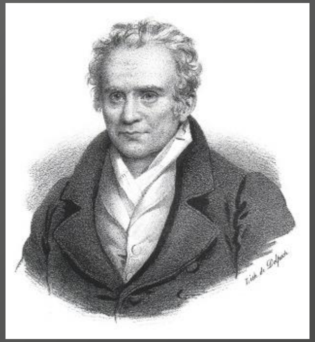
ETH zürich

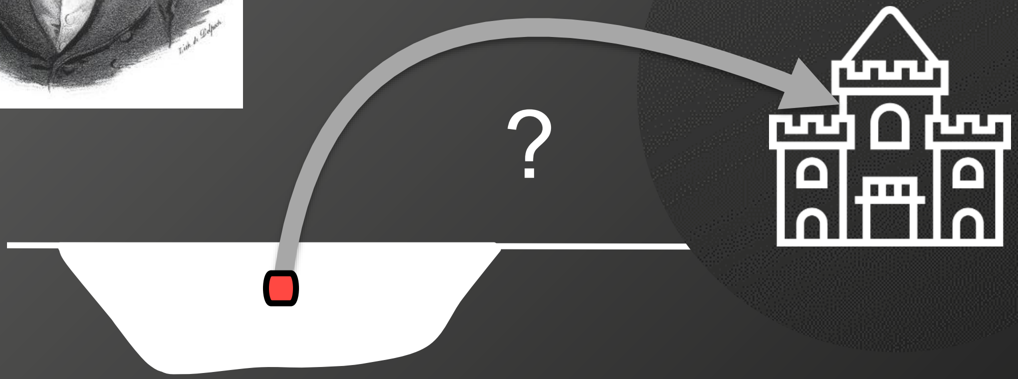
Assyr Abdulle Lecture, Prof. Dr. Alessio Figalli

Optimal transport: From A to B…and beyond

History of optimal transport (OT)

Monge’s problem

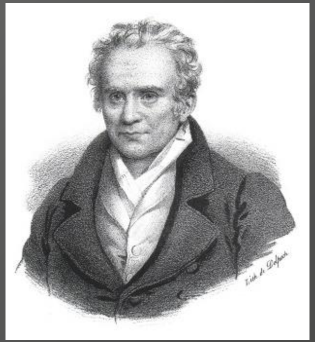


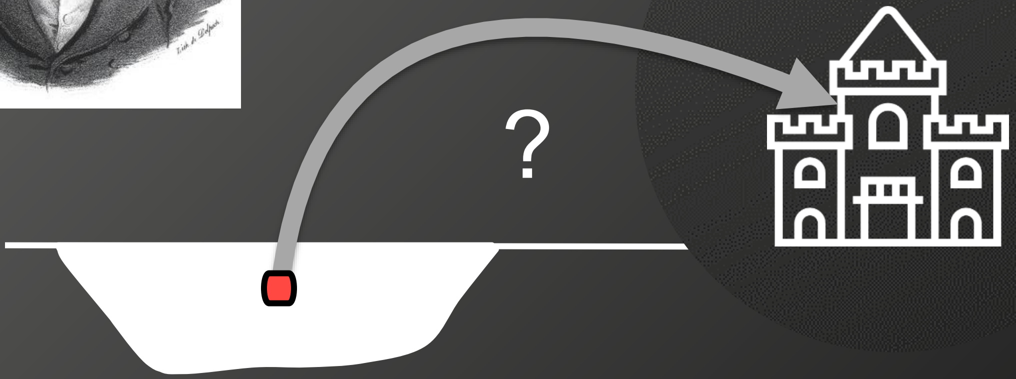


You have a certain amount of soil to extract from the ground, and transport to different places, to build a fortification.

History of optimal transport (OT)

Monge’s problem





You have a certain amount of soil to extract from the ground, and transport to different places, to build a fortification.

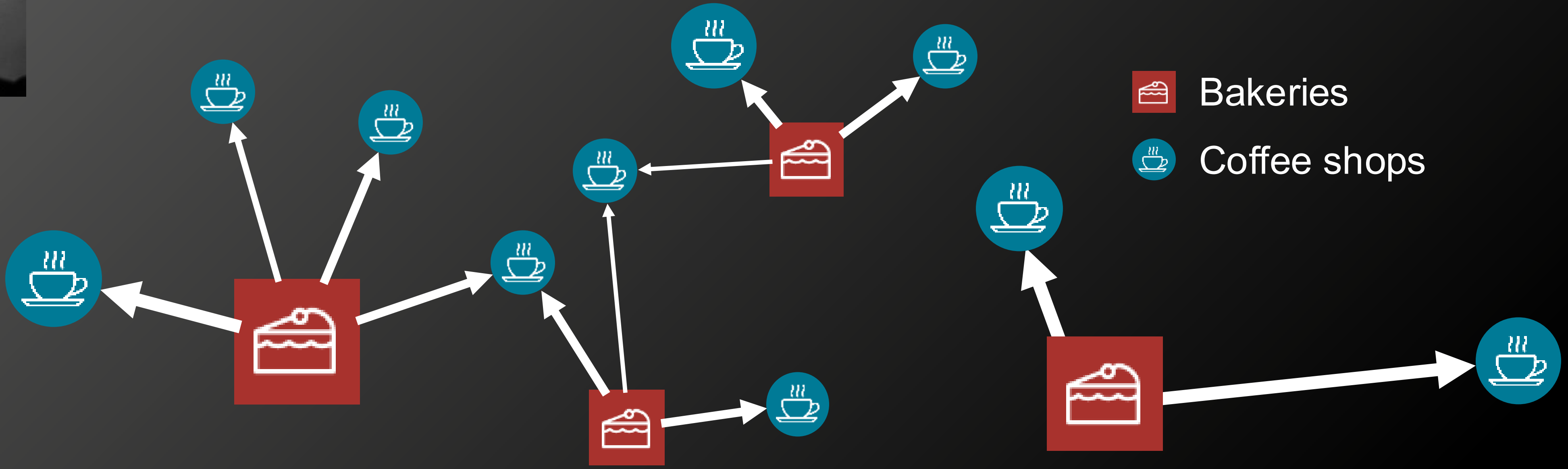
Consider each element of soil: to which destination should it be sent to guarantee the construction can be achieved while minimizing the total *transportation* cost?

Monge’s cost = total traveled distance

History of optimal transport (OT)

Leonid Kantorovich

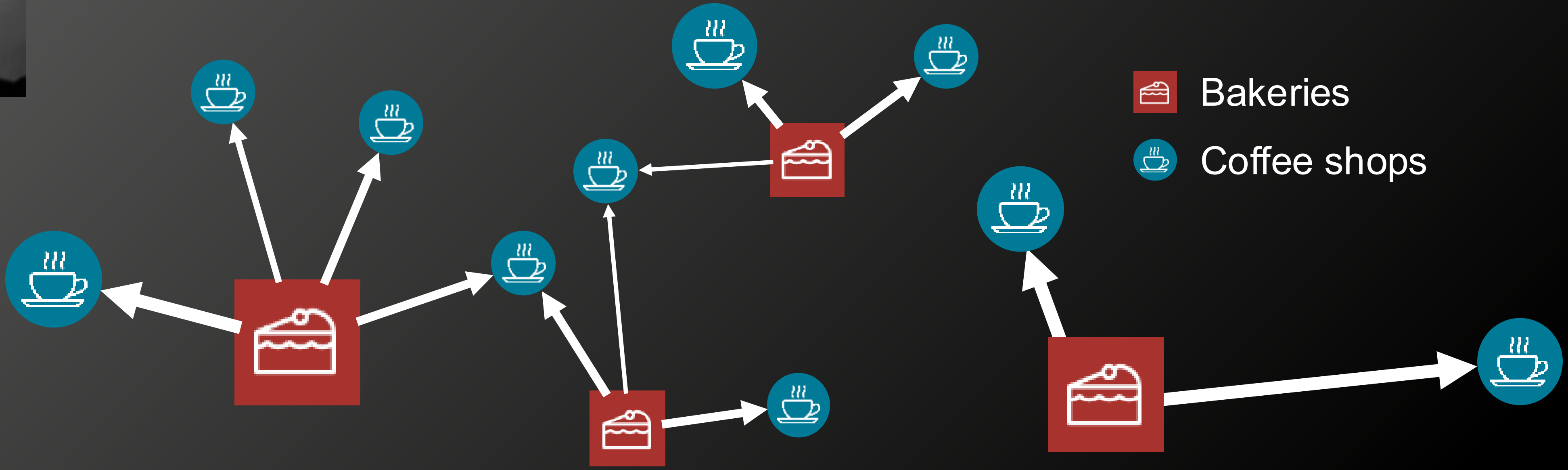
In the 1940s, Leonid Kantorovich studied Monge's problem and found a new formulation that allows for "non-deterministic" transports.



