Statistical Physics of Assignment and Optimal Transport Problems: Some applications

Henri Orland
IPhT, CEA, CNRS
Université Paris-Saclay

Patrice Koehl, UC Davis

Marc Delarue, Institut Pasteur

Applications of Statistical Physics to Optimal Transport Theory

- Assignment and Matching problems: balanced and unbalanced
- Optimal Transport (balanced)
- Optimal Transport (unbalanced) with variable masses
- Gromov-Wasserstein Optimal Transport

A long history…

Gaspard Monge
(1746-1818)

Erwin Schroedinger
(1887-1961)

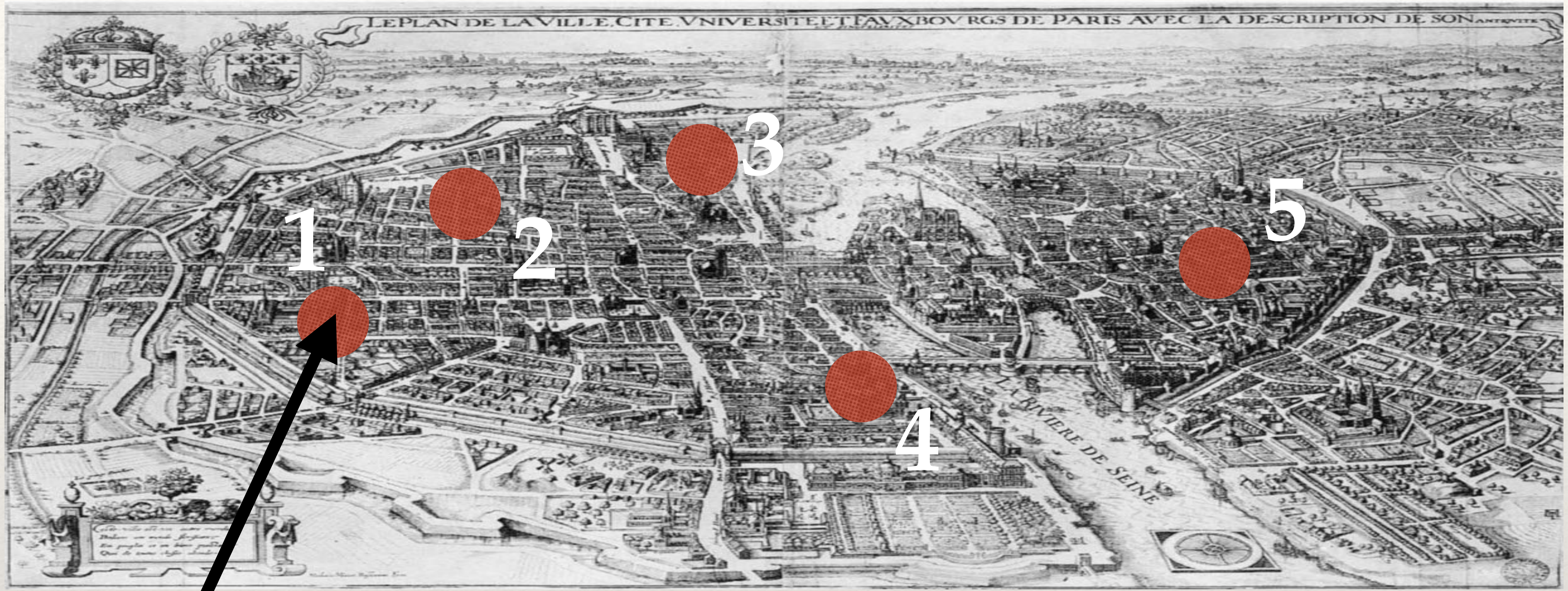
Leonid Kantorovich
(1912-1986)

