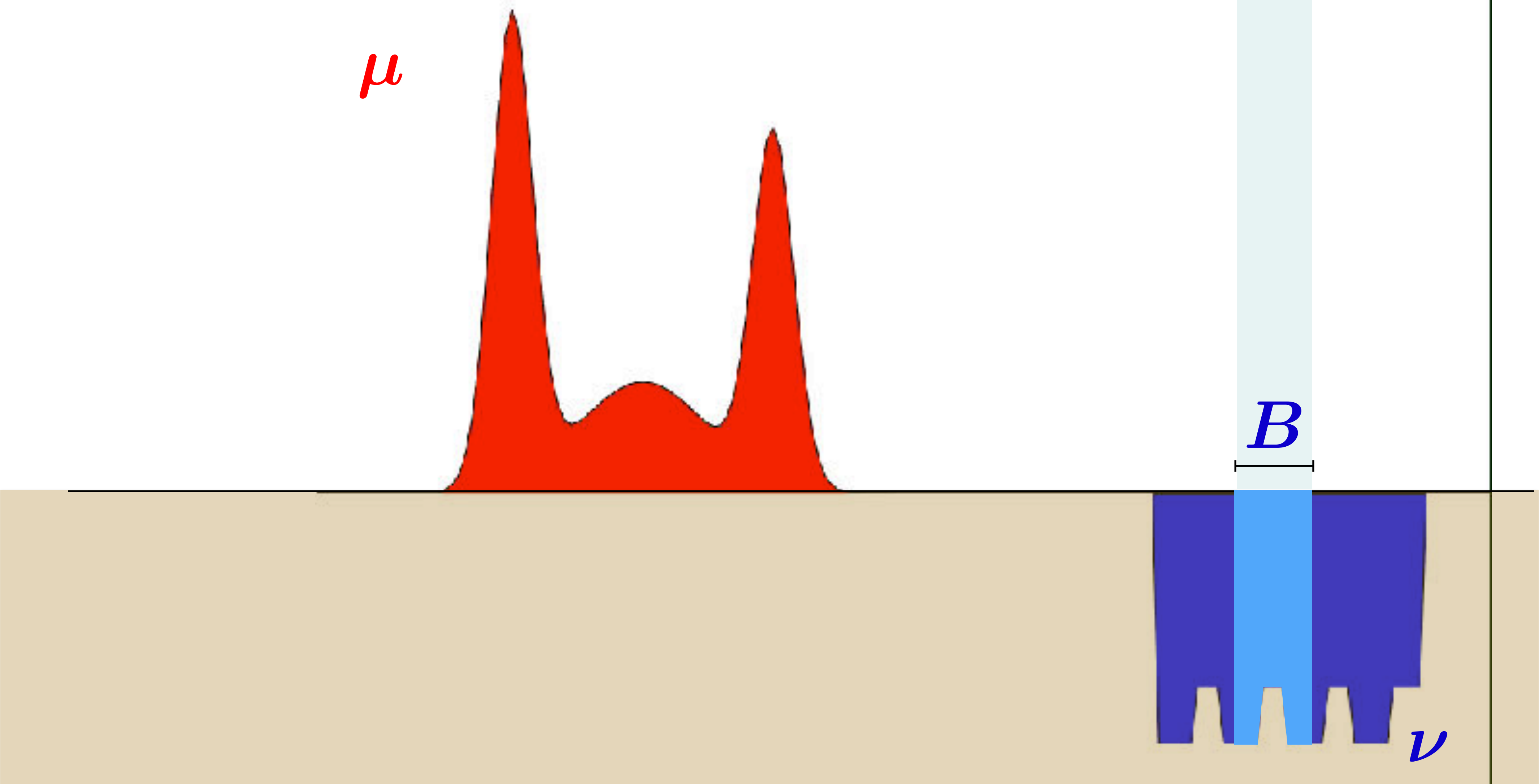
$$T^{-1}(B) = \{x \mid T(x) \in B\}$$



$$\mu(A_1) + \mu(A_2) + \mu(A_3) = \nu(B)$$

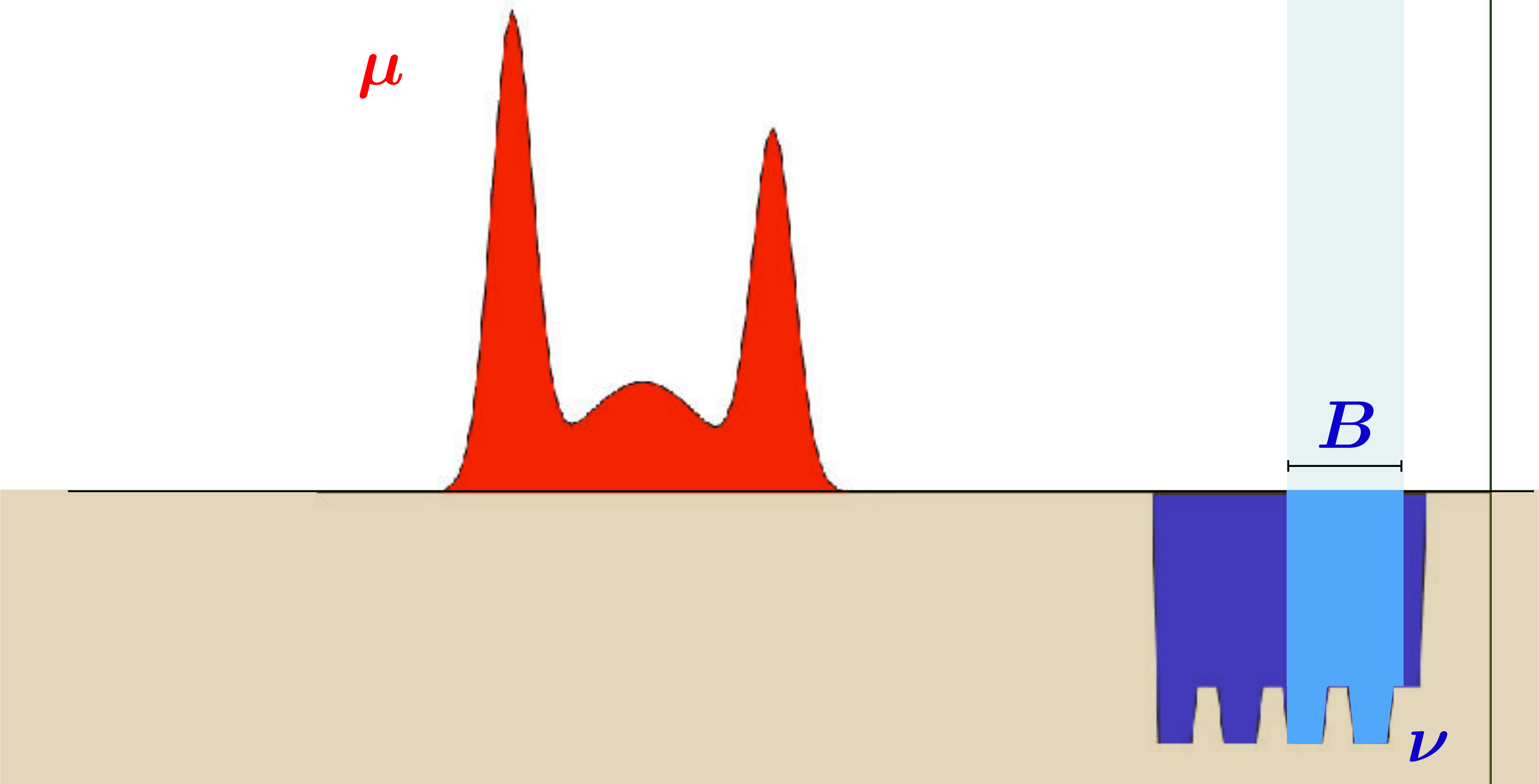
Origins: Monge's Probler

T must map red to blue.



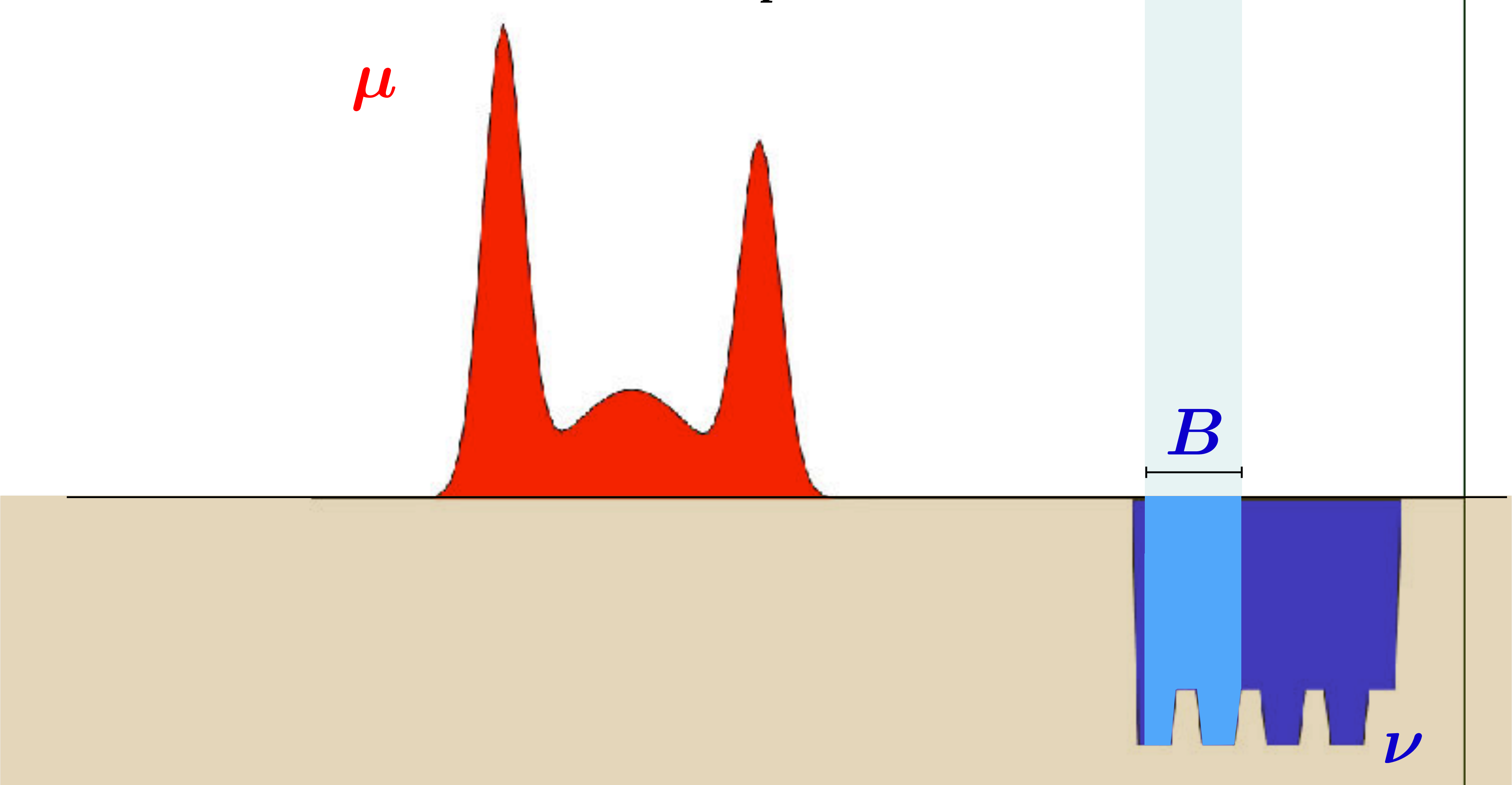
Origins: Monge's Problem

T must map red to blue.



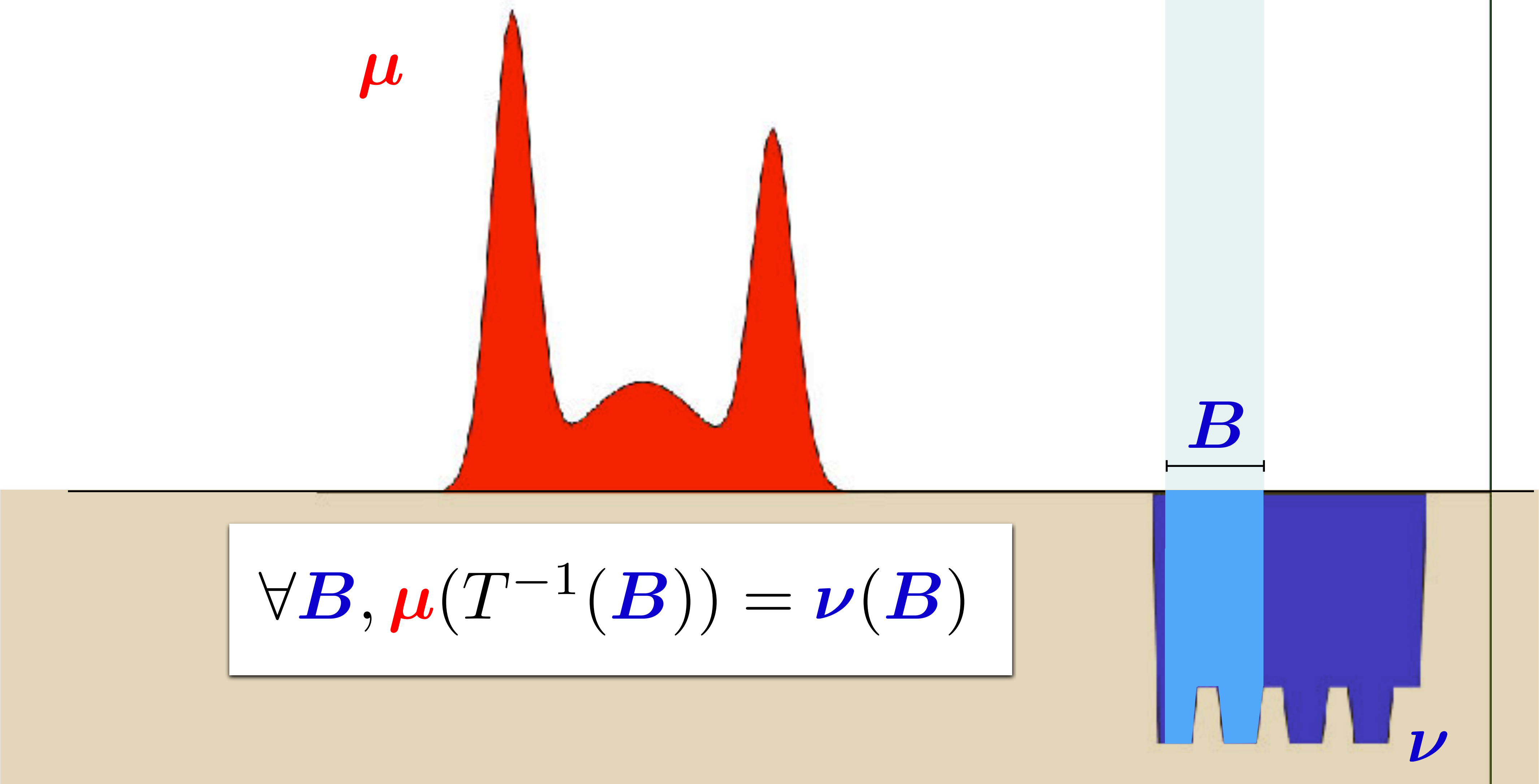
Origins: Monge's Problem

T must map red to blue.



Origins: Monge's Problem

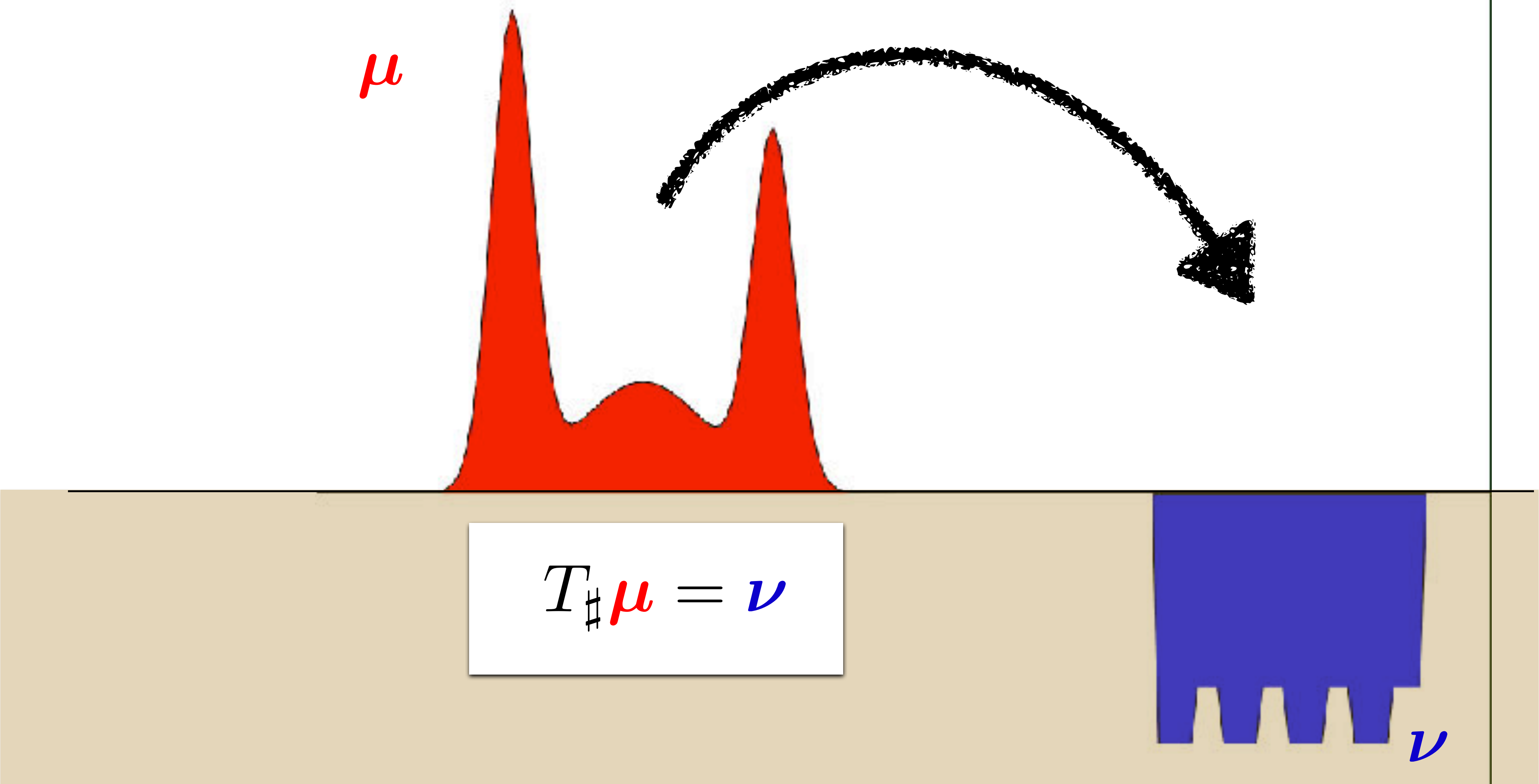
T must map red to blue.



$$\forall B, \mu(T^{-1}(B)) = \nu(B)$$

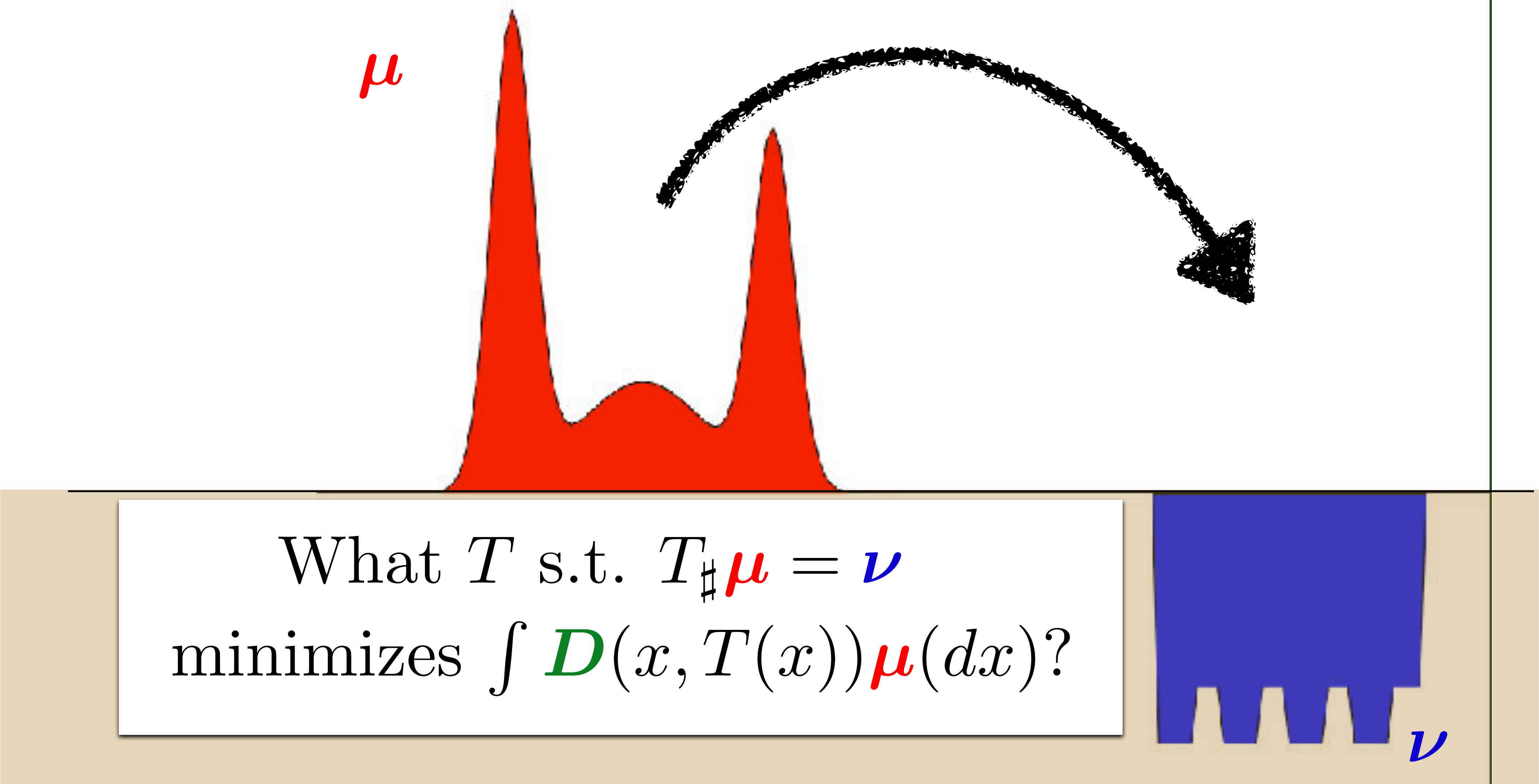
Origins: Monge's Problem

T must **push-forward** *the red measure towards the blue*



Origins: Monge's Problem

T must **push-forward** *the red measure towards the blue*



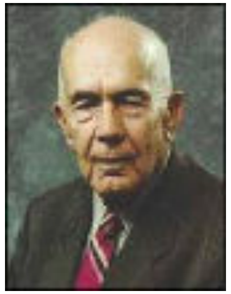
What T s.t. $T_\sharp \mu = \nu$

minimizes $\int D(x, T(x))\mu(dx)$?

