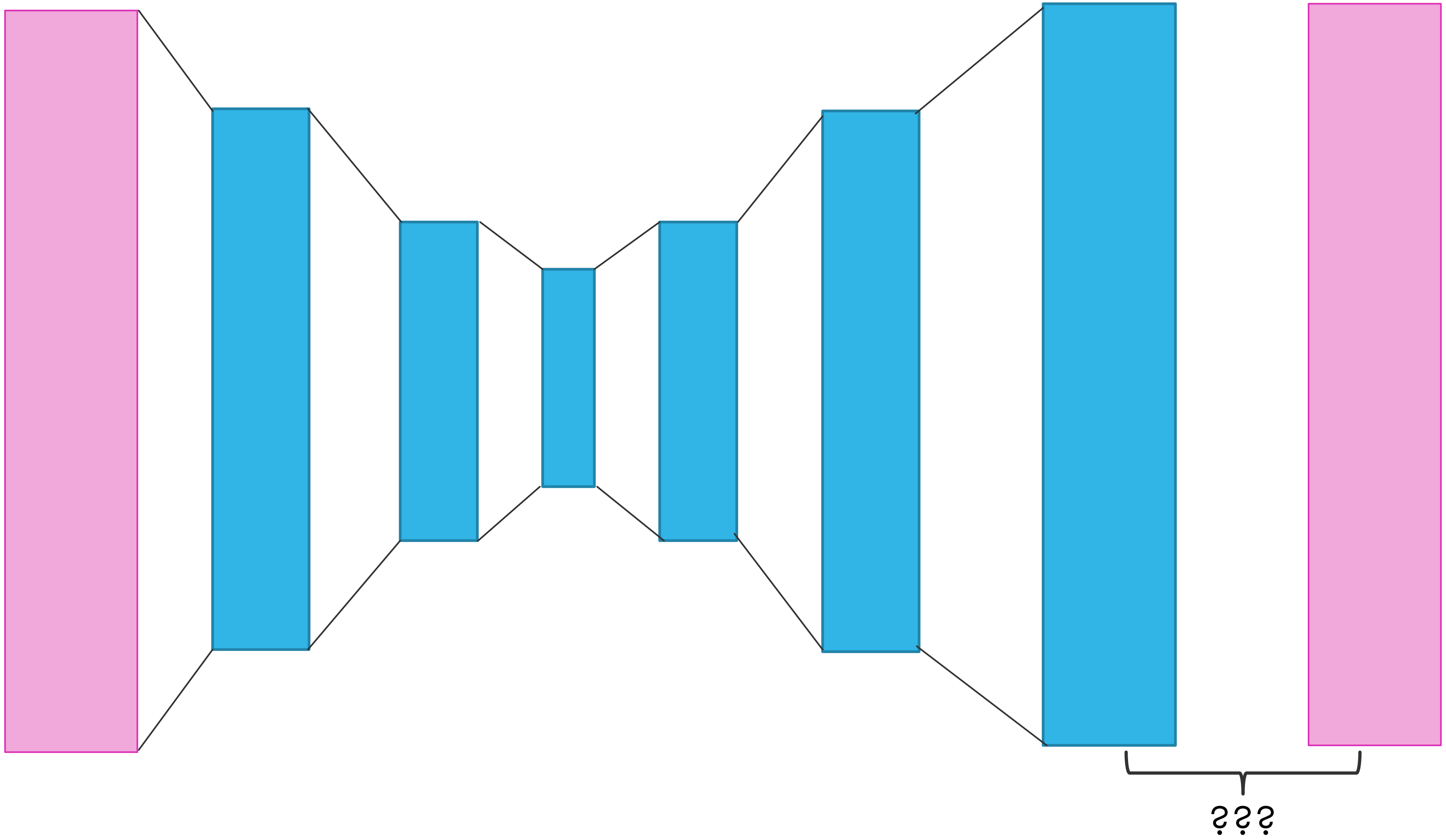
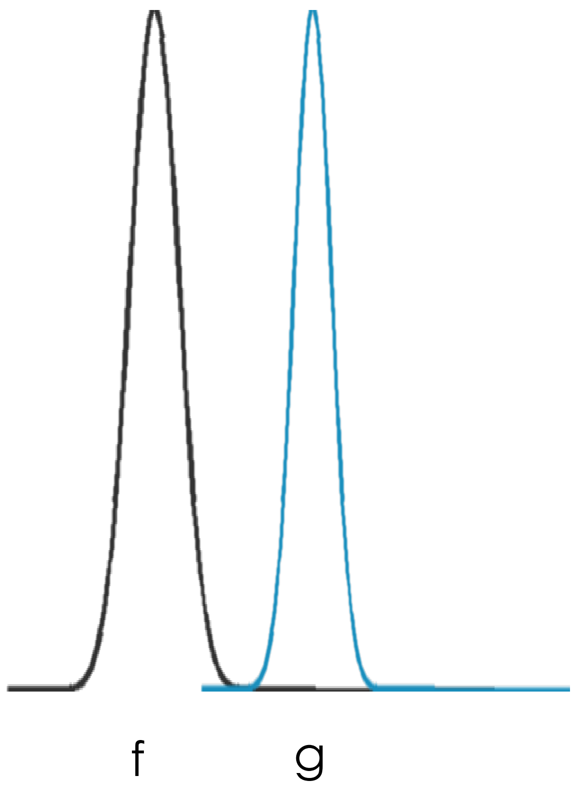
Optimal Transport Principles for Computer Graphics and Machine Learning

Nicolas Bonneel

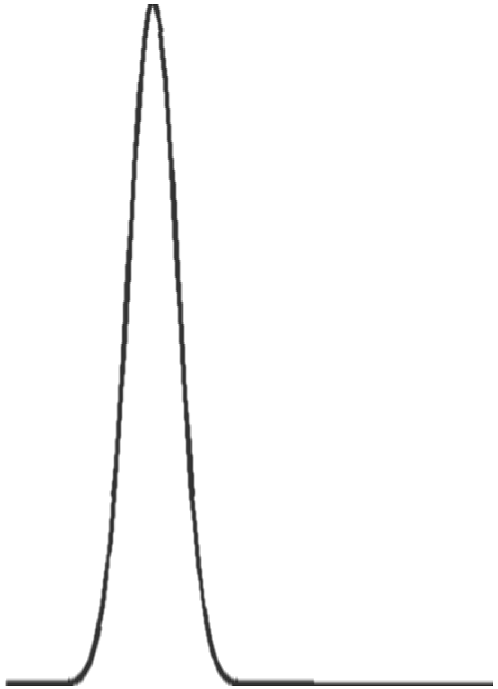
Motivation in neural networks



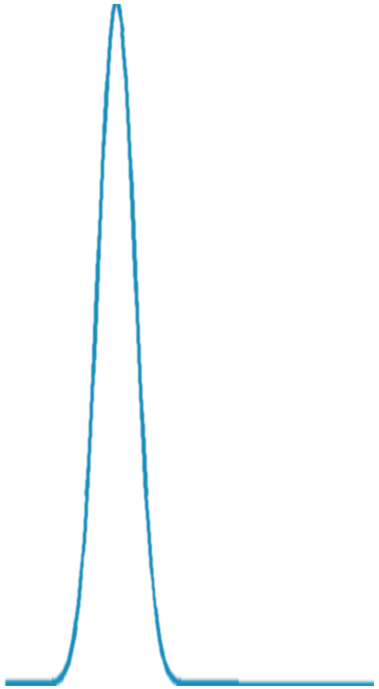
How to compare functions ?



How to compare functions ?

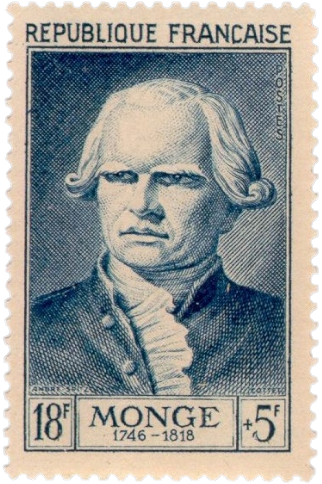


f



g

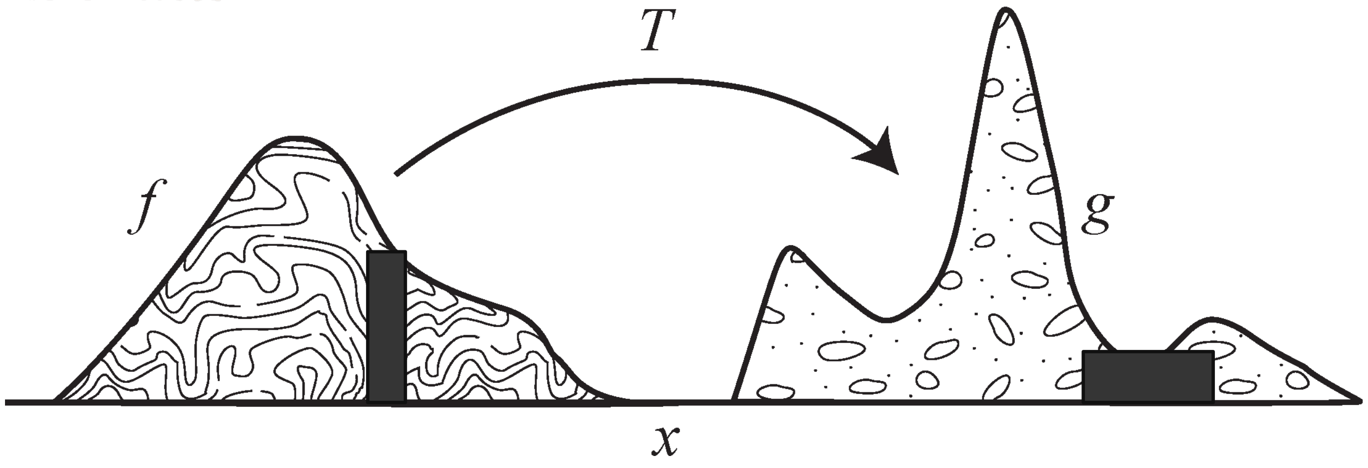
Monge formulation



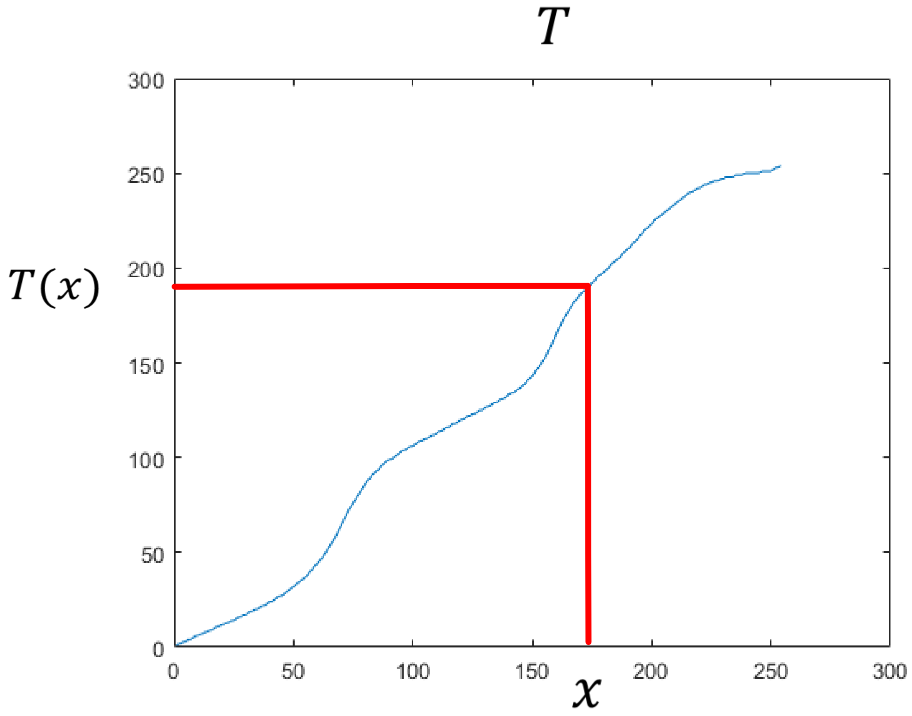
$$W(f,g) = \min_{T} \int_{X} c\,(x, T(x))\, f(x)\, \mathrm{dx}$$

s.t. $\quad f(x) = g(T(x))\, |\det J_T(x)|$

Monge used $c(x,y) = |x - y|$



Same total mass



"find a good warping between f and g with the change of variable formula"

Kantorovich formulation

Cost

$$\min_m \sum_i \sum_j c_{i,j}\, m_{i\to j}$$

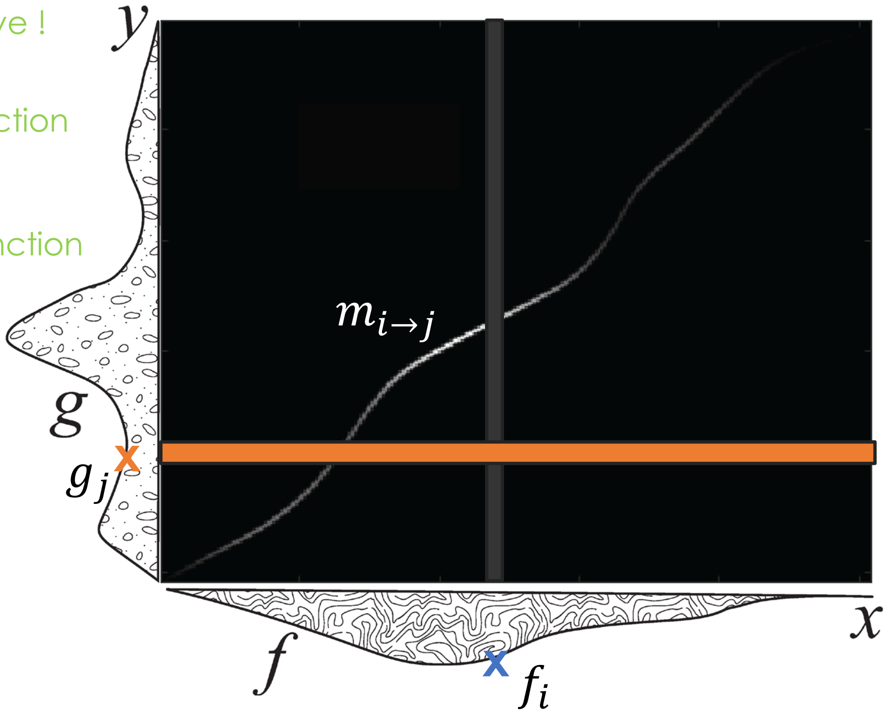
$m_{i\to j}$ particles will move from i to j

such that: $m_{i\to j} \geq 0$ Nb of particles is positive !

$\sum_i m_{i\to j} = g_j$ Reconstruct target function

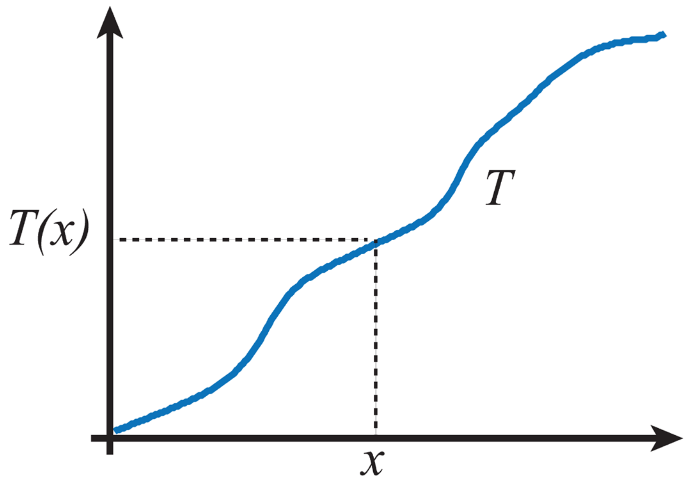
$\sum_j m_{i\to j} = f_i$ Reconstruct source function

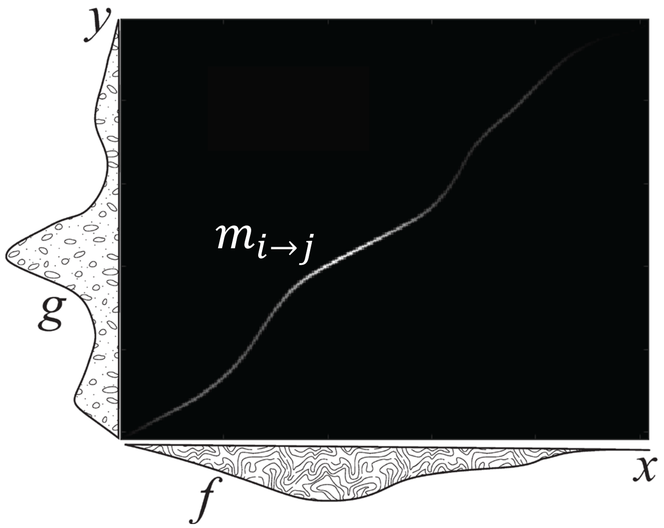
Work for transforming f into g



Discretization of the Kantorovich problem
Earth Mover's Distance

Intuition: comparison





- Finds a "transport map"
- Difficult non-linear problem
- May have no solution
  (e.g., a Dirac splitting in two)
- Leads to PDEs

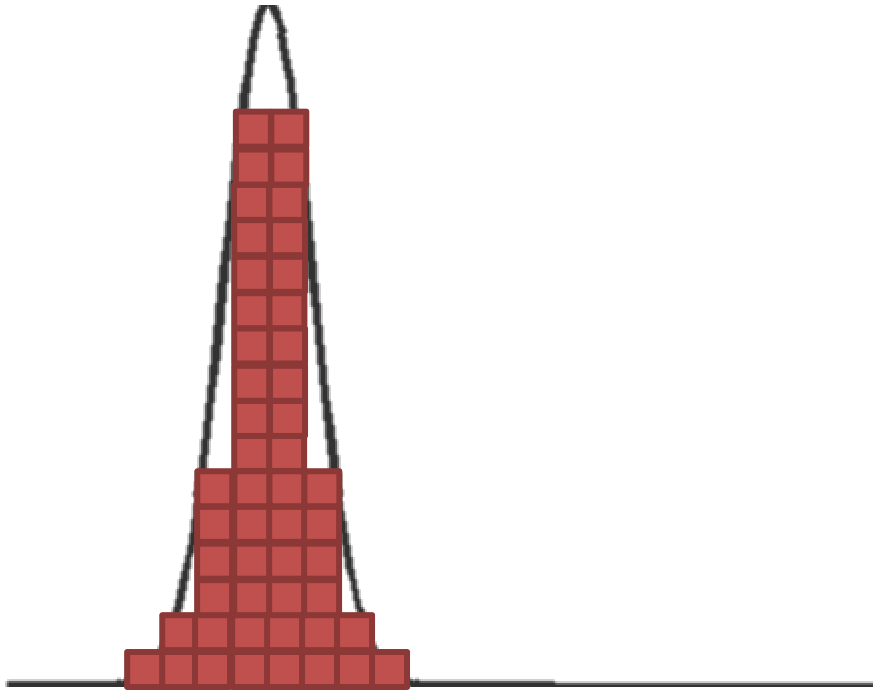
- Finds a "transport plan"
- Linear program
- "Always" has solution
  (i.e., under reasonable assumptions)
- Also has dual formulation

When it exists, the solution is the same.