Cedric Villani

Alessio Figalli

The study of optimal transportation
and allocation of resources

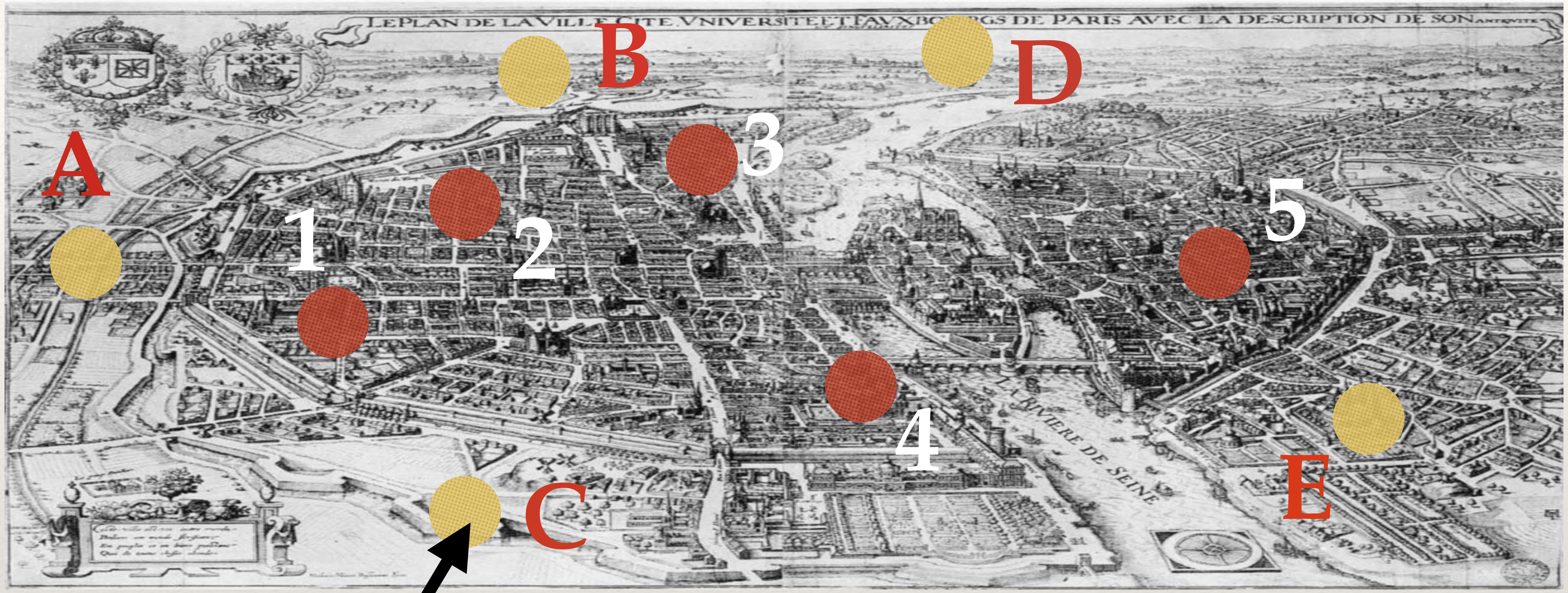
Optimal Transport (Monge): Assignment



Bakeries need flour

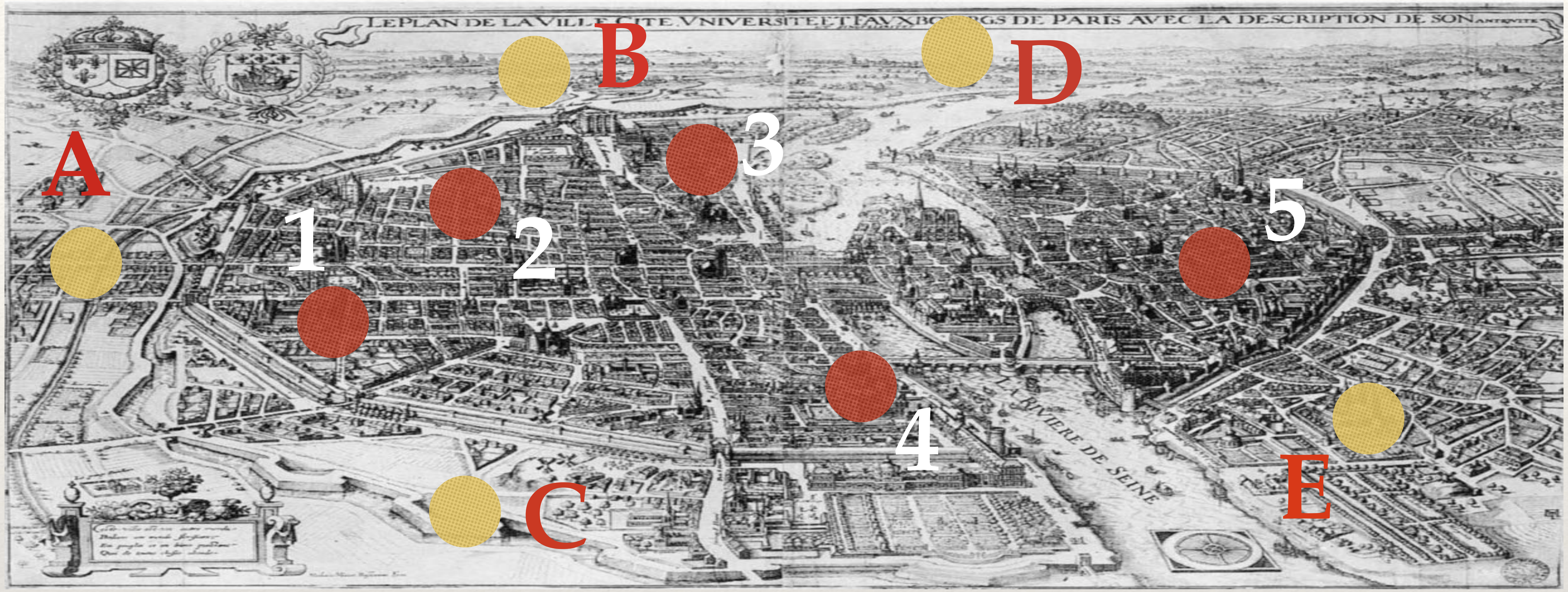
Monge, 1781