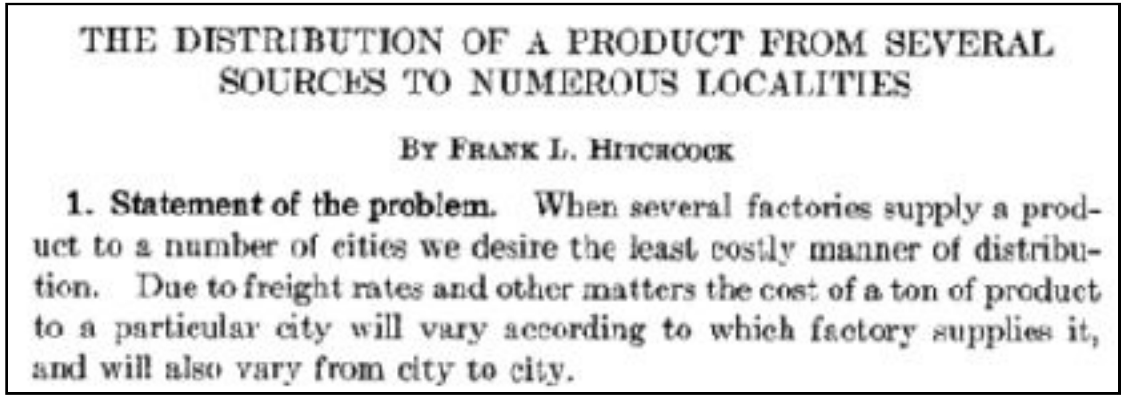
Kantorovich Problem

Kantorovich

1939



Hitchcock



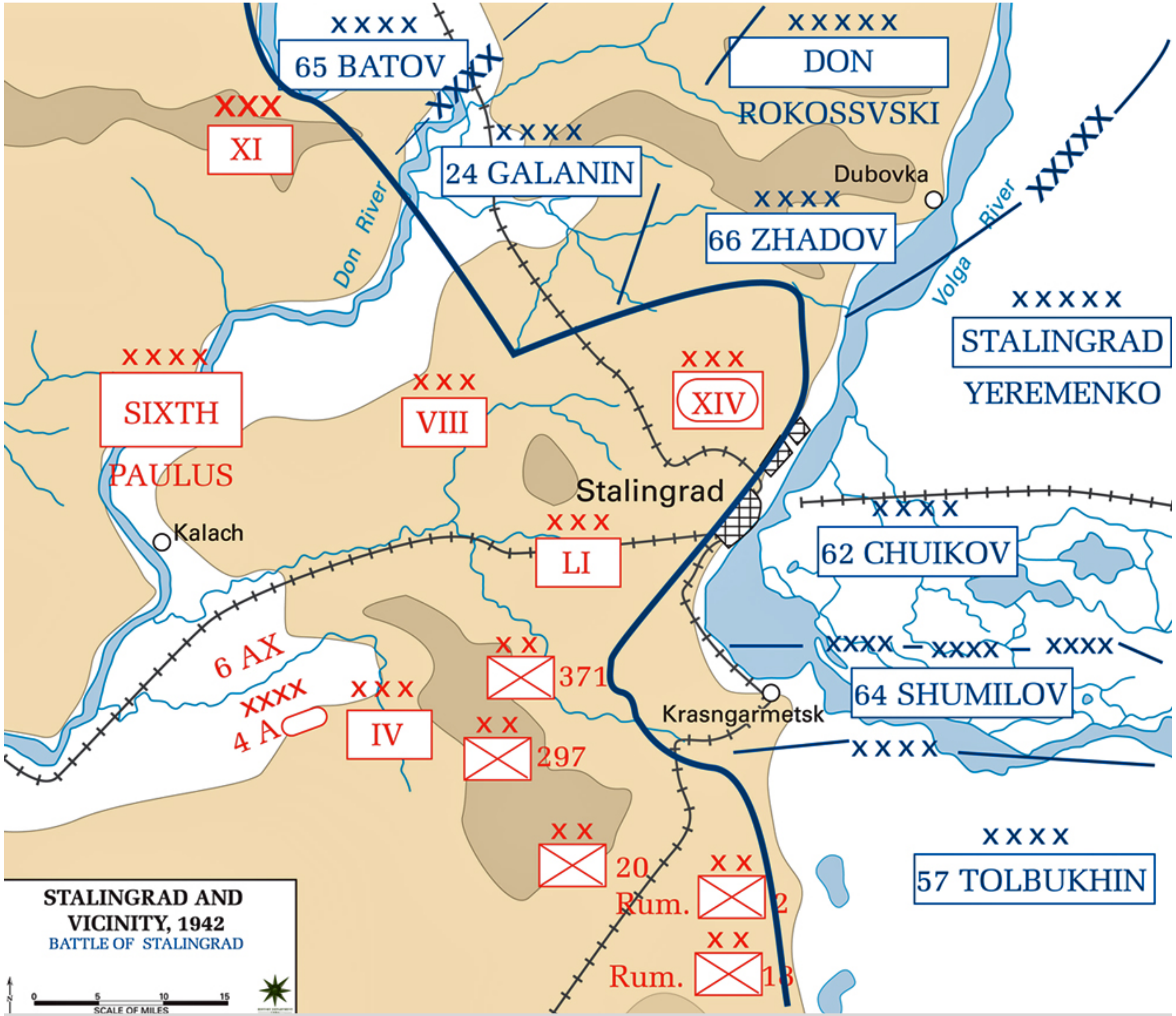
THE DISTRIBUTION OF A PRODUCT FROM SEVERAL SOURCES TO NUMEROUS LOCALITIES

BY FRANK L. HITCHCOCK

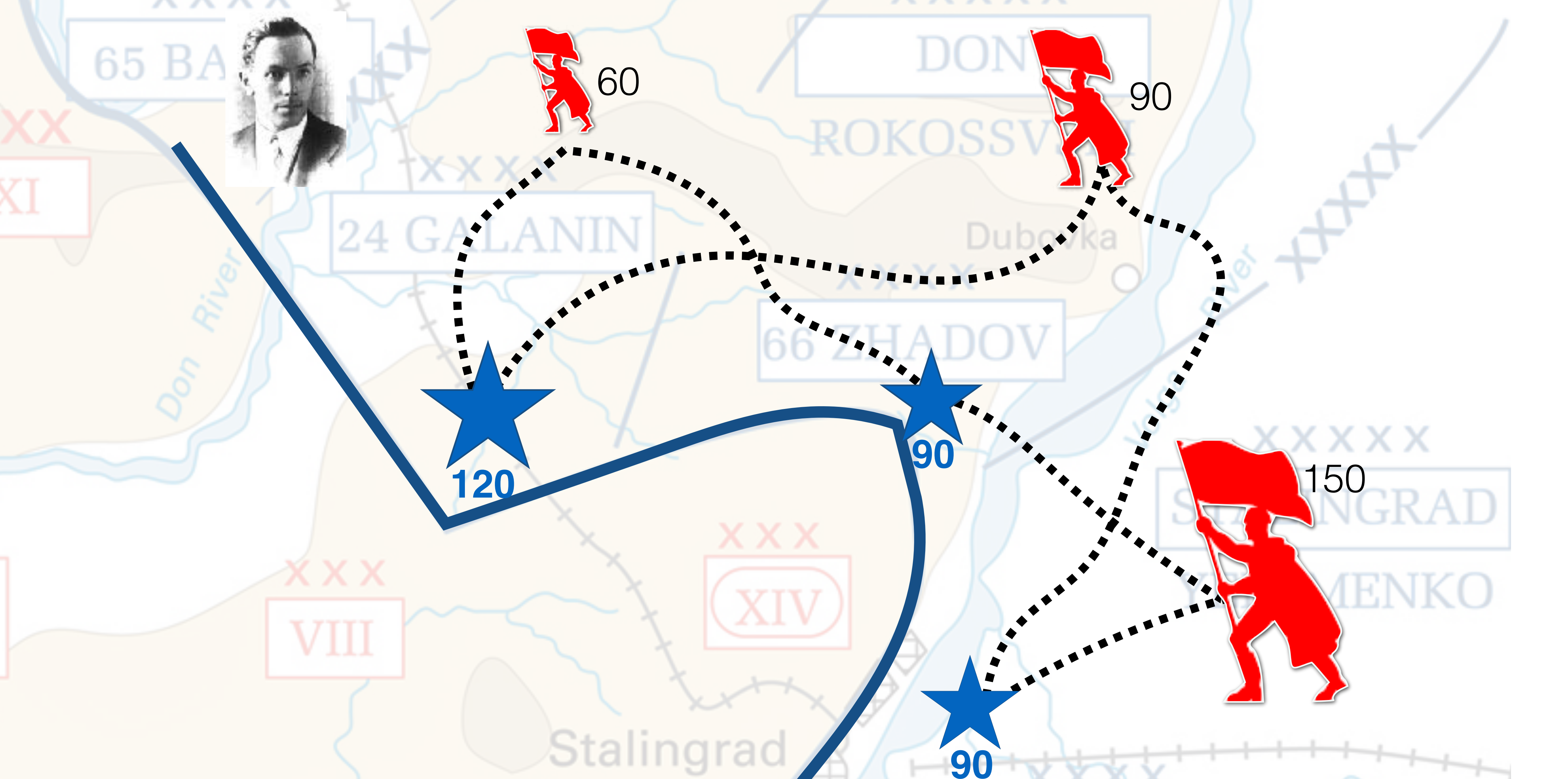
1. Statement of the problem. When several factories supply a product to a number of cities we desire the least costly manner of distribution. Due to freight rates and other matters the cost of a ton of product to a particular city will vary according to which factory supplies it, and will also vary from city to city.

1941

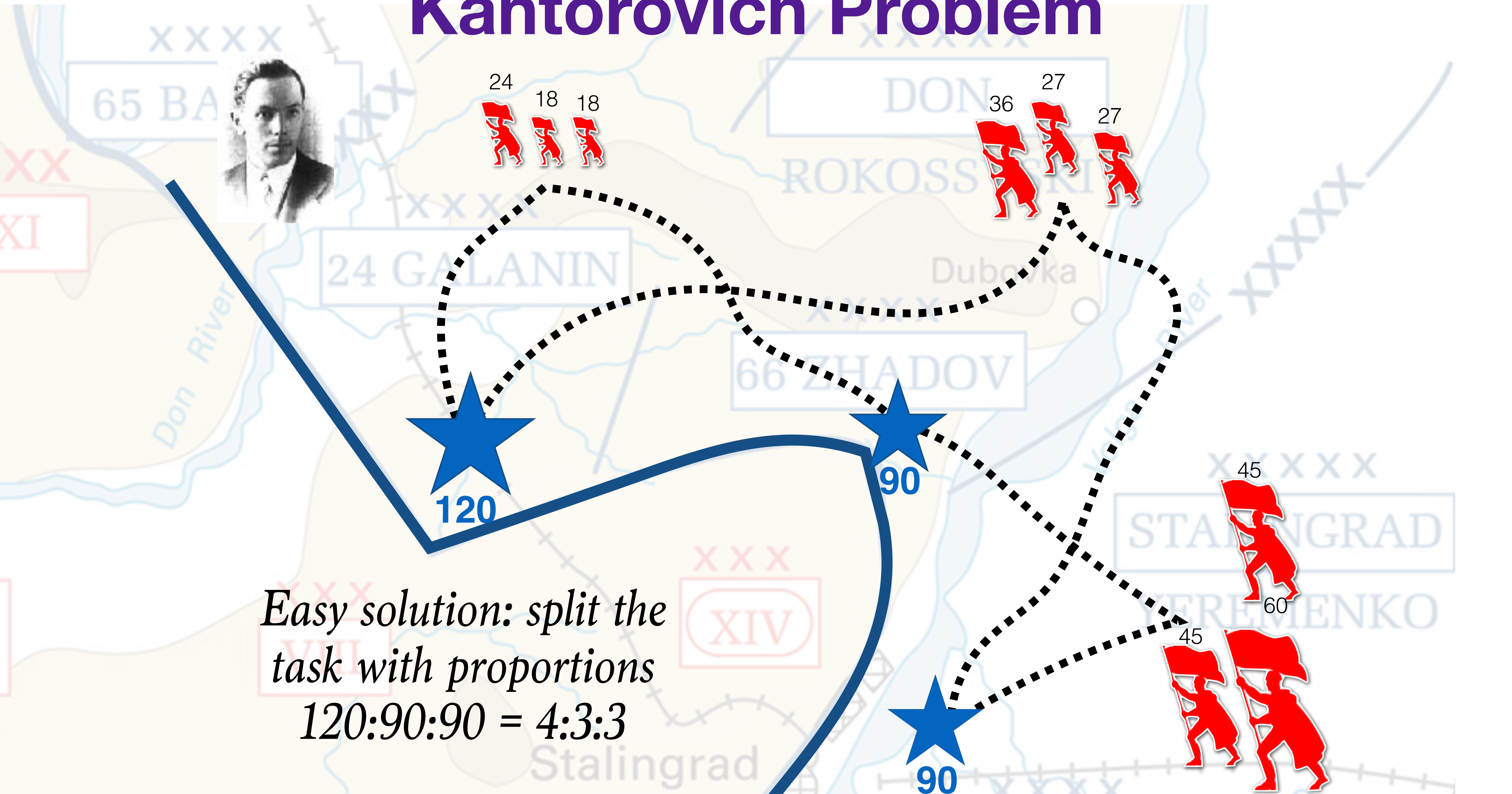
Kantorovich Problem



Kantorovich Problem

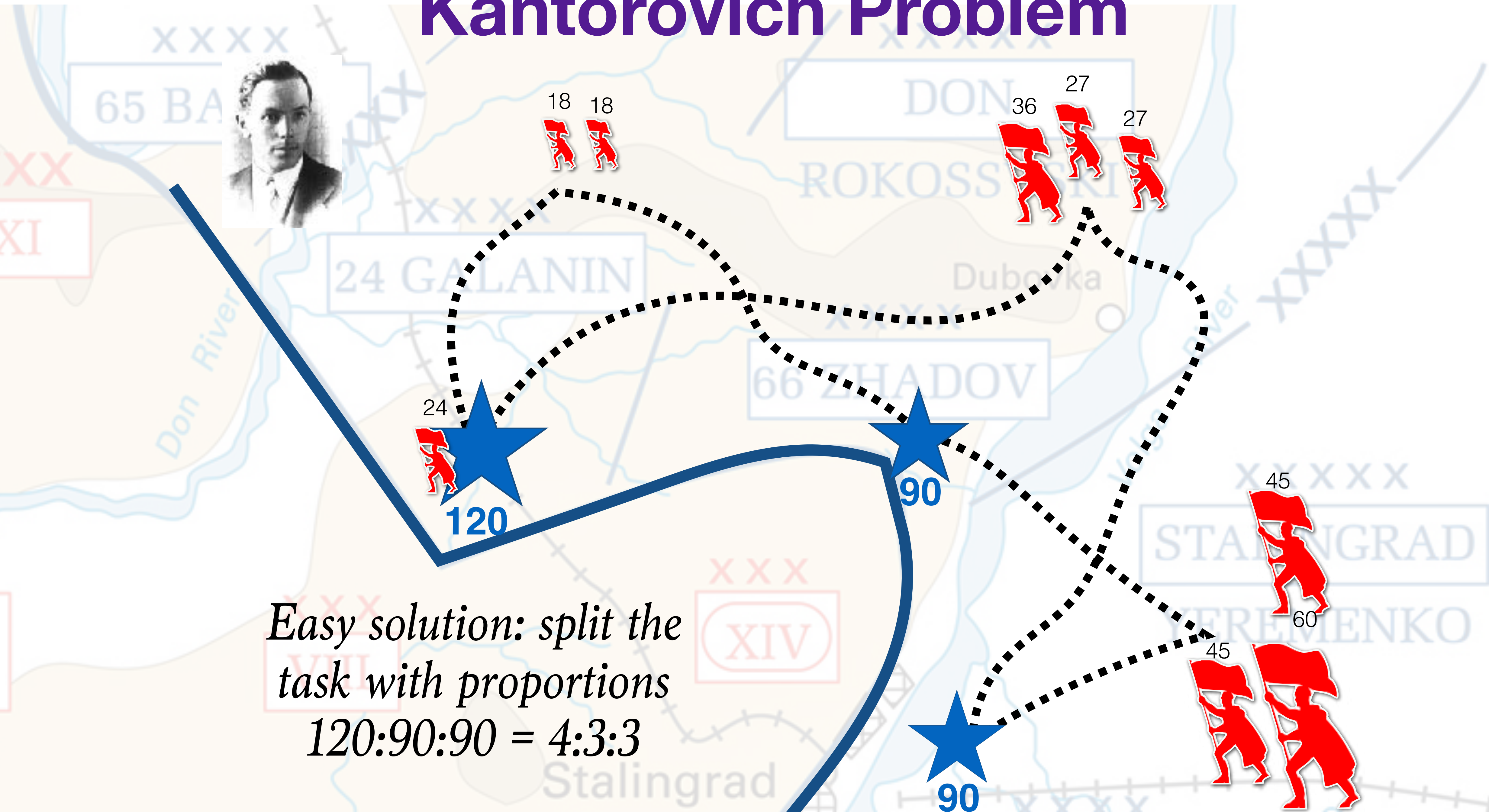


Kantorovich Problem



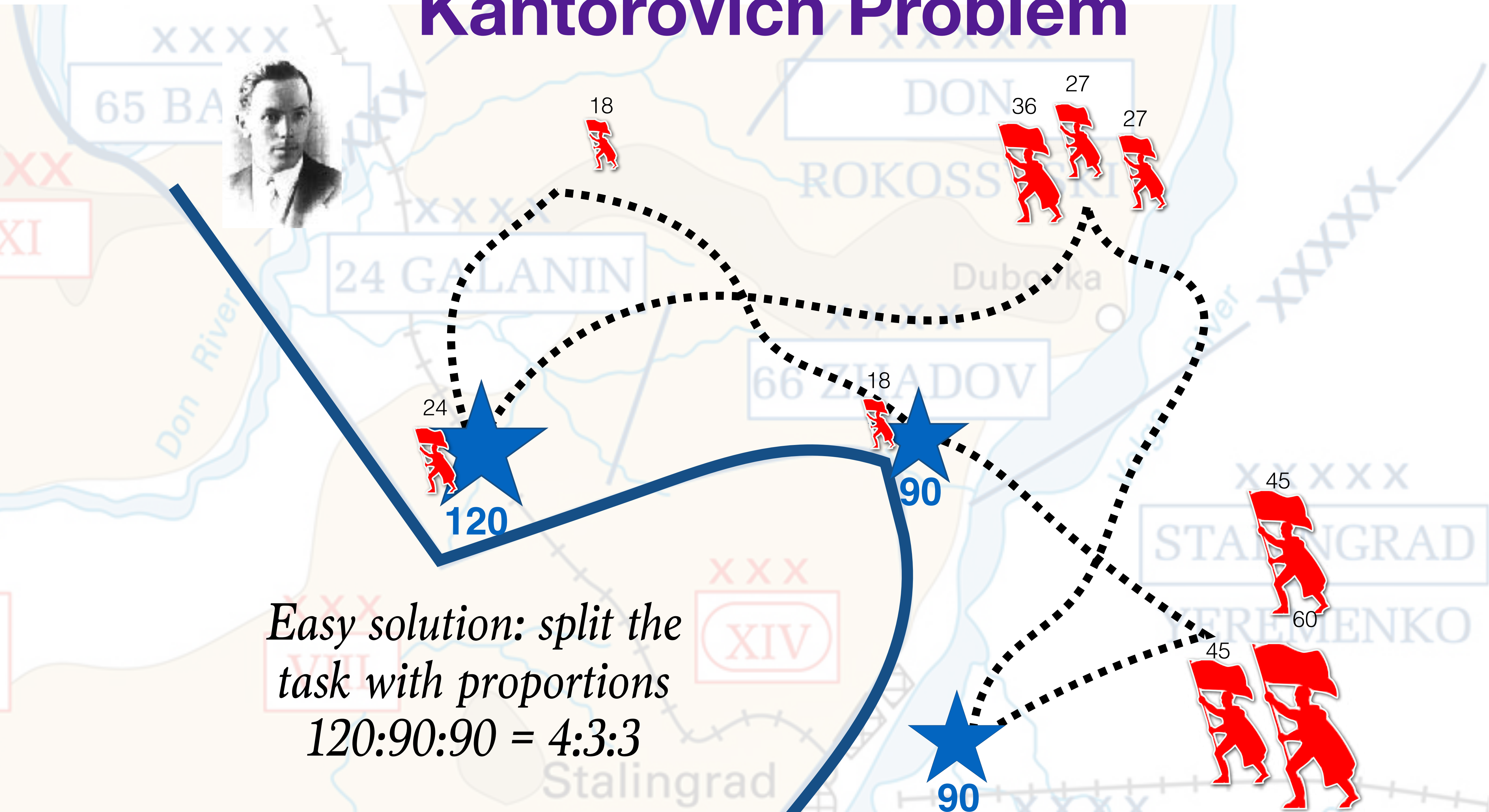
Easy solution: split the task with proportions 120:90:90 = 4:3:3

Kantorovich Problem



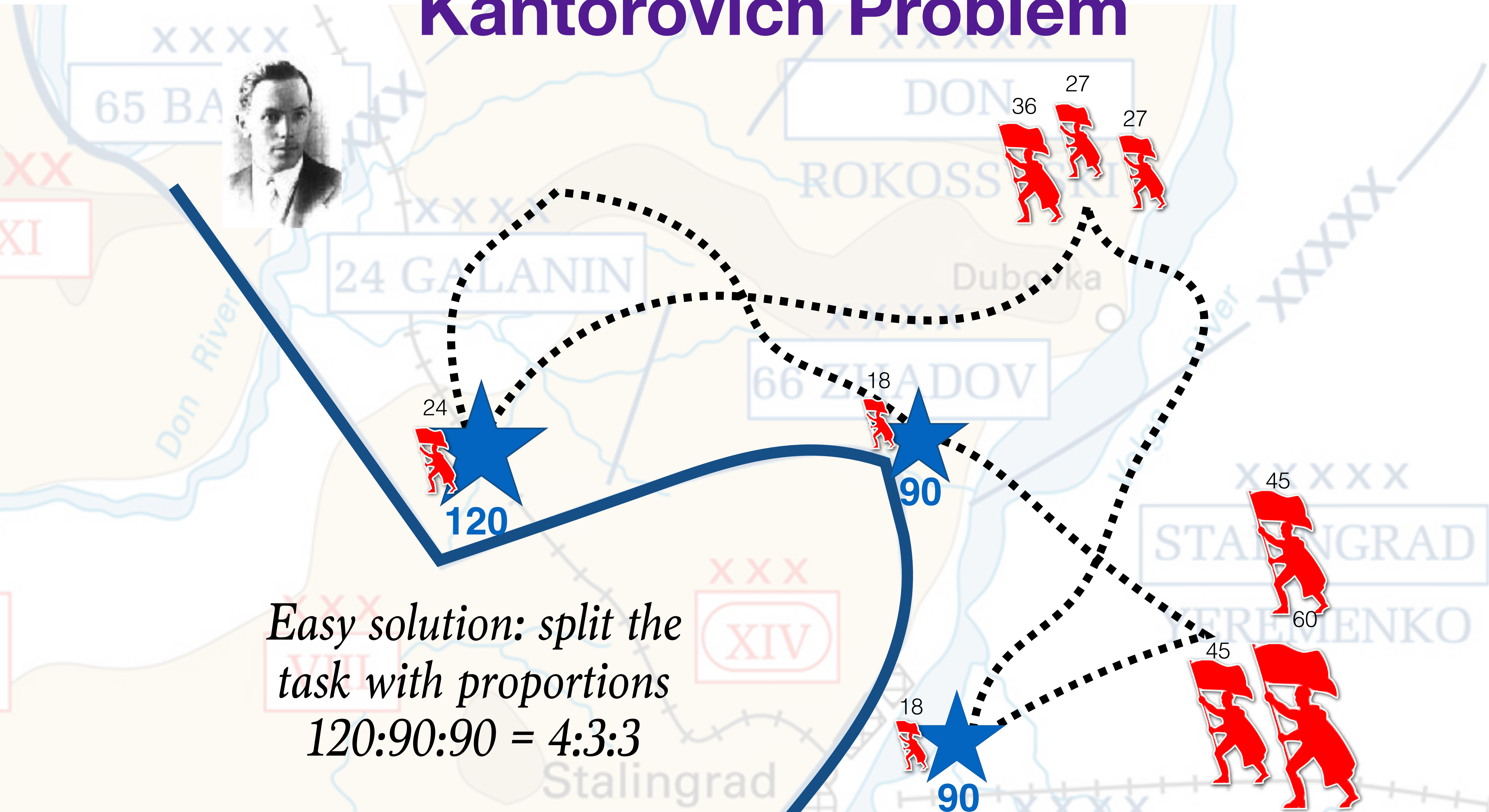
Easy solution: split the task with proportions 120:90:90 = 4:3:3