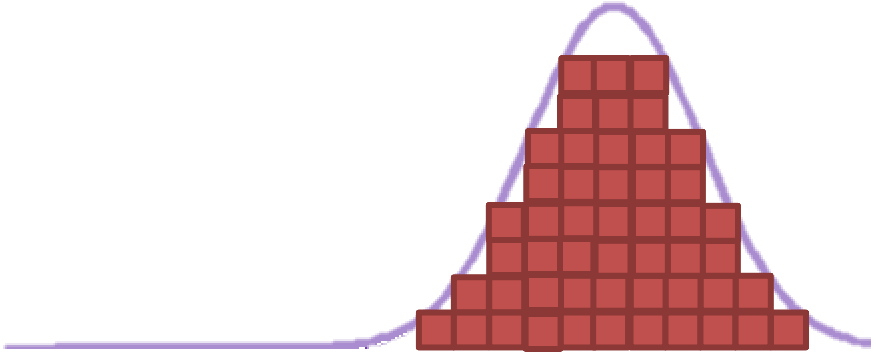
Often, $c(x, y) = \|x - y\|_p^p \Rightarrow W_p^p$
W_p is a distance

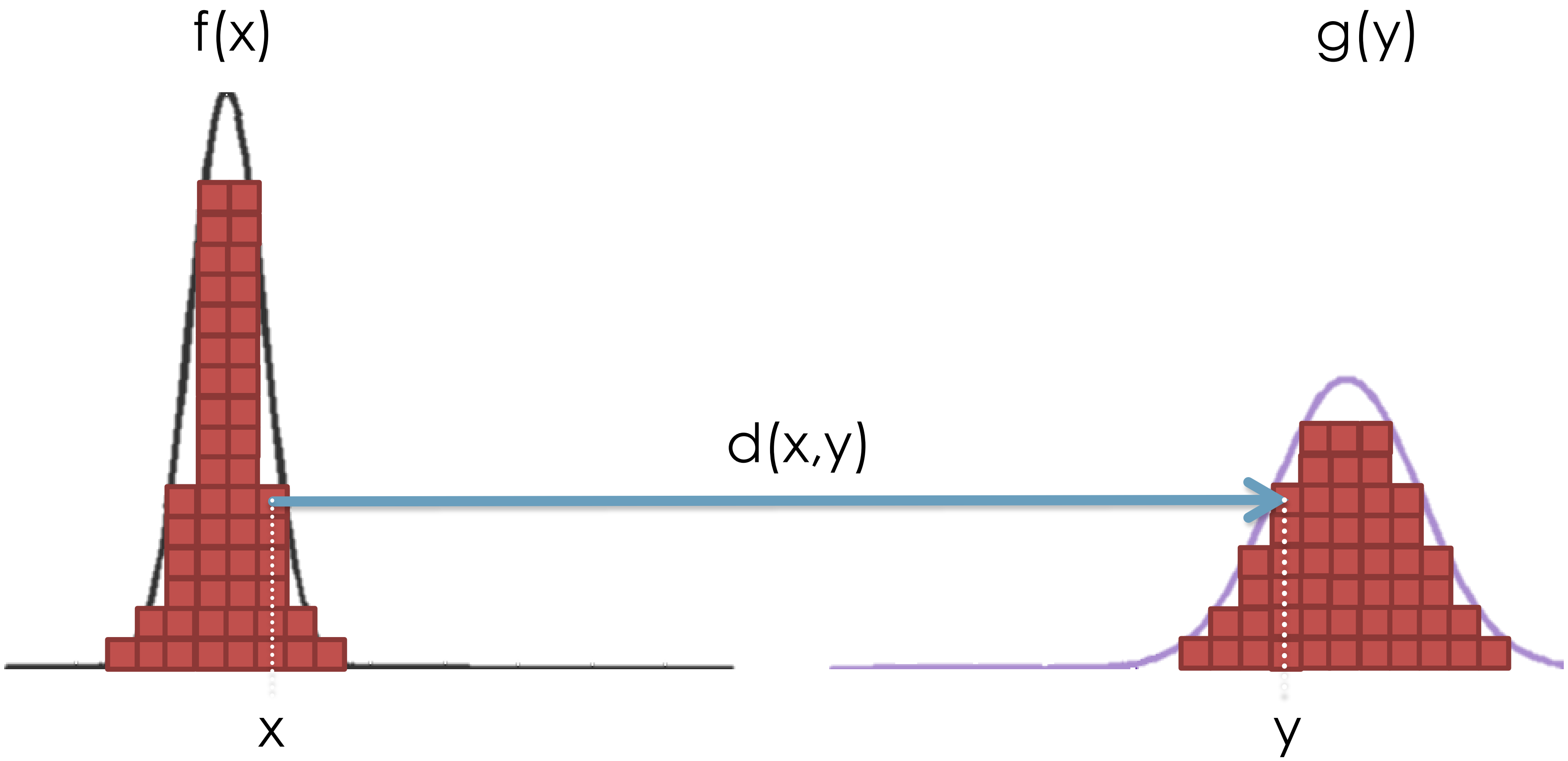

f(x)
g(y)

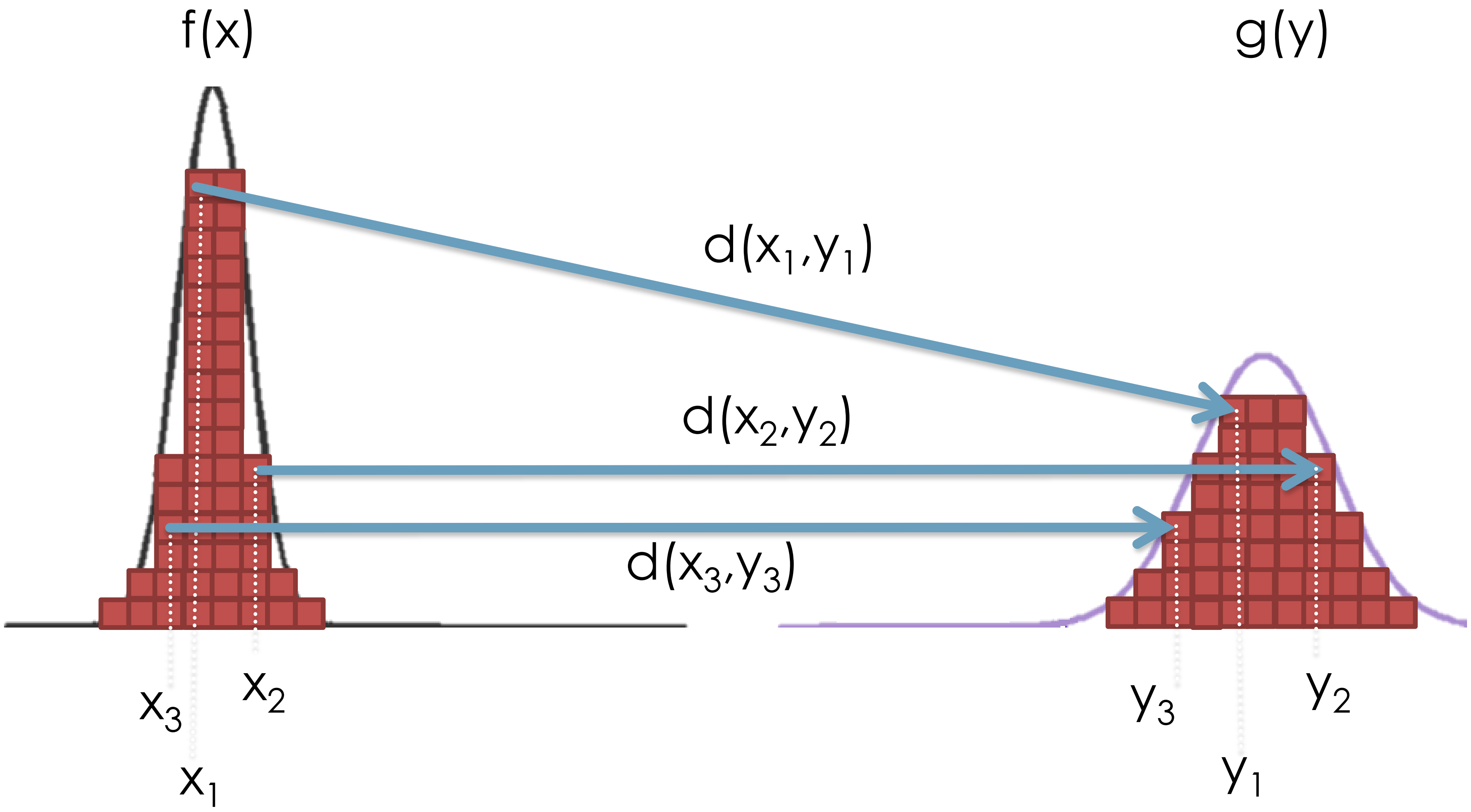
f(x)

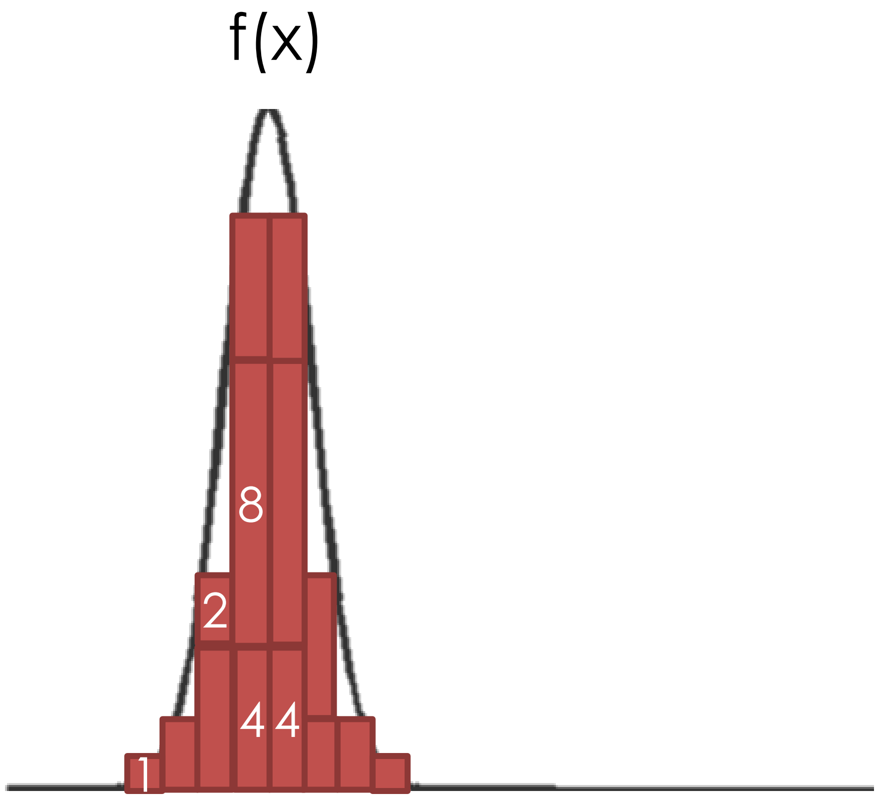


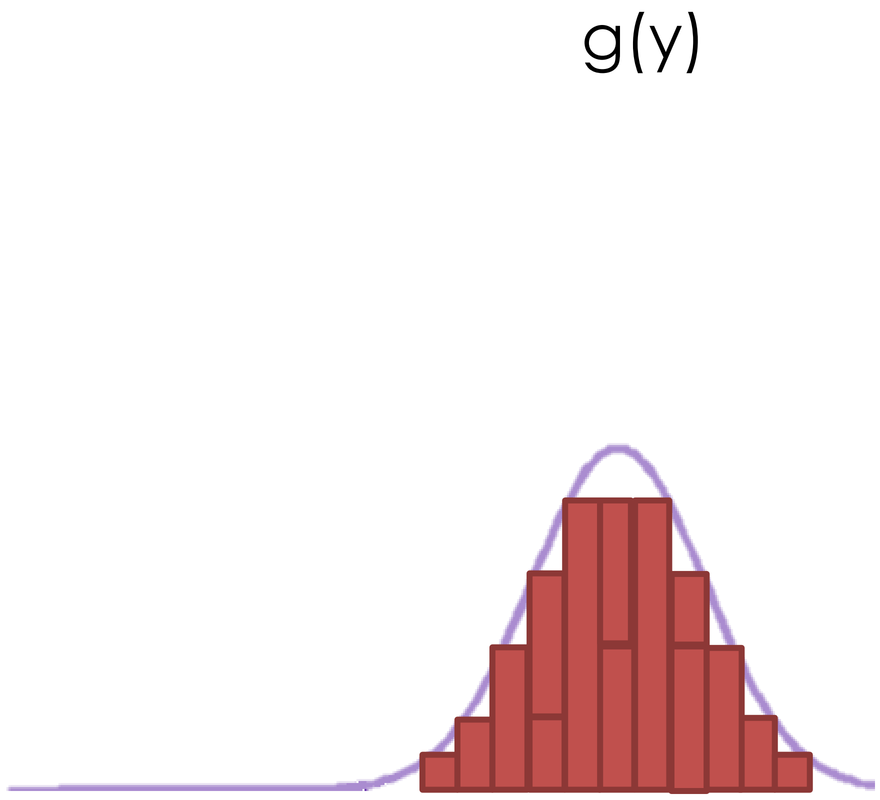
g(y)