History of optimal transport (OT)

Leonid Kantorovich

In the 1940s, Leonid Kantorovich studied Monge's problem and found a new formulation that allows for "non-deterministic" transports.



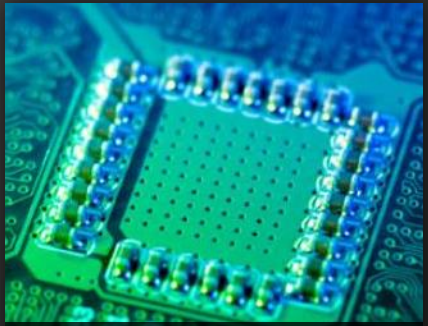
For his work, Kantorovich received the Nobel prize in economics in 1975 (joint with Koopmans).

Ubiquity of OT

OT has applications to a variety of fields


urban planning


meteorology

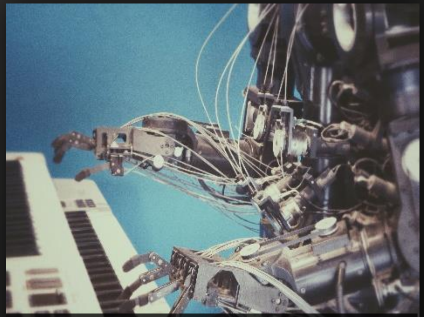

engineering design


statistics


hydrodynamics


biology


image processing


machine learning

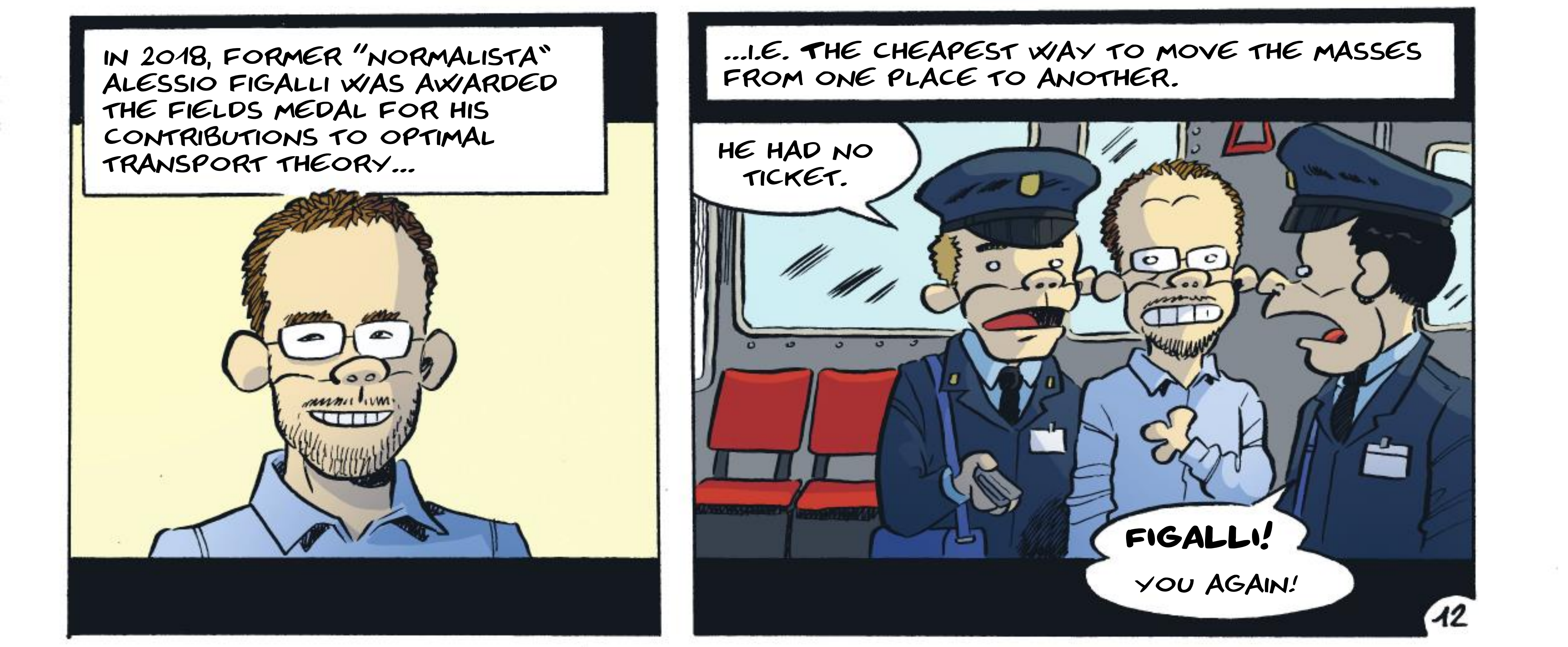

finance

...

more…

A Fields Medal solution to the optimal transport problem

What's the best way to move from one place to another?



ETH zürich

Problem I

Dido’s legend

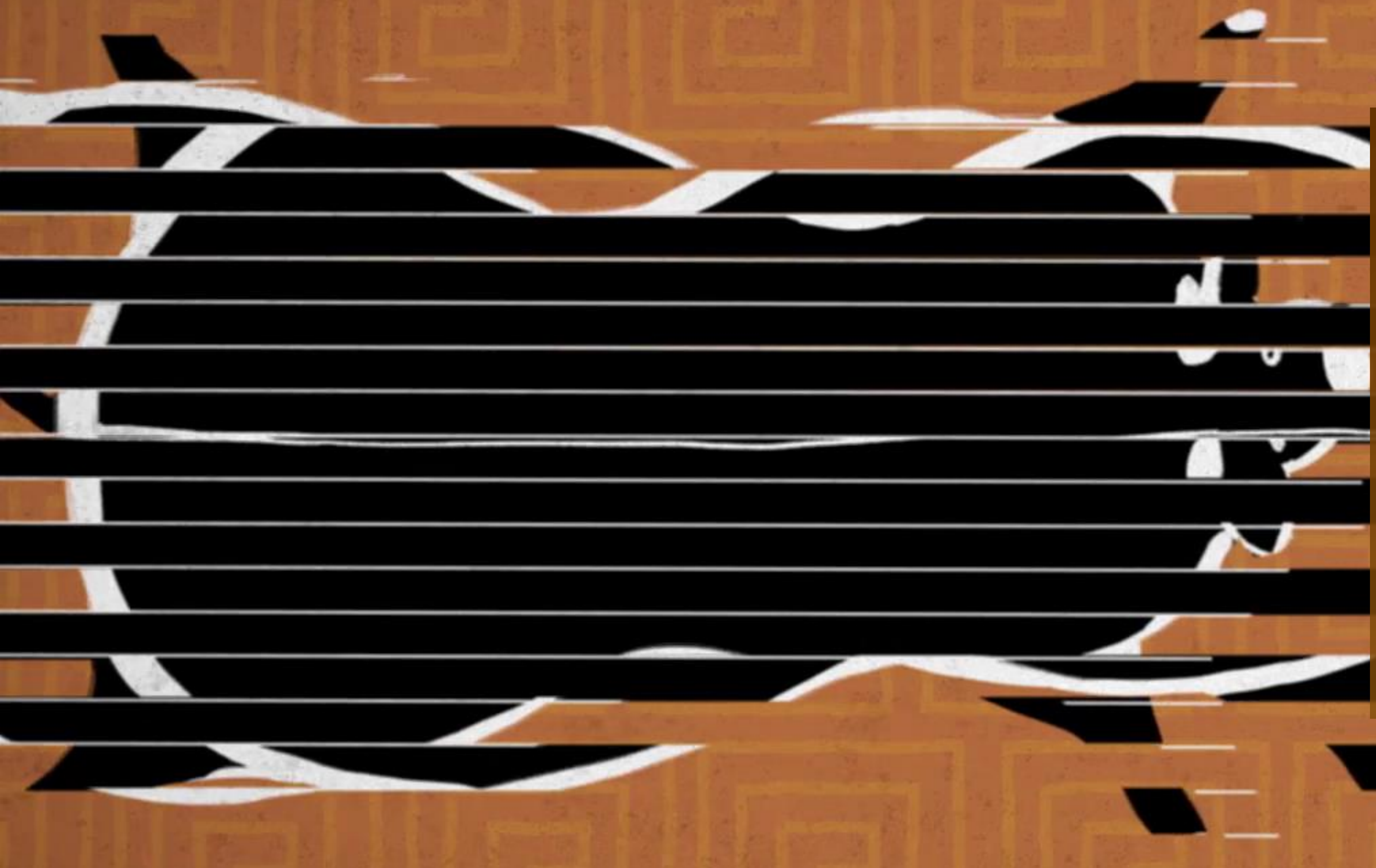
Around the 8th century BC., after Dido’s brother (king of Tyre) murdered her husband, Dido fled the city of Tyre and escaped with a few of her servants.