Kantorovich Problem



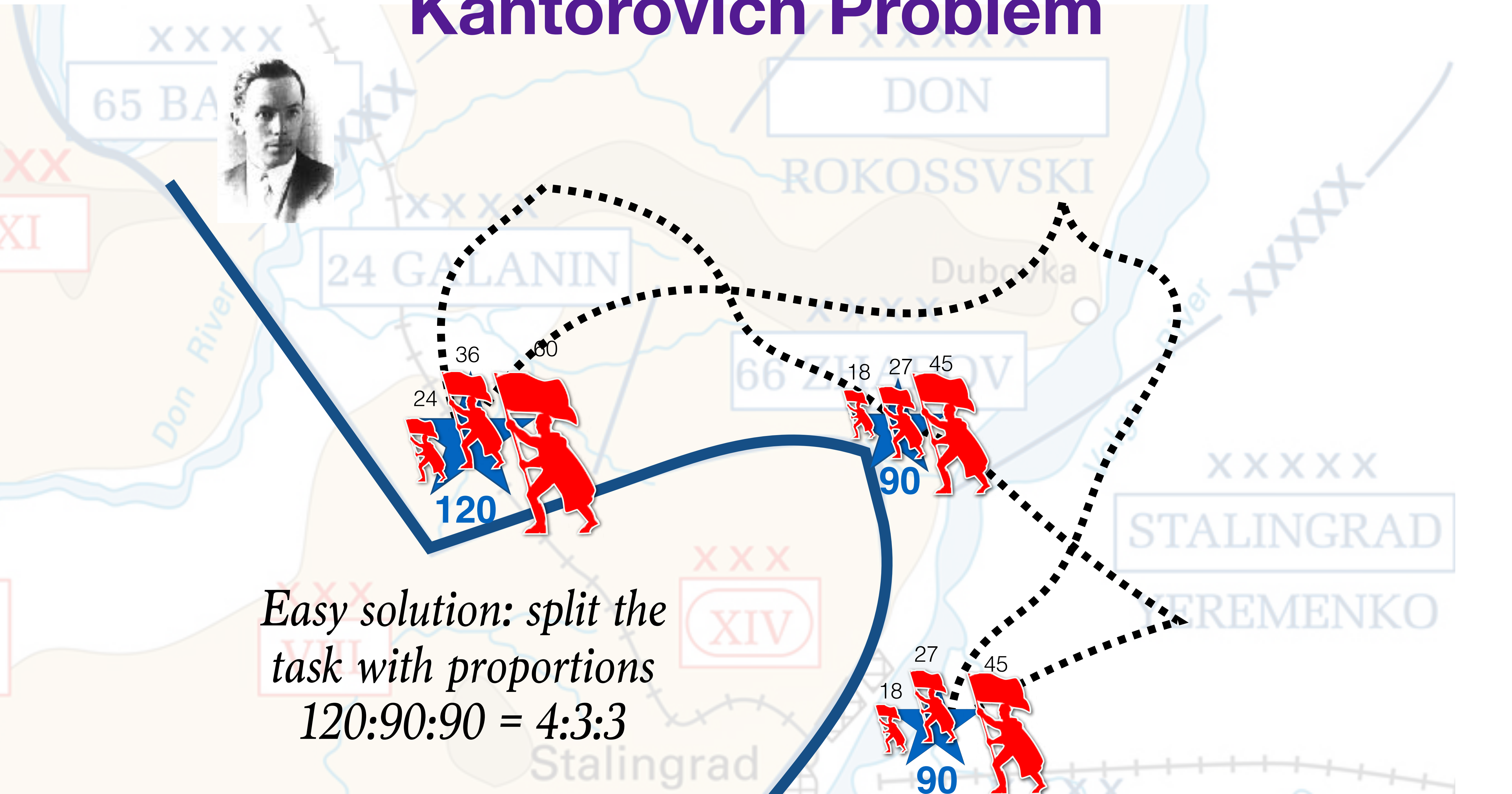
Easy solution: split the task with proportions 120:90:90 = 4:3:3

Kantorovich Problem



Easy solution: split the task with proportions 120:90:90 = 4:3:3

Kantorovich Problem



Easy solution: split the task with proportions 120:90:90 = 4:3:3

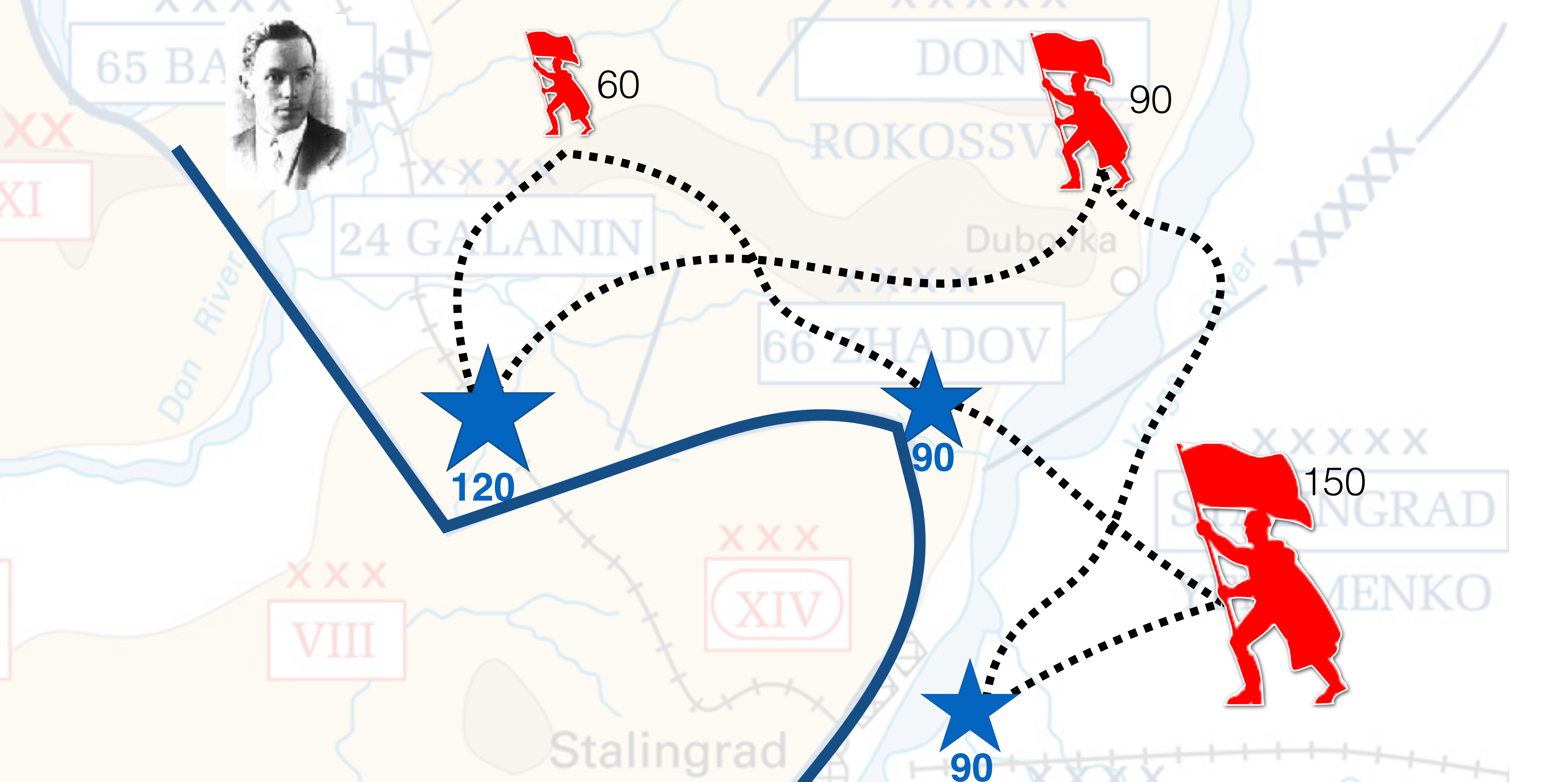
Kantorovich Problem

Naive approach results in too many displacements.

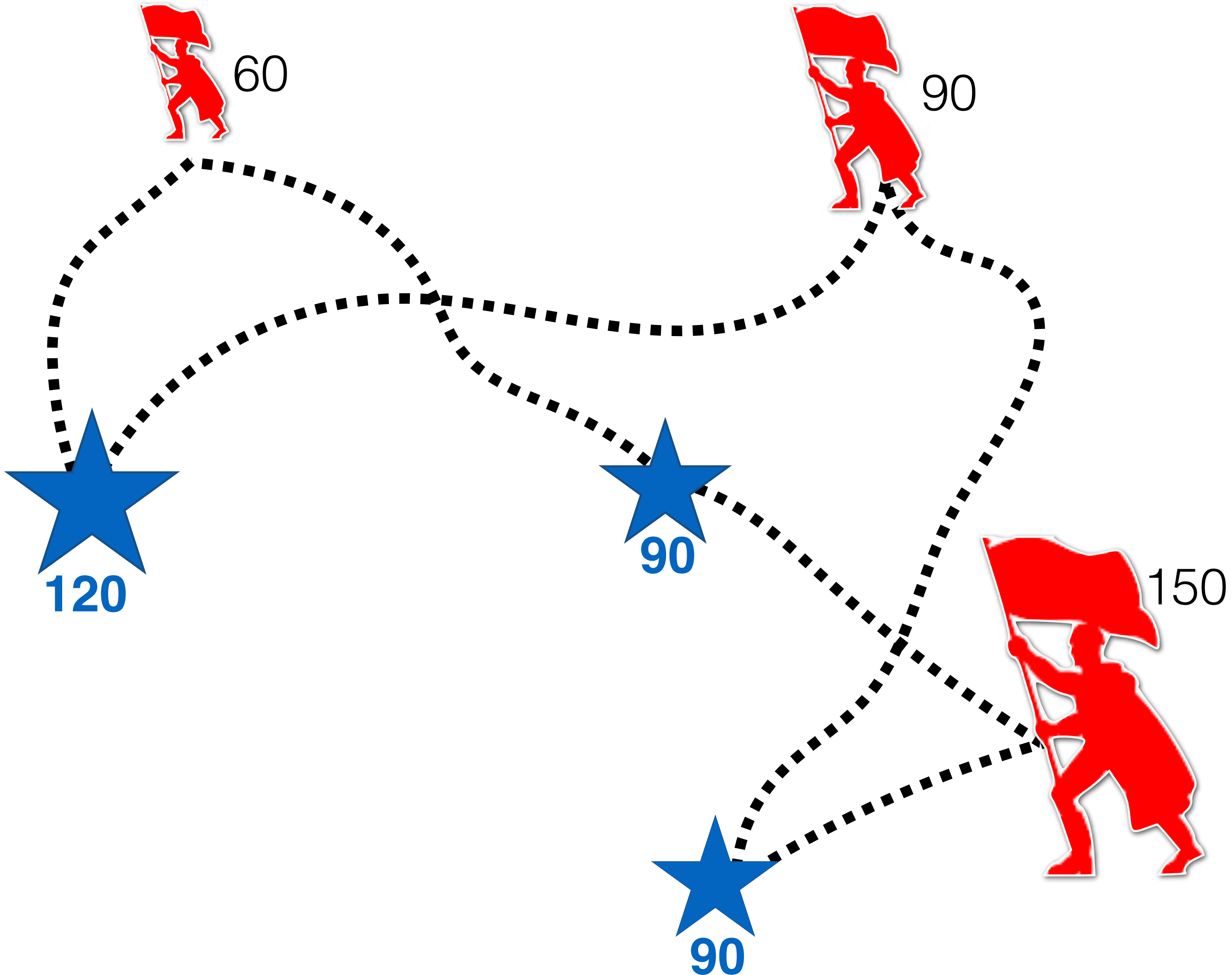
Goal: find a cheaper alternative

Easy solution: split the task with proportions 120:90:90 = 4:3:3

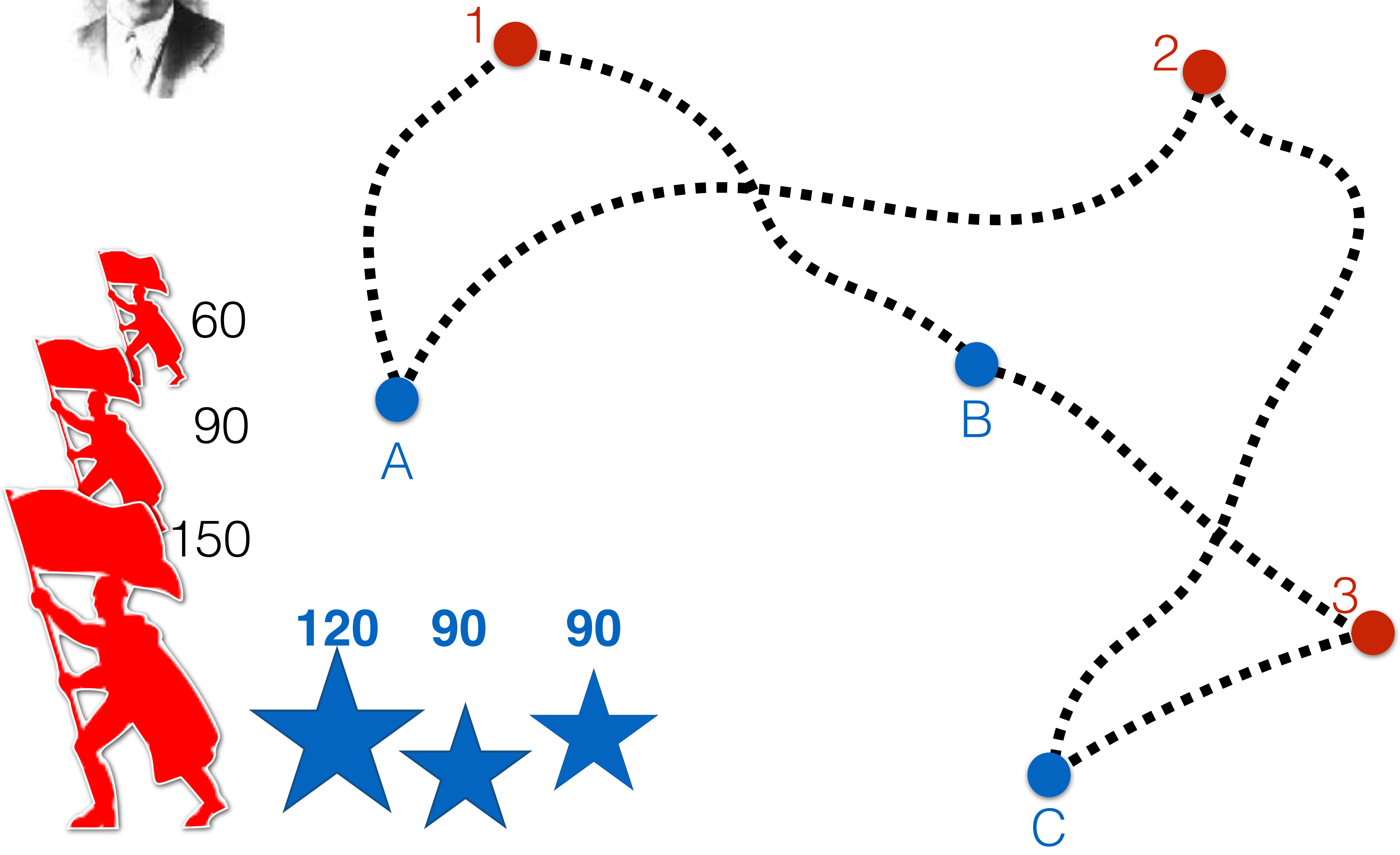
Kantorovich Problem



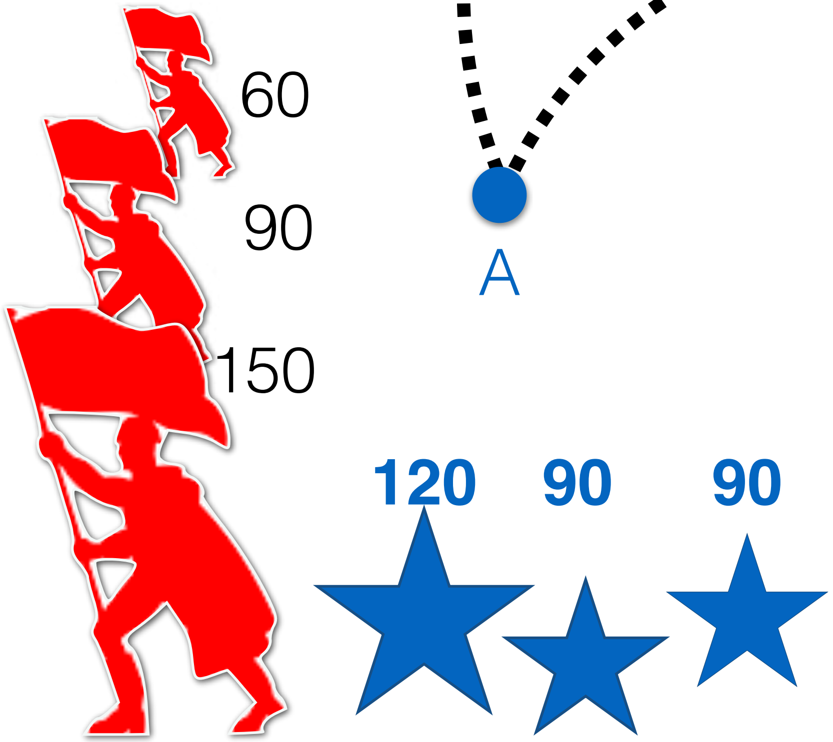
Kantorovich Problem

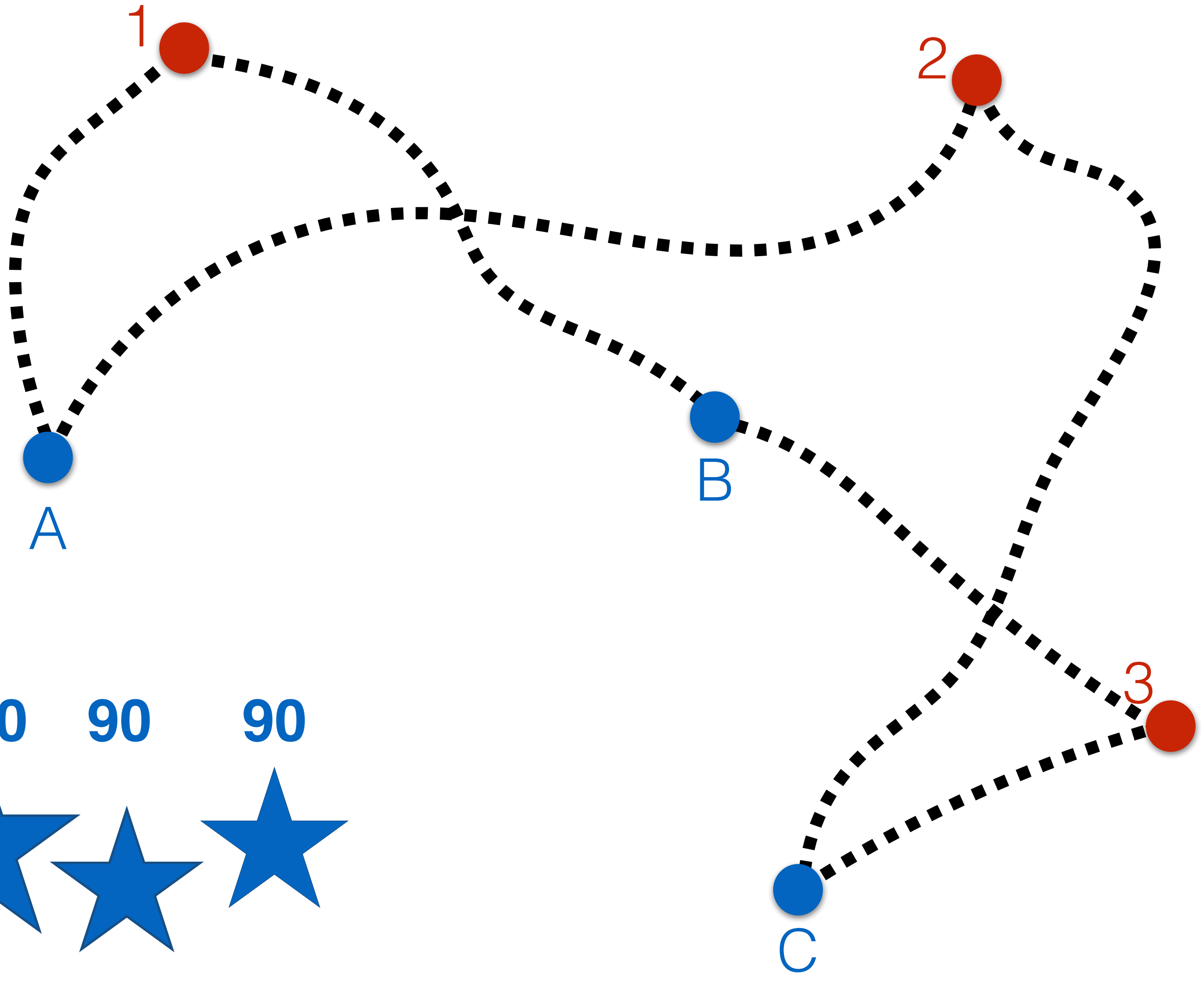


Kantorovich Problem



Kantorovich Problem





Kantorovich Problem

	120	90	90
60	?	?	?
90	?	?	?
150	?	?	?

	A	B	C
1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}

Kantorovich Problem

Transportation matrix

60	?	?	?
90	?	?	?
150	?	?	?
	120	90	90

Distance matrix

1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}
	A	B	C

Kantorovich Problem

The problem is entirely described by **counts** *and a* **cost/distance matrix**

Transportation matrix

60	?	?	?
90	?	?	?
150	?	?	?
	120	90	90

Distance matrix

	A	B	C
1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}

Kantorovich Problem

Transportation matrix

60	?	?	?
90	?	?	?
150	?	?	?
	120	**90**	**90**

Distance matrix

1	$d_{1\mathrm{A}}$	$d_{1\mathrm{B}}$	$d_{1\mathrm{C}}$
2	$d_{2\mathrm{A}}$	$d_{2\mathrm{B}}$	$d_{2\mathrm{C}}$
3	$d_{3\mathrm{A}}$	$d_{3\mathrm{B}}$	$d_{3\mathrm{C}}$
	A	B	C

Kantorovich Problem

Transportation matrix

	A	B	C
60	p_{1A}	p_{1B}	p_{1C}
90	p_{2A}	p_{2B}	p_{2C}
150	p_{3A}	p_{3B}	p_{3C}
	120	**90**	**90**

Distance matrix

1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}
	A	B	C

Kantorovich Problem

Transportation matrix

$$\begin{array}{c|c|c|c|}
\hline
a_1 & p_{1A} & p_{1B} & p_{1C} \\
\hline
a_2 & p_{2A} & p_{2B} & p_{2C} \\
\hline
a_3 & p_{3A} & p_{3B} & p_{3C} \\
\hline
\multicolumn{1}{c}{} & \multicolumn{1}{c}{b_A} & \multicolumn{1}{c}{b_B} & \multicolumn{1}{c}{b_C}
\end{array}$$

Distance matrix

$$\begin{array}{c|c|c|c|}
\hline
1 & d_{1A} & d_{1B} & d_{1C} \\
\hline
2 & d_{2A} & d_{2B} & d_{2C} \\
\hline
3 & d_{3A} & d_{3B} & d_{3C} \\
\hline
\multicolumn{1}{c}{} & \multicolumn{1}{c}{A} & \multicolumn{1}{c}{B} & \multicolumn{1}{c}{C}
\end{array}$$

Kantorovich Problem

Transportation matrix

$$\begin{array}{c|c|c|c|}
\hline
a_1 & p_{1A} & p_{1B} & p_{1C} \\
\hline
a_2 & p_{2A} & p_{2B} & p_{2C} \\
\hline
a_3 & p_{3A} & p_{3B} & p_{3C} \\
\hline
\multicolumn{1}{c}{} & \multicolumn{1}{c}{b_A} & \multicolumn{1}{c}{b_B} & \multicolumn{1}{c}{b_C}
\end{array}$$

Distance matrix

$$\begin{array}{c|c|c|c|}
\hline
1 & d_{1A} & d_{1B} & d_{1C} \\
\hline
2 & d_{2A} & d_{2B} & d_{2C} \\
\hline
3 & d_{3A} & d_{3B} & d_{3C} \\
\hline
\multicolumn{1}{c}{} & \multicolumn{1}{c}{A} & \multicolumn{1}{c}{B} & \multicolumn{1}{c}{C}
\end{array}$$