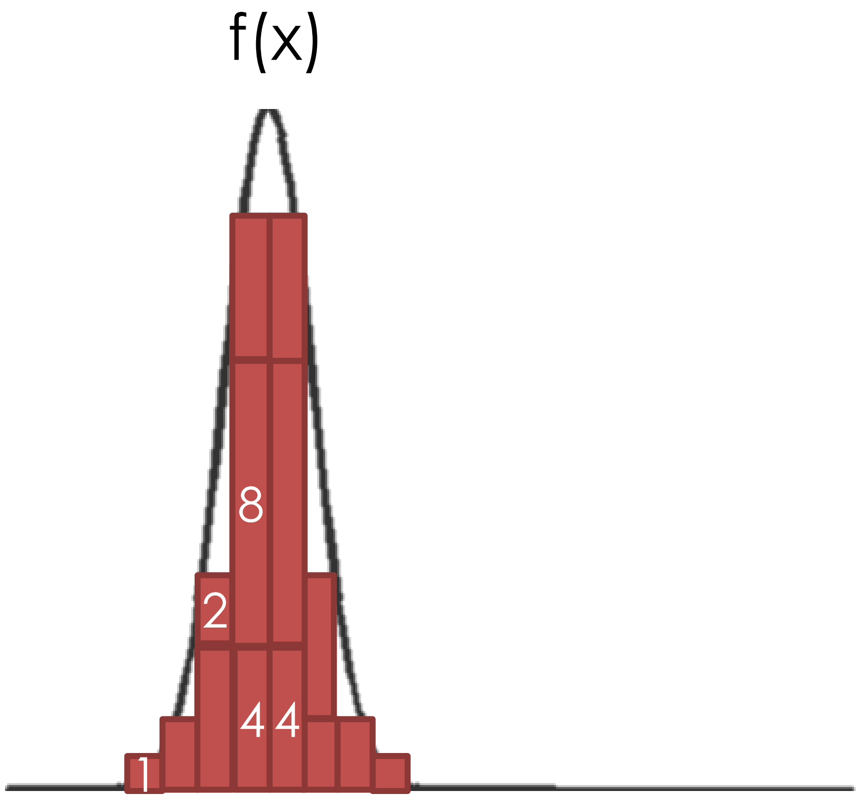


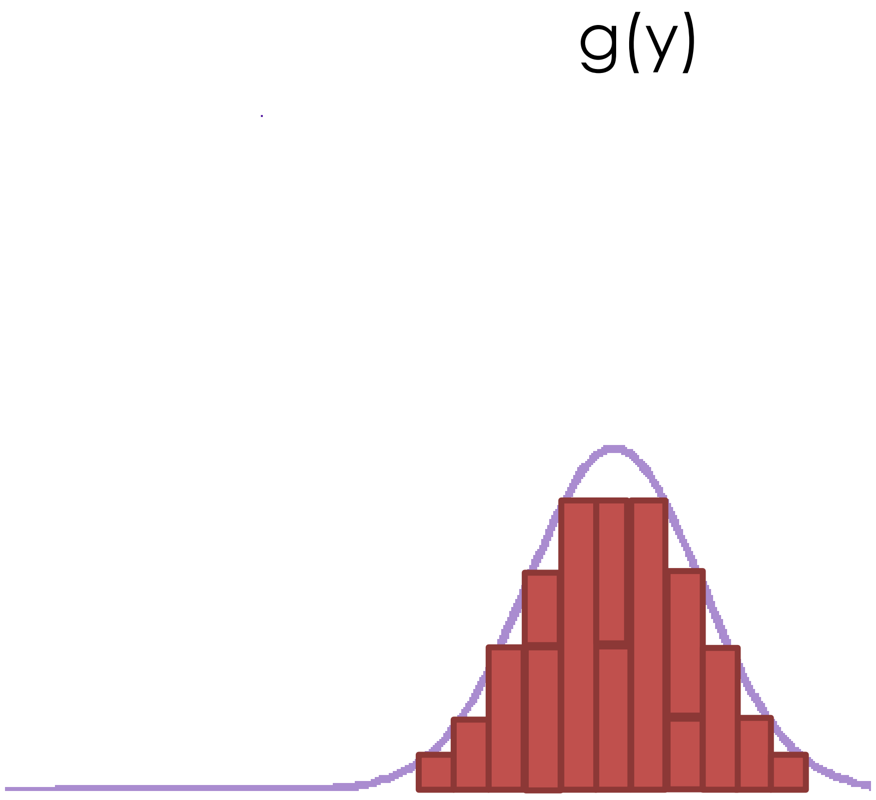

f(x)
g(y)
d(x,y)
x
y

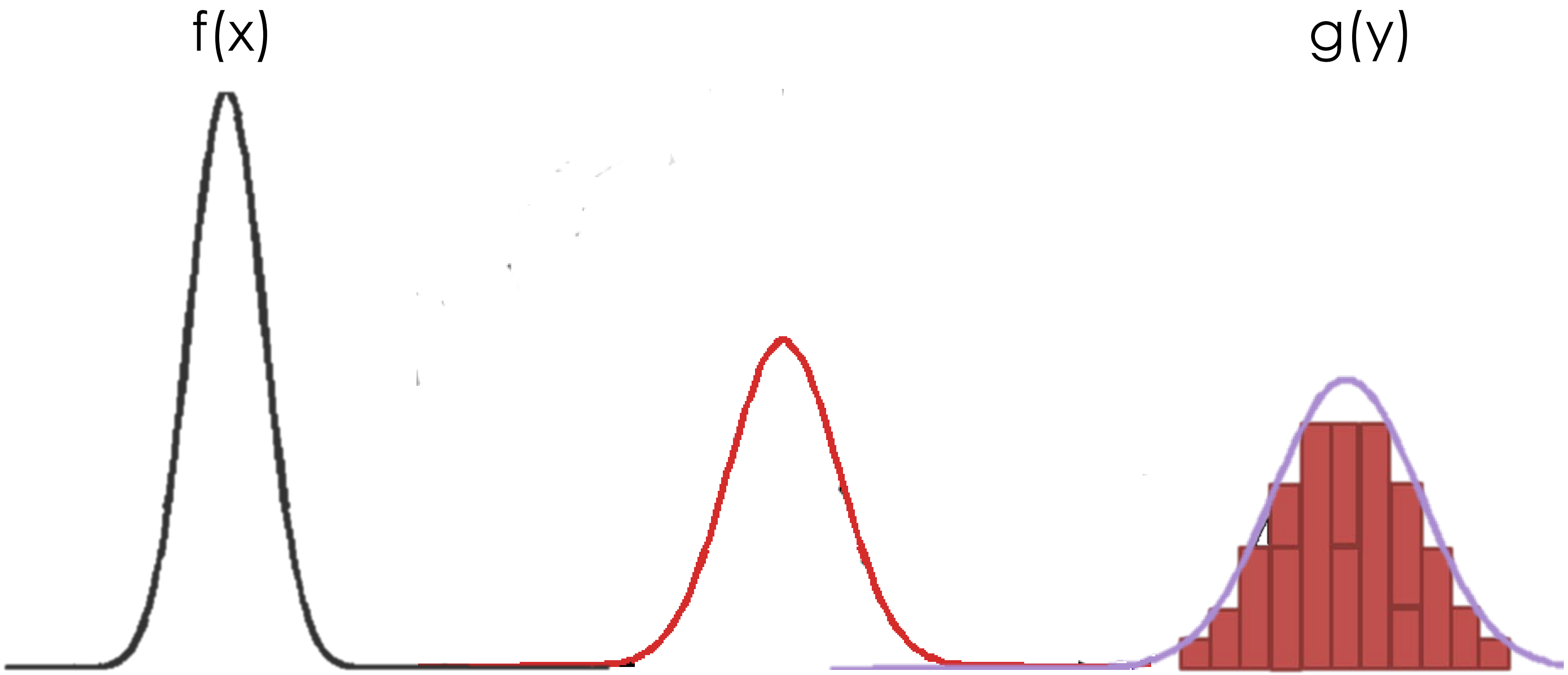

f(x)
g(y)
d(x1,y1)
d(x2,y2)
d(x3,y3)
x3
x2
x1
y3
y2
y1


f(x)
8
2
4 4
1

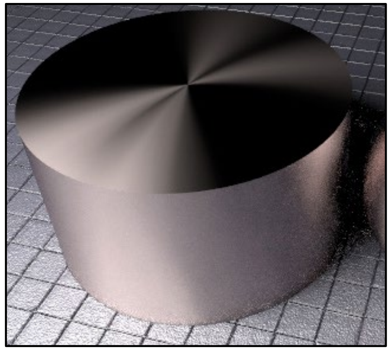

g(y)


f(x)
8
2
4 4
1

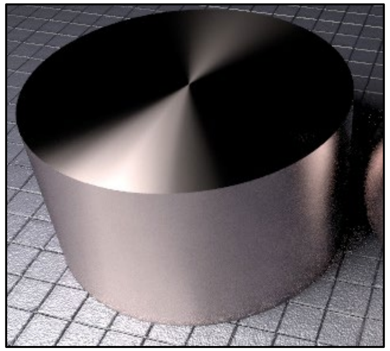

g(y)


f(x)
g(y)

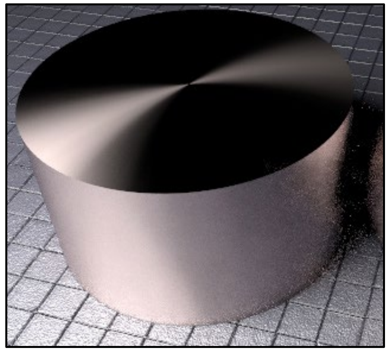
Application: BRDF



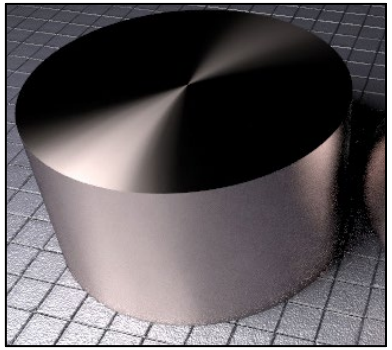
Linear interpolation



Function A



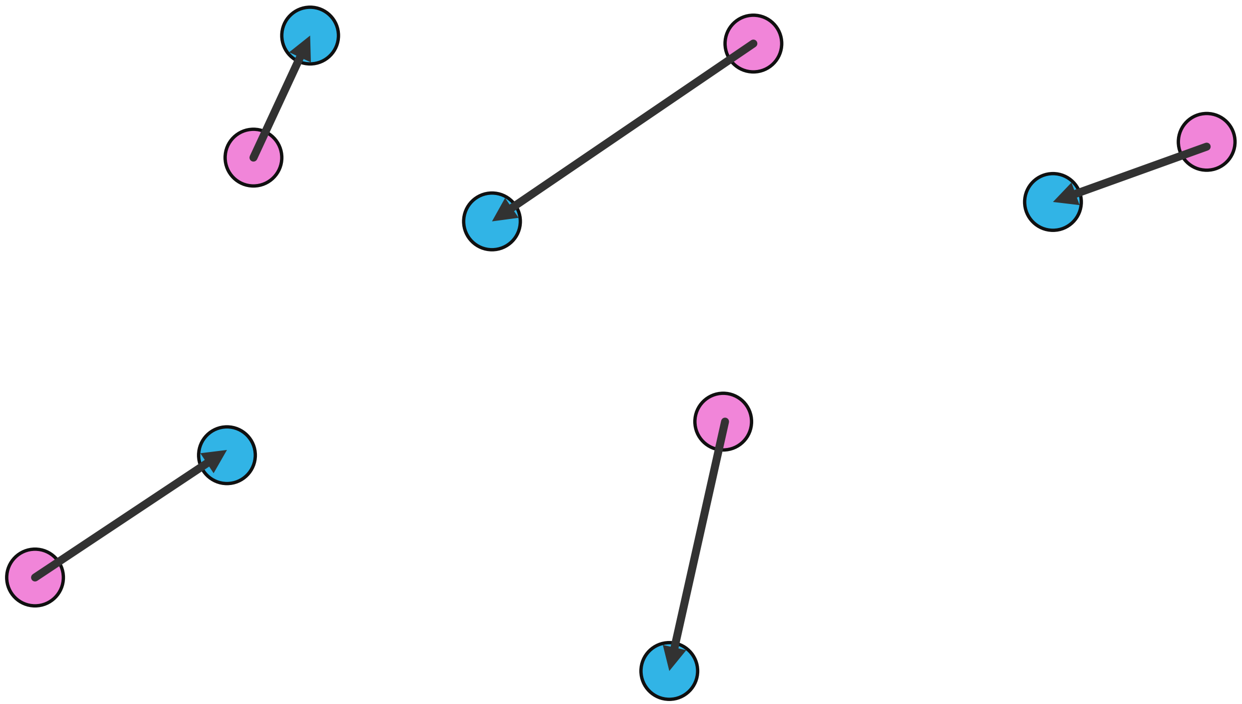
Function B



Displacement interpolation

[Bonneel et al. 2011] *Displacement Interpolation using Lagrangian Mass Transport*, Siggraph Asia

The assignment problem



Applications to Color Grading

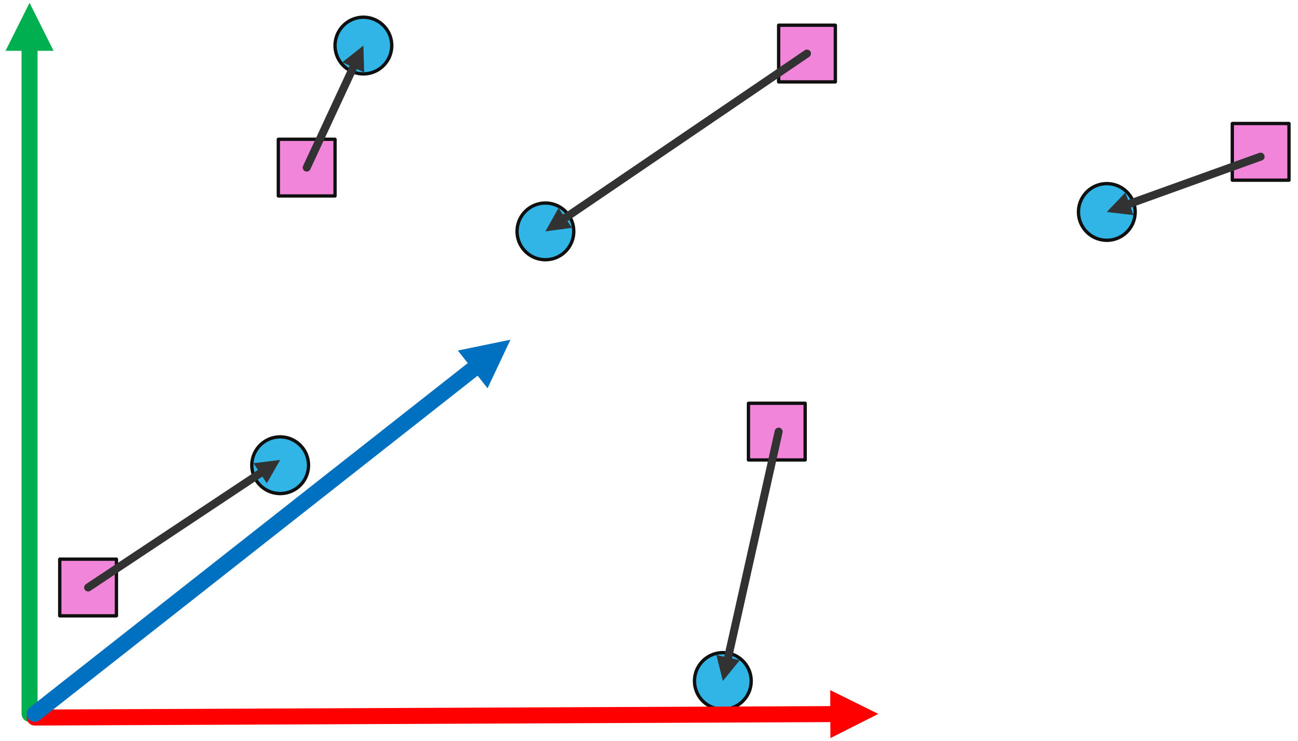
Input photo

Target style


Input photo


Target style

The assignment problem

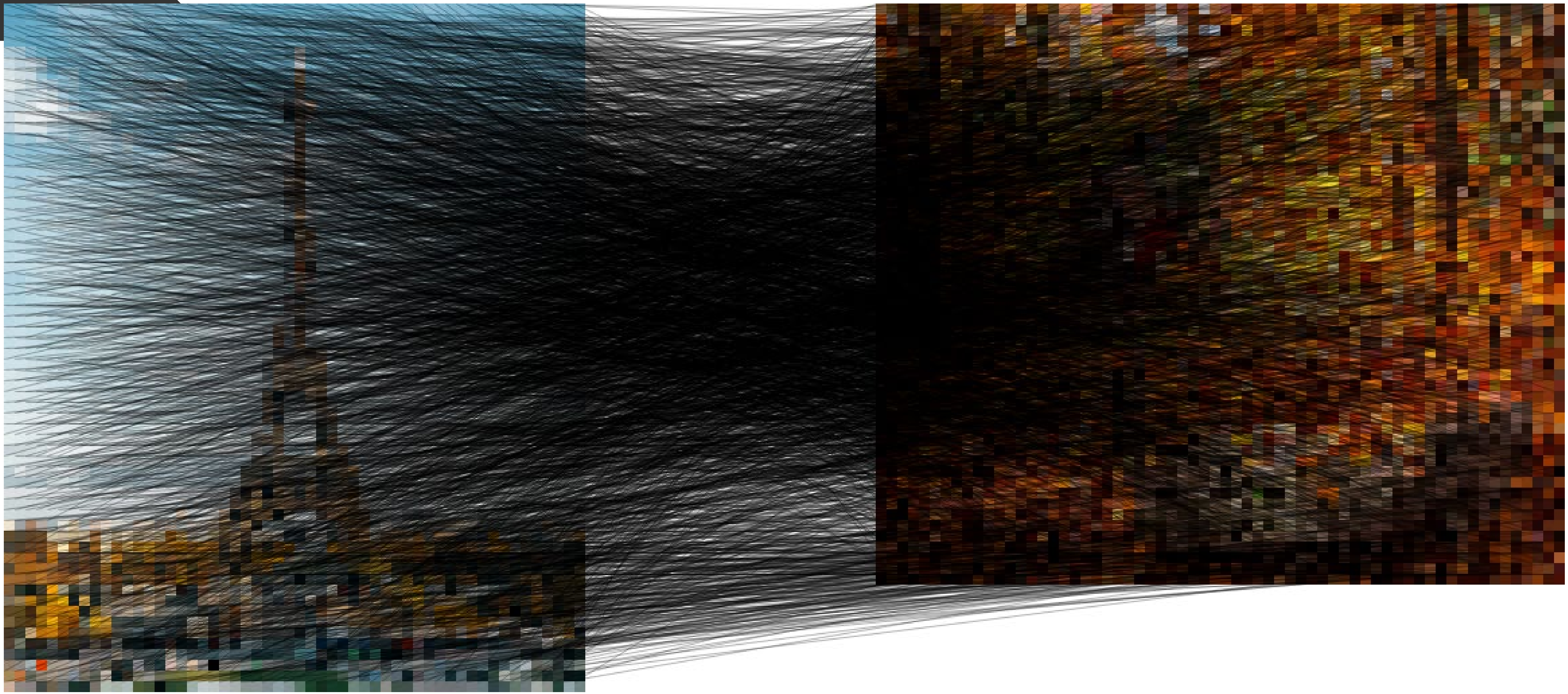


Pixels

Pixels

Input photo

Target style



Input photo

Target style

Input photo

Target style


WORMSLOE

Sliced and Radon Wasserstein Barycenters of Measures

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, Hanspeter Pfister

Journal of Mathematical Imaging and Vision (2014)