She sailed across the Mediterranean and landed in the realm of King Iarbas, from whom she sought to buy a piece of land where she and her servants could live.

ETH zürich

Problem I

Dido’s legend

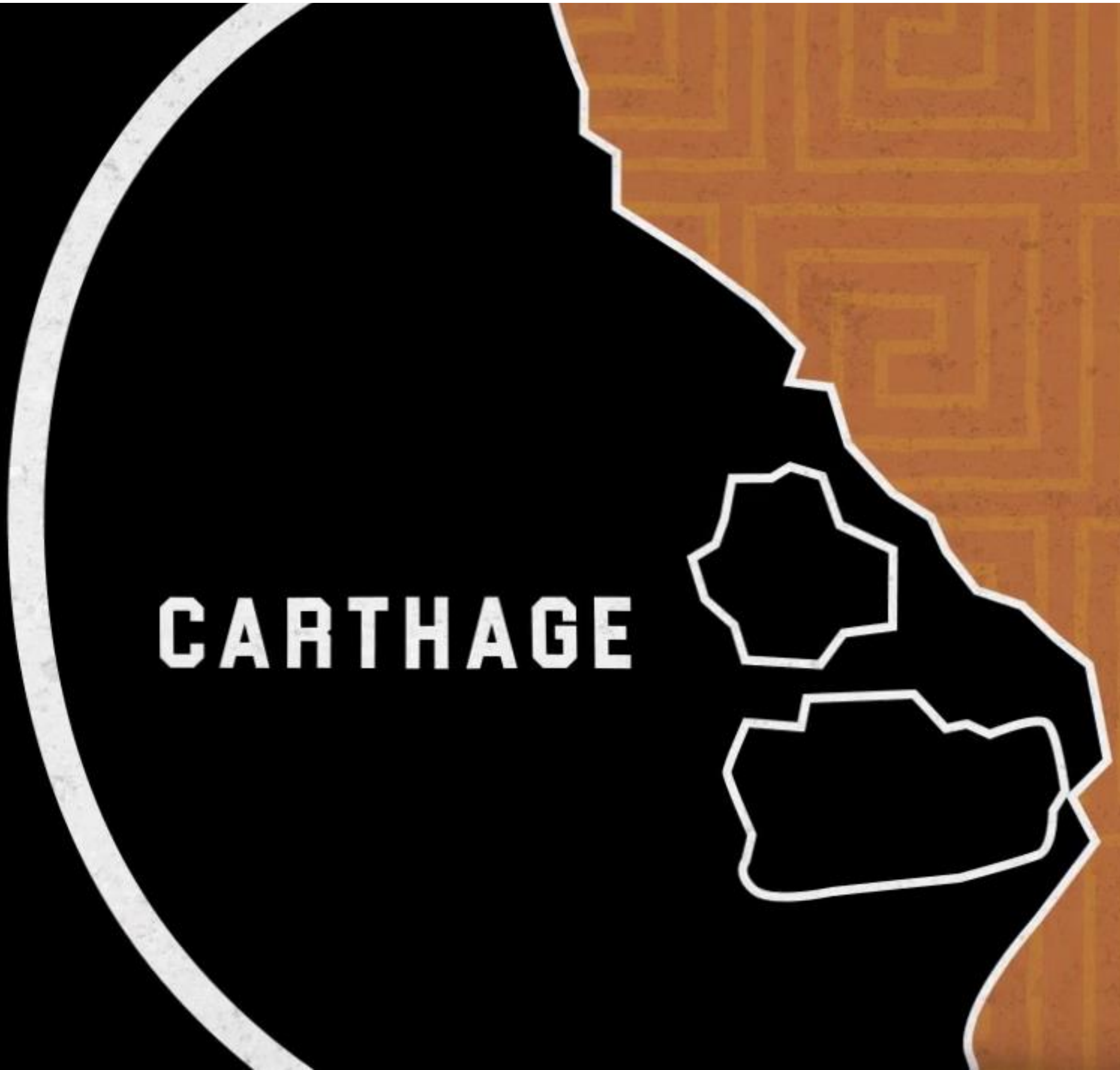


The king offered her what he thought was a raw deal where, for her money, she could have as much land as she could enclose with a single oxhide.

Dido cut the oxhide into incredibly thin strips.

Problem I

Dido’s legend



By sewing the strips together, she was able to enclose the largest possible area by laying the tied-together strips in a semicircle up against the coast.

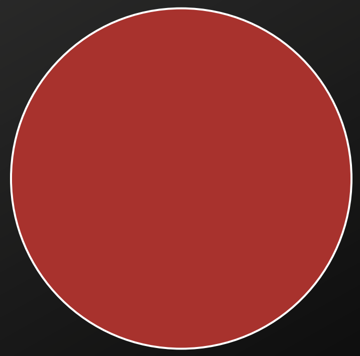
This became the heart of Carthage.

Problem I

Dido's legend is commemorated in mathematics

Isoperimetric problem:

Given a curve of a certain length,
enclose the maximum amount of area



Problem I

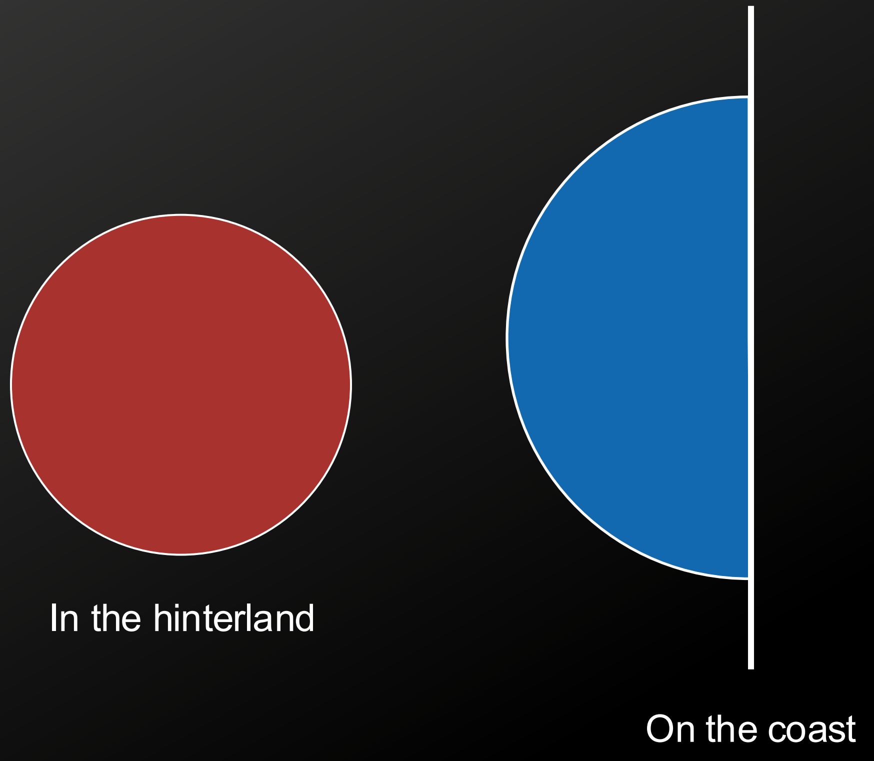
Dido’s legend is commemorated in mathematics

Isoperimetric problem:

Given a curve of a certain length, enclose the maximum amount of area

or equivalently

Minimize length among curves enclosing a fixed area



Problem I

What happens in three dimensions?

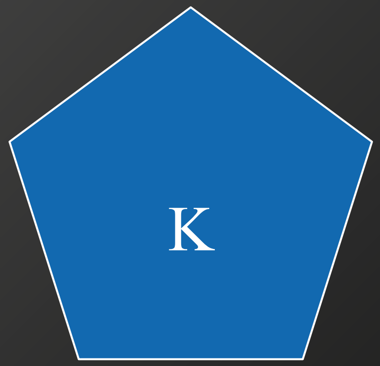
Isoperimetric problem:
Given a certain volume, what is the least surface area need to enclose it?