Constraints

$$\forall i \in \{1, 2, 3\}, \quad \sum_{j \in \{A,B,C\}} p_{ij} = a_i$$

$$\forall j \in \{A, B, C\}, \quad \sum_{i \in \{1,2,3\}} p_{ij} = b_j$$

$$p_{ij} \geq 0$$

Kantorovich Problem

Transportation matrix

$$\begin{array}{c|c|c|c|} \cline{2-4} a_1 & p_{1\mathrm{A}} & p_{1\mathrm{B}} & p_{1\mathrm{C}} \\ \cline{2-4} a_2 & p_{2\mathrm{A}} & p_{2\mathrm{B}} & p_{2\mathrm{C}} \\ \cline{2-4} a_3 & p_{3\mathrm{A}} & p_{3\mathrm{B}} & p_{3\mathrm{C}} \\ \cline{2-4} \multicolumn{1}{c}{} & \multicolumn{1}{c}{b_\mathrm{A}} & \multicolumn{1}{c}{b_\mathrm{B}} & \multicolumn{1}{c}{b_\mathrm{C}} \end{array}$$

Distance matrix

$$\begin{array}{c|c|c|c|} \cline{2-4} 1 & d_{1\mathrm{A}} & d_{1\mathrm{B}} & d_{1\mathrm{C}} \\ \cline{2-4} 2 & d_{2\mathrm{A}} & d_{2\mathrm{B}} & d_{2\mathrm{C}} \\ \cline{2-4} 3 & d_{3\mathrm{A}} & d_{3\mathrm{B}} & d_{3\mathrm{C}} \\ \cline{2-4} \multicolumn{1}{c}{} & \multicolumn{1}{c}{\mathrm{A}} & \multicolumn{1}{c}{\mathrm{B}} & \multicolumn{1}{c}{\mathrm{C}} \end{array}$$

Constraints

$$\forall i \in \{1, 2, 3\}, \quad \sum_{j \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}} p_{ij} = a_i$$

$$\forall j \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}, \quad \sum_{i \in \{1,2,3\}} p_{ij} = b_j$$

$$p_{ij} \geq 0$$

Cost function

$$C(P) = \sum_{j \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}} \sum_{i \in \{1,2,3\}} p_{ij} d_{ij}$$

Kantorovich Problem

Transportation matrix

a_1	$p_{1\mathrm{A}}$	$p_{1\mathrm{B}}$	$p_{1\mathrm{C}}$
a_2	$p_{2\mathrm{A}}$	$p_{2\mathrm{B}}$	$p_{2\mathrm{C}}$
a_3	$p_{3\mathrm{A}}$	$p_{3\mathrm{B}}$	$p_{3\mathrm{C}}$
	b_A	b_B	b_C

Distance matrix

1	$d_{1\mathrm{A}}$	$d_{1\mathrm{B}}$	$d_{1\mathrm{C}}$
2	$d_{2\mathrm{A}}$	$d_{2\mathrm{B}}$	$d_{2\mathrm{C}}$
3	$d_{3\mathrm{A}}$	$d_{3\mathrm{B}}$	$d_{3\mathrm{C}}$
	A	B	C

Constraints

$$\forall i \in \{1,2,3\}, \quad \sum_{j \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}} p_{ij} = a_i$$

$$\forall j \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}, \quad \sum_{i \in \{1,2,3\}} p_{ij} = b_j$$

$$p_{ij} \geq 0$$

Cost function

$$C(\boldsymbol{P}) = \sum_{j \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}} \sum_{i \in \{1,2,3\}} p_{ij} d_{ij}$$

Problem

$$\min_{\text{all valid } \boldsymbol{P}} C(\boldsymbol{P})$$

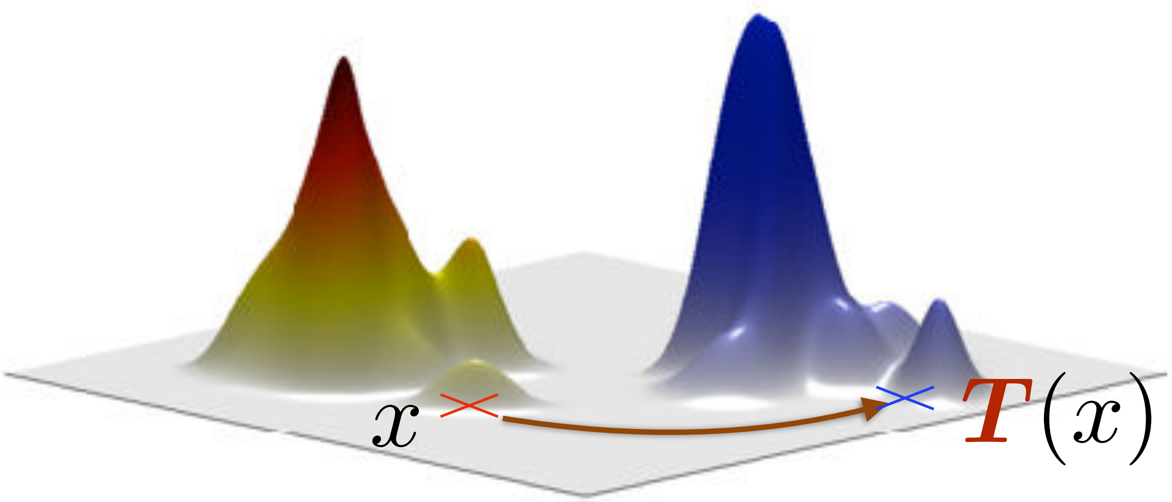
II. Mathematical Formalism of OT

Monge Problem

Ω a probability space, $\boldsymbol{c} : \Omega \times \Omega \to \mathbb{R}$.
$\boldsymbol{\mu}, \boldsymbol{\nu}$ two probability measures in $\mathcal{P}(\Omega)$.

[Monge'81] problem: find a map $\boldsymbol{T} : \Omega \to \Omega$

$$\inf_{\boldsymbol{T}_\sharp \boldsymbol{\mu} = \boldsymbol{\nu}} \int_\Omega \boldsymbol{c}(x, \boldsymbol{T}(x)) \boldsymbol{\mu}(dx)$$

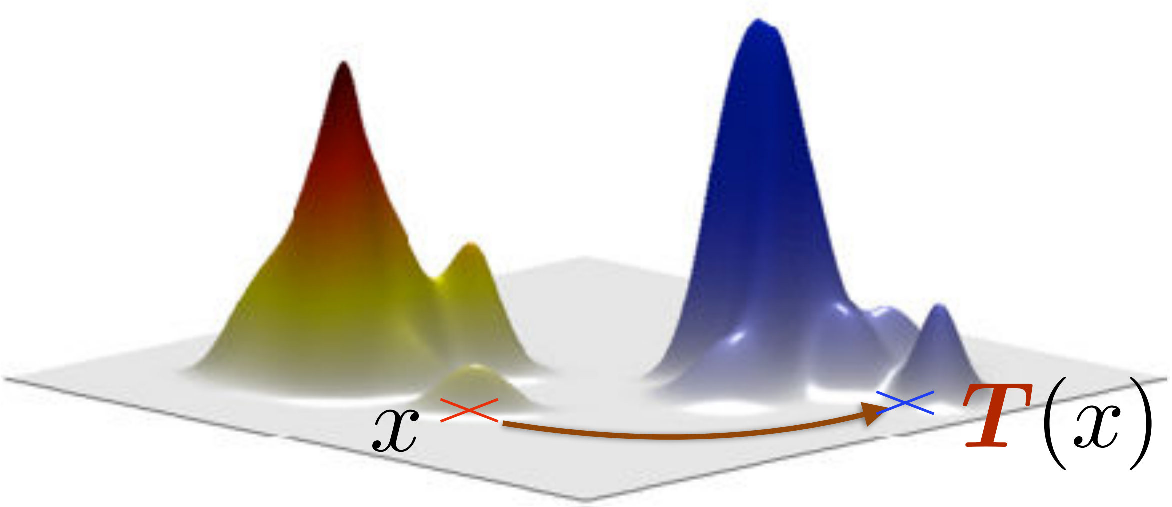


Monge Problem

Ω a probability space, $\boldsymbol{c} : \Omega \times \Omega \to \mathbb{R}$.
$\boldsymbol{\mu}, \boldsymbol{\nu}$ two probability measures in $\mathcal{P}(\Omega)$.

[Monge'81] problem: find a map $\boldsymbol{T} : \Omega \to \Omega$

[Brenier'87] If $\Omega = \mathbb{R}^d$, $\boldsymbol{c} = \|\cdot - \cdot\|^2$,
$\boldsymbol{\mu}, \boldsymbol{\nu}$ a.c., then $\boldsymbol{T} = \nabla \boldsymbol{u}$, $\boldsymbol{u}$ **convex**.

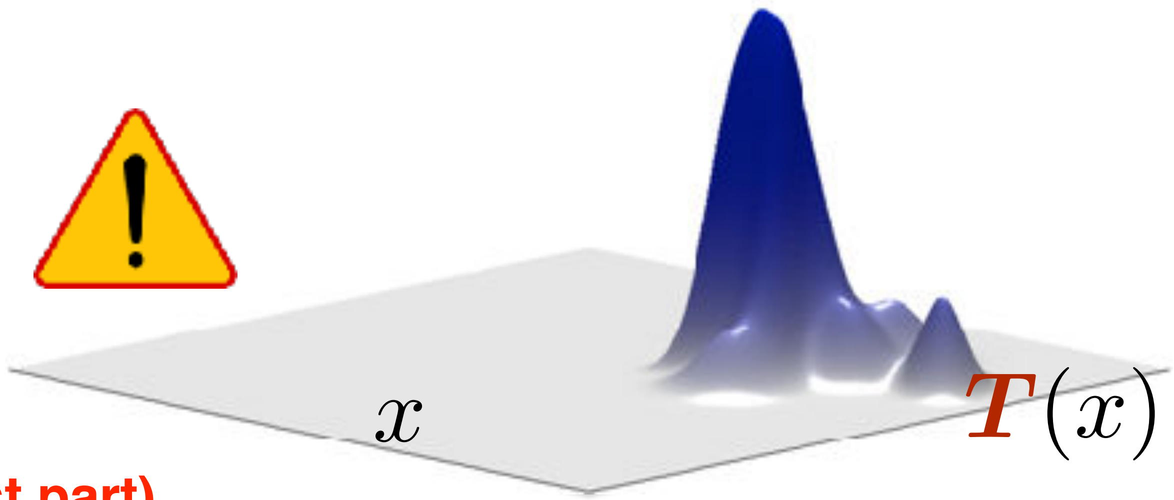


Monge Problem

Ω a probability space, $c : \Omega \times \Omega \to \mathbb{R}$.
μ, ν two probability measures in $\mathcal{P}(\Omega)$.

[**Monge'81**] problem: find a map $T : \Omega \to \Omega$

$$\inf_{T_\sharp \mu = \nu} \int_\Omega c(x, T(x)) \mu(dx)$$



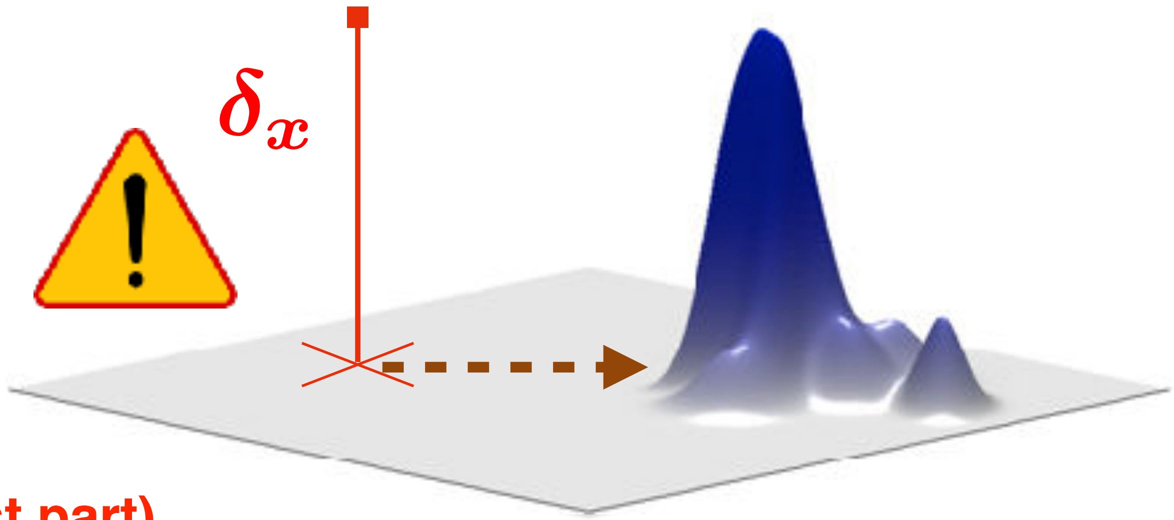
Homework Exercise #2 (last part)

Monge Problem

Ω a probability space, $c : \Omega \times \Omega \to \mathbb{R}$.
μ, ν two probability measures in $\mathcal{P}(\Omega)$.

[Monge'81] problem: find a map $T : \Omega \to \Omega$

$$\inf_{T_\sharp \mu = \nu} \int_\Omega c(x, T(x))\mu(dx)$$



Homework Exercise #2 (last part)

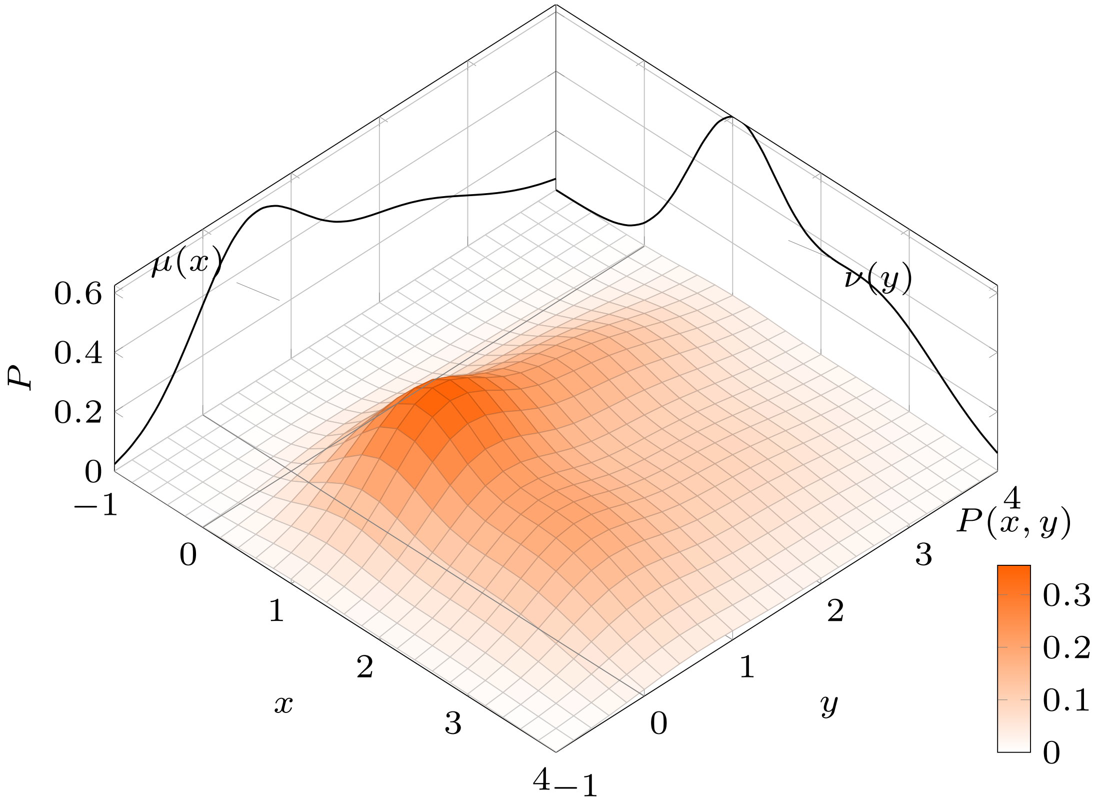
Kantorovich Relaxation

- Instead of maps $T : \Omega \to \Omega$, consider probabilistic maps, i.e. **couplings** $P \in \mathcal{P}(\Omega \times \Omega)$

$$\Pi(\mu, \nu) \stackrel{\text{def}}{=} \{P \in \mathcal{P}(\Omega \times \Omega) | \forall A, B \subset \Omega, \\ P(A \times \Omega) = \mu(A), \\ P(\Omega \times B) = \nu(B)\}$$

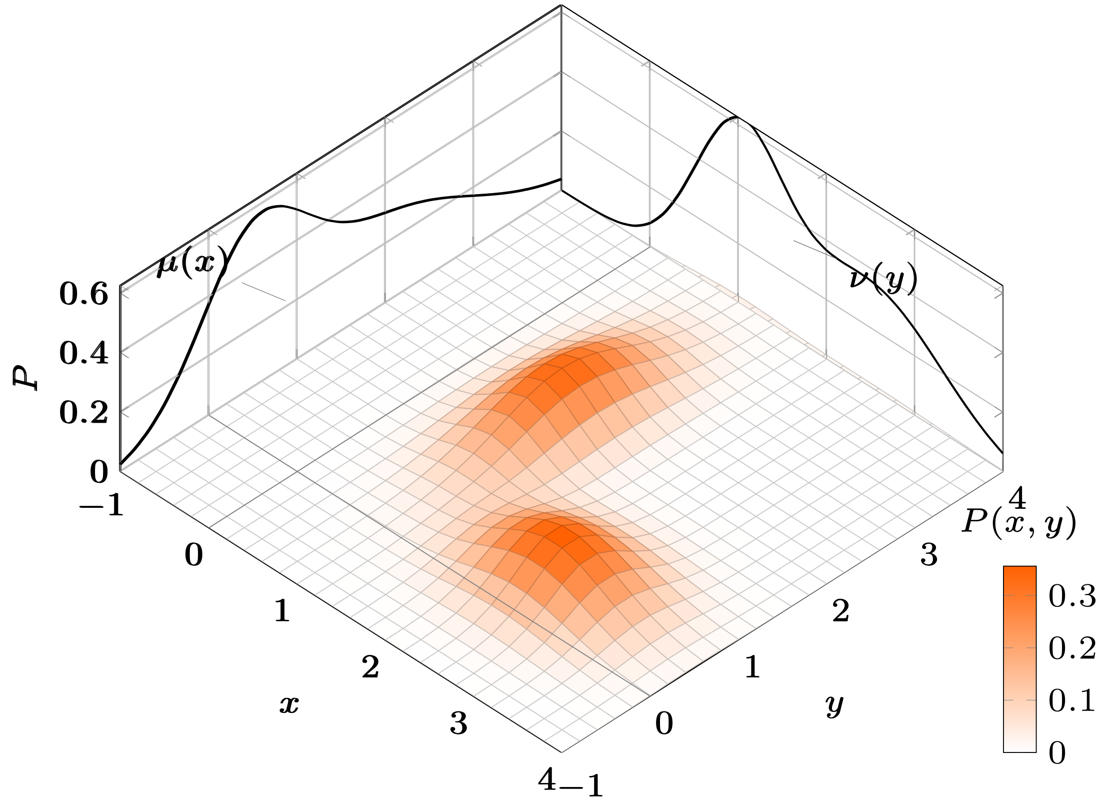
Kantorovich Relaxation

$$\Pi(\mu, \nu) \stackrel{\text{def}}{=} \{P \in \mathcal{P}(\Omega \times \Omega) \,|\, \forall A, B \subset \Omega, \\ P(A \times \Omega) = \mu(A), P(\Omega \times B) = \nu(B)\}$$



Kantorovich Relaxation

$$\Pi(\mu,\nu) \overset{\text{def}}{=} \{P \in \mathcal{P}(\Omega \times \Omega) \,|\, \forall A, B \subset \Omega,$$
$$P(A \times \Omega) = \mu(A), P(\Omega \times B) = \nu(B)\}$$



Kantorovich Problem

Def. Given $\boldsymbol{\mu}, \boldsymbol{\nu}$ in $\mathcal{P}(\Omega)$; a cost function $\boldsymbol{c}$ on $\Omega \times \Omega$, the Kantorovich problem is

$$\inf_{\boldsymbol{P} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \iint \boldsymbol{c}(x, y) \boldsymbol{P}(dx, dy).$$

Kantorovich Problem

Def. Given $\boldsymbol{\mu}, \boldsymbol{\nu}$ in $\mathcal{P}(\Omega)$; a cost function $\boldsymbol{c}$ on $\Omega \times \Omega$, the Kantorovich problem is

$$\inf_{\boldsymbol{P} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \iint \boldsymbol{c}(x, y) \boldsymbol{P}(dx, dy).$$

$$\inf_{\boldsymbol{P} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \mathbb{E}_{\boldsymbol{P}}[\boldsymbol{c}(X, Y)]$$

Links between Monge & Kantorovich

Prop. For "well behaved" costs $\boldsymbol{c}$, if $\boldsymbol{\mu}$ has a density then an *optimal* Monge map $T^\star$ between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ must exist.

Prop. In that case

$$\boldsymbol{P}^\star := (\mathrm{Id}, T^\star)_\sharp \boldsymbol{\mu} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$$

is also *optimal* for the Kantorovich problem.

[Brenier'91] [Smith&Knott'87] [McCann'01]

(Kantorovich) Wasserstein Distances

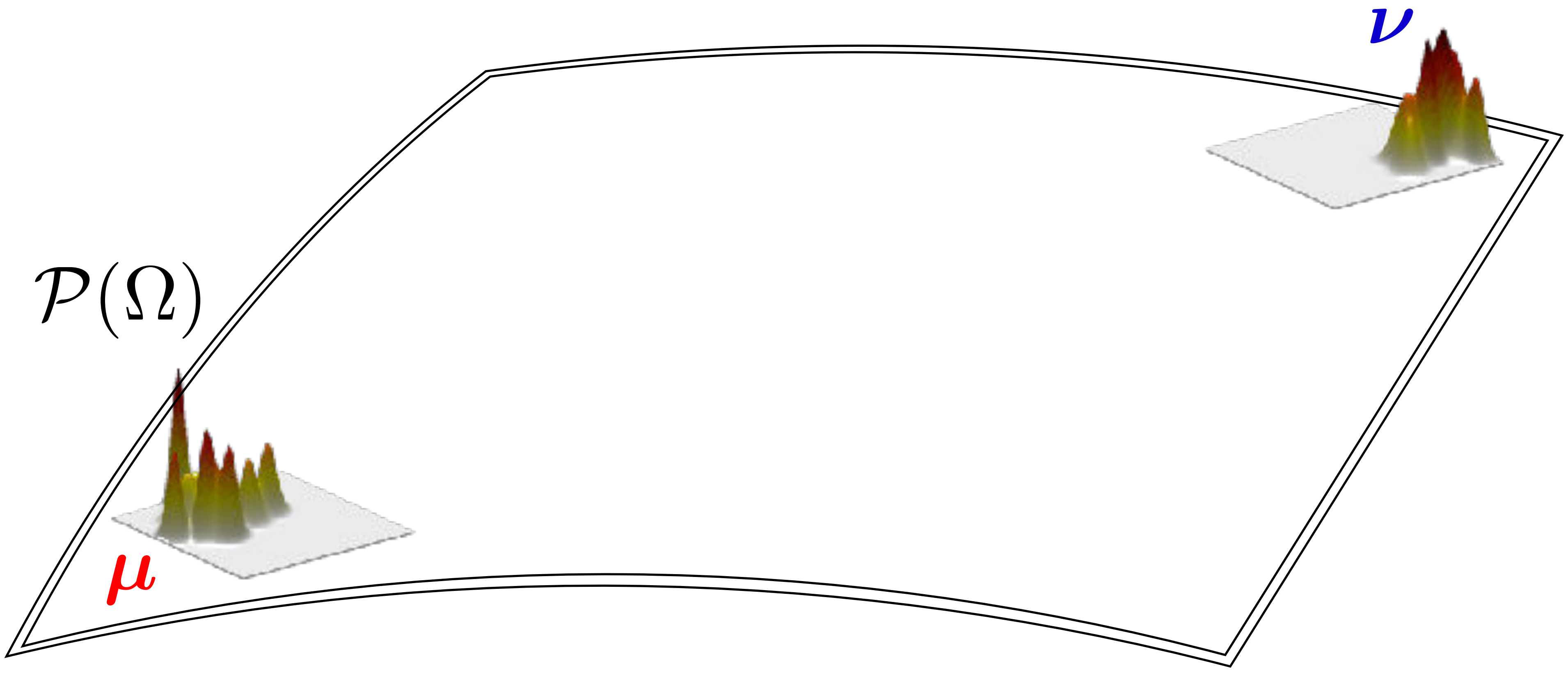
Let $p \geq 1$. Let $\boldsymbol{c} := \boldsymbol{D}$, a metric.