Simple cases

- Transport 1 Gaussian ↔ 1 Gaussian
- Transport = translation + scaling
- Transport 1D function ↔ 1D function

Optimal transport is simple for Gaussians

- Optimal transport and barycenters trivially solved for
  - Gaussian distributions with $c(x, y) = \|x - y\|^2$
    - $W_2^2(\mathcal{N}_0, \mathcal{N}_1) = \mathrm{tr}\left(\Sigma_0 + \Sigma_1 - 2\Sigma_{0,1}\right) + \|\mu_0 - \mu_1\|$ with $\Sigma_{0,1} = \left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{1/2}$
    - $T(x) = \Sigma_{0,1}x$
    - Barycenter: $\mathcal{N}(\mu, \Sigma)$ with $\mu = \sum_k \lambda_k \mu_k$ and iterations

$$\Sigma^{(n+1)} = \sum_k \lambda_k \left(\sqrt{\Sigma^{(n+1)}}\Sigma_k\sqrt{\Sigma^{(n+1)}}\right)^{1/2}$$

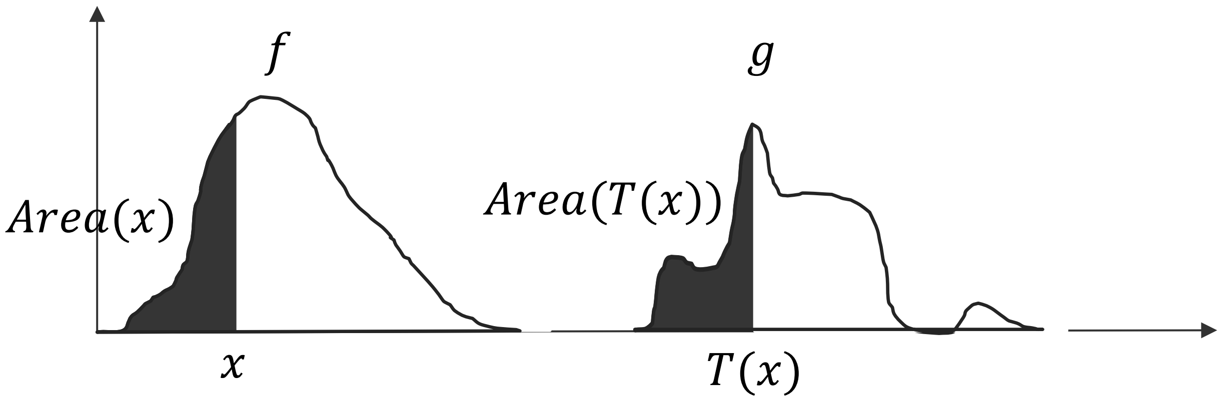
Optimal transport is simple in 1D

- Continuous case with density, convex cost, $\mu = f\,dx$, $\nu = g\,dy$
  - Need: $\int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{T(x)} g(x)dx$

$$T = G^{-1} \circ F$$

with $F(x) = \int_{-\infty}^{x} f(x)dx$ and $G(x) = \int_{-\infty}^{x} g(x)dx$

Generalize G^{-1} : $G^{-1}(y) = \min_x \{y = G(x)\}$ → Quantile function: e.g.: "what salary corresponds to the first percentile"



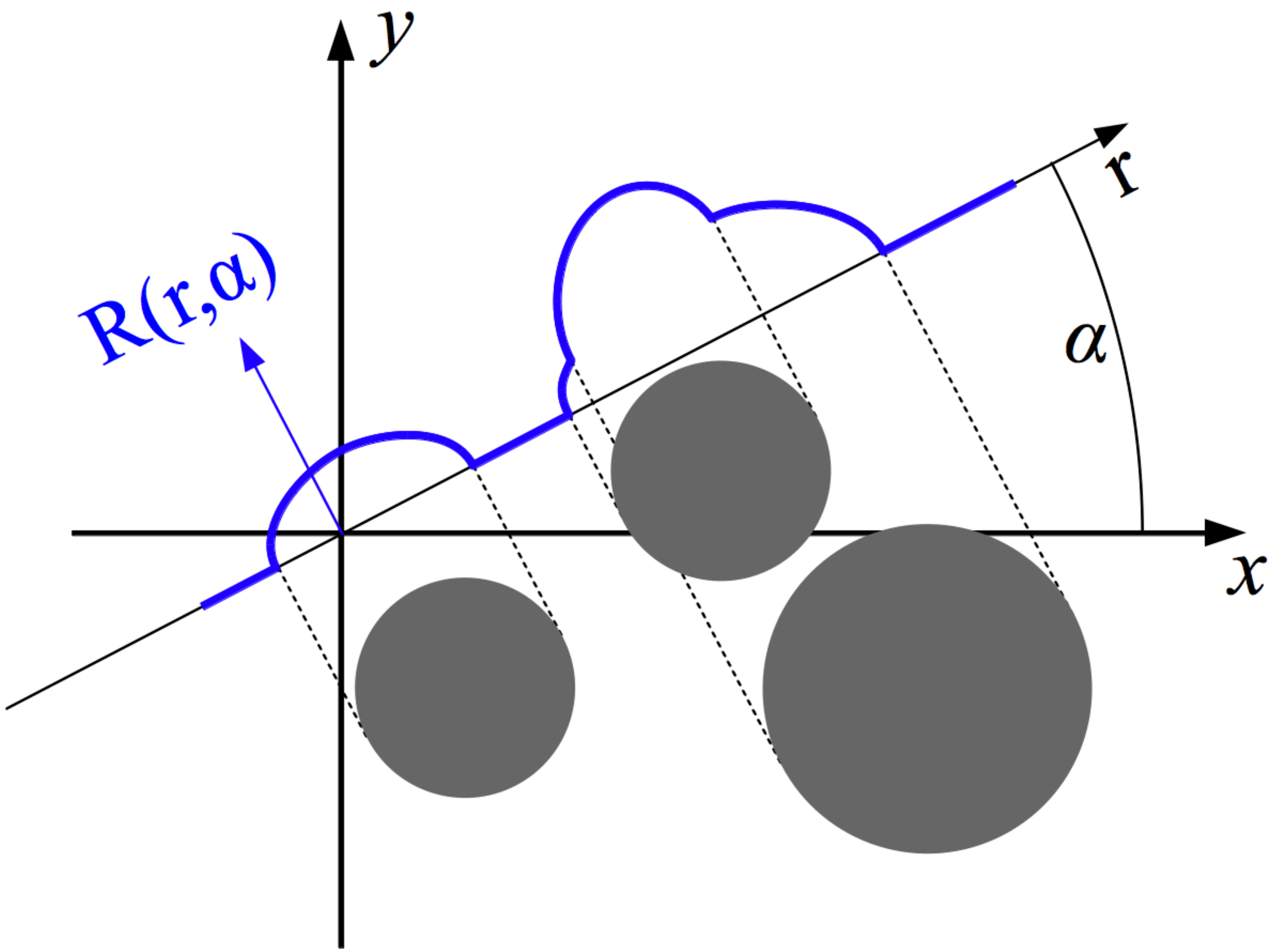
1D Case

OT Map: $T = G^{-1} \circ F$

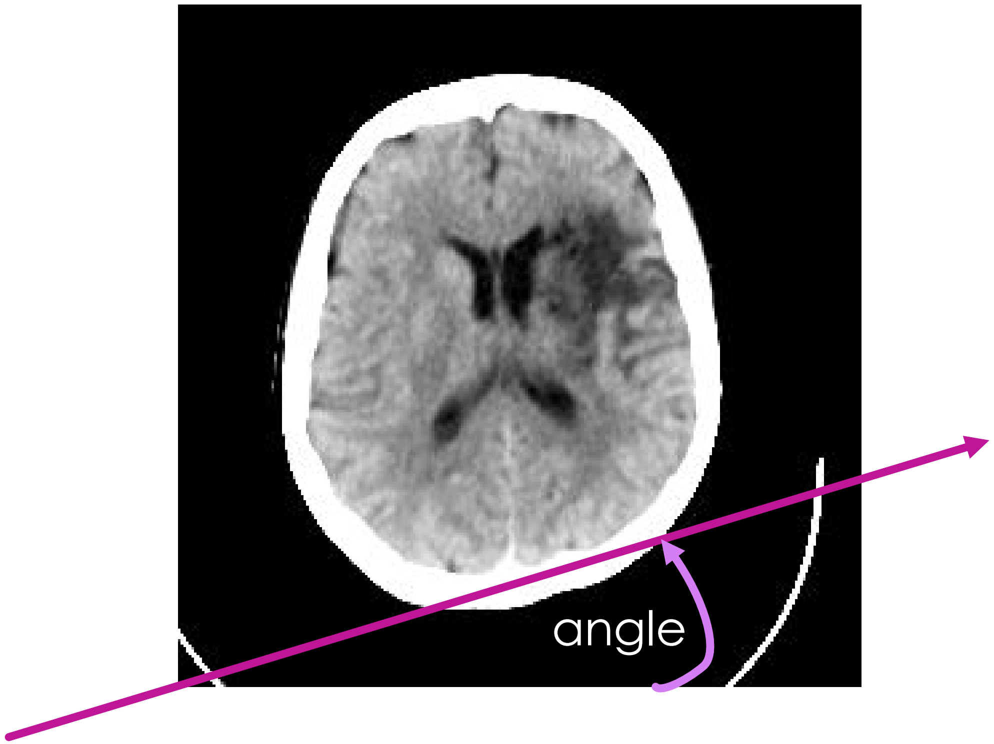
OT cost: $\int_0^1 c\big(F^{-1}(t) - G^{-1}(t)\big)dt$

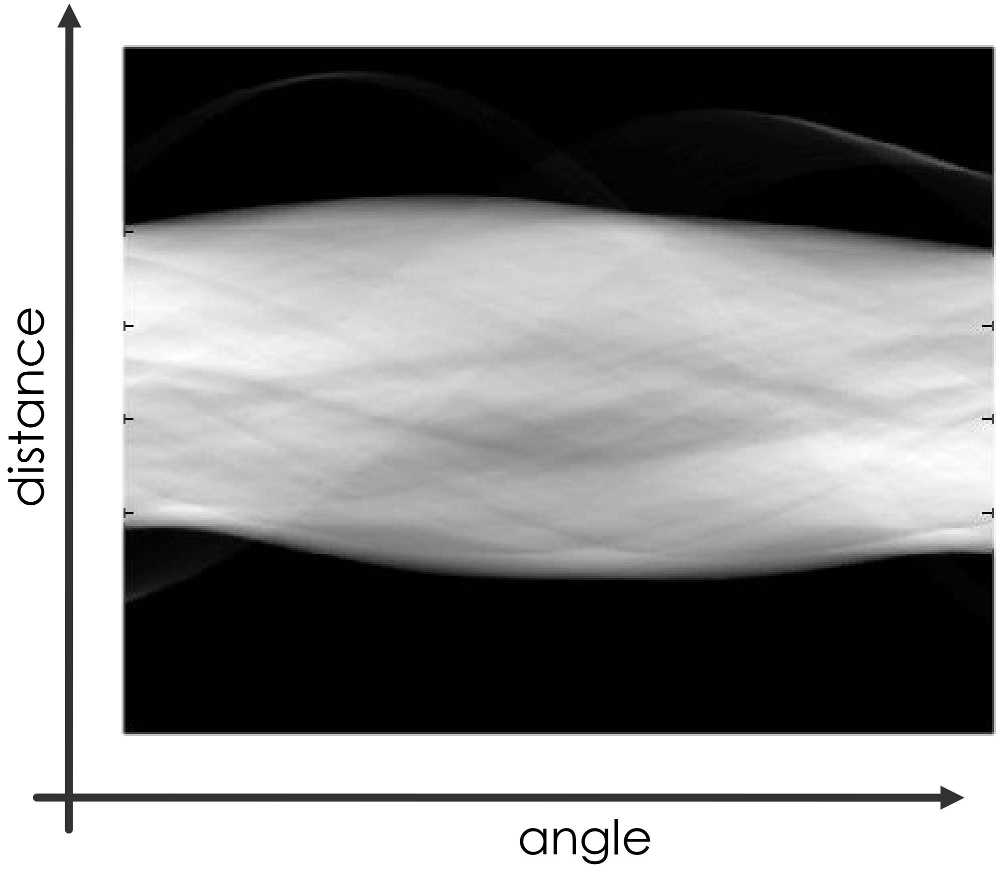
Interpolation: $F_{interp}^{-1}(x) = \sum_i \alpha_i \, F_i^{-1}(x)$