Optimal Transport: Assignment



Mills producing flour

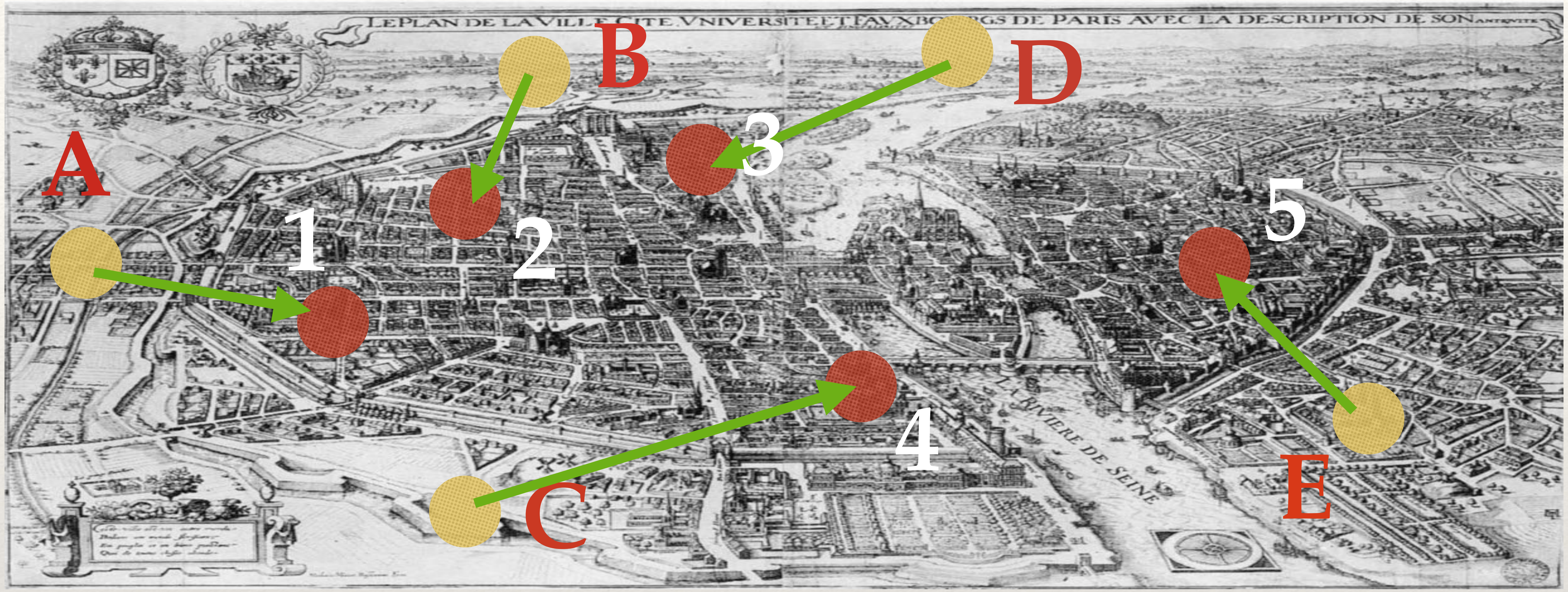
Optimal Transport: Assignment



	A	B	C	D	E
1	2	3	2,5	5	8
2	3	1,5	3	4	7,5
3	4	1,5	3,5	2	4,5
4	6	4	3	3	4
5	9	6,5	7	3	2

Transportation Cost