Def. The p-**Wasserstein distance** between $\boldsymbol{\mu}, \boldsymbol{\nu}$ in $\mathcal{P}(\Omega)$ is

$$W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) \stackrel{\text{def}}{=} \left(\inf_{\boldsymbol{P} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \iint \boldsymbol{D}(\boldsymbol{x}, \boldsymbol{y})^p \boldsymbol{P}(dx, dy) \right)^{1/p}.$$

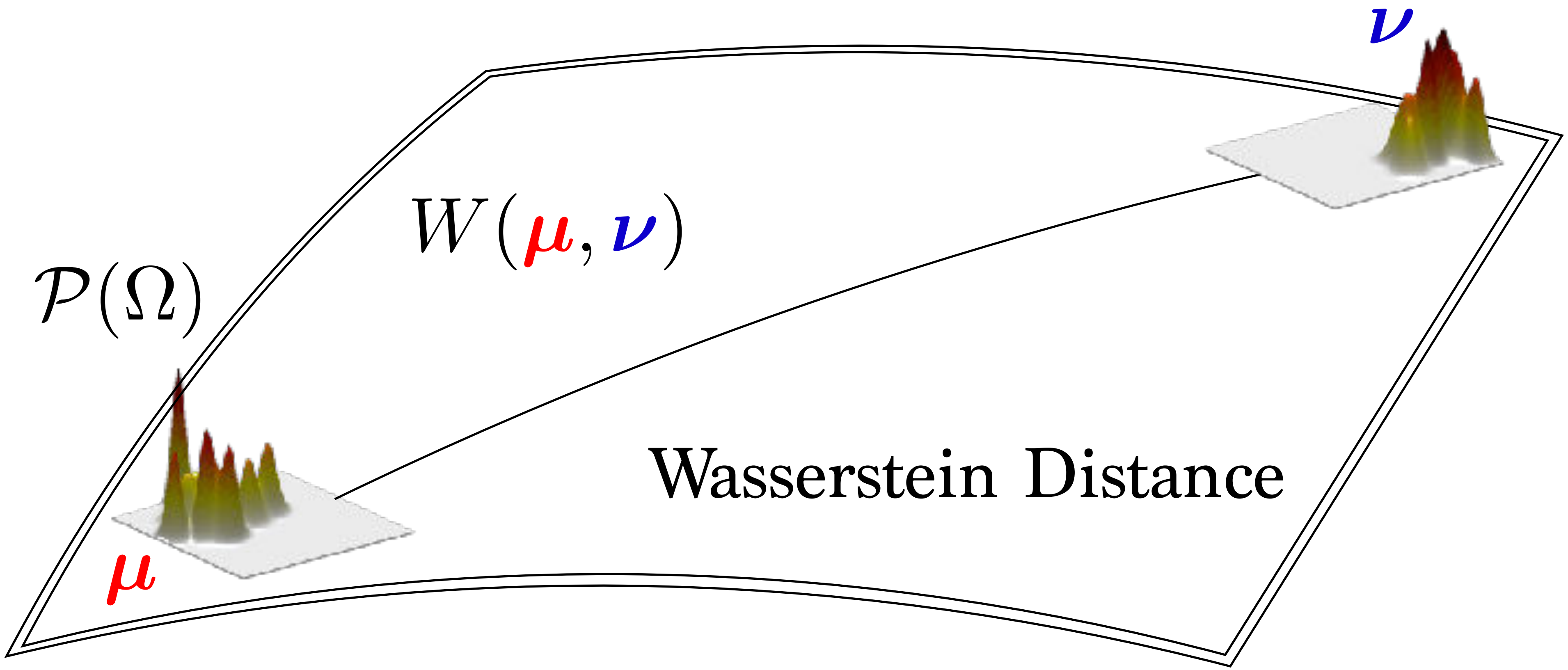
Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)



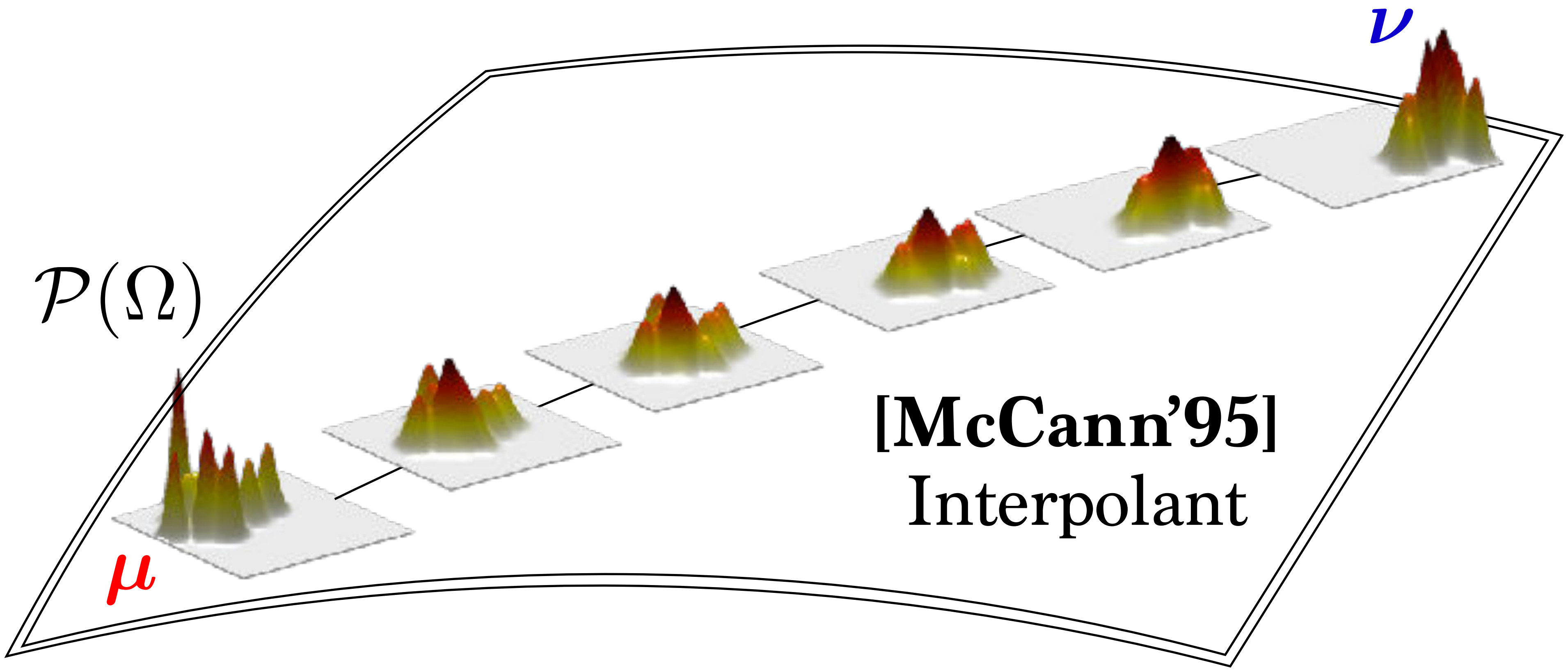
Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)



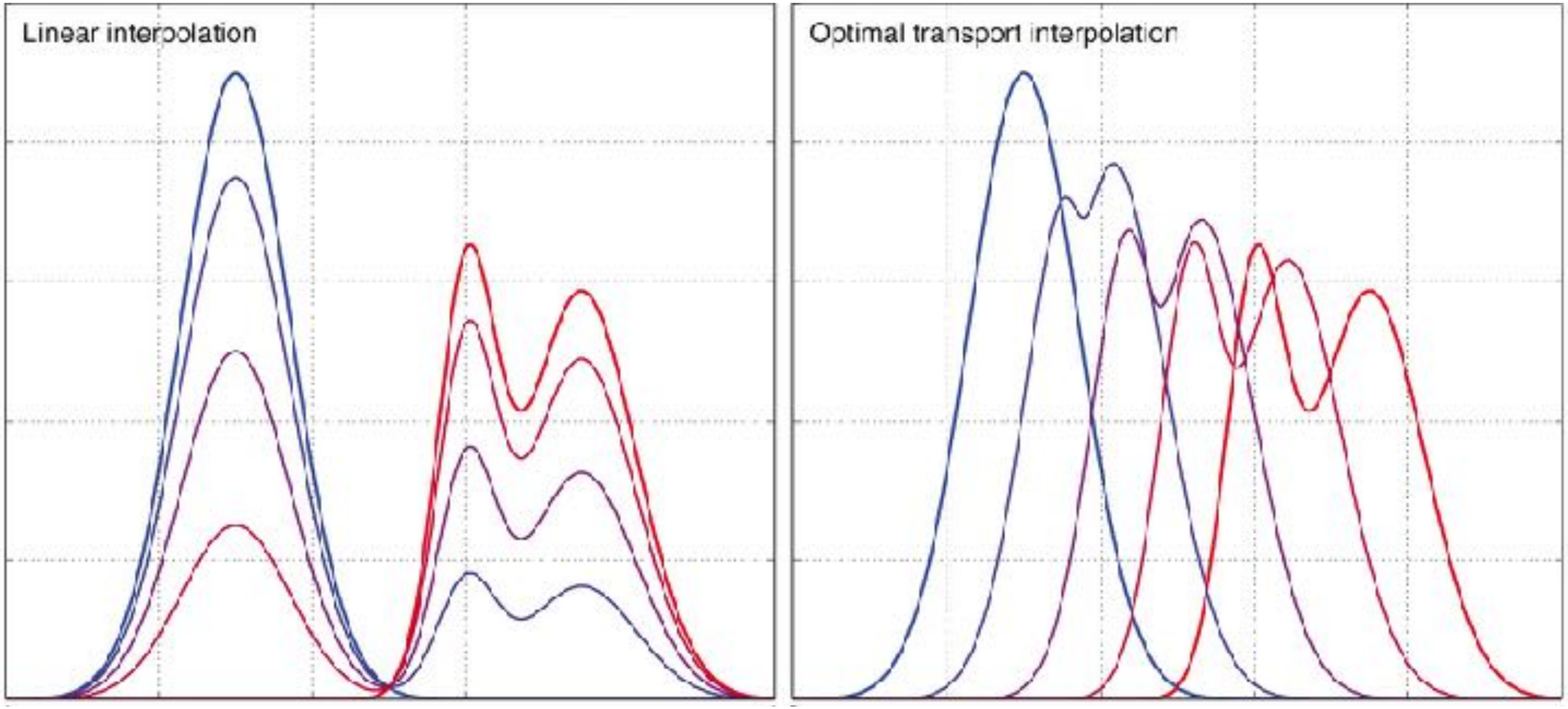
Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)



Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)



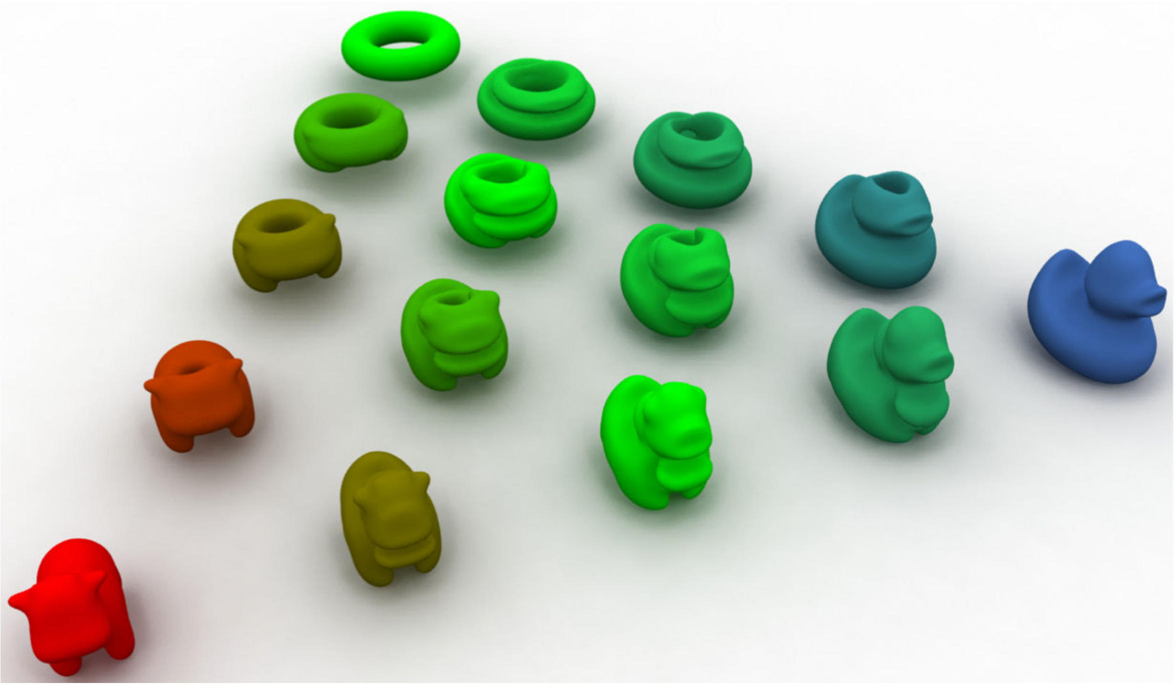
Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)

[Solomon'15]

Optimal Transport Geometry

Very different geometry than standard information divergences (Euclidean)



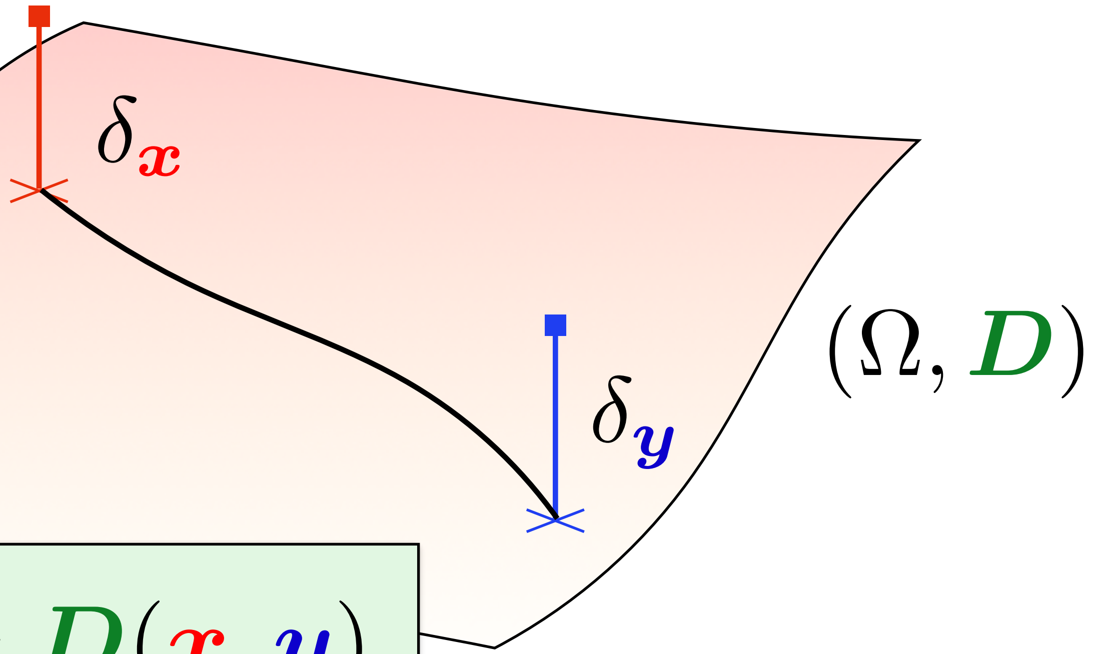
[Solomon'15]

III. Computational Optimal Transport

Wasserstein is hard to compute

- Wasserstein distances are generally intractable
- There are notable exceptions:
  - Distances between histograms (1D)
  - Gaussian distributions, when cost is quadratic
  - Discrete distributions

Wasserstein Between Two Diracs

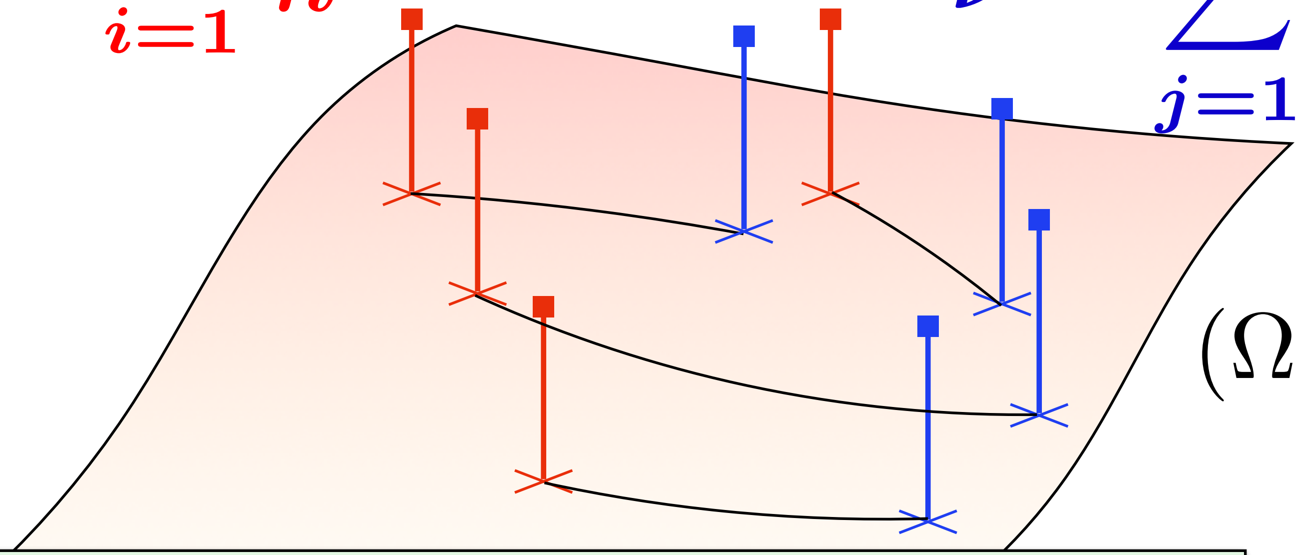


$$W_p(\delta_{\boldsymbol{x}}, \delta_{\boldsymbol{y}}) = \boldsymbol{D}(\boldsymbol{x}, \boldsymbol{y})$$

Linear Assignment ⊂ Wasserstein

$$\mu = \sum_{i=1}^{n} \frac{1}{n} \delta_{x_i}$$

$$\nu = \sum_{j=1}^{n} \frac{1}{n} \delta_{y_j}$$

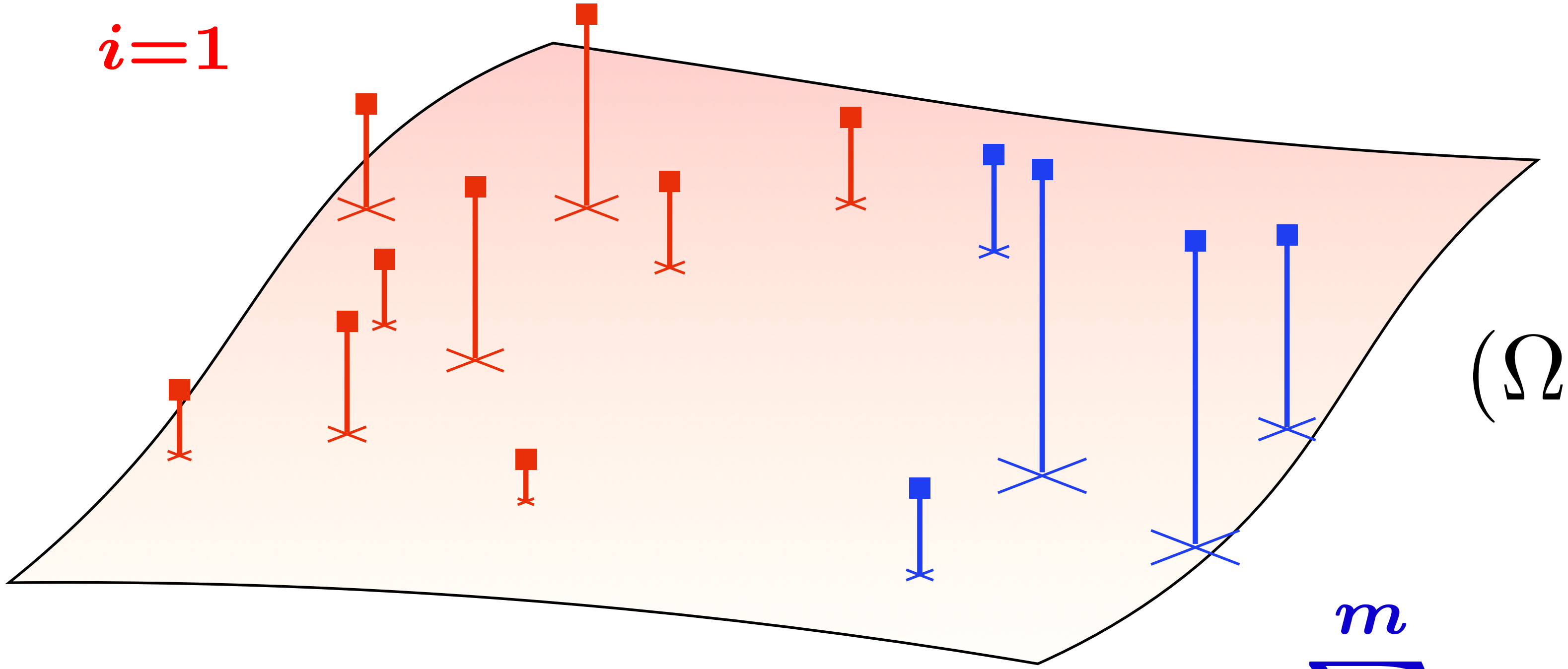


(Ω, D)

$$W_p^p(\mu, \nu) = \min_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^{n} D(x_i, y_{\sigma_i})^p$$

OT on Two Empirical Measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$$

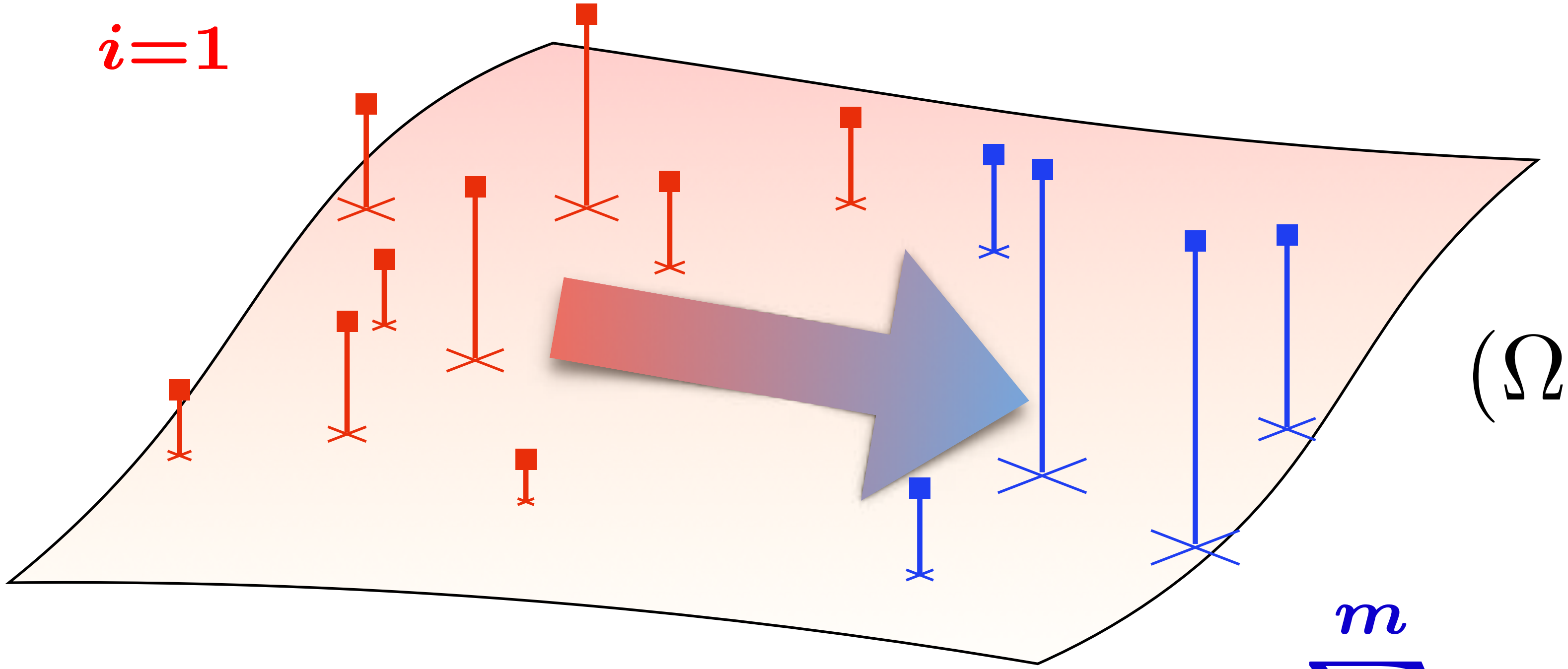


$$(\Omega, D)$$

$$\nu = \sum_{j=1}^{m} b_j \delta_{y_j}$$

OT on Two Empirical Measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$$



$$(\Omega, D)$$

$$\nu = \sum_{j=1}^{m} b_j \delta_{y_j}$$

Wasserstein on Empirical Measures

Consider $\boldsymbol{\mu} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{a_i \delta_{x_i}}$ and $\boldsymbol{\nu} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{m}} \boldsymbol{b_j \delta_{y_j}}$.