Radon transform

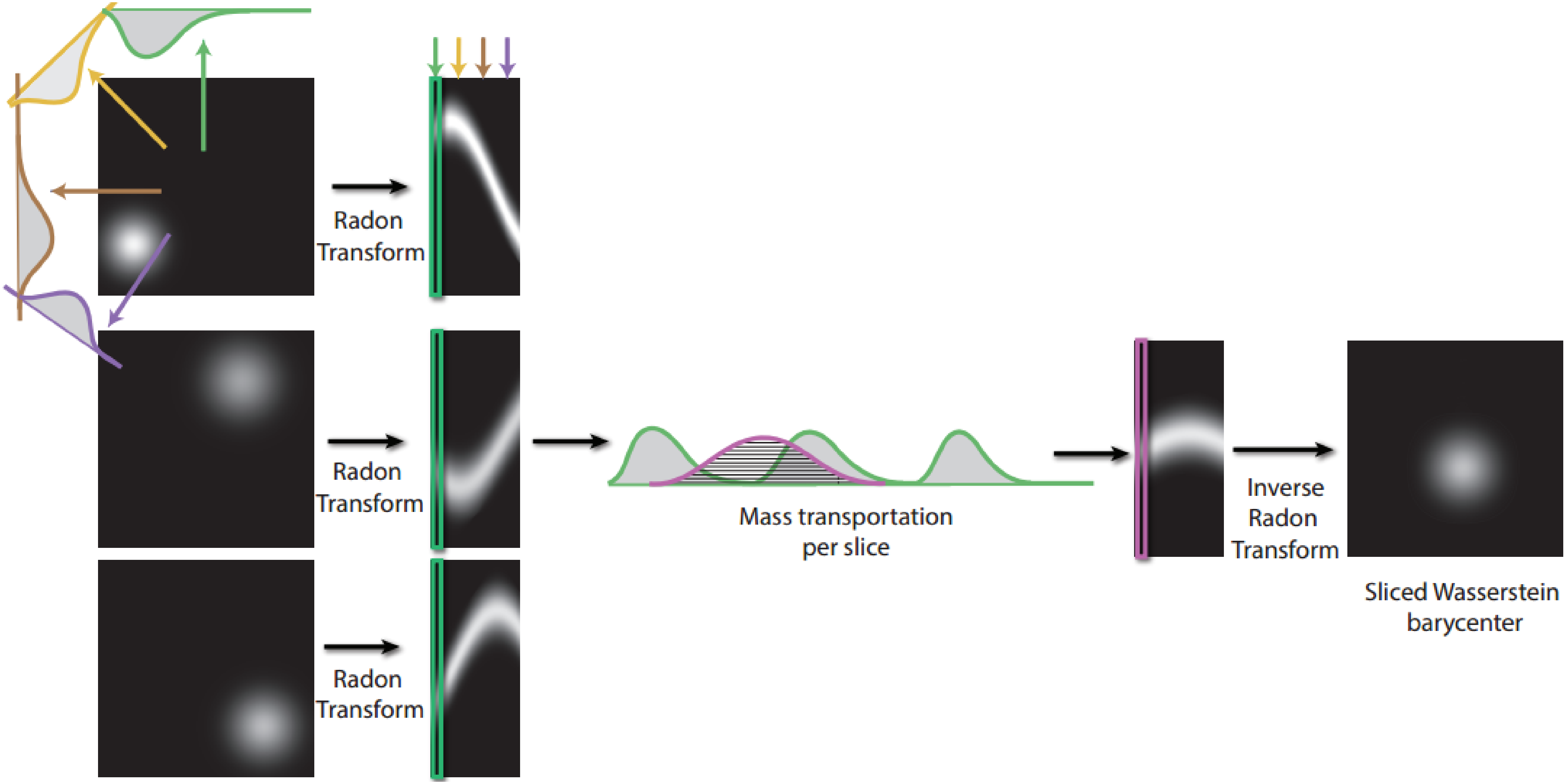


Radon transform

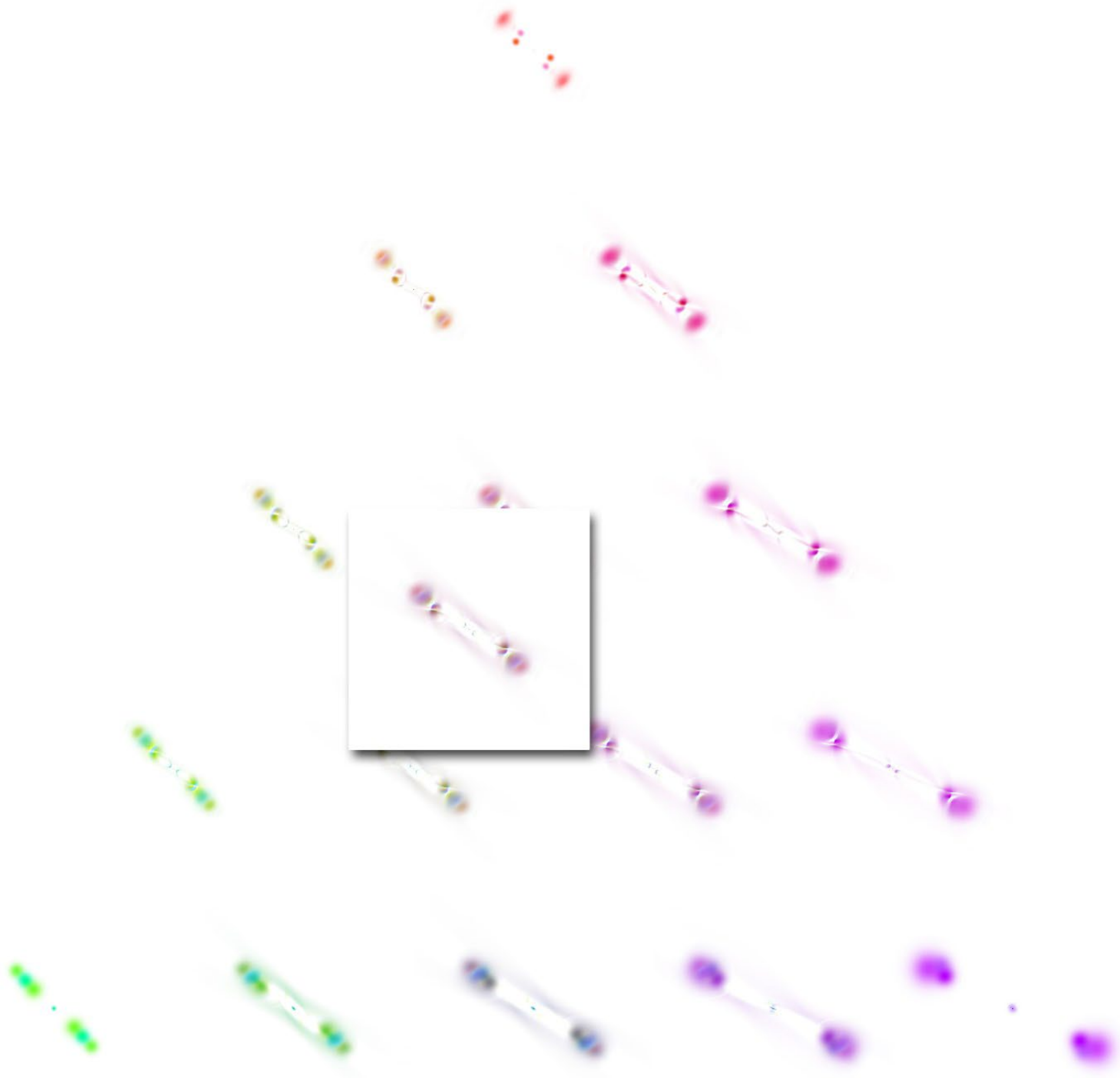
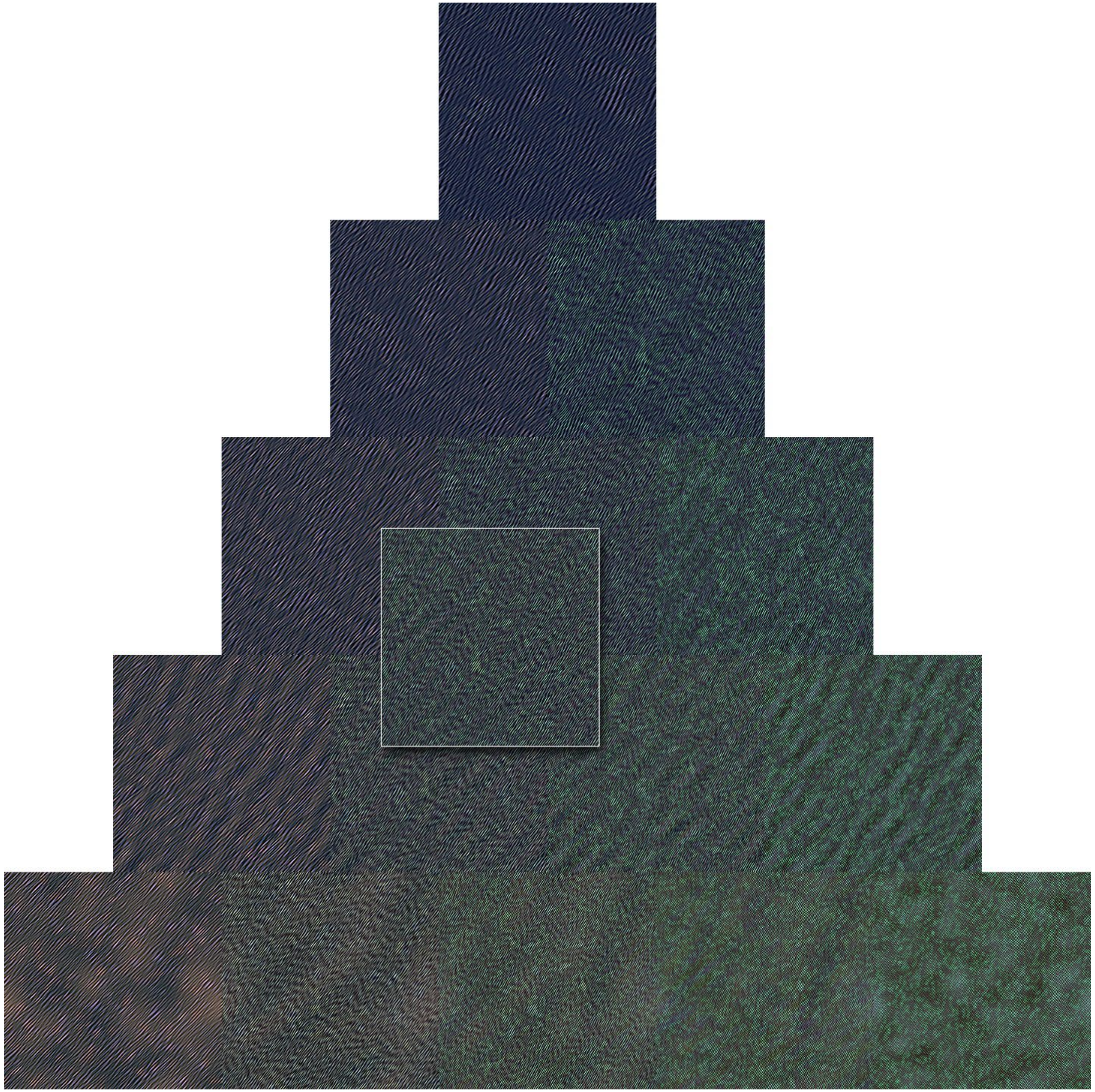




Method

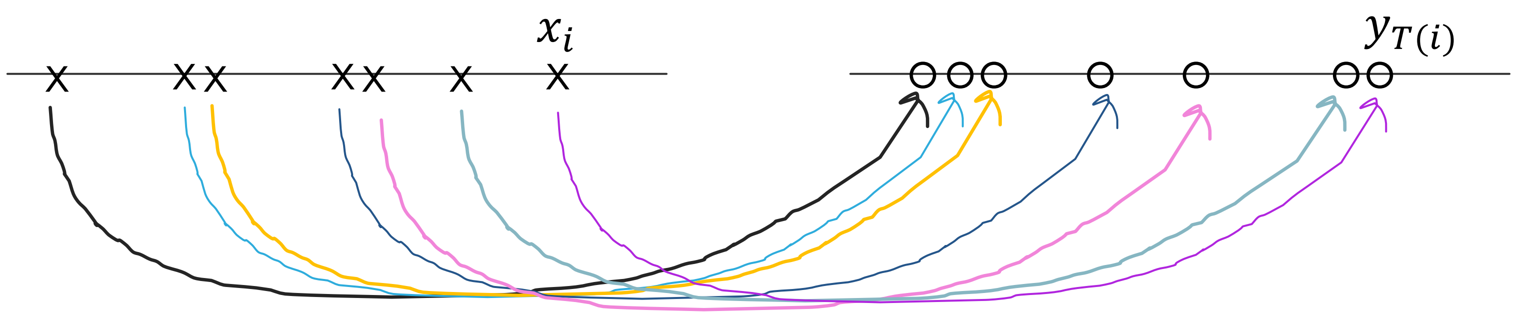


Radon
Transform
Radon
Transform
Radon
Transform
Mass transportation
per slice
Inverse
Radon
Transform
Sliced Wasserstein
barycenter

1D Case, discrete

- Discrete case, $\mu = \sum_{i=1}^{n} \delta_{x_i}$, $\nu = \sum_{i=1}^{n} \delta_{y_i}$ (same for interpolating between more than 2 measures)
- Optimal transport for convex cost = pairing sorted samples



Sliced Wasserstein Distance

- For discrete high-dimensional distributions $\mu = \sum_{i=1}^{n} \delta_{x_i}$ and $\nu = \sum_{i=1}^{n} \delta_{y_i}$

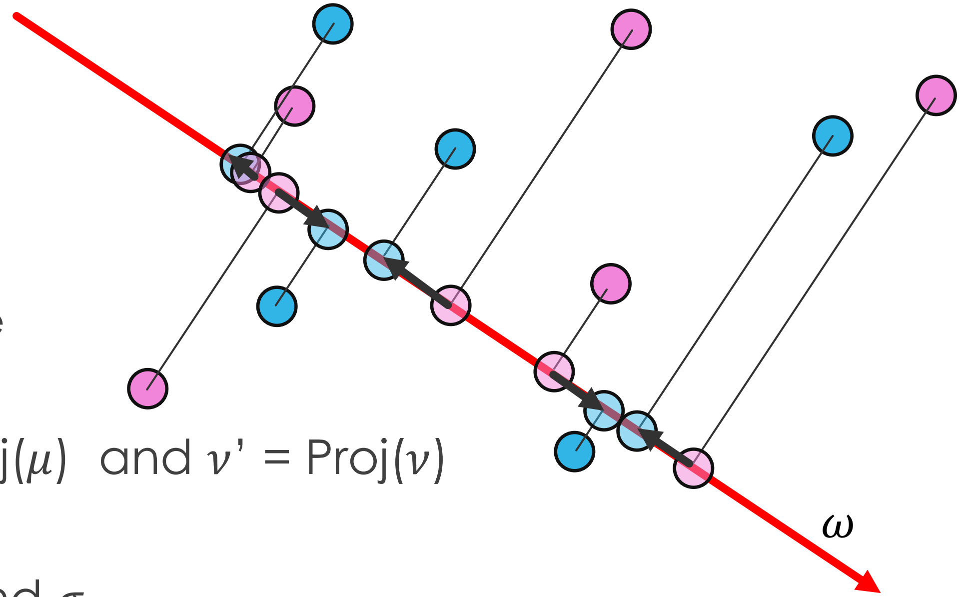
Consider energy

$$SW(\mu,\nu) = \int_S W_2^2\big(\mathrm{proj}(\mu,\omega), \mathrm{proj}(\nu,\omega)\big)\, \mathrm{d}\omega$$

Where $\mathrm{proj}(\mu,\omega)$ is the 1-d distribution : $\mathrm{proj}(\mu,\omega) = \sum_i \delta_{\langle x_i,\omega\rangle}$ (same for ν)

And W_2^2 computes the 1-d squared Wasserstein distance

Sliced Wasserstein Distance



- Take a uniform random direction ω
  - $\omega \leftarrow (\mathcal{N}(0,1), \mathcal{N}(0,1), \mathcal{N}(0,1))$ and normalize
- Project samples of μ and ν on ω : μ' = Proj(μ) and ν' = Proj(ν)
- Sort μ' and ν', i.e, find permutations σ_μ and σ_ν
- To compute the Sliced Wasserstein Distance:

$$d^2 \leftarrow d^2 + \sum_i \left| \left\langle x_{\sigma_\mu(i)}, \omega \right\rangle - \langle y_{\sigma_\nu(i)}, \omega \rangle \right|^2$$

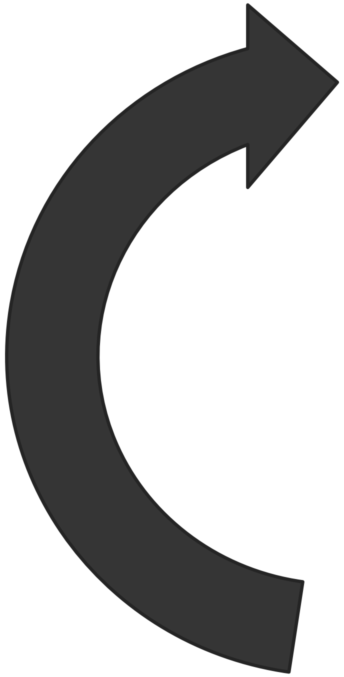
Wasserstein Barycenter and its Application to Texture Mixing. [Rabin et al. 2011]

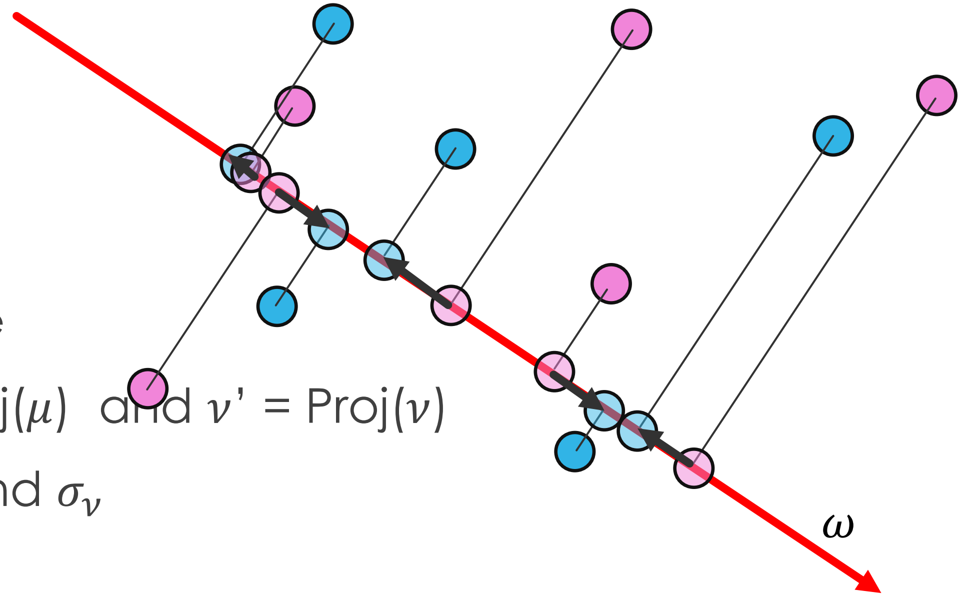
Gradient flow

- Given an energy (e.g., the Sliced Wasserstein Distance), you may try to minimize it with gradient descent
- Each step of such a gradient descent is a "gradient flow"

- This "moves" the input distribution towards the target (the target is such that the energy is 0)