From Mill (column) to Bakery (line)

Optimal Transport: Assignment

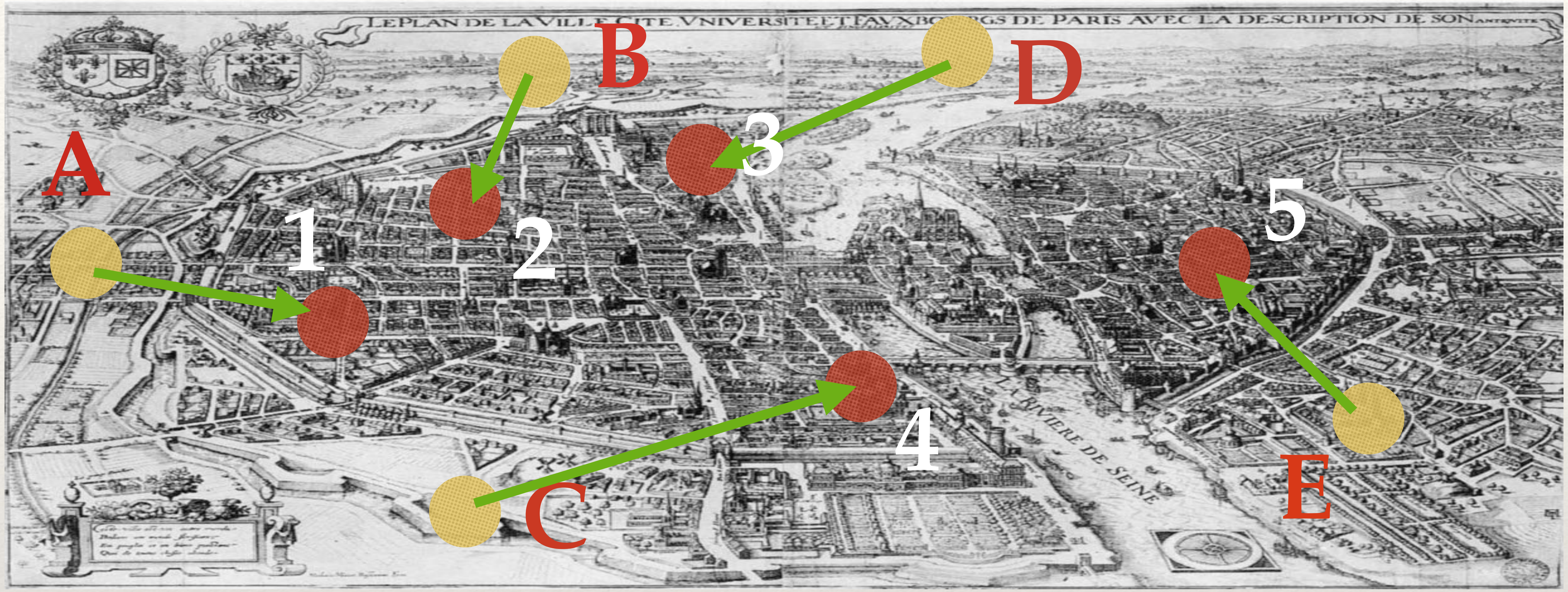


	A	B	C	D	E
1	**2**	3	2,5	5	8
2	3	**1,5**	3	4	7,5
3	4	1,5	3,5	**2**	4,5
4	6	4	**3**	3	4
5	9	6,5	7	3	**2**

How to minimise Total Transportation Cost?