$$M_{\mathbf{XY}} \stackrel{\text{def}}{=} [D(\boldsymbol{x}_i, \boldsymbol{y}_j)^p]_{ij}$$

$$U(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$

Def. Optimal Transport Problem

$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle$$

Dual Kantorovich Problem

$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\substack{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} \\ \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle$$

Homework Exercise #4

Dual Kantorovich Problem

$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\substack{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} \\ \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_n = \boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle$$

Def. Dual OT problem

$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \max_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^m \\ \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j \leq D(\boldsymbol{x}_i, \boldsymbol{y}_j)^p}} \alpha^T \boldsymbol{a} + \beta^T \boldsymbol{b}$$

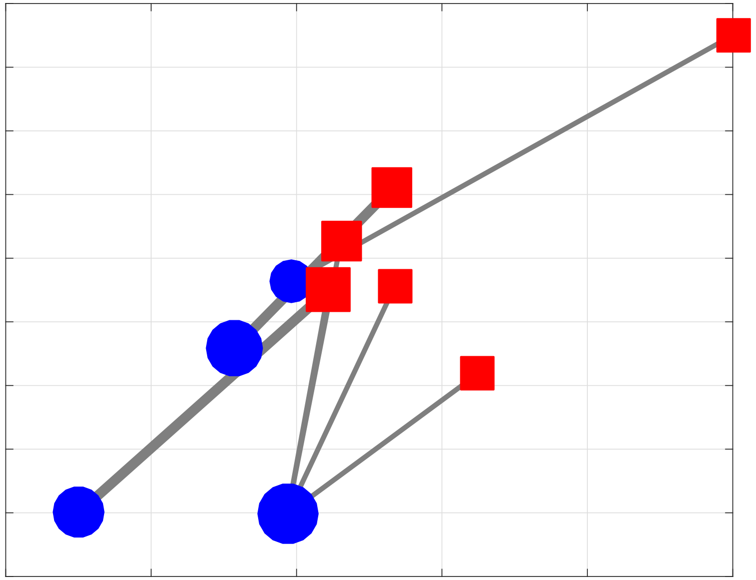
Homework
Exercise #4

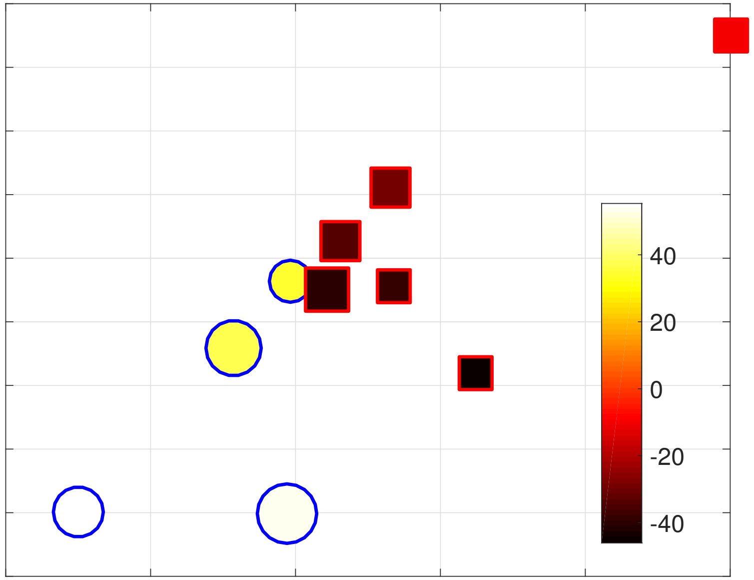
Dual Kantorovich Problem

$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\substack{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} \\ \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_n = \boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle$$

Def. Dual OT problem

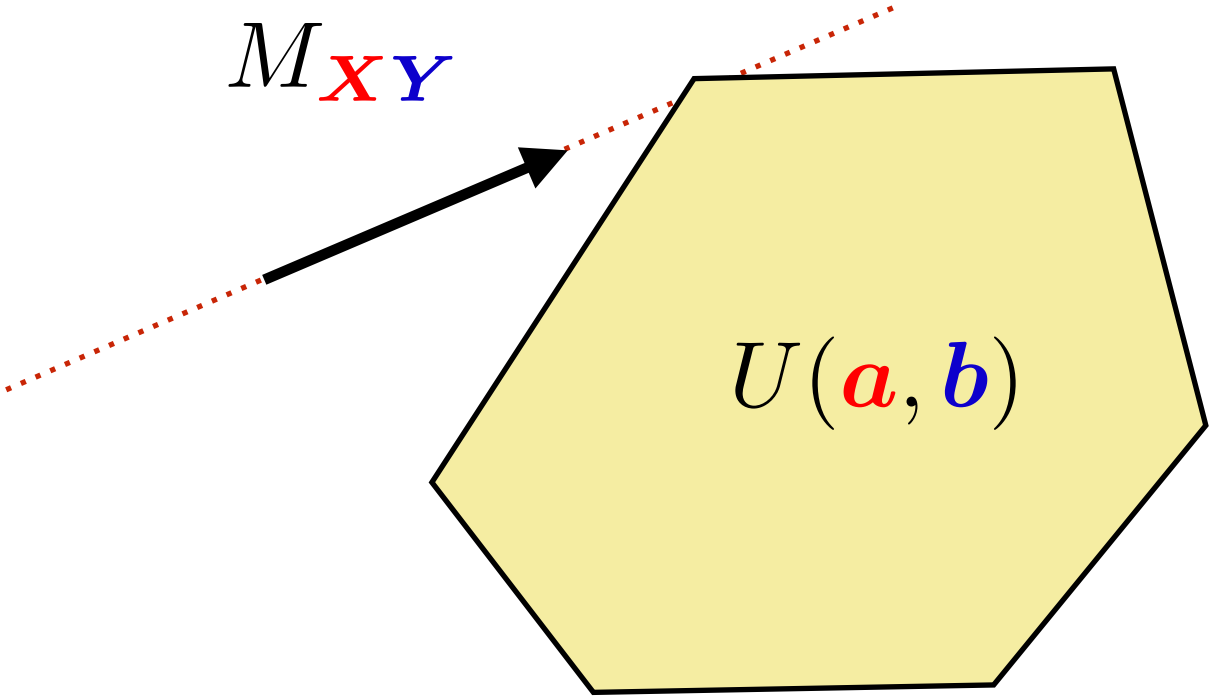
$$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \max_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^m \\ \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j \leq D(\boldsymbol{x}_i, \boldsymbol{y}_j)^p}} \alpha^T \boldsymbol{a} + \beta^T \boldsymbol{b}$$



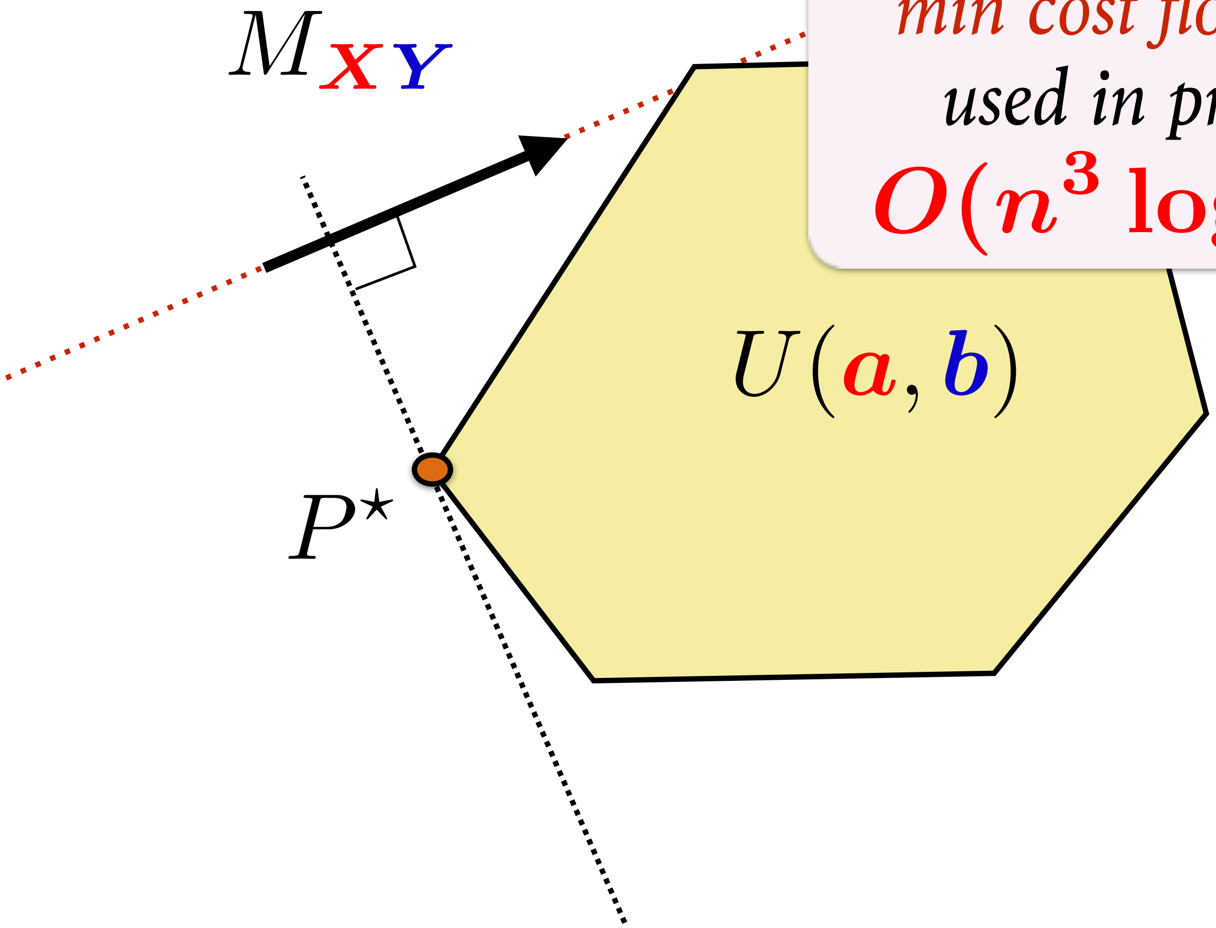


Homework Exercise #4

Solving the OT Problem

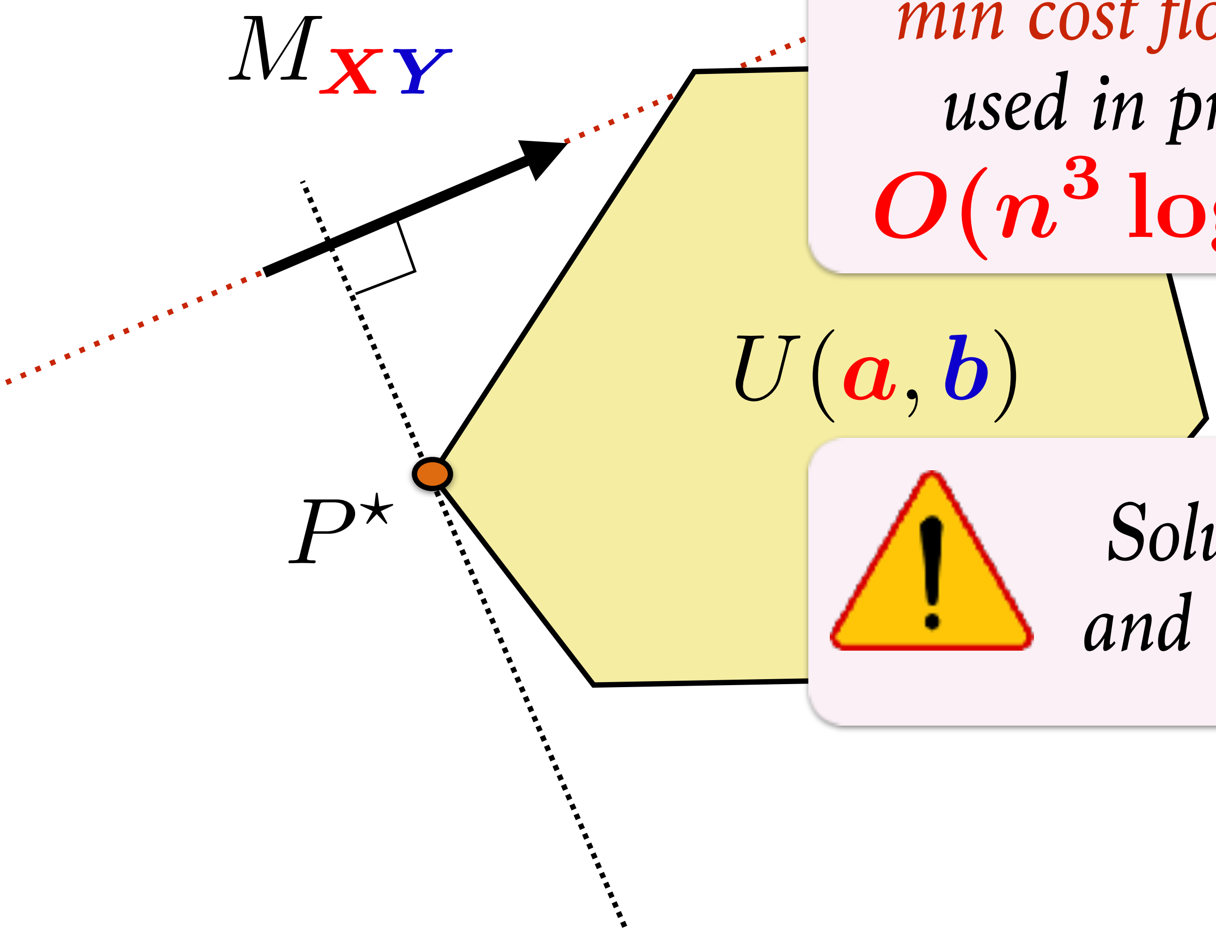


Solving the OT Problem



min cost flow solver
used in practice.
$O(n^3 \log(n))$

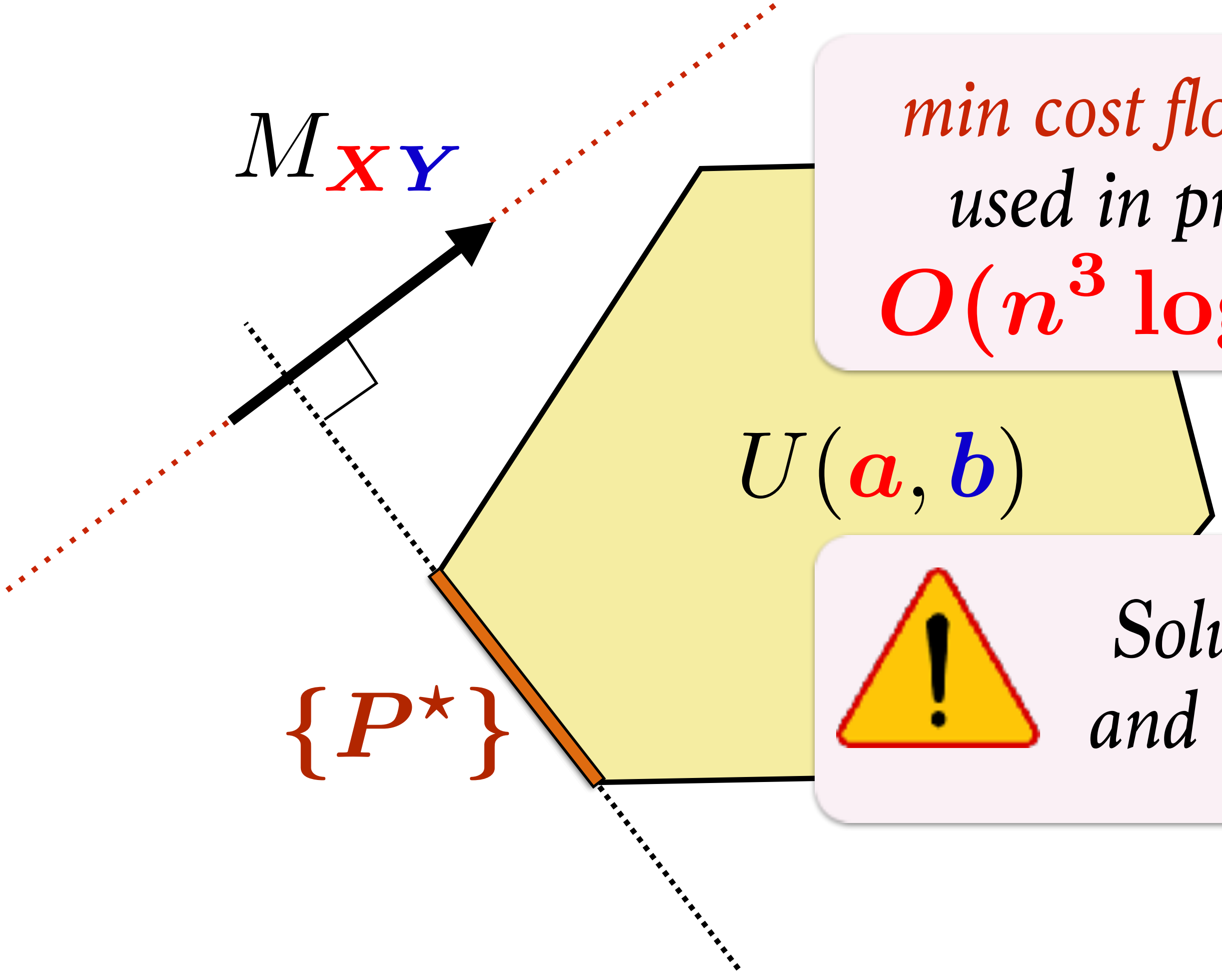
Solving the OT Problem



min cost flow solver
used in practice.
$O(n^3 \log(n))$

Solution $P^\star$ unstable
and not always unique.

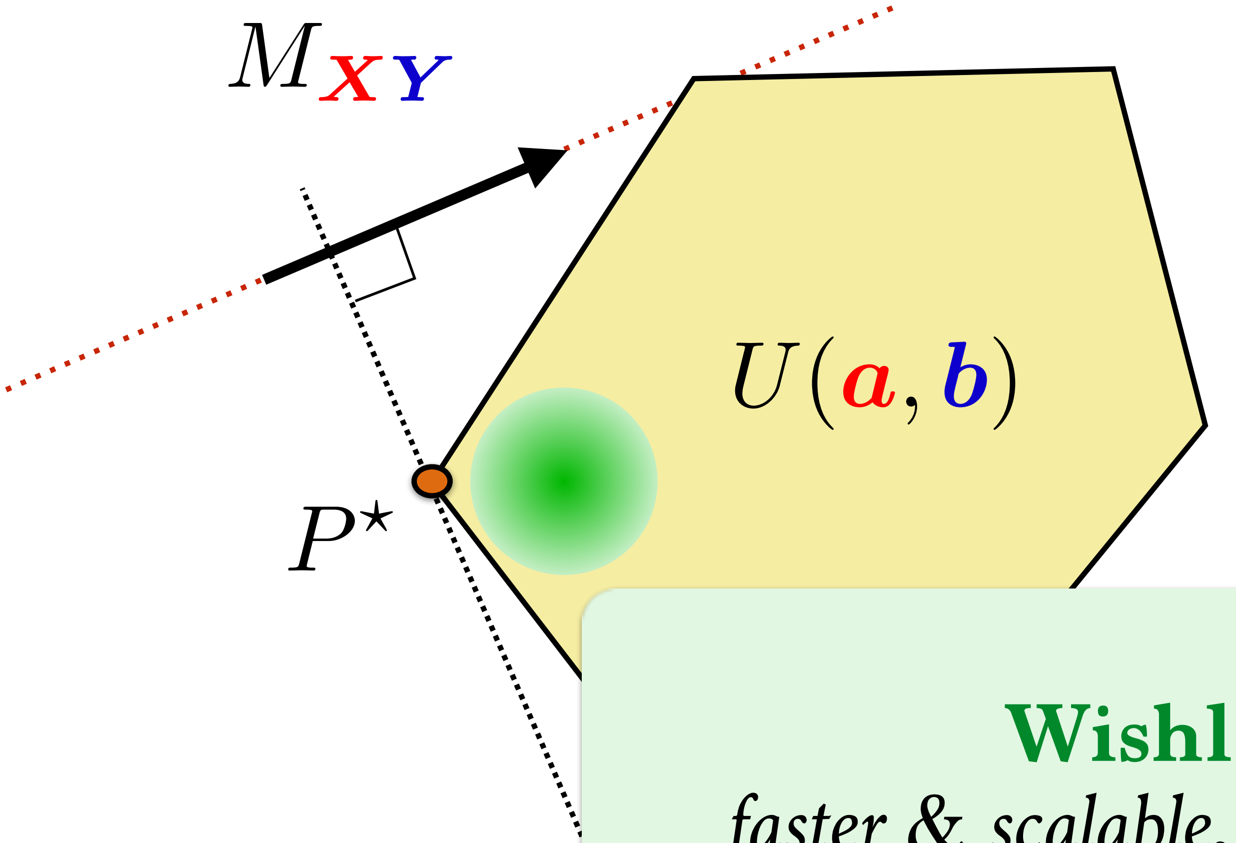
Solving the OT Problem



min cost flow solver *used in practice.*
$O(n^3 \log(n))$

Solution $P^\star$ unstable and not always unique.

Solution: Regularization



Wishlist:
faster & *scalable*, more *stable*

Entropic Regularization [Wilson'62]

Def. Regularized Wasserstein, $\gamma \geq 0$

$$W_\gamma(\boldsymbol{\mu}, \boldsymbol{\nu}) \stackrel{\text{def}}{=} \min_{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle - \gamma E(\boldsymbol{P})$$

$$E(P) \stackrel{\text{def}}{=} - \sum_{i,j=1}^{nm} P_{ij} (\log P_{ij} - 1)$$

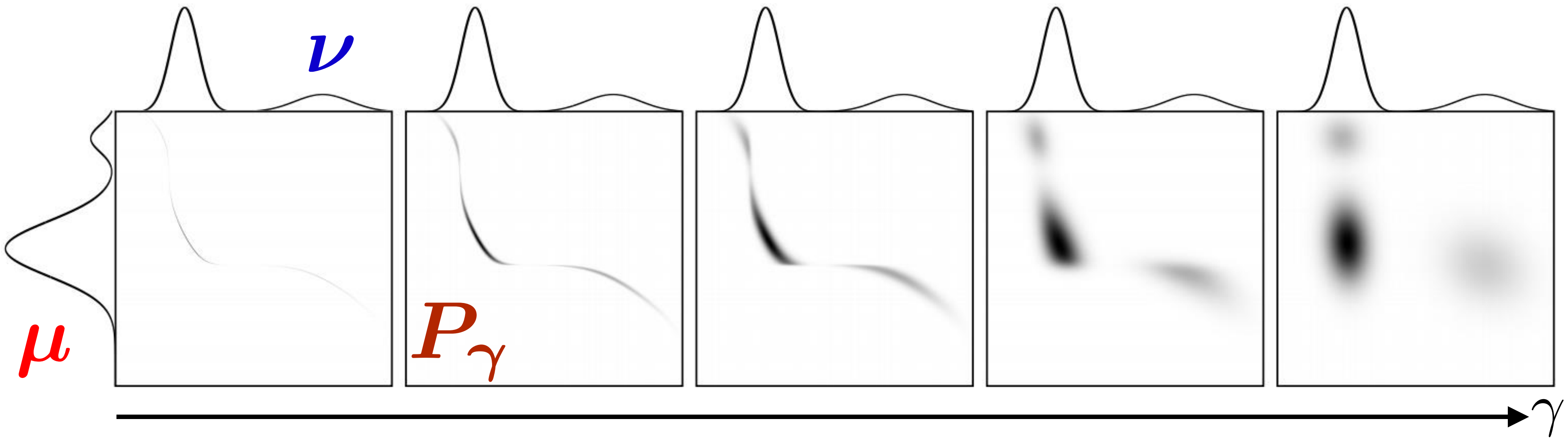
Note: Unique optimal solution because of strong concavity of entropy

Homework Exercise #5

Entropic Regularization [Wilson'62]

Def. Regularized Wasserstein, $\gamma \geq 0$

$$W_\gamma(\boldsymbol{\mu}, \boldsymbol{\nu}) \overset{\text{def}}{=} \min_{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{X}\mathbf{Y}} \rangle - \gamma E(\boldsymbol{P})$$



Note: Unique optimal solution because of strong concavity of entropy

Homework Exercise #5

Dual Formulation

Def. Regularized Wasserstein, $\gamma \geq 0$

$$W_\gamma(\boldsymbol{\mu}, \boldsymbol{\nu}) \stackrel{\text{def}}{=} \min_{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle - \gamma E(\boldsymbol{P})$$

REGULARIZED DISCRETE PRIMAL

$$W_\gamma(\boldsymbol{\mu}, \boldsymbol{\nu}) = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \boldsymbol{\alpha}^T \boldsymbol{a} + \boldsymbol{\beta}^T \boldsymbol{b} - \gamma (e^{\boldsymbol{\alpha}/\gamma})^T \; K(e^{\boldsymbol{\beta}/\gamma})$$

REGULARIZED DISCRETE DUAL

The dual is decoupled in its parameters!