Sliced Wasserstein Gradient Flow





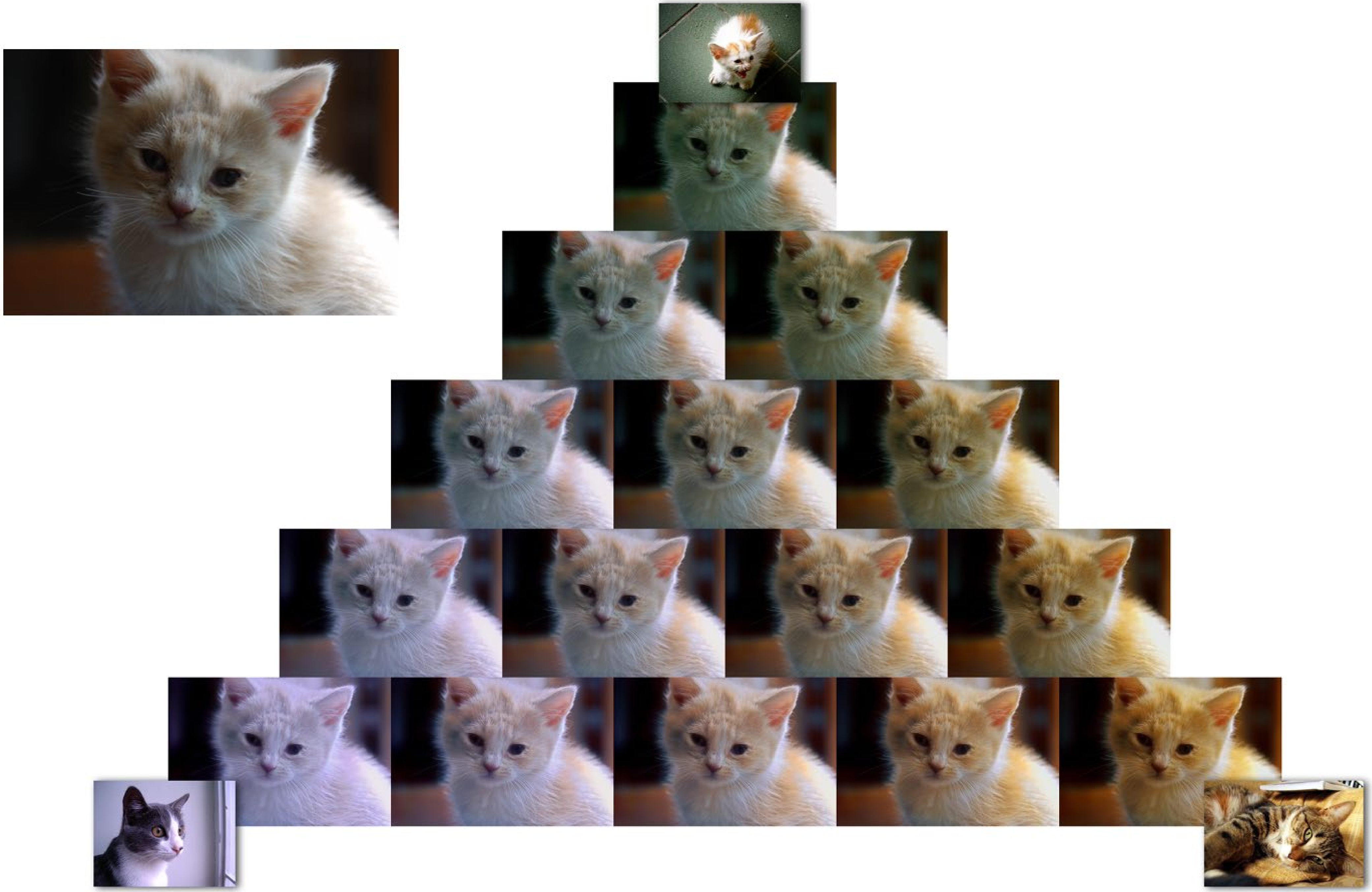
- Take a uniform random direction ω
  - $\omega \leftarrow (\mathcal{N}(0,1), \mathcal{N}(0,1), \mathcal{N}(0,1))$ and normalize
- Project samples of μ and ν on ω : μ' = Proj(μ) and ν' = Proj(ν)
- Sort μ' and ν', i.e, find permutations σ_μ and σ_ν
- Update μ by $x_{\sigma_\mu(i)} \leftarrow x_{\sigma_\mu(i)} + \left(\langle y_{\sigma_\nu(i)}, \omega\rangle - \left\langle x_{\sigma_\mu(i)}, \omega\right\rangle\right) \omega$
  - Corresponds to moving particles in the direction ω

Wasserstein Barycenter and its Application to Texture Mixing. [Rabin et al. 2011]

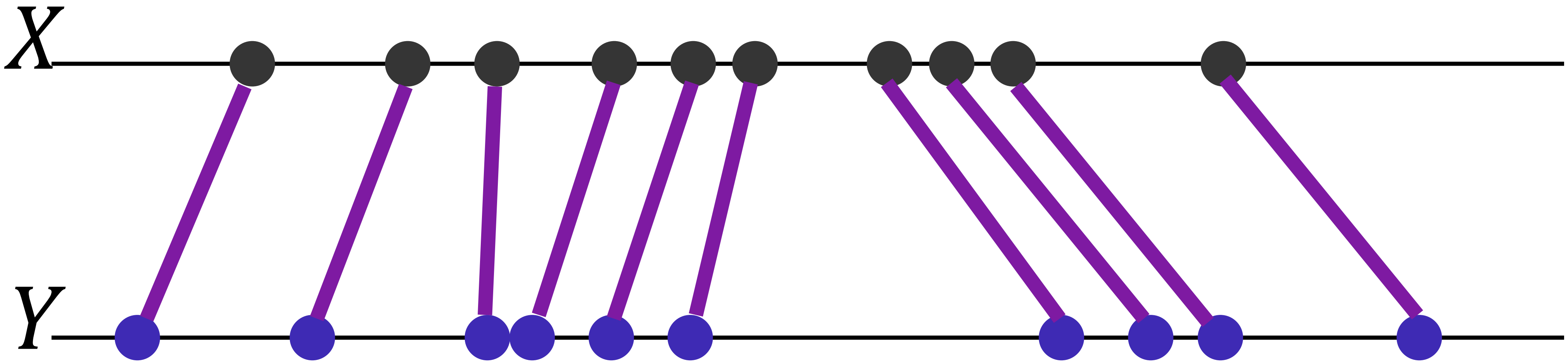
Sliced Wasserstein Distance

f

g



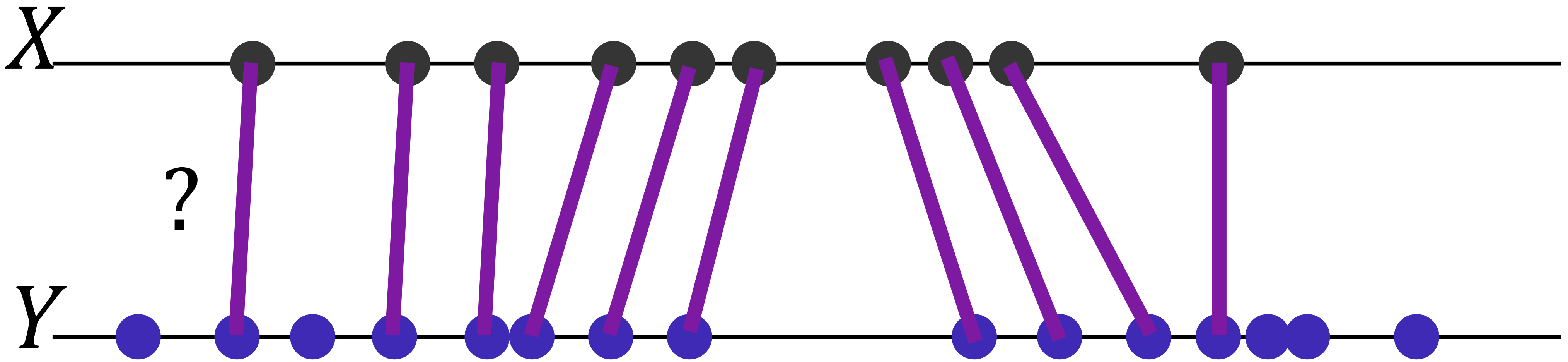
1-d Linear Assignment Problem is trivial*



*assuming the cost c is a convex function of |x-y|

Partial optimal assignment ?

=> Sliced Partial Optimal Transport, [Bonneel and Coeurjolly 2019]

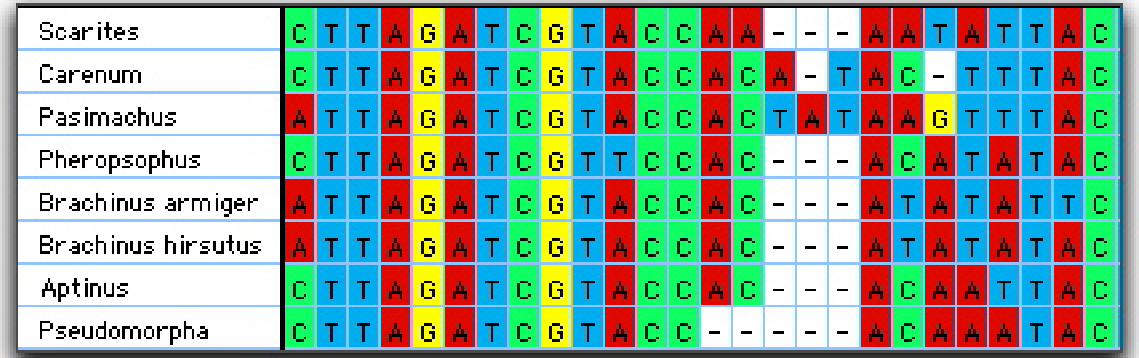


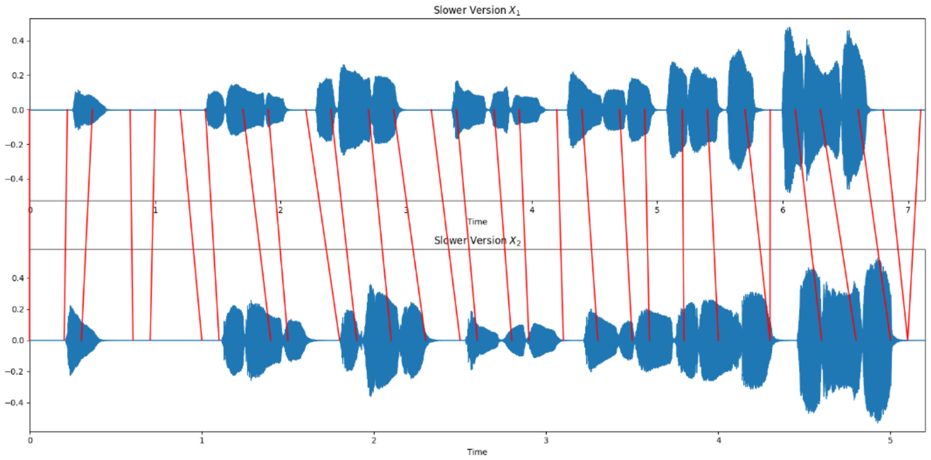
$$W(f,g) = \min \sum_{i,j} c_{i,j}\pi_{i,j} \quad \text{s.t.} \quad \sum_j \pi_{i,j} = 1, \quad \sum_i \pi_{i,j} \leq 1, \quad \pi_{i,j} \geq 0$$

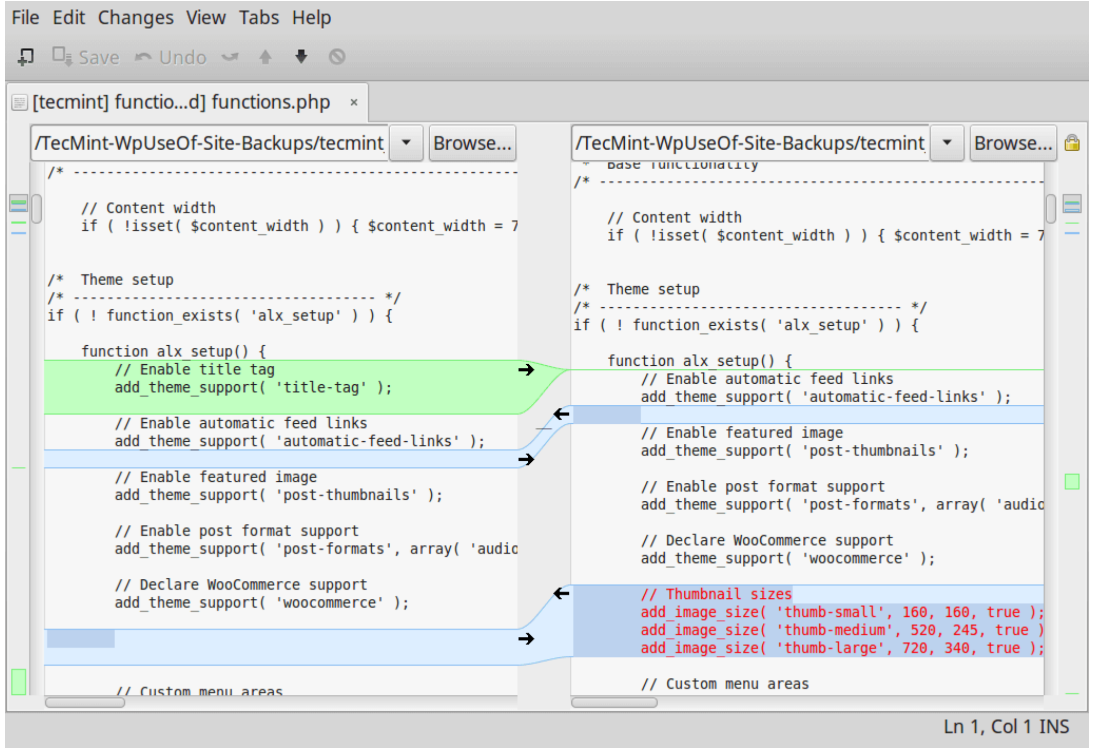
$$\min_{T \text{ injective}} \sum_i c(x_i, y_{T(i)})$$

Similar problems

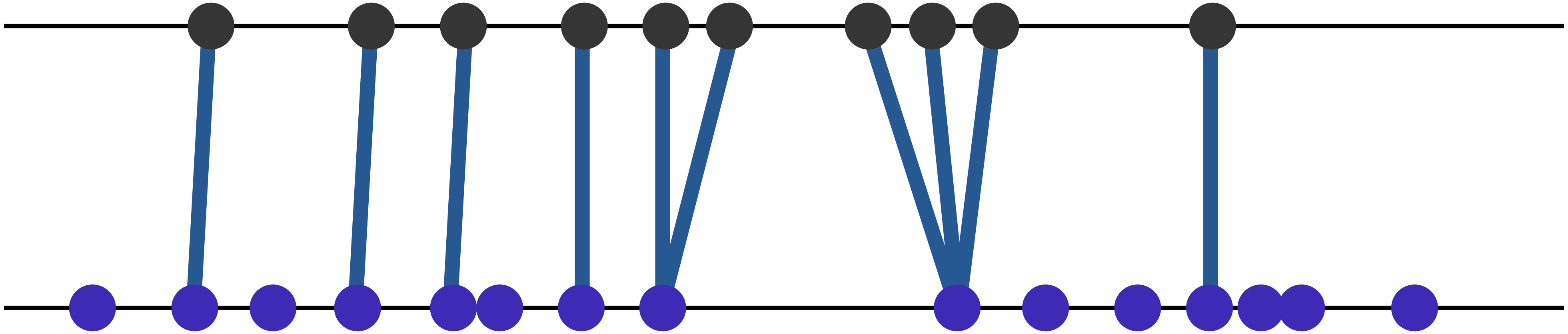
- DNA sequence alignment
- Text alignment
- Music synchronization
- …





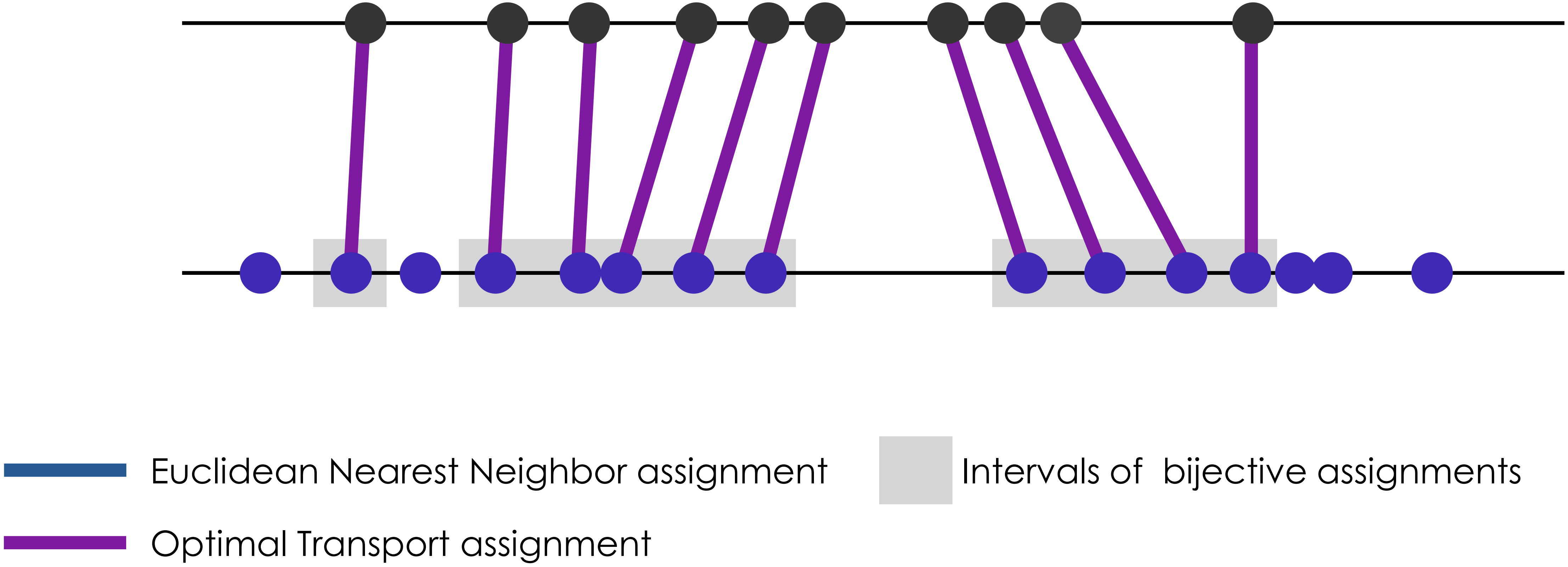


Quadratic time complexity algorithm (linear space)