Operational research
Economics
Data Science
Physics

Optimal Transport: Assignment



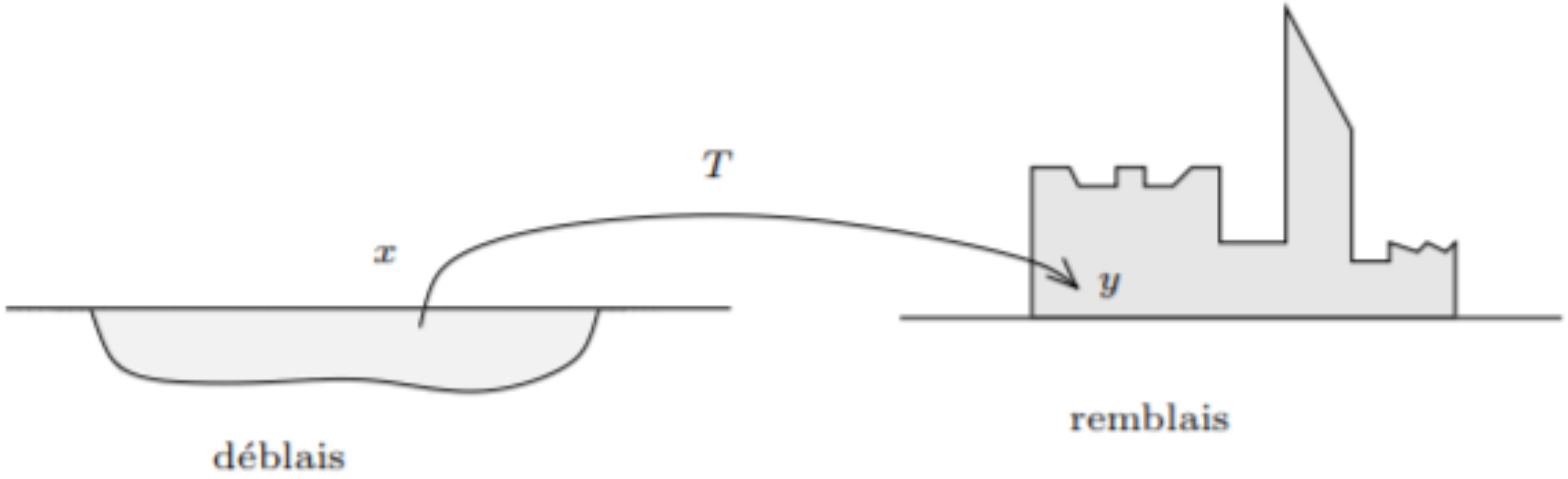
	A	B	C	D	E
1	2	3	2,5	5	8
2	3	1,5	3	4	7,5
3	4	1,5	3,5	2	4,5
4	6	4	3	3	4
5	9	6,5	7	3	2

$\odot$

	A	B	C	D	E
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	0	1	0
4	0	0	1	0	0
5	0	0	0	0	1