Fast & Scalable Algorithm

Prop. If $P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \boldsymbol{P}, M_{\mathbf{X}\mathbf{Y}} \rangle - \gamma E(\boldsymbol{P})$

then $\exists! \boldsymbol{u} \in \mathbb{R}_+^n, \boldsymbol{v} \in \mathbb{R}_+^m$, such that

$$P_\gamma = \mathbf{diag}(\boldsymbol{u}) K \mathbf{diag}(\boldsymbol{v}), \quad K \stackrel{\text{def}}{=} e^{-M_{\mathbf{X}\mathbf{Y}}/\gamma}$$

Fast & Scalable Algorithm

Prop. If $P_\gamma \overset{\text{def}}{=} \underset{P \in U(a,b)}{\operatorname{argmin}} \langle P, M_{XY} \rangle - \gamma E(P)$

then $\exists! u \in \mathbb{R}^n_+, v \in \mathbb{R}^m_+$, such that

$$P_\gamma = \mathbf{diag}(u) K \mathbf{diag}(v), \quad K \overset{\text{def}}{=} e^{-M_{XY}/\gamma}$$

- **[Sinkhorn'64]** fixed-point iterations for (u, v)

$$u \leftarrow a / Kv, \quad v \leftarrow b / K^T u$$

- $O(nm)$ complexity.

IV. Dynamic Optimal Transport

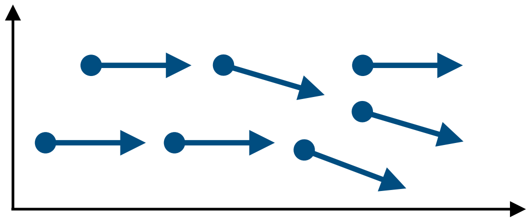
Vector Fields & Friends

Homework Exercise #2

Vector Fields & Friends

Vector Field:

$$u_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$

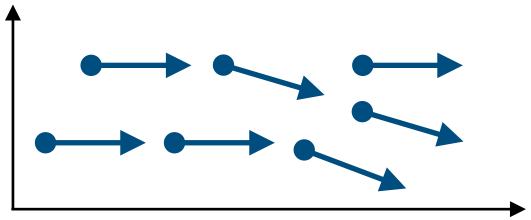


Homework Exercise #2

Vector Fields & Friends

Vector Field:

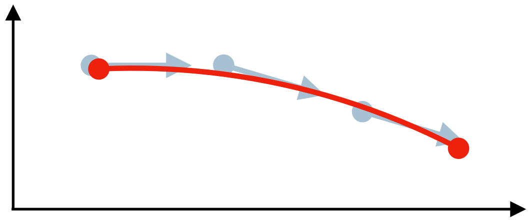
$$u_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$



Flow:

$$\phi_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$

$$\frac{d}{dt}\phi_t(x) = u_t(\phi_t(x)),$$



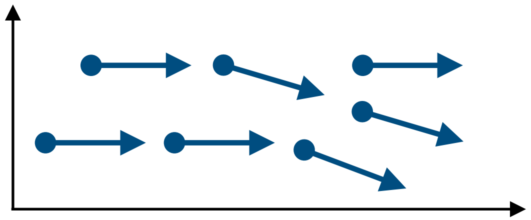
$$\phi_0(x) = x$$

Homework Exercise #2

Vector Fields & Friends

Vector Field:

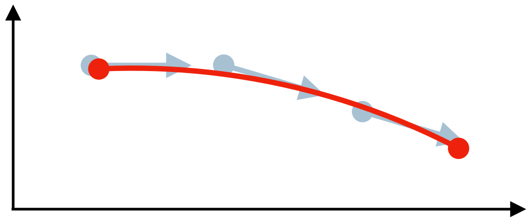
$$u_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$



Flow:

$$\phi_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$

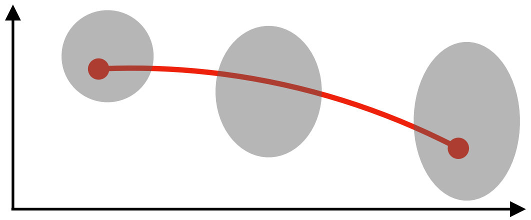
$$\frac{d}{dt}\phi_t(x) = u_t(\phi_t(x)),$$



$$\phi_0(x) = x$$

Probability Density Path:

$$p_t(x) = [\phi_t]_{\#} p_0(x)$$



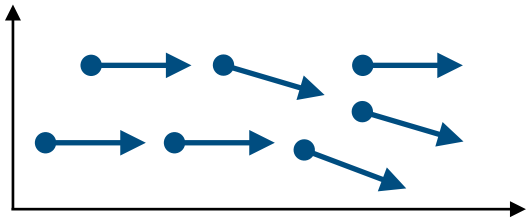
Homework Exercise #2

Vector Fields & Friends

(Numerical) Integration

Vector Field:

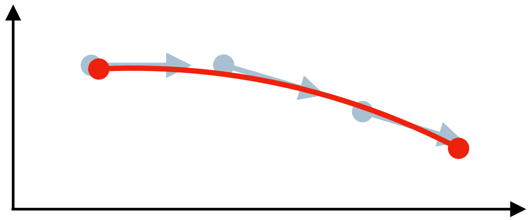
$$u_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$



Flow:

$$\phi_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$

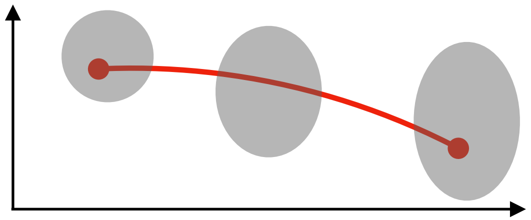
$$\frac{d}{dt}\phi_t(x) = u_t(\phi_t(x)),$$



$$\phi_0(x) = x$$

Probability Density Path:

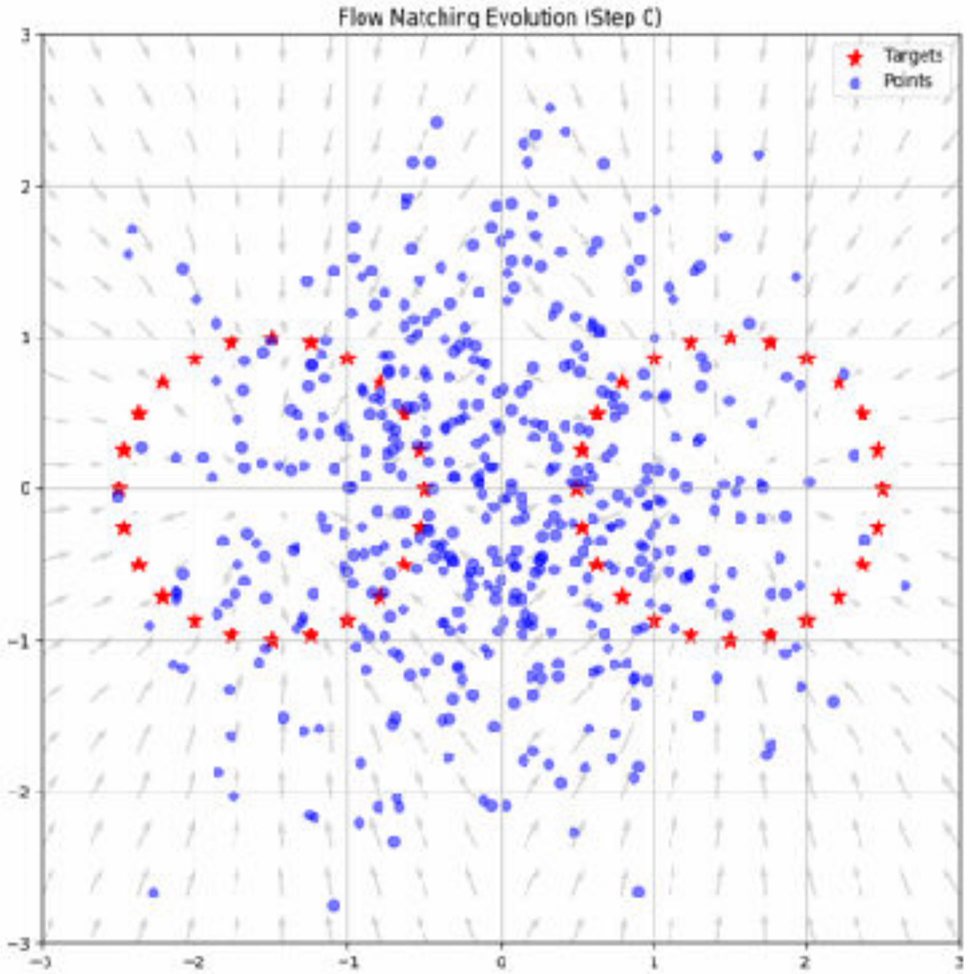
$$p_t(x) = [\phi_t]_{\#}p_0(x)$$



Differentiation

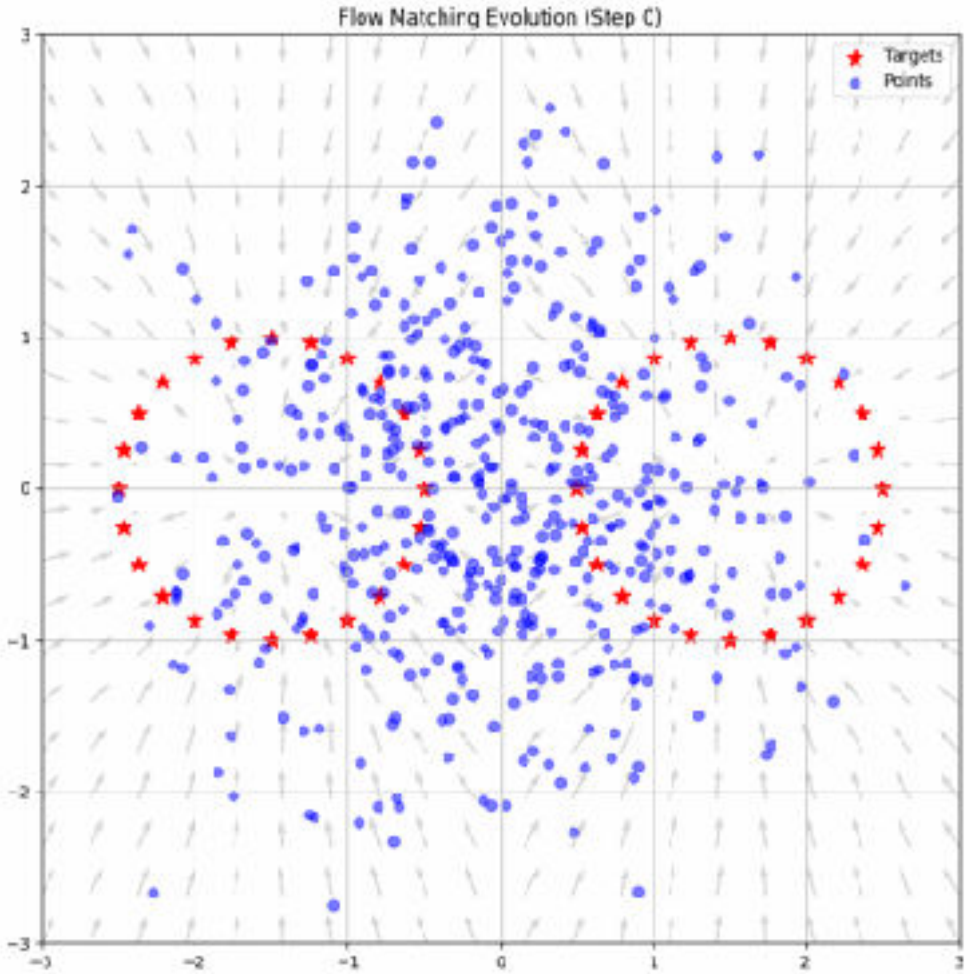
Homework Exercise #2

Particles flowing down a vector field



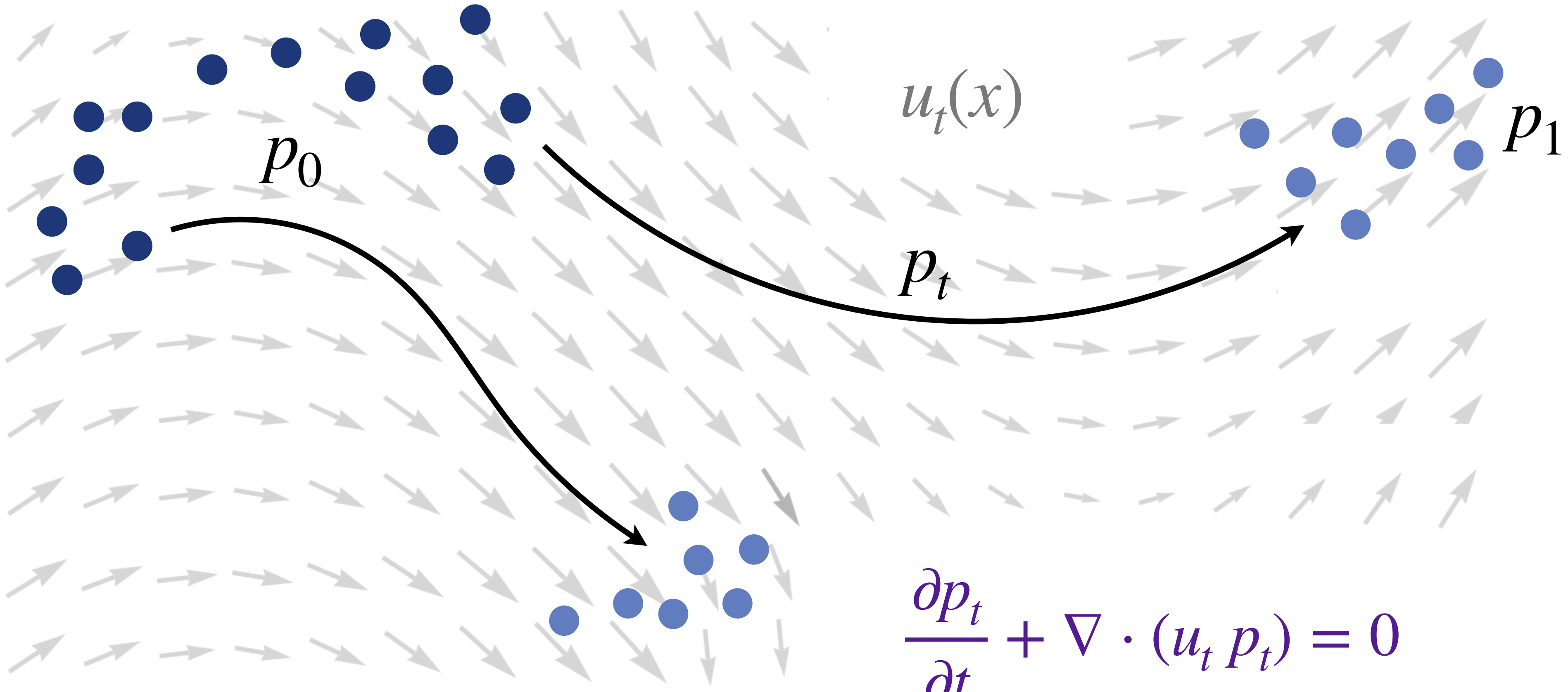
Federico Sarrocco

Particles flowing down a vector field



Federico Sarrocco

The continuity equation



$$\frac{\partial p_t}{\partial t} + \nabla \cdot (u_t\, p_t) = 0$$

Homework Exercise #3 (first part)

Dynamic characterization of OT

Theorem (Benamou–Brenier). For p_0, p_1 with finite second moments,

$$W_2^2(p_0, p_1) = \inf_{(p_t, u_t)} \int_0^1 \int_{\mathbb{R}^d} |\, u_t(x) \,|^2 \, \mathrm{d}p_t(x) \, \mathrm{d}t \qquad \text{subject to} \qquad \partial_t p_t + \nabla \cdot (u_t \, p_t) = 0$$

the infimum being over all pairs (p_t, u_t) solving the continuity equation, with p_0 and p_1 as endpoints.

Dynamic characterization of OT

Theorem (Benamou–Brenier). For p_0, p_1 with finite second moments,

$$W_2^2(p_0, p_1) = \inf_{(p_t, u_t)} \int_0^1 \int_{\mathbb{R}^d} |u_t(x)|^2 \, \mathrm{d}p_t(x) \, \mathrm{d}t \qquad \text{subject to} \qquad \partial_t p_t + \nabla \cdot (u_t \, p_t) = 0$$

the infimum being over all pairs (p_t, u_t) solving the continuity equation, with p_0 and p_1 as endpoints.

- The OT distance is the minimal kinetic energy cost of continuously morphing p_0 into p_1. Additionally, the optimal vector field is fully described from the Monge map T pushing p_0 into p_1.

Dynamic characterization of OT

Theorem (Benamou–Brenier). For p_0, p_1 with finite second moments,

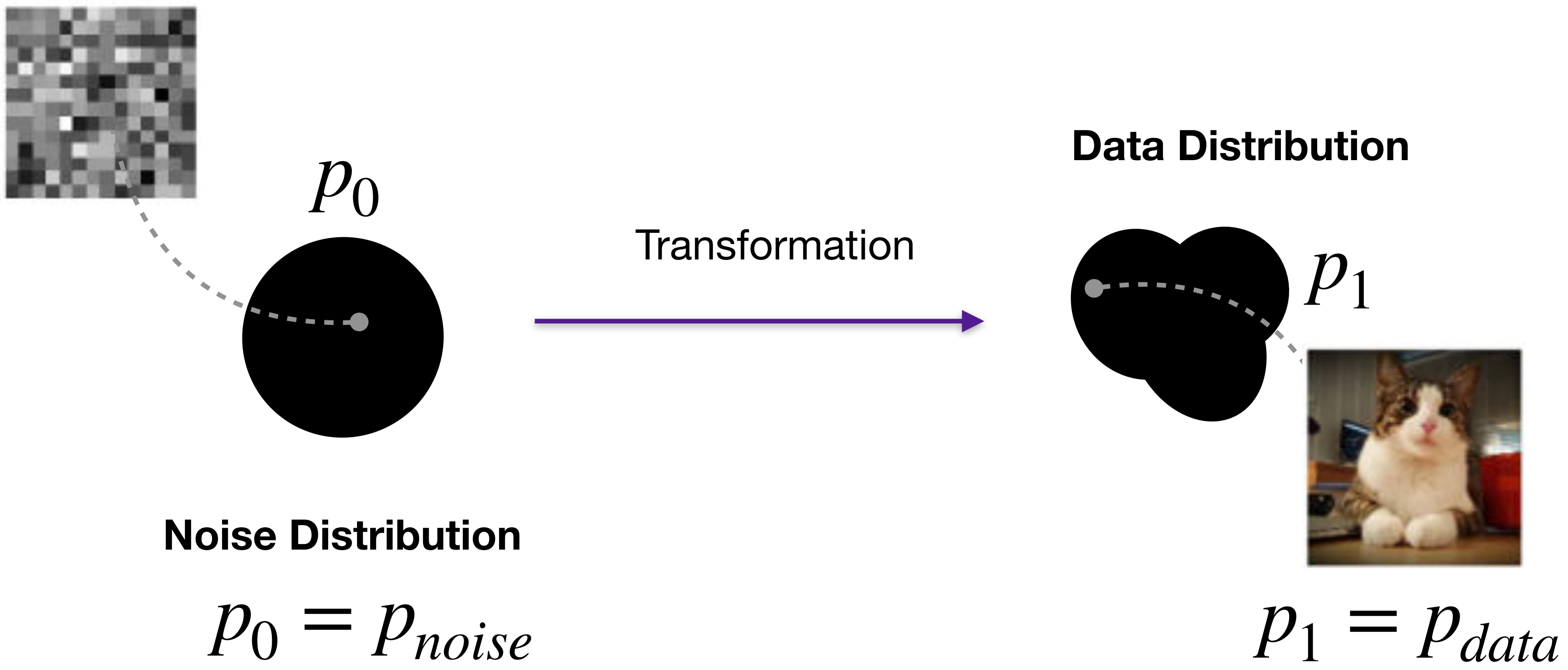
$$W_2^2(p_0, p_1) = \inf_{(p_t, u_t)} \int_0^1 \int_{\mathbb{R}^d} |u_t(x)|^2 \, \mathrm{d}p_t(x) \, \mathrm{d}t \qquad \text{subject to} \qquad \partial_t p_t + \nabla \cdot (u_t \, p_t) = 0$$

the infimum being over all pairs (p_t, u_t) solving the continuity equation, with p_0 and p_1 as endpoints.

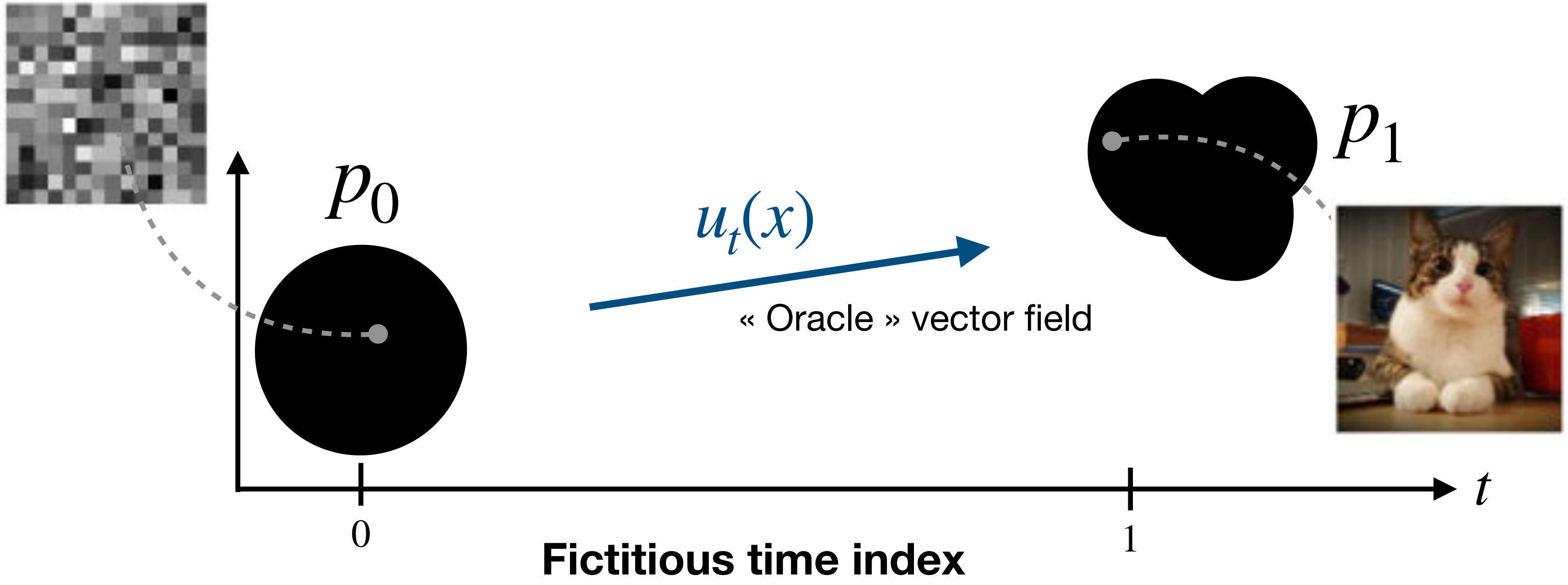
- The OT distance is the minimal kinetic energy cost of continuously morphing p_0 into p_1. Additionally, the optimal vector field is fully described from the Monge map T pushing p_0 into p_1.
- **OT still remains hard to scale to millions of points…**
  Can we come up with another velocity field that is not « optimal » but is simpler to learn?