Euclidean Nearest Neighbor assignment

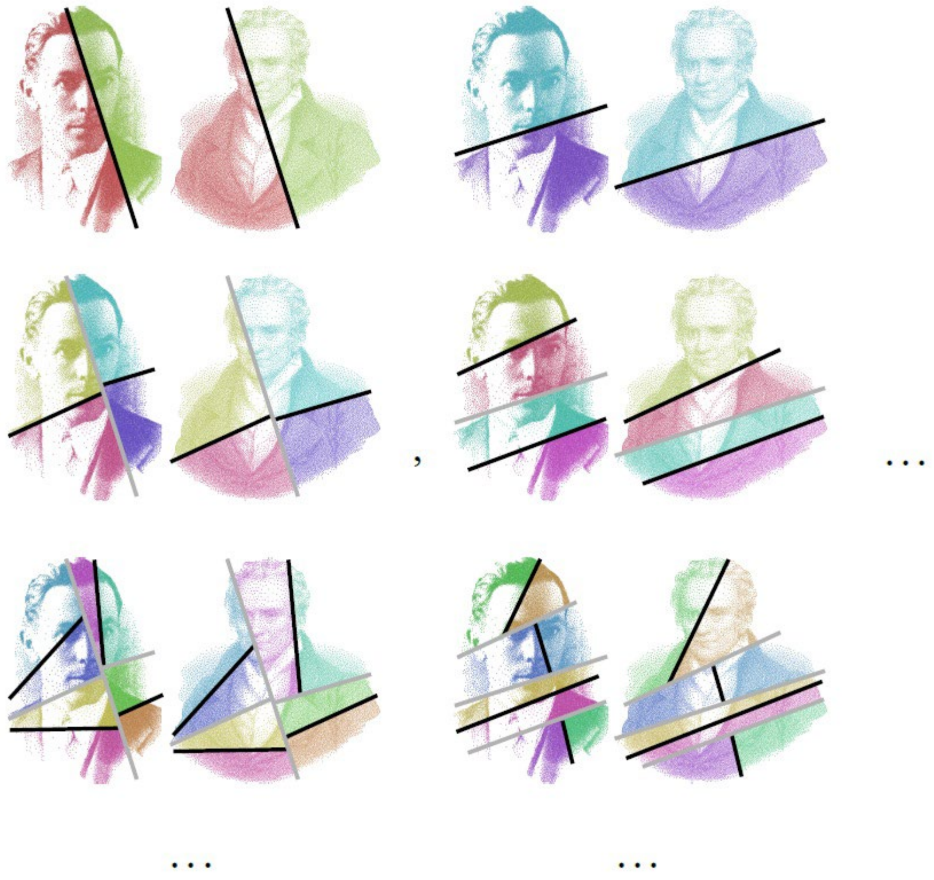
Quadratic time complexity algorithm (linear space)



BSP OT

=> BSP-OT: Sparse transport plans between discrete measures in loglinear time, [Genest et al. 2025]

- Replaces the sorting step by a variant of QuickSort
- Each partition step of QuickSort splits the two input measures in two parts along one random direction
  - Each side of both point clouds has the same number of points
  - Results in a BSP tree



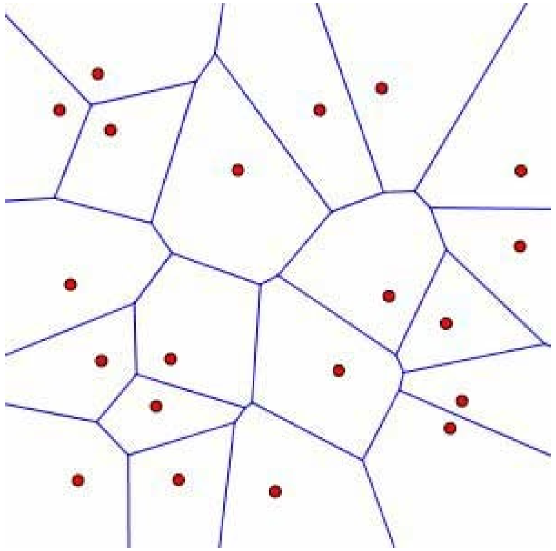
Semi-discrete optimal transport

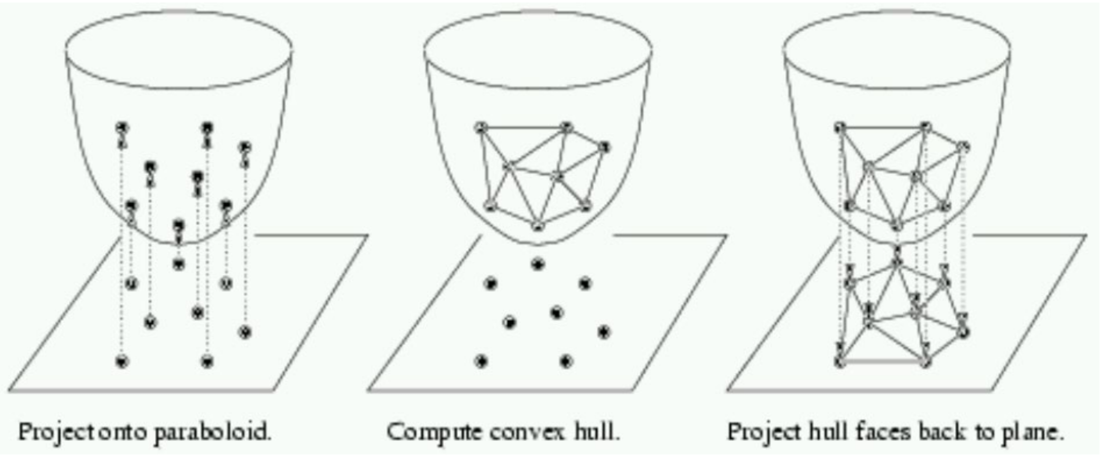
Voronoi diagram

- A partition such that each point x is assigned to its closest site x_i

$$\|x - x_i\|^2 \leq \|x - x_j\|^2 \quad \forall j$$

- The dual of a Delaunay triangulation: a triangulation of the sites such that no other site is encompassed by the circumcircle of a triangle
  - Also: convex hull of a parabolic lifting





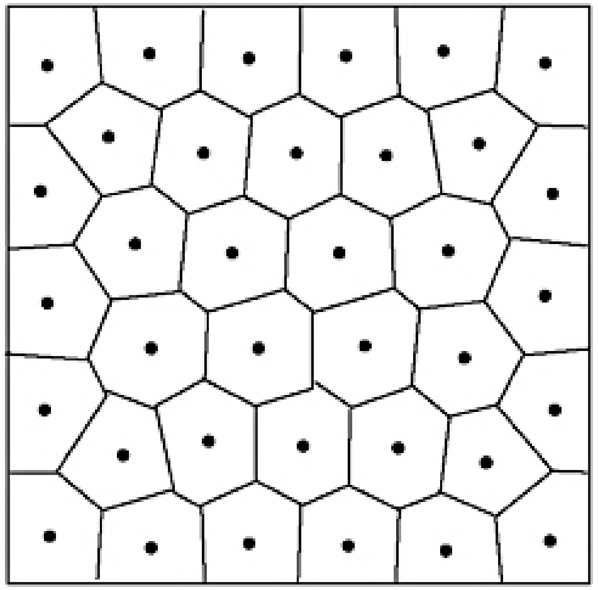
Centroidal Voronoi Diagram

- Can be defined as the solution to a least-square problem

$$\min \int_{Vor_i} \sum_i \|x - x_i\|^2 dx$$

Also says that the centroid of Vor_i is the site x_i

- Can be computed by:
  - A Lloyd clustering algorithm
  - A descent approach on the above energy

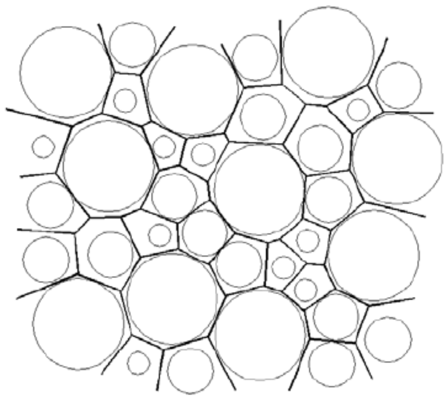


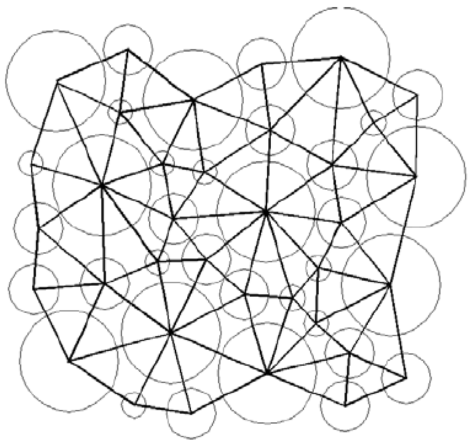
Power diagram (Laguerre diagram)

- A partition s.t. each point x is assigned to its closest site x_i with weight w_i

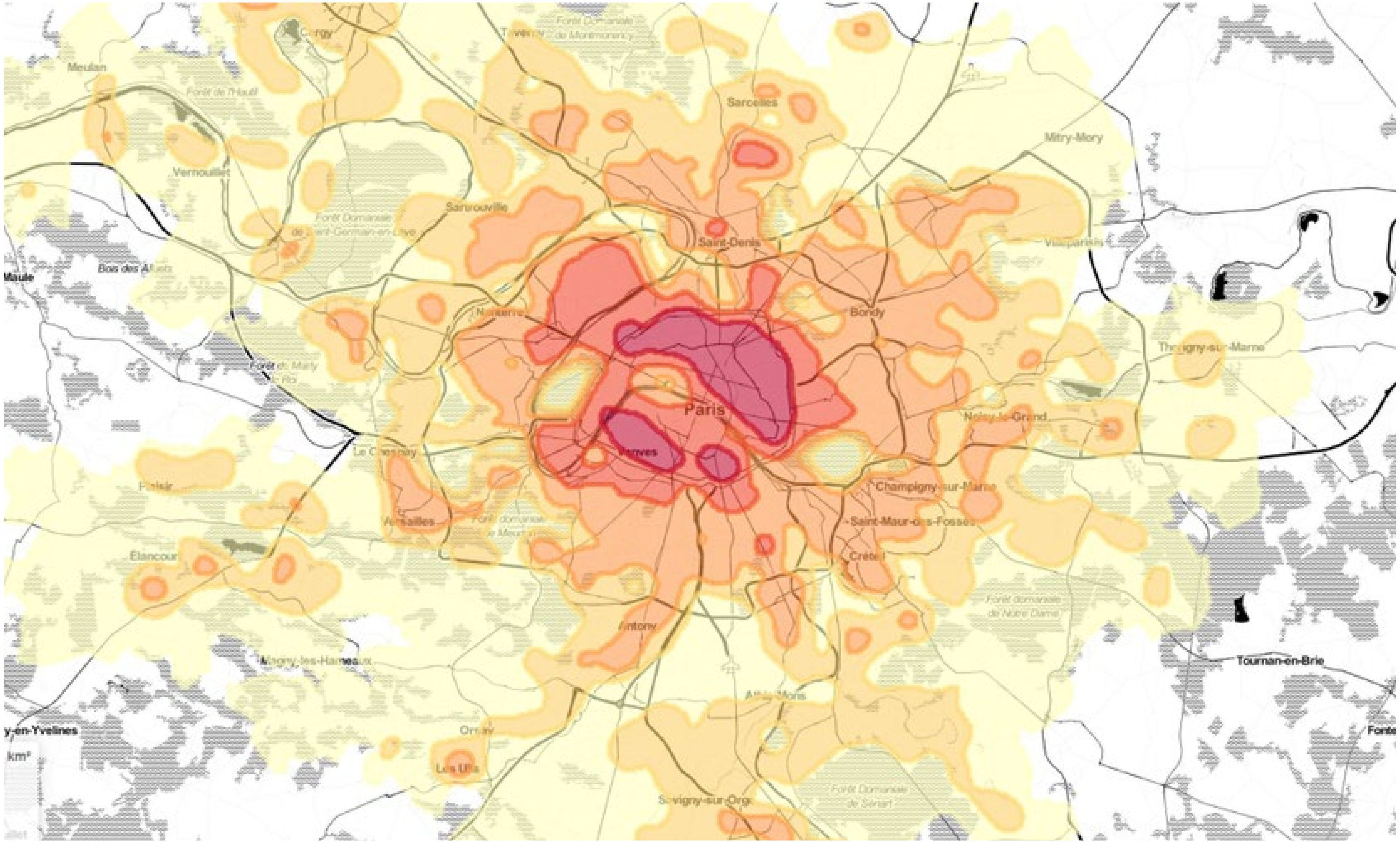
$$\|x - x_i\|^2 - w_i \leq \|x - x_j\|^2 - w_j \quad \forall j$$

- Can be computed by lifting a Voronoi diagram
  - Consider site coordinates $x_i' = \left(x_i \,;\, \sqrt{c - w_i} \right)$ for large constant c ; $x' = (x\,;0)$
  - Then $\|x' - x_i'\|^2 \leq \|x' - x_j'\|^2 \;\; \forall j$
- Any partition into convex polyhedral cells is a power diagram of some sites