= 10.5

Monge formulation: déblais et remblais



$\rho_0(x)$ $\rho_1(x)$

T is a transport of ρ_0 to ρ_1 iff

$\forall A \subset \mathbb{R}^d$

$$\int_{x\in A} \rho_1(x)dx = \int_{x\in T^{-1}(A)} \rho(x)dx$$

$$\mathrm{Det}\big(\nabla T(x)\big)\,.\,\rho_1(T(x)) = \rho_0(x)$$

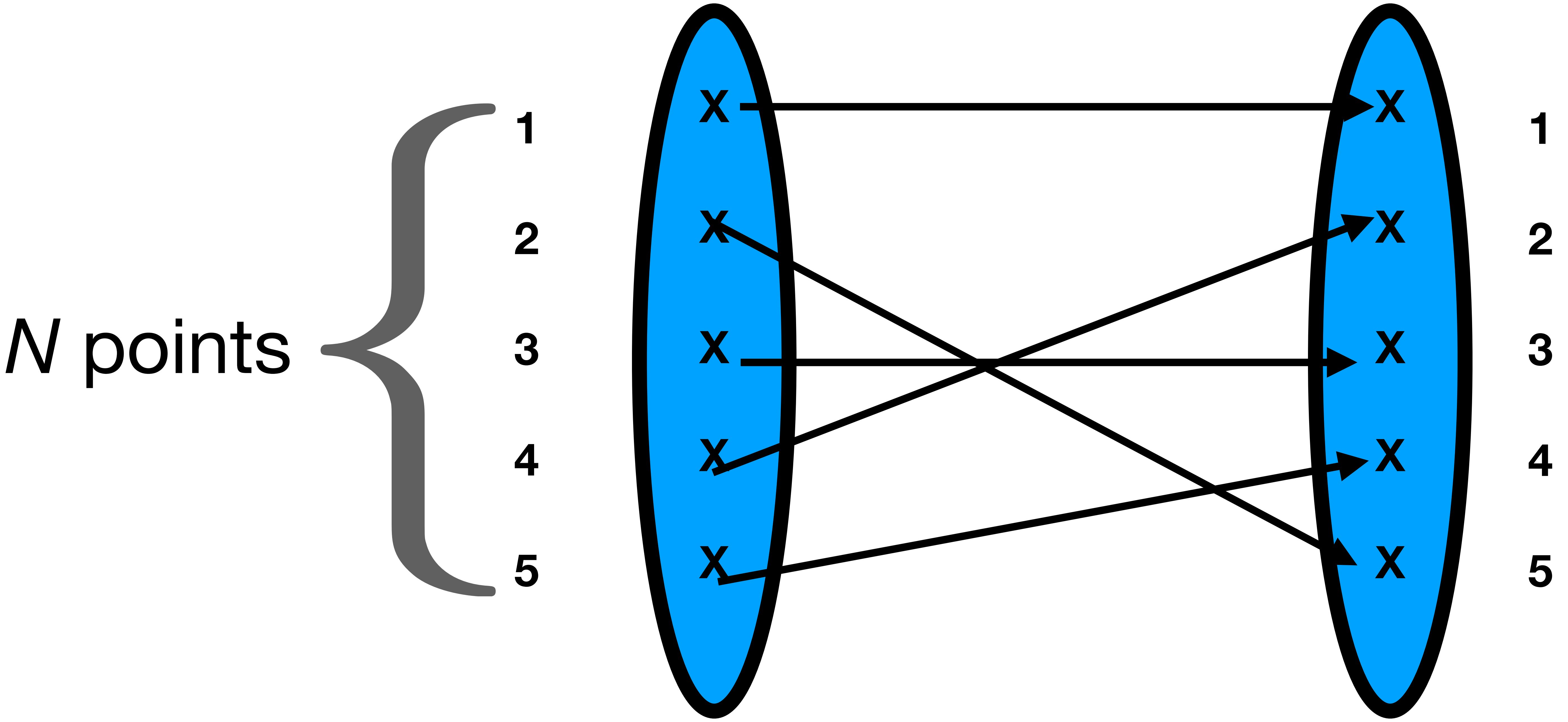
- $C(x,y) \geq 0$: cost of transporting unit mass from x to y
- Optimal transport = Monge distance

$$d(\rho_0, \rho_1) = \inf_{\{T \text{ transport}\}} \int dx \; C(x, T(x)) \; \rho_0(x)$$

- Examples :
  - Monge: $C(x,y) = |x-y|$
  - p-Wasserstein: $C(x,y) = |x-y|^p$ Wasserstein distance

Assignment = Bipartite Matching Problem = Marriage Problem.....