The answer is given to us by soap bubbles:
the sphere.

Given a certain amount of air, the bubble adjusts its shape to minimize its energy tension.

Problem I
Droplets and crystals

Gibbs 1878: crystals/droplets arrange themselves by assuming the shape with the lowest surface energy.

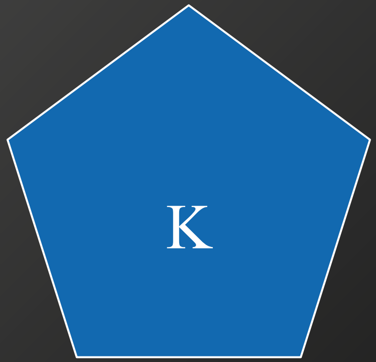


Problem I

Droplets and crystals

Gibbs 1878: crystals/droplets arrange themselves by assuming the shape with the lowest surface energy.

Assume we give energy to these crystals by raising the temperature. How much will the shape change, based on the amount of energy given to the crystal?

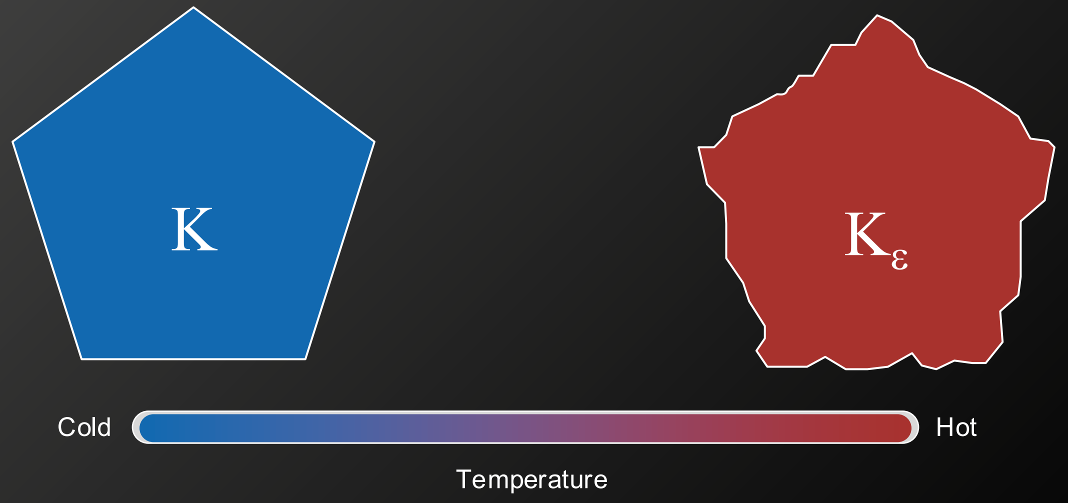


Problem I

Droplets and crystals

Gibbs 1878: crystals/droplets arrange themselves by assuming the shape with the lowest surface energy.

Assume we give energy to these crystals by raising the temperature. How much will the shape change, based on the amount of energy given to the crystal?

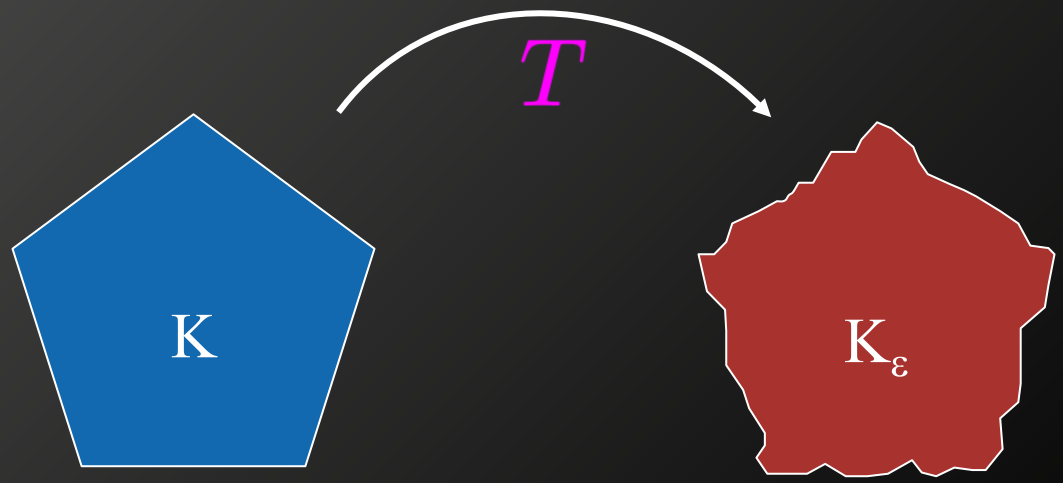


ε = amount of energy added to the system

Problem I

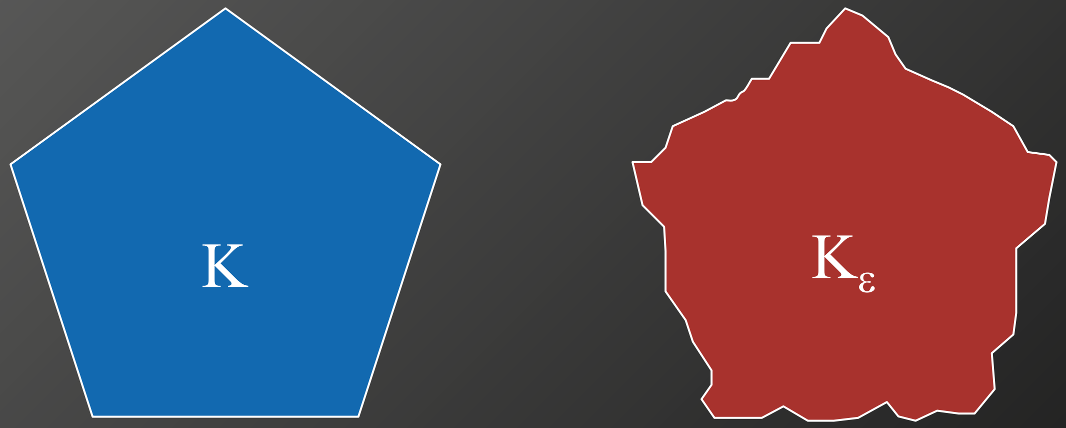
Droplets and crystals

We can think of OT as a way to follow how the individual particles move along the process of heating the crystal, so to understand the change in shape.



Problem I

Droplets and crystals



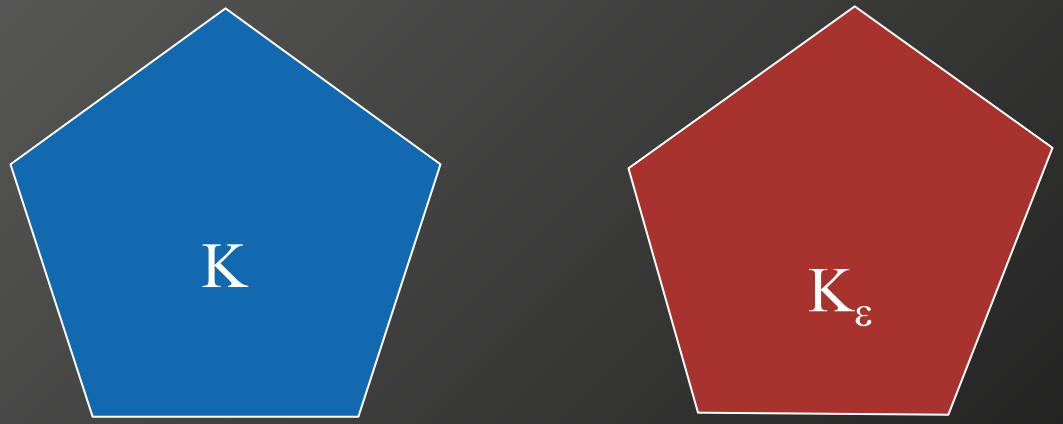
Theorem (Figalli, Maggi, Pratelli - 2010**)**

The shape of the crystal can change at most by $\sqrt{\varepsilon}$.