V. Flow Matching

Basic idea behind Generative Modeling



Generative Modeling via Flow Matching



Lipman et al. (2023)

Flow Matching amounts to Supervised Learning

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1],\ x\sim p_t(x)}\, ||v_t(x;\theta) - u_t(x)||_2^2$$

Lipman et al. (2023)

Flow Matching amounts to Supervised Learning

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t \sim \mathcal{U}[0,1],\, x \sim p_t(x)} \, || \underbrace{v_t(x;\theta)}_{\text{Neural Net}} - \underbrace{u_t(x)}_{\text{Target}} ||_2^2$$

Lipman et al. (2023)

Flow Matching amounts to Supervised Learning

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t \sim \mathcal{U}[0,1],\, x \sim \boxed{p_t(x)}} ||v_t(x;\theta) - \boxed{u_t(x)}||_2^2$$

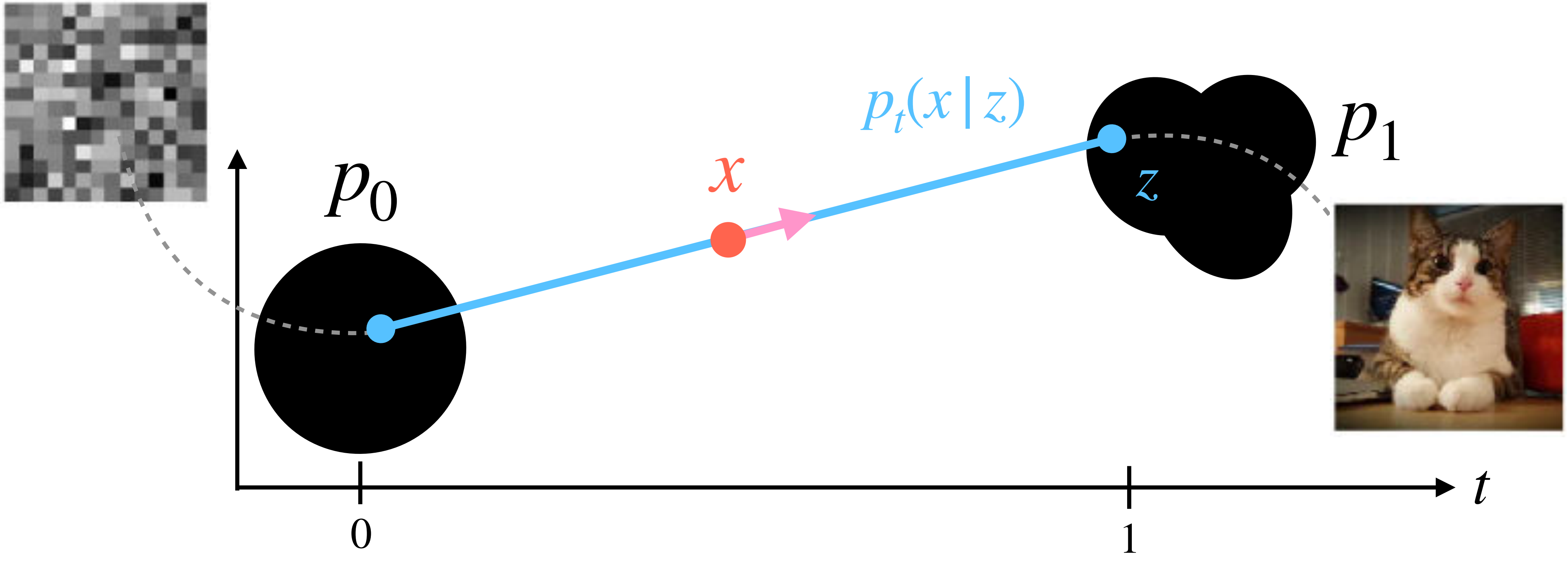
Key challenge: how to define these?

Main answer: define them implicitly

Lipman et al. (2023)

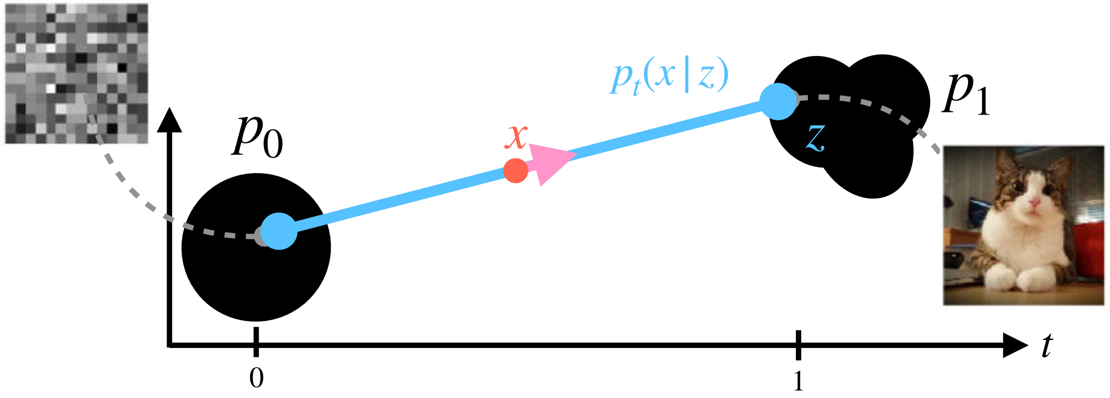
Conditional Flow Matching

Conditional Flow Matching

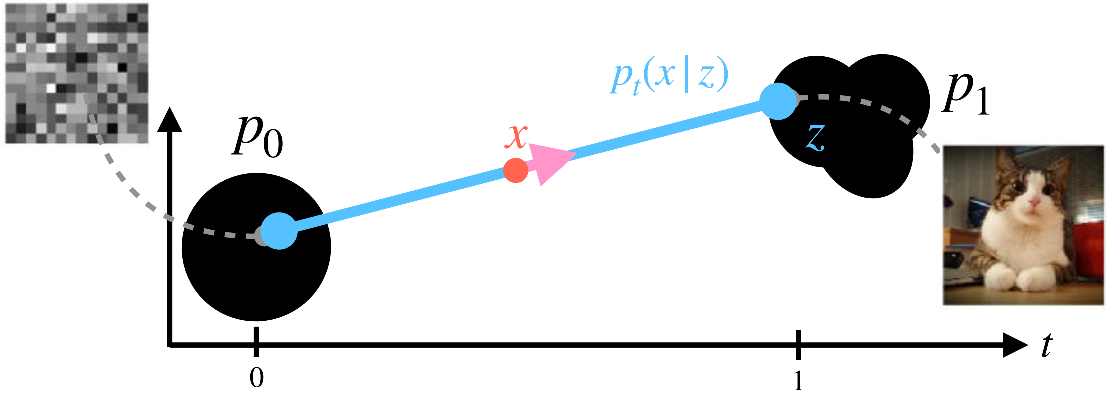


If we decide on a end point,
can we then settle on a « simple » vector field?

Conditional Probability Path



Conditional Probability Path

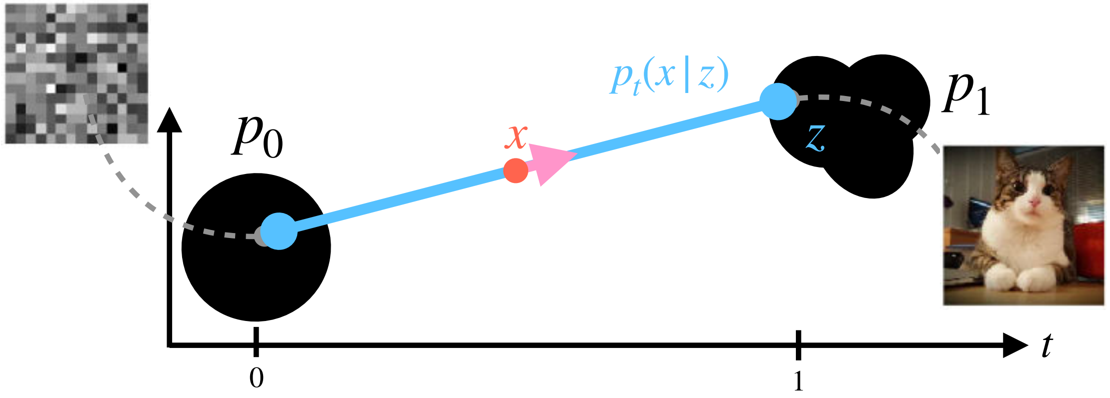


$$p_t(x \mid z) = \mathcal{N}(x \mid \mu_t(z), \sigma_t(z)^2 I)$$

$$\mu_t(z) = tz$$

$$\sigma_t(z) = 1 - t$$

Conditional Probability Path



$$p_t(x \mid z) = \mathcal{N}(x \mid \mu_t(z), \sigma_t(z)^2 I)$$

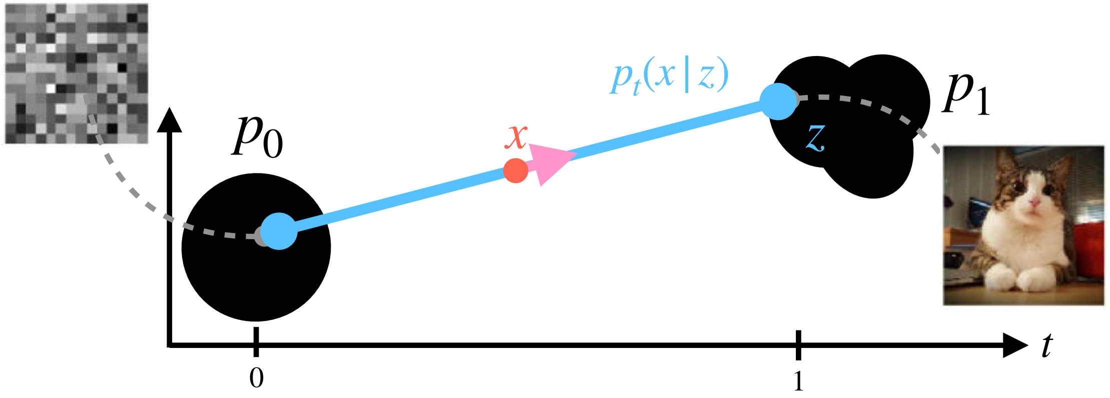
$$\mu_t(z) = tz$$

$$\sigma_t(z) = 1 - t$$

With this specification, we have the right boundary conditions for the **marginal probability path**:

$$p_t = \int p_t(\,\cdot \mid z)\, \mathrm{d}p_{\text{data}}(z) \qquad p_0 = \int_{\mathbb{R}^d} p_{\text{noise}}\, \mathrm{d}p_{\text{data}}(z) = p_{\text{noise}} \qquad p_1 = \int_{\mathbb{R}^d} \delta_z\, \mathrm{d}p_{\text{data}}(z) = p_{\text{data}}$$

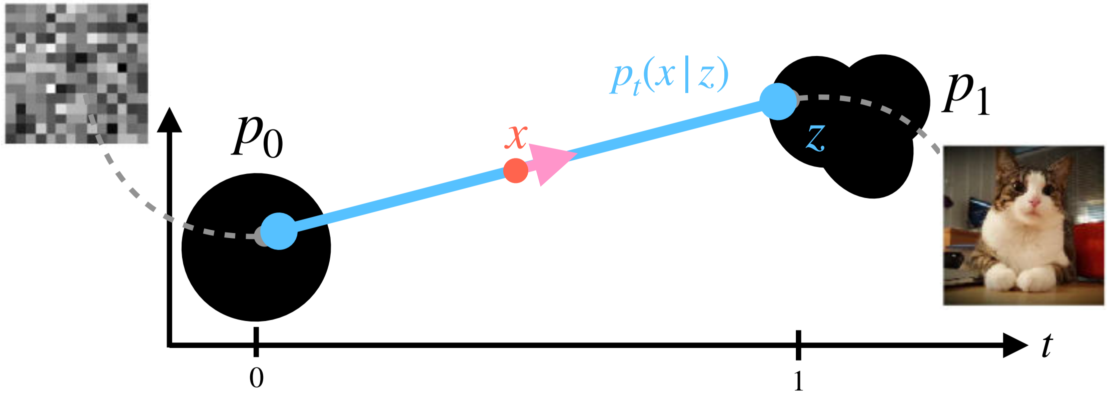
Conditional Vector Field



Homework Exercise #3 (second part)

Lipman et al. (2023)

Conditional Vector Field



$$u_t(x \mid z) = \frac{\sigma_t'(z)}{\sigma_t(z)}(x - \mu_t(z)) + \mu_t'(z)$$

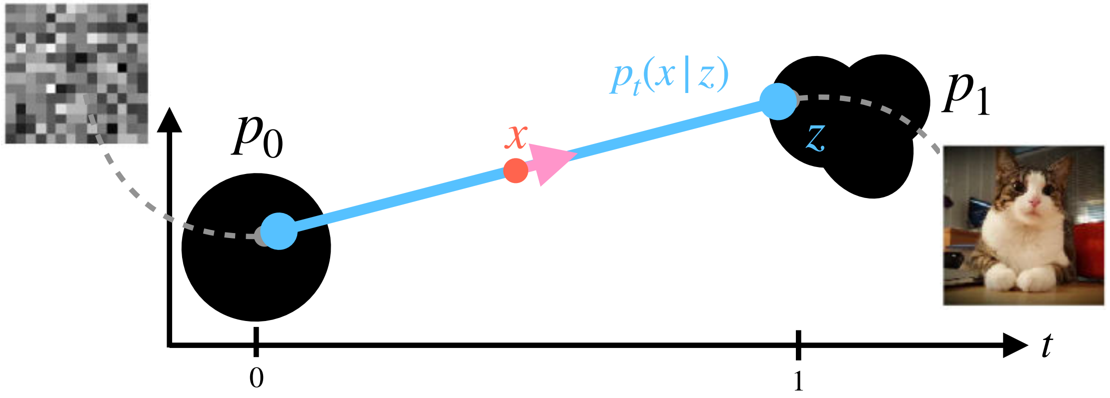
With our specific design:

$$u_t(x \mid z) = \frac{z - x}{1 - t}$$

Homework Exercise #3 (second part)

Lipman et al. (2023)

Conditional Vector Field



$$u_t(x \mid z) = \frac{\sigma_t'(z)}{\sigma_t(z)}(x - \mu_t(z)) + \mu_t'(z)$$

With our specific design:

$$u_t(x \mid z) = \frac{z - x}{1 - t}$$

This implicitly defines a **marginal vector field** (which is itself intractable):

$$u_t(x) := \int u_t(x \mid z) \frac{p_t(x \mid z)\, p_{\text{data}}(z)}{p_t(x)} \, \mathrm{d}z$$

Homework Exercise #3 (second part)

Lipman et al. (2023)

Conditional Flow Matching Loss

$$\mathscr{L}_{CFM}(\theta) = \mathbb{E}_{t \sim \mathscr{U}_{[0,1]}, z \sim p_{data},\ x \sim p_t(x|z)} \,||\, v_t(x;\theta) - u_t(x \,|\, z) \,||_2^2$$

Lipman et al. (2023)

Conditional Flow Matching Loss

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t \sim \mathcal{U}_{[0,1]}, z \sim p_{data},\, x \sim p_t(x|z)} \, || \underbrace{v_t(x;\theta)}_{\textbf{Neural Net}} - \underbrace{u_t(x\,|\,z)}_{\textbf{(Conditional) Target}} ||_2^2$$

Lipman et al. (2023)

The miracle of conditional flow matching

$$\mathscr{L}_{CFM}(\theta) = \mathbb{E}_{t \sim \mathscr{U}_{[0,1]}, z \sim p_{data},\, x \sim p_t(x|z)} \,||\, v_t(x;\theta) - u_t(x\,|\,z)\,||_2^2$$

$$\mathscr{L}_{FM}(\theta) = \mathbb{E}_{t \sim \mathscr{U}[0,1],\, x \sim p_t(x)} \,||\, v_t(x;\theta) - u_t(x)\,||_2^2$$

Homework Exercise #1

Lipman et al. (2023)

The miracle of conditional flow matching

The two following objective functions:

$$\mathscr{L}_{CFM}(\theta) = \mathbb{E}_{t \sim \mathscr{U}_{[0,1]}, z \sim p_{data},\ x \sim p_t(x|z)} \, ||v_t(x;\theta) - u_t(x|z)||_2^2$$

$$\mathscr{L}_{FM}(\theta) = \mathbb{E}_{t \sim \mathscr{U}[0,1],\ x \sim p_t(x)} \, ||v_t(x;\theta) - u_t(x)||_2^2$$

have the same gradients with respect to the neural net parameters

$$\boxed{\nabla_\theta \mathscr{L}_{\text{FM}}(\theta) = \nabla_\theta \mathscr{L}_{\text{CFM}}(\theta)}$$

Homework Exercise #1

Lipman et al. (2023)

Example of problems we will read about in the seminar

Find the full paper list on the course website!
https://romain-lopez.github.io/learning-to-match-distributions/

Better controllable generation with flow matching

Everything so far learns to sample from a distribution.

Classifier-free guidance, what everyone uses to sample from a « prompt », turns out to be systematically miscalibrated.

Fan et al., CFG-Zero*, 2025

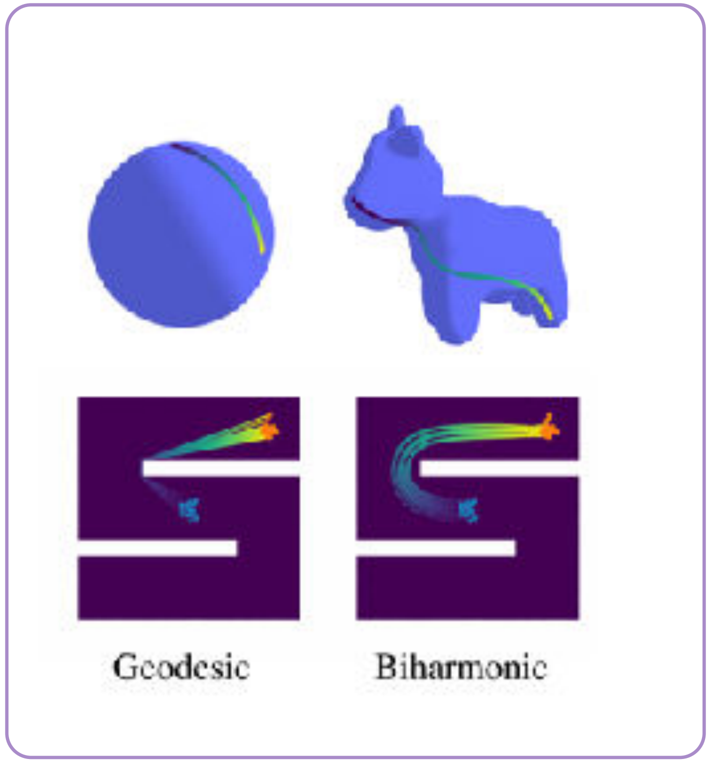
Guided Flows (Zheng et al., 2023)
CFG-Zero* (Fan et al., 2025)

Does any of this survive when the data is not a vector?

Flow matching was derived for Euclidean space with Gaussian paths.

Text is discrete. Proteins and orientations live on (non-flat) manifolds.

How to use this for certain data types (e.g., graphs, counts) is not figured out.



Chen & Lipman, Riemannian Flow Matching, 2023

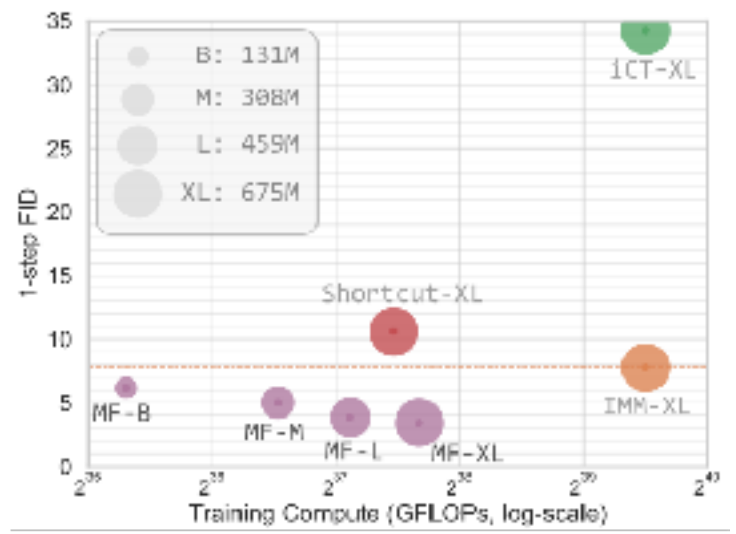
Riemannian FM (Chen & Lipman, 2023)
Discrete FM (Gat et al., 2024)

Can generation be faster?

Sampling from the generative model means numerically solving a differential equation.

It requires many full passes through the neural network. Can we get it in one step?

One of the most competitive problems in generative modeling right now!



Geng et al., Mean Flow, 2025

Flow Map Matching (Boffi et al., 2024)
Mean Flow (Geng et al., 2025)

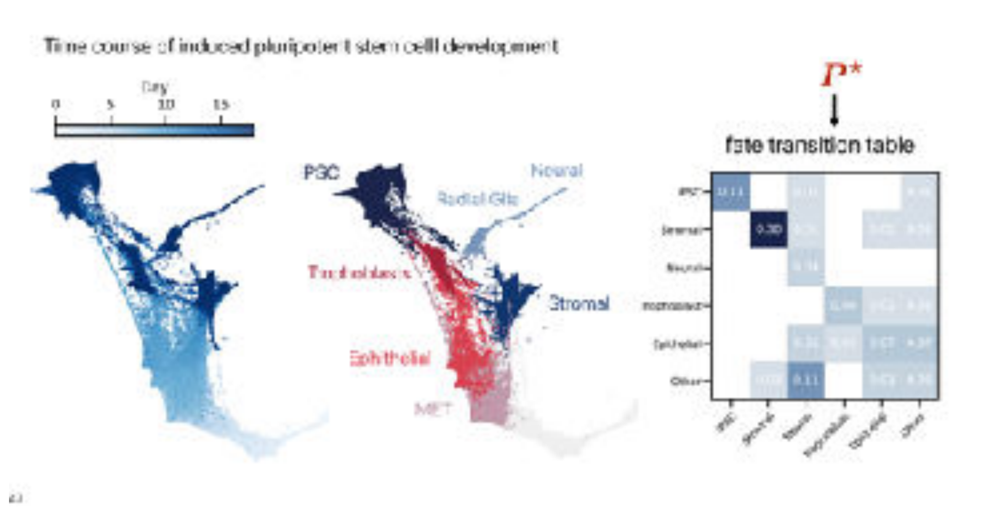
Monday Nov 9

Does it work on real problems?

In biology the difficulty is that measuring a cell destroys it, so you only ever see populations before and after, never a trajectory.

Optimal transport and flow matching is used to reconstruct cellular dynamics, and learn causality.

How to effectively reconstruct causal trajectories from snapshots?



Bunne & Cuturi, ICML 2023 tutorial

We will also explore applications to music generation!

GENOT (Klein et al., 2024)
MusicFlow (Prajwal et al., 2024)

Monday Nov 30

Wrapping up

Our promise to you for December

You will be able to open any optimal transport or flow matching paper on arXiv and understand their core contribution.

And you will have deeper knowledge from the project, as well as your panel appearance.

Coming up

Tonight	Join Discord. The invite is on Brightspace.
Wed Sep 16	Paper sign-up begins (Brightspace).
Mon Sep 21	**Homework due at the start of class.** Whiteboard session: bring pen and paper.
Mon Sep 28	**Paper sign-up is due.** Whiteboard session: bring pen and paper.
Mon Oct 5	**Project team registration is due.** Role-play seminar starts!

Office hours:
Romain: Mondays 3:30-4:30, 60FA#304
Vaibhavi: Fridays 12:30-1:30, 60FA#502.