C.P. Bachas, 1985: pairing of defects lines in crystals



Bakeries	←	**Mills**
Factories	←	**Mines**
Women	←	**Men**
...	←	**...**

$$C_{ij} = \text{transport cost from } i \text{ to } j$$
$$N \times N \text{ positive matrix}$$

Assignment = Permutation

Find permutation *P* which minimises cost

$$\mathcal{C} = \inf_{P \in \mathcal{S}_N} \left(\sum_{i=1}^{N} C_{iP(i)} \right)$$

The Hungarian Algorithm

- Published by Harold Kuhn in 1955
- Based on work by Hungarian mathematicians Dénes Kőnig and Jenő Egerváry in 1931
- In fact (2006), solved by Jacobi and published posthumously in 1890 in latin
- IDEA: Shift row values / column values to get a “simpler” cost matrix
- Complexity is of order $O(N^3)$: unpractical for large problems. Note that $\geq 3-$ matching is NP-complete!
- Algorithm serial in nature and difficult to parallelize

Finite Temperature Assignment

- Finite temperature introduced in combinatorial optimisation (TSP) by Kirkpatrick in 1981, 1983

- Statistical physics of the random assignment problem by Vannimenus and Mézard (1984) and HO (1985), Mézard and Parisi (1985): Quenched average over random cost function

- The study of disordered systems in the lab, was initiated by Cirano De Dominicis, who performed seminal work in the theory of Spin-Glasses. Followed at that time by Edouard Brézin, Claude Itzykson, Jean Zinn-Justin, Bernard Derrida, Thomas Garel, Jean-Marc Luck…

Disordered Assignment

$$Z = \sum_{P \in \mathcal{S}_N} e^{-\beta \sum_{i=1]}^{N} C_{iP(i)}}$$

$$U_{ij} = e^{-\beta C_{ij}}$$

Take C_{ij} as random quenched variables and introduce replicas.

With no replica symmetry breaking, infinite number of order parameters.

One obtains an integral equation for an order parameter function ϕ (MP, HO).

In the case of an exponential distribution for C_{ij} , the ground state energy has been calculated by Mézard and Parisi (1985):

$$E_0/N = \frac{\pi^2}{6}$$

$N \times N$ Cost Matrix C_{ij}

Assignment matrix G_{ij} with

$$\begin{cases} G_{ij} = 0 \text{ or } 1 \\ \forall j \in \{1,\ldots,N\} \quad \sum_{i=1}^{N} G_{ij} = 1 \\ \forall i \in \{1,\ldots,N\} \quad \sum_{j=1}^{N} G_{ij} = 1 \end{cases}$$

Cost Function $E = \sum_{i,j} C_{ij} G_{ij}$

Partition function at temperature $T = \frac{1}{\beta}$

$$Z = \sum_{\text{all } C \in \mathscr{C}} e^{-\beta E(C)}$$

$$Z = \sum_{\{G_{ij}=0,1\}} \prod_{i=1}^{N} \delta\left(\sum_{j=1}^{N} G_{ij} - 1\right) \prod_{j=1}^{N} \delta\left(\sum_{i=1}^{N} G_{ij} - 1\right) e^{-\beta \sum_{ij} C_{ij} G_{ij}}$$

P Koehl, H Orland

Physical Review E 103 (4), 042101, 2021

Fourier representation of δ-functions

$$\delta\left(\sum_j G_{ij} - 1\right) = \int \frac{\beta d\lambda(i)}{2\pi} e^{i\beta\lambda(i)\left(1 - \sum_j G_{ij}\right)}$$

$$Z(\beta) = \int_{-\infty}^{+\infty} \prod_k d\lambda(k) \int_{-\infty}^{+\infty} \prod_l d\mu(l) e^{\beta\left(\sum_k i\lambda(k) + \sum_l i\mu_l\right)} \sum_{G(k,l)\in\{0,1\}} e^{-\beta \sum_{k,l} G(k,l)\left(C(k,l) + i\lambda(k) + i\mu(l)\right)}$$

Do summation over the G

$$Z(\beta) = \int_{-\infty}^{+\infty} \prod_k d\lambda(k) \int_{-\infty}^{+\infty} \prod_l d\mu_l e^{-\beta F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu})}$$