ε = amount of energy added to the system

Problem I

Droplets and crystals



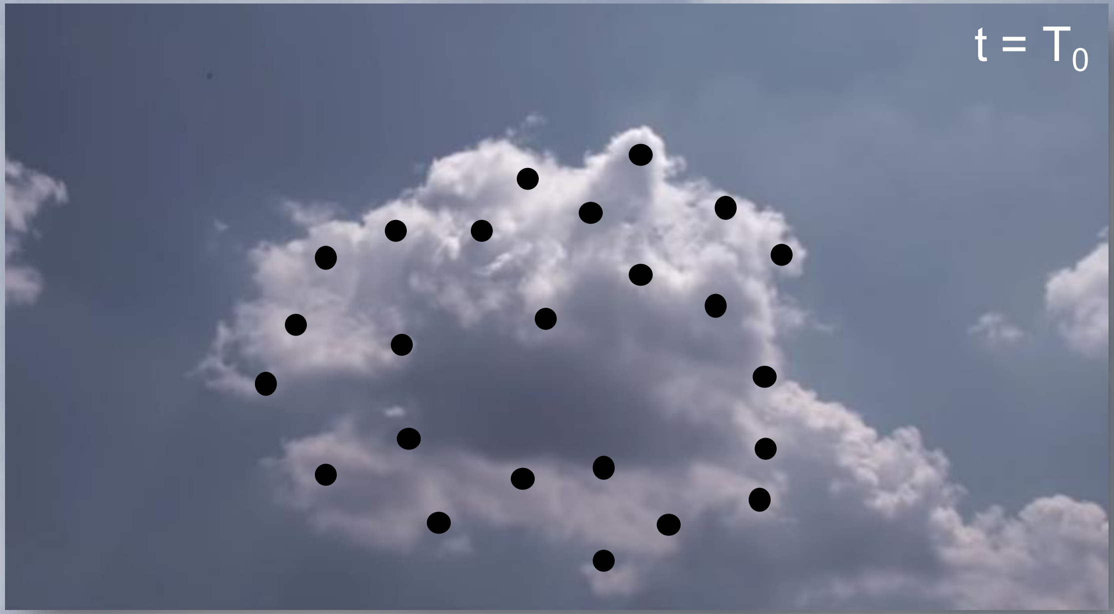
Theorem (Figalli, Zhang - 2020)

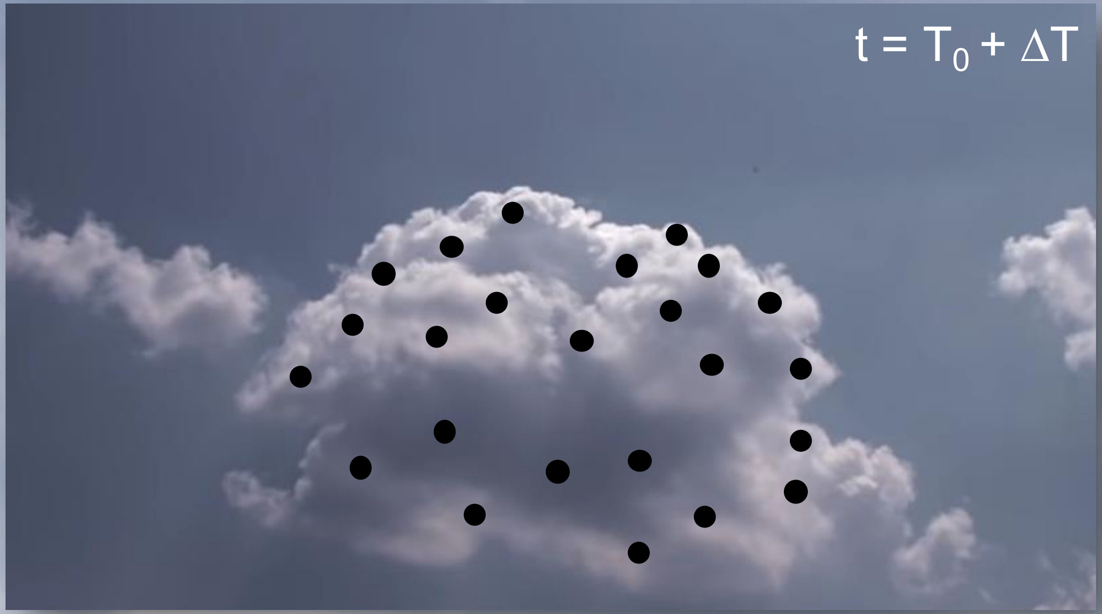
The crystal cannot change its structure.

ε = amount of energy added to the system

Problem II

Meteorology – Semigeostrophic equations





Who went where?

Problem II

Meteorology – Semigeostrophic equations

Who went where?

Answer: Particles move in an optimal way (Cullen, 1990's)

Problem II

Meteorology – Semigeostrophic equations

Connection OT ←→ SG

known for more than 20 years…

…but there was lack of results on the regularity theory of optimal transport.

Problem II

Meteorology – Semigeostrophic equations

Connection OT ←→ SG

known for more than 20 years…

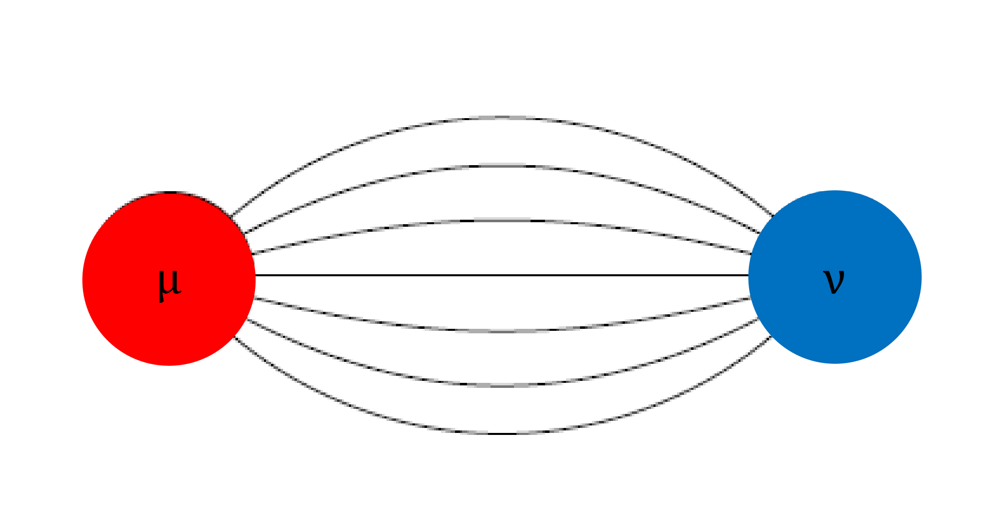
…but there was lack of results on the regularity theory of optimal transport.

In 2013, with De Philippis, we developed it.

Then we used it with Ambrosio and Colombo to solve the SG equations.

Optimal transport and Wasserstein distance

A way to measure distance between distributions



Optimal transport can be used to define a “distance” between probability distributions:

$$W_c(\mu, \nu) := \inf_{X\sim\mu,\, Y\sim\nu} \mathbb{E}[c(X, Y)]$$

What for? Image transfer (Rabin-Delon-Gousseau 2010)