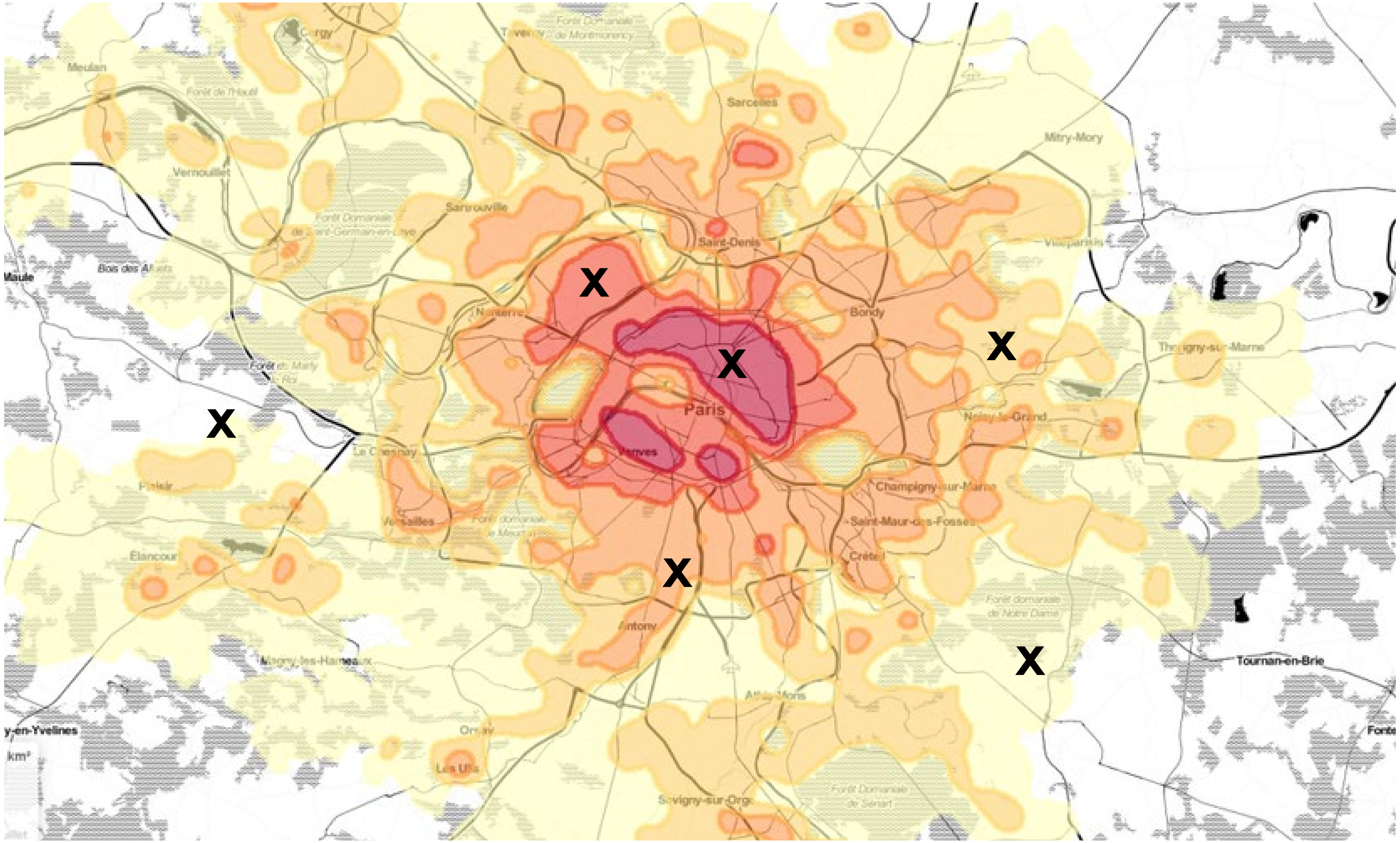


Semi-discrete Optimal Transport



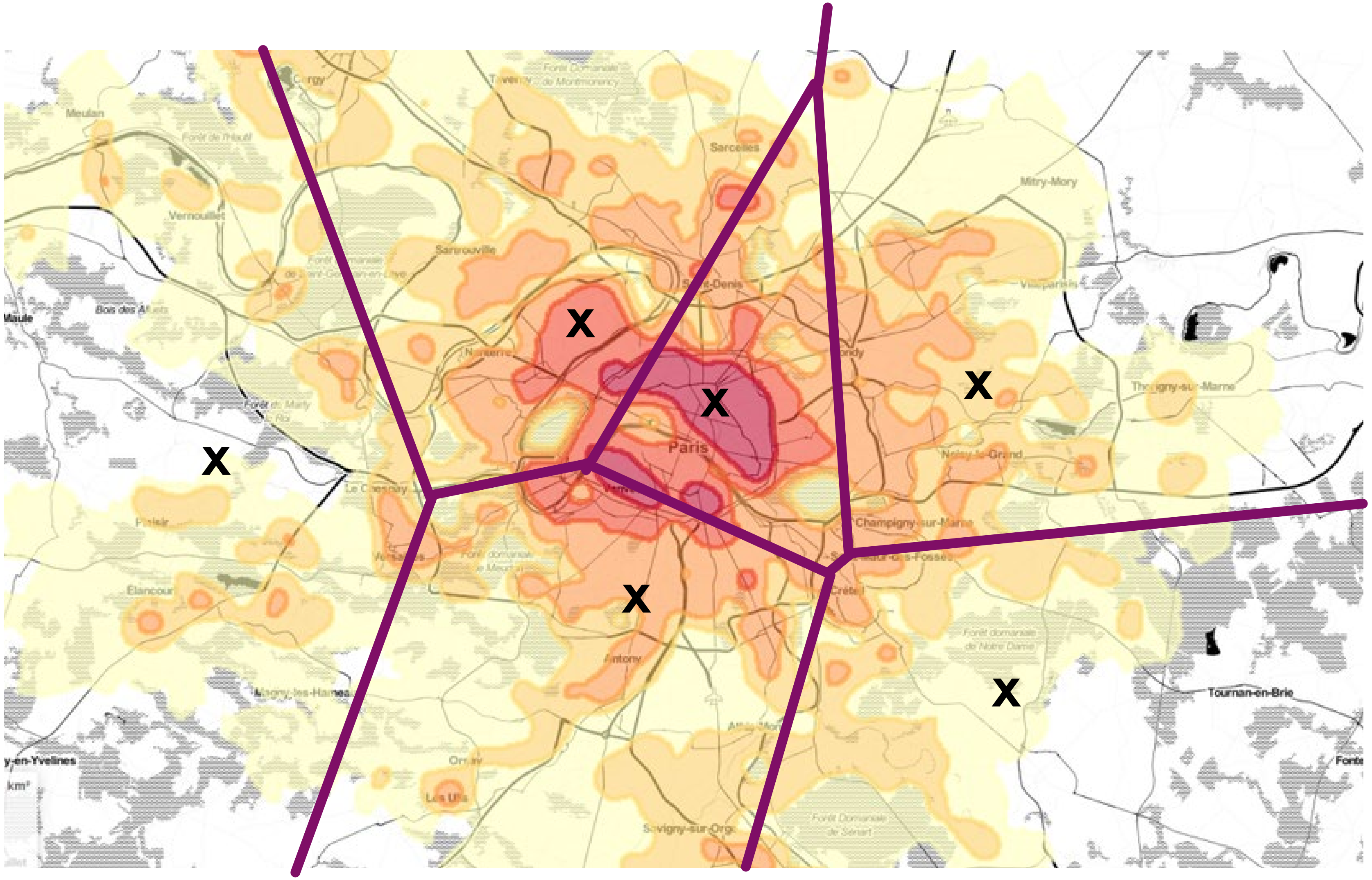
Population density f

Semi-discrete Optimal Transport



Set of bakeries, factories, …?

Semi-discrete Optimal Transport



No constraint on production: population go to their nearest bakery/factory/… regardless of populat

Semi-discrete Optimal Transport

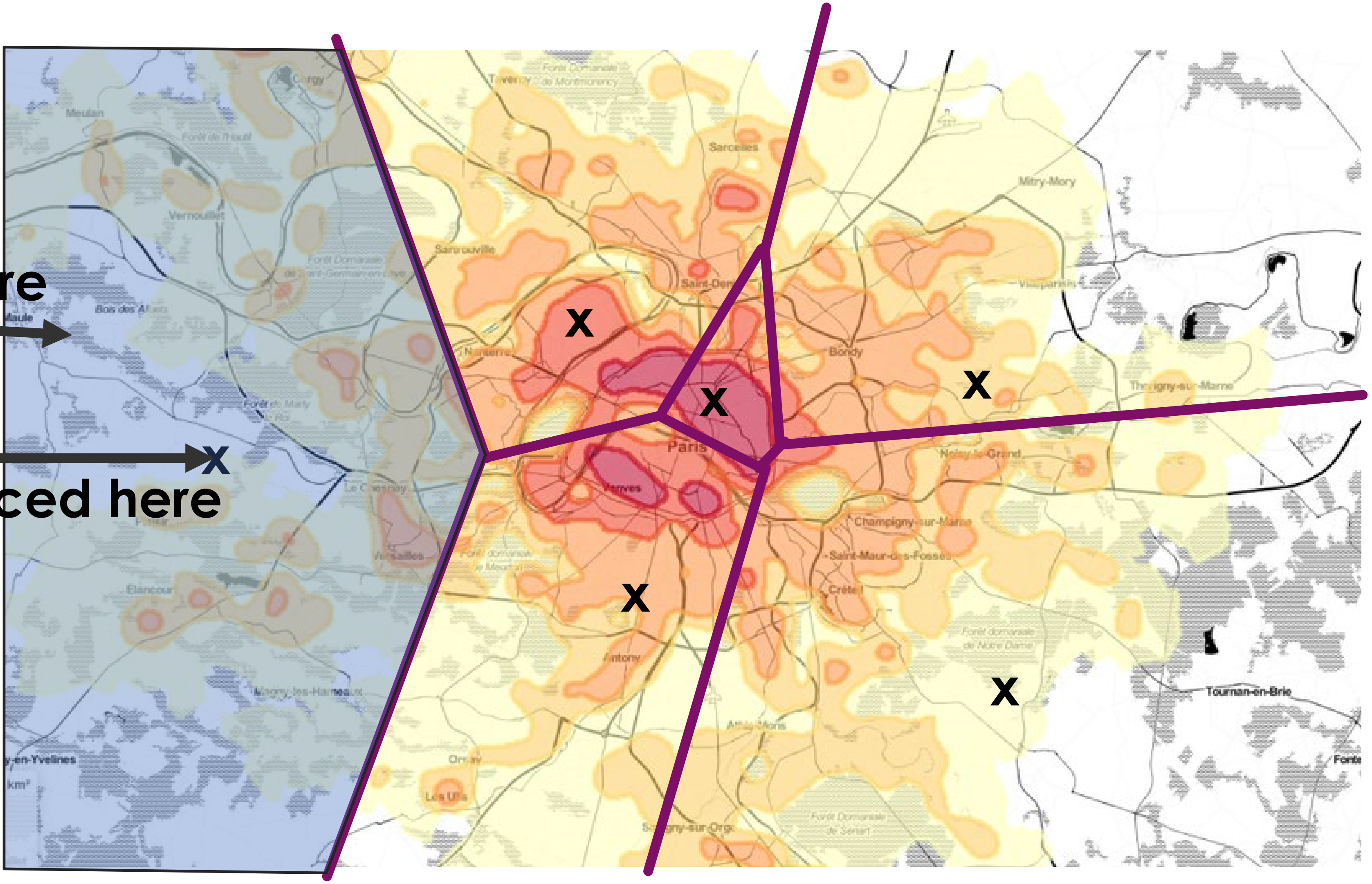
Limited production: population go to the nearest bakery/factory **with sufficient production**!

Semi-discrete Optimal Transport

(needs for)
population here

=

quantity produced here



Limited production: population go to the nearest bakery/factory **with sufficient production**!

Back to optimal transport

- Optimal transport (Monge version) :

$$\min \int \|x - T(x)\|^2 \, d\mu(x)$$

Considering μ is continuous with density ρ

$$\min \int \|x - T(x)\|^2 \, \rho(x) dx$$

Considering ν (the target measure) discrete: $\nu = \sum \lambda_p \delta_p$

The mass preservation constraint is:

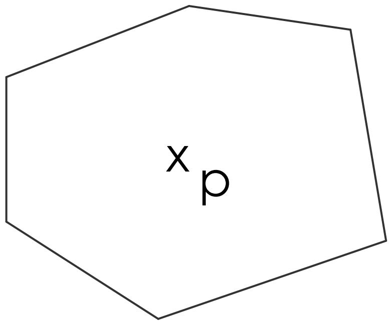
$$\lambda_p = \int_{T^{-1}(\{p\})} \rho(x) dx$$

A Multiscale Approach to Optimal Transport [Mérigot 2011]
Minkowski-Type Theorems and Least-Squares Clustering [Aurenhammer et al. 98]

Back to optimal transport

- In this case : $T^{-1}(\{p\}) = Vor^W(p)$

a power cell for some weight w_p



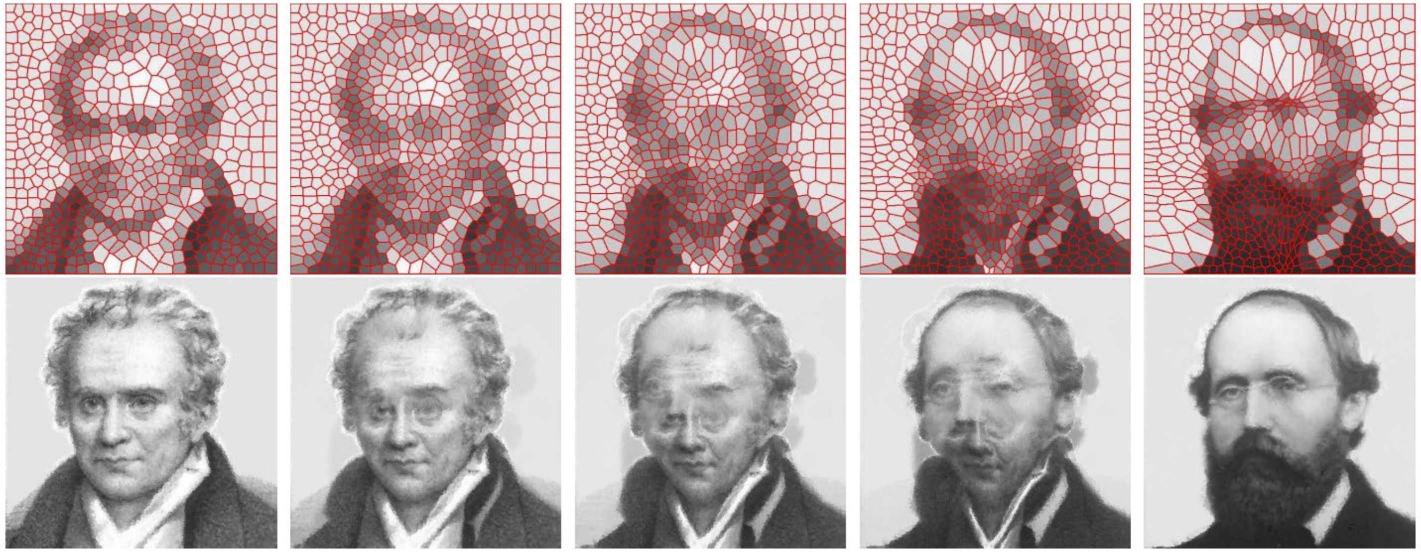
- This determines a partition, so Monge problem is:

$$\min \sum_p \int_{Vor^W(p)} \|x - p\|^2 \, \rho(x) \, dx$$

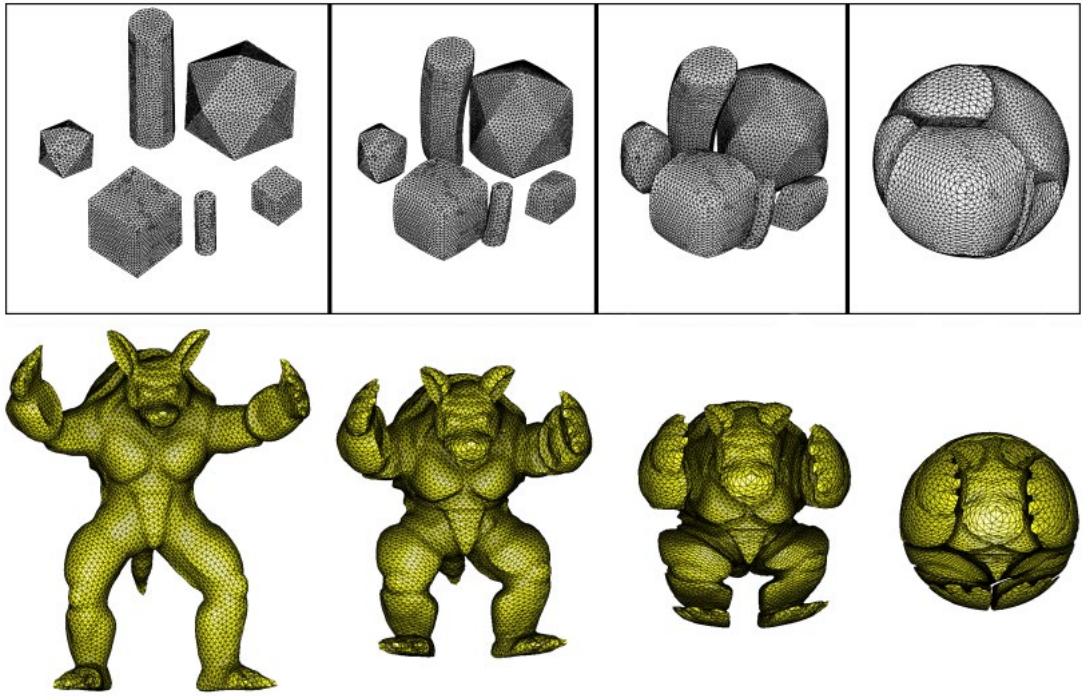
- Idea: optimize weights w for each site to grow/shrink power cells until $\lambda_p = \int_{T^{-1}(\{p\})} \rho(x) dx$

- Gradient of appropriate functional given by $\frac{\partial \phi}{\partial w(p)}(w) = \lambda_p - \int_{Vor^W(p)} \rho(x) \, dx$

Back to optimal transport



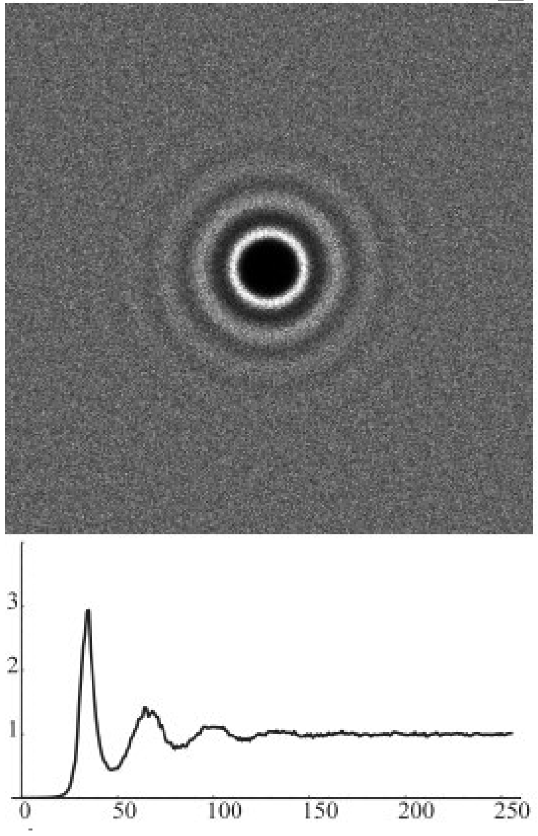
A Multiscale Approach to Optimal Transport [Mérigot 2011]



A Numerical Algorithm for L2 Semi-discrete Optimal Transport in 3D [Lévy 2015]

Application

- Also optimizes for the locations p



Blue Noise through Optimal Transport [de Goes et al. 2012]