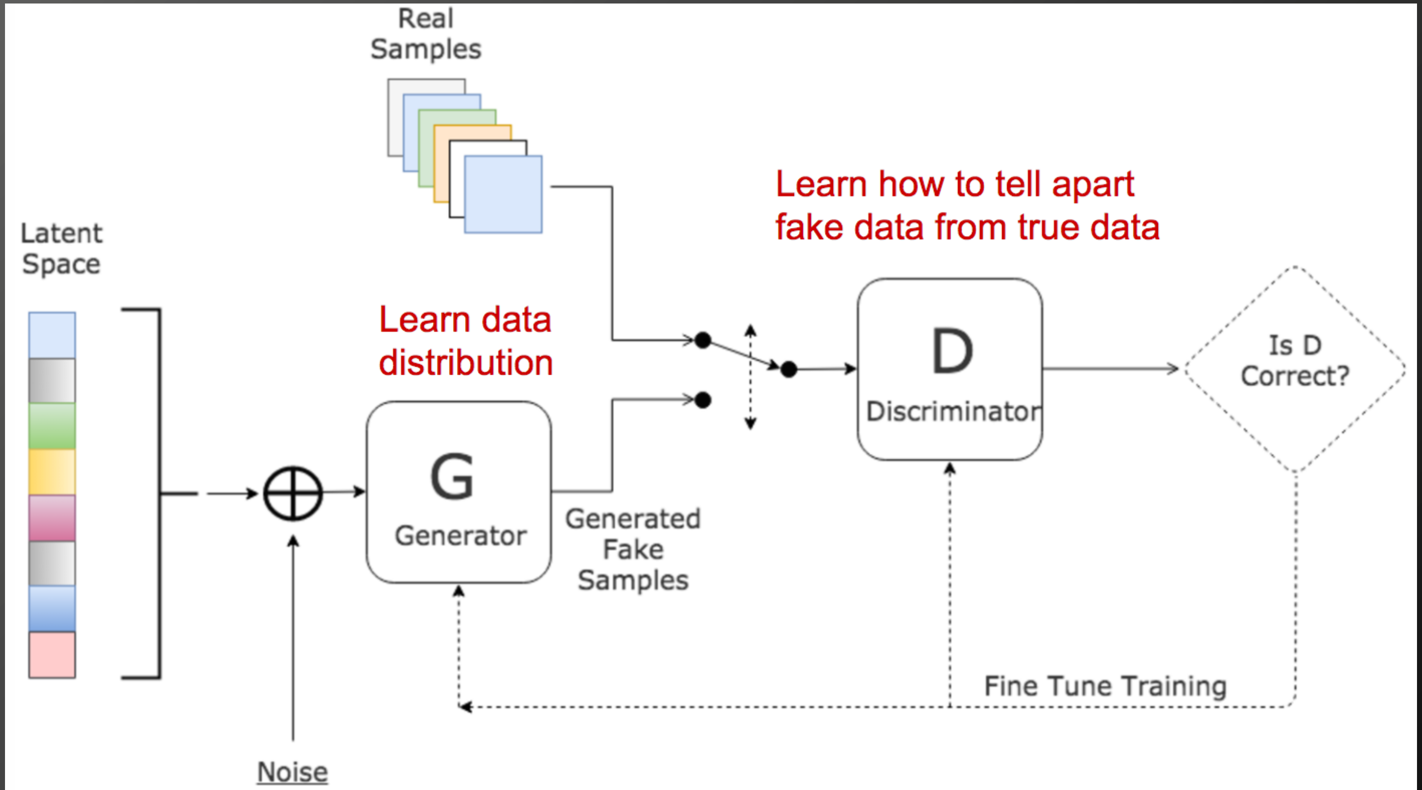
(a) Source image (Auguste Renoir, *Le déjeuner des Canotiers*, 1881). (b) Style image (Paul Gauguin, *Mahana no atua – le jour de Dieu*, 1894). (c) Color transfer. (d) Iterated TRM filter.

(e) Source image. (f) Style image. (g) Color transfer. (h) Iterated TRM filter.

Optimal transport can be used to efficiently “transport” colors from one image to another.

OT: what else?

1) Generative Adversarial Networks (GANs)



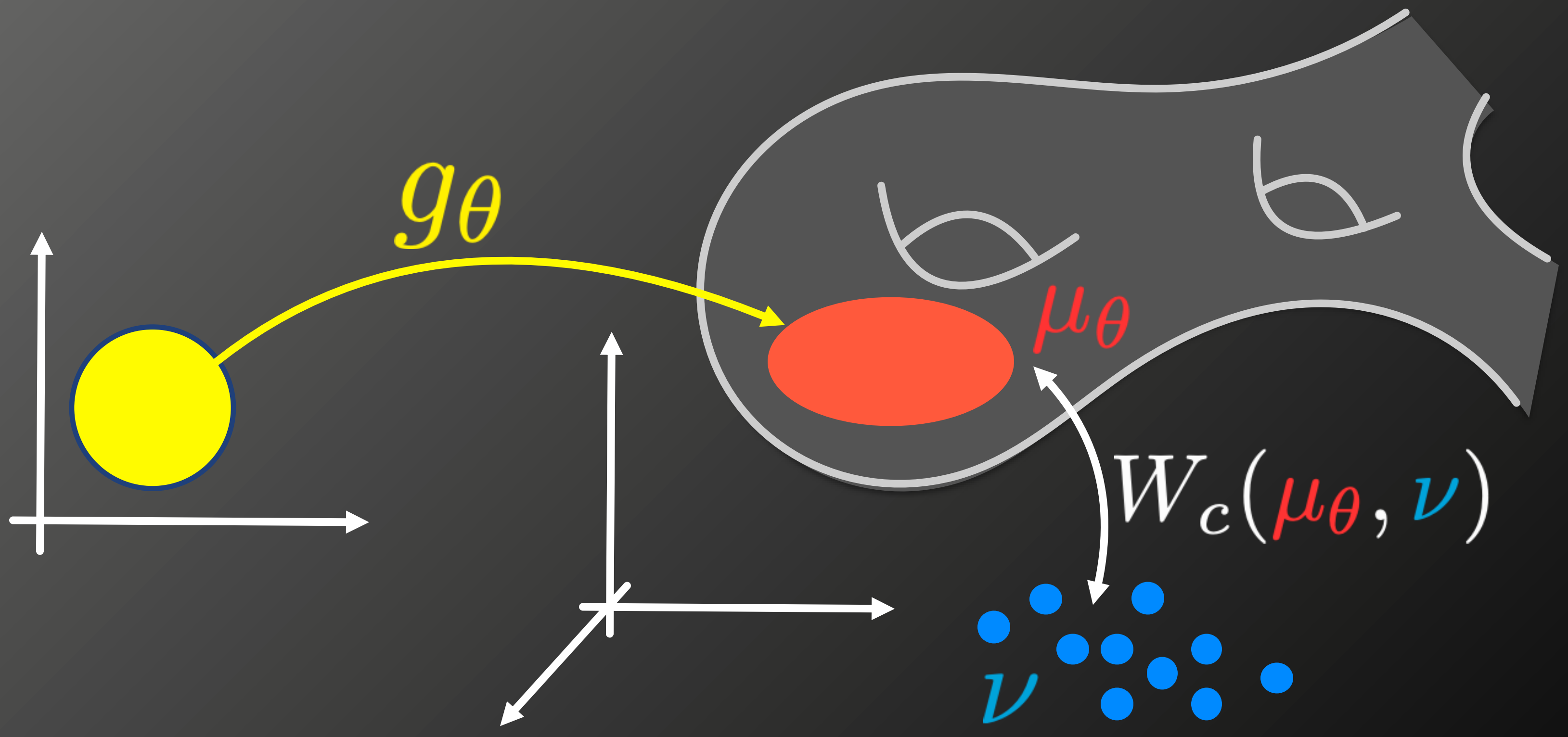
GAN (Goodfellow et al. 2014):

- a discriminator D estimates the probability of a given sample coming from the real dataset
- a generator G captures real data to generate samples as real as possible

Goal: optimize G and D

OT: what else?

1) Wasserstein GAN (Arjovsky et al. 2017, Lei et al. 2019)

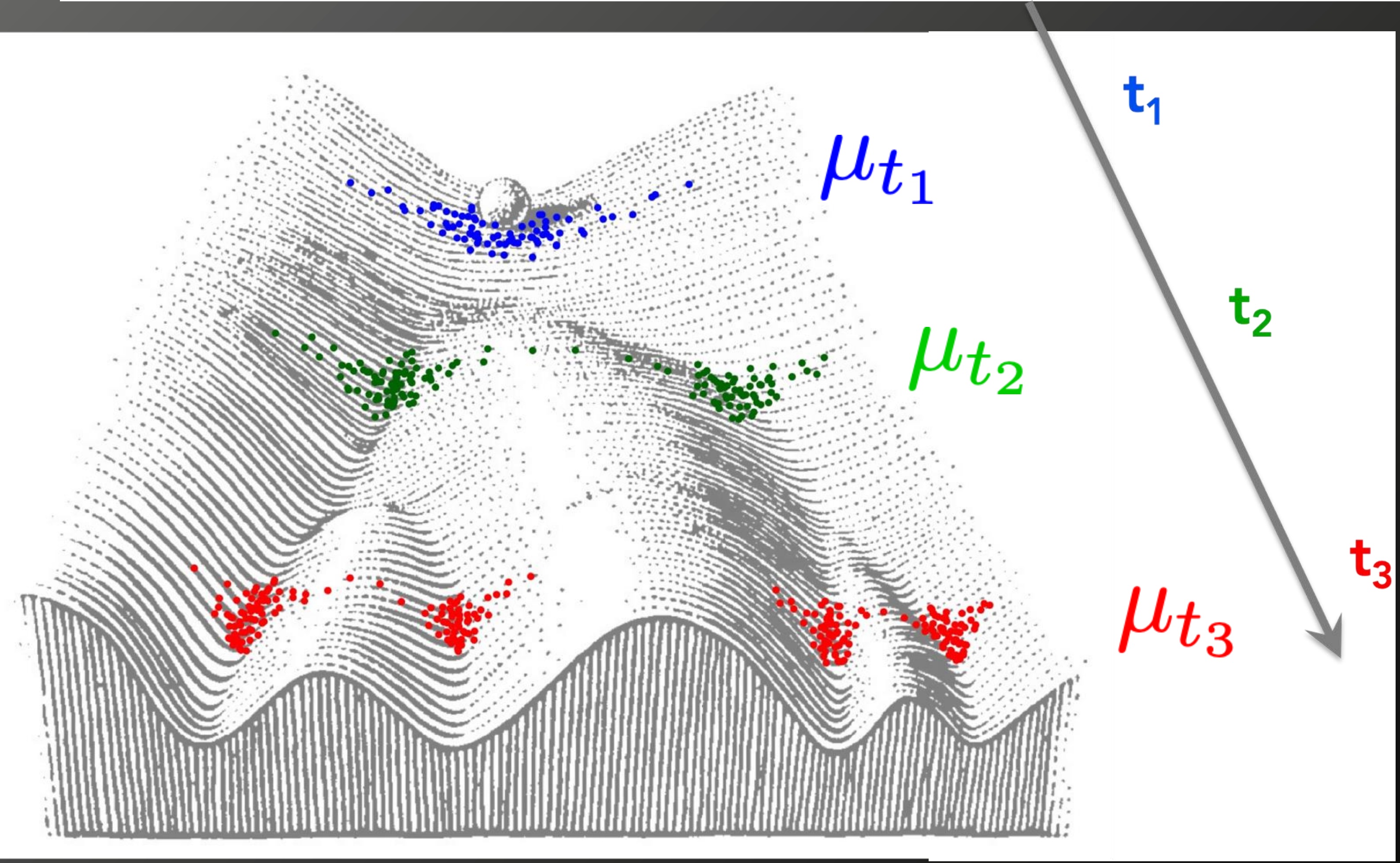


W-GAN: apply OT distances to decide whether an image is real or fake.

$$\min_\theta W_c(\mu_\theta, \nu)$$

OT: what else?

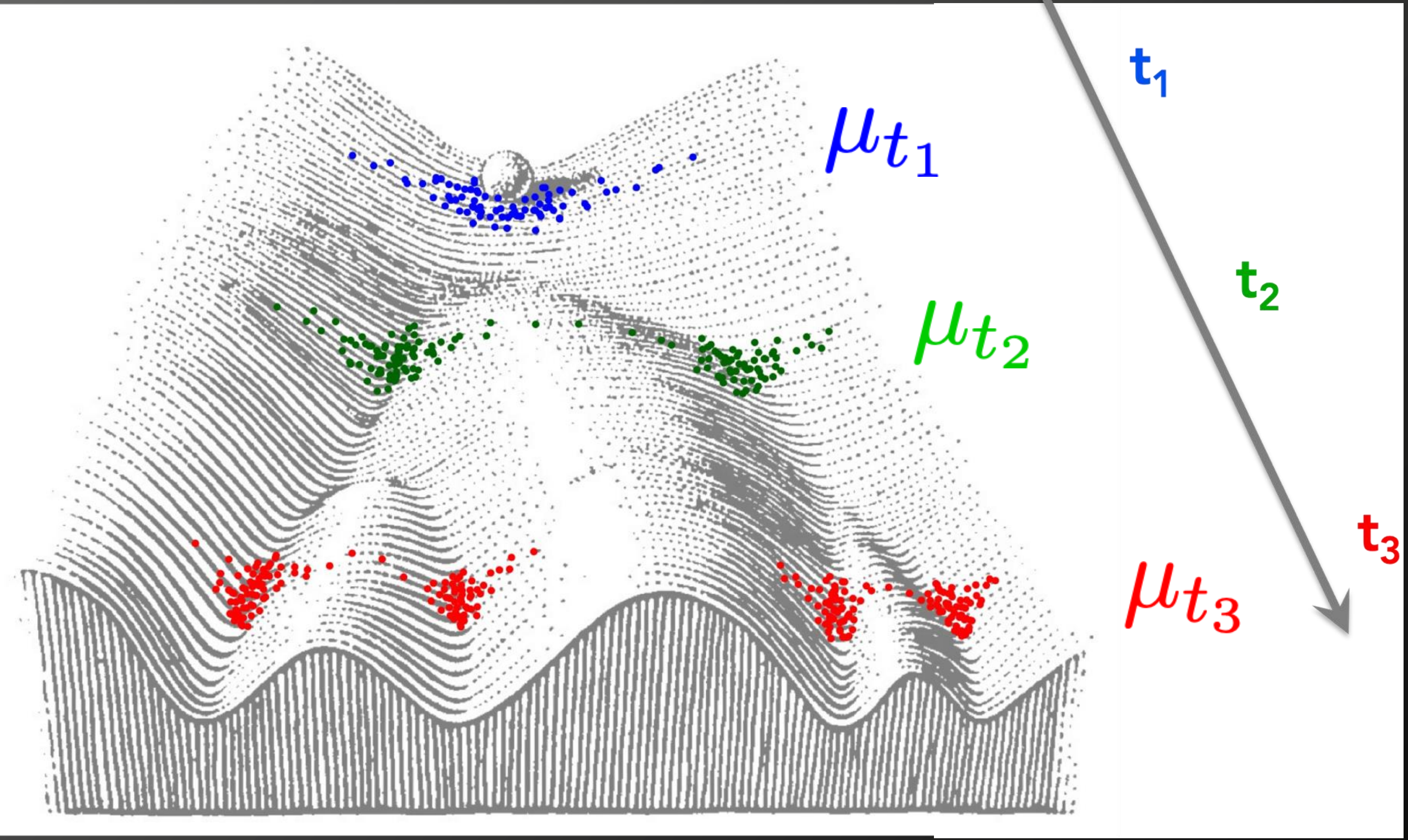
2) Single-cell genomics (Schiebinger et al. 2019)



Goal: understand the molecular programs that guide differentiation of cells during development

OT: what else?

2) Single-cell genomics (Schiebinger et al. 2019)



Goal: understand the molecular programs that guide differentiation of cells during development

Waddington-OT: combine Waddington landscape model with OT to identify the trajectories of the cells from samples collected independently at various times.

OT

Recent questions: numerical complexity

A fundamental question: how to solve OT computationally?
If one considers discrete measures with N points, standard algorithms require N^3.

OT

Recent questions: numerical complexity

A fundamental question: how to solve OT computationally?
If one considers discrete measures with N points, standard algorithms require N^3.

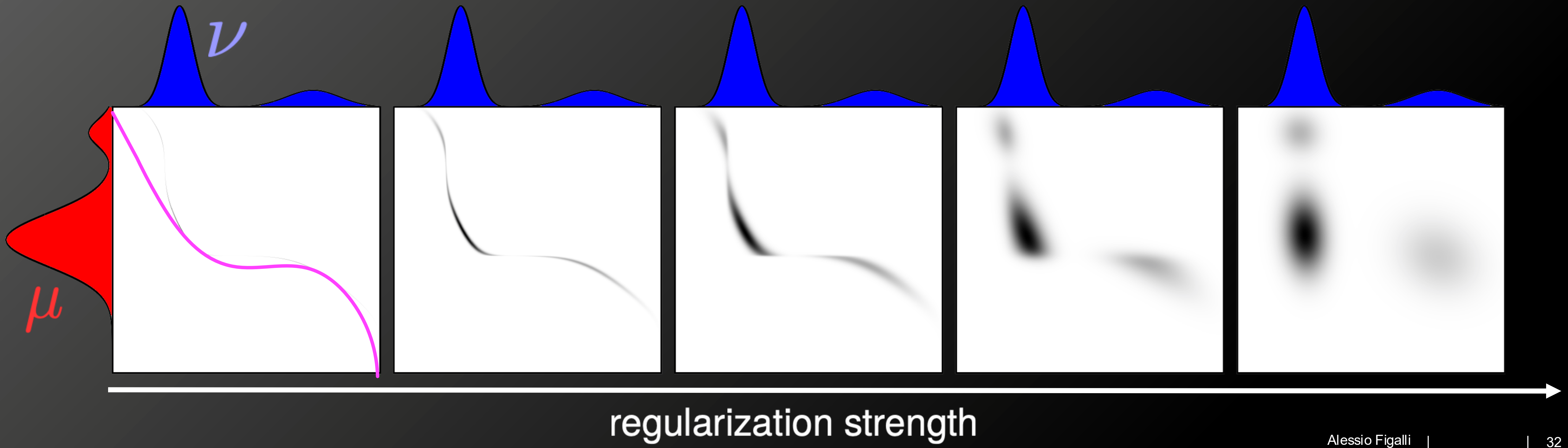
Entropic regularization: improves computational cost (cf. Cuturi-Peyré 2018).

OT

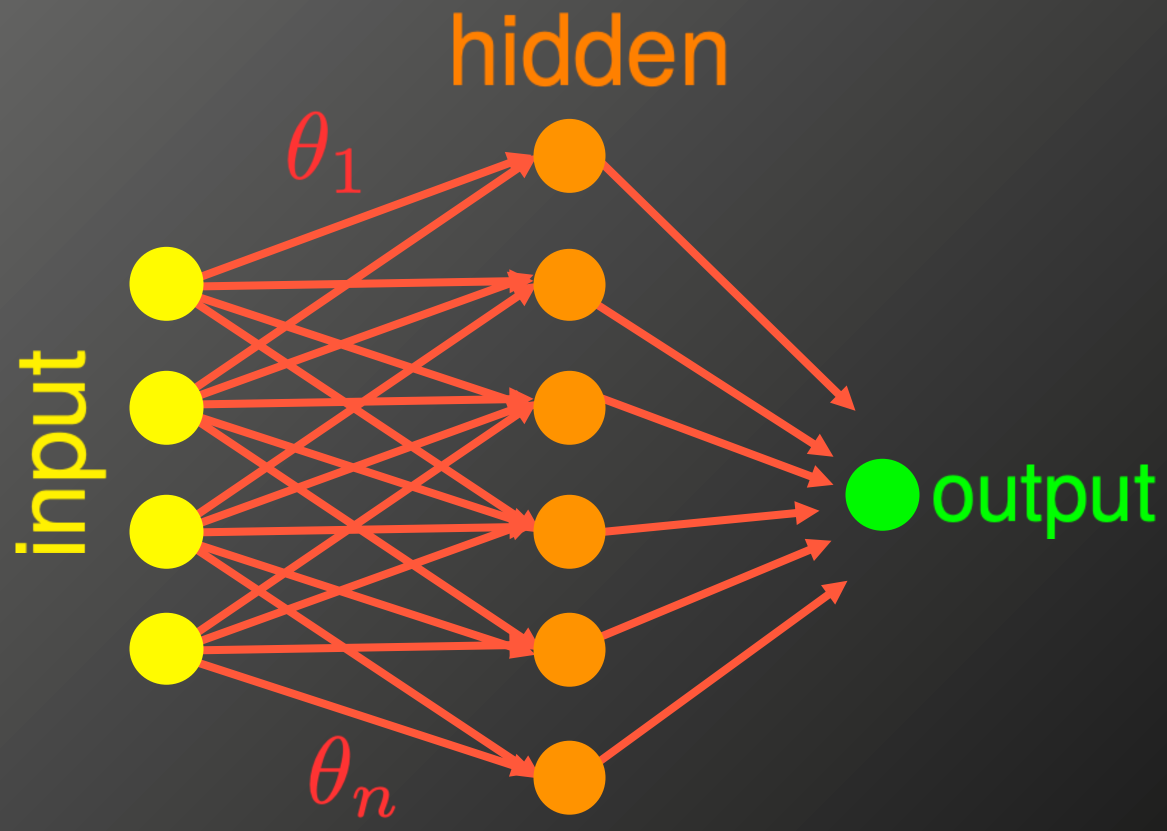
Recent questions: numerical complexity

A fundamental question: how to solve OT computationally?
If one considers discrete measures with N points, standard algorithms require N^3.

Entropic regularization: improves computational cost (cf. Cuturi-Peyré 2018).



OT
Recent questions: neural networks and gradient flows

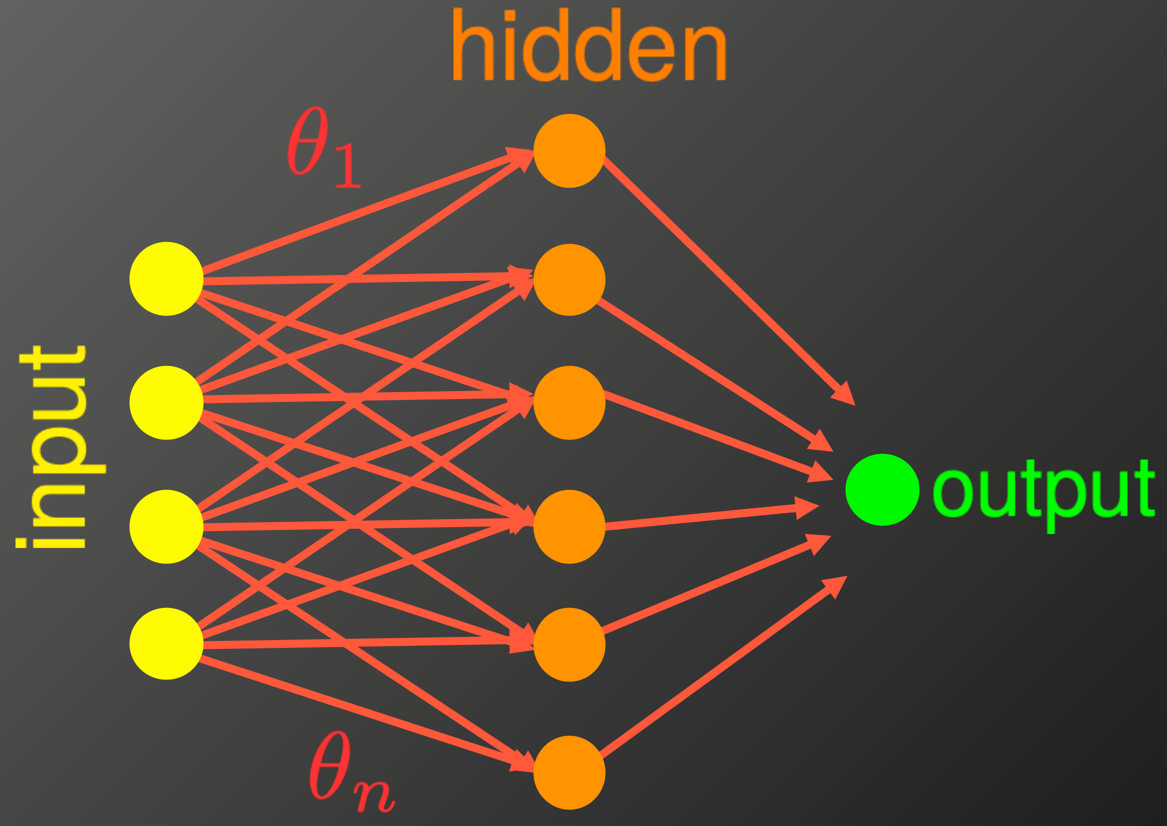


$$\mu := \frac{1}{n}\sum_i \delta_{\theta_i}$$

Training of a *single* layer neural network can be reinterpreted as an «OT gradient flow» for the weights.

OT

Recent questions: neural networks and gradient flows



$$\mu := \frac{1}{n}\sum_i \delta_{\theta_i}$$

Training of a *single* layer neural network can be reinterpreted as an «OT gradient flow» for the weights.

Gives rise to equations of the form:

$$\partial_t \mu_t = \mathrm{div}(\nabla \partial \mathrm{Loss}(\mu_t))$$

For multi layers NN, no such interpretation is available.

OT

Recent questions: OT and loss functions

Natural problem in machine learning:

- ν models data distribution;
- learned probability distribution (from which one samples) $\mu_\theta := (g_\theta)_\sharp \mu$

Training is an optimization problem: $\min_\theta W_c(\mu_\theta, \nu)$

Non-convex problem: makes it highly challenging.

OT

Recent questions: OT and loss functions

Natural problem in machine learning:
- ν models data distribution;
- learned probability distribution (from which one samples) $\mu_\theta := (g_\theta)_\sharp \mu$

Training is an optimization problem: $\min_\theta W_c(\mu_\theta, \nu)$

Non-convex problem: makes it highly challenging.

In practice: OT seems to « convexify » some problems.

Understanding why, could lead to important breakthroughs.

Connections to Wasserstein GANs, Wasserstein linear regression, global optimization networks, inverse problems, etc.

Assyr Abdulle Lecture, Prof. Dr. Alessio Figalli

Thank you for your attention.