Duality

with

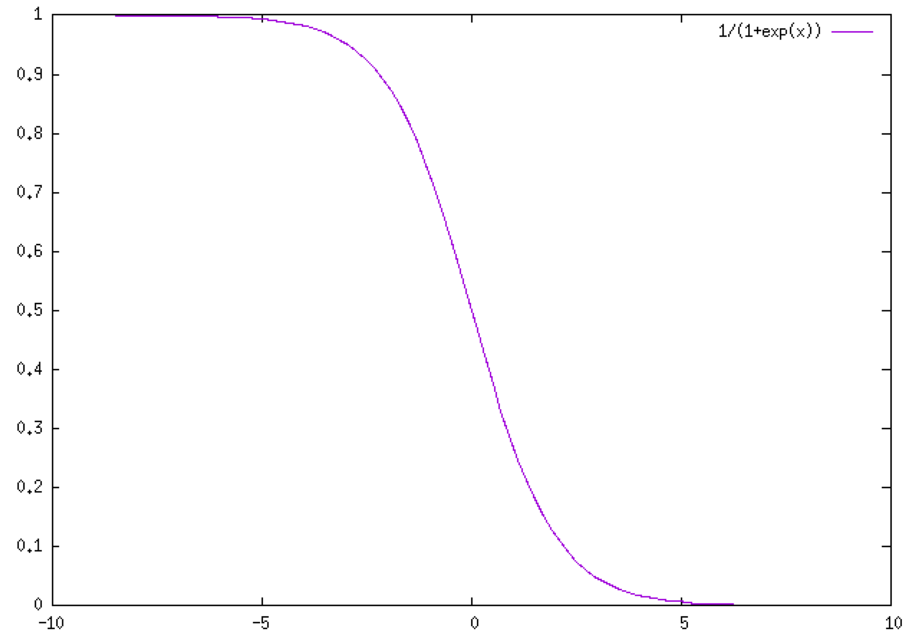
$$F_\beta\left(\boldsymbol{\lambda}, \boldsymbol{\mu}\right) = -\left(\sum_k i\lambda(k) + \sum_l i\mu(l)\right) - \frac{1}{\beta}\sum_{kl} \ln\left[1 + e^{-\beta(C(k,l) + i\lambda(k) + i\mu(l))}\right]$$

Saddle-Point Approximation

Change $$i\lambda(k) \to \lambda(k)$$ $$i\mu(l) \to \mu(l)$$ Saddle point is pure imaginary

SP equations $$\frac{\partial F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \lambda_k} = 0 \quad \text{and} \quad \frac{\partial F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu_l} = 0$$

Define $X(k,l) = h\left[\beta\left(C_{kl} + \lambda(k) + \mu(l)\right)\right]$ where $h(x) = \dfrac{1}{e^x + 1}$



$X(k,l) = \bar{G}(k,l)$ Thermal average of G

$$\left.\begin{aligned} \forall k, \quad & \sum_l X(k,l) = 1 \\ \forall l, \quad & \sum_k X(k,l) = 1 \end{aligned}\right\}$$ Set of $2N$ non-linear equations with $2N$ variables

Some properties of the SP

- The Hessian of F is negative, except for one trivial zero mode → maximum is unique

- Free energy F ↗ Internal energy U ↗ as a function of T

$$F = U - TS$$

$$U = \sum_{kl} C(k,l) X(k,l)$$

Lattice gaz entropy

$$S = -\sum_{kl} \left(X(k,l) \ln X(k,l) + (1 - X(k,l)) \ln(1 - X(k,l)) \right)$$

- $$F^* = \lim_{\beta \to +\infty} F(\beta) \quad U^* = \lim_{\beta \to +\infty} U(\beta)$$

At zero temperature, the SP free and internal energy converge to exact solutions

- Solve equations by Newton-Raphson. Very efficient.

- Start at high temperature and decrease temperature: annealing procedure.
When do we stop?

- When T decreases to zero, one may reach solutions where the $X(k, l)$ do not converge to 0 or 1: problem is degenerate.

- If the problem is degenerate, the entropy does not vanish at $T = 0$.

- In that case, how to obtain the optimal assignment?

Solving Degenerate Assignment Problems

- **Theorem:** if the problem is degenerate, there are ground state solutions with integer $G^*(k,l)$ with same total cost as the fractional solution (Gartner and Matousek, 2006).

- If Δ is the gap between the optimal solution and the second best solution, the optimal solution of the assignment problem with cost function

$$C'(k,l) = C(k,l) + \alpha \times \mathrm{random}(k,l) \quad \text{random}(k,l) \in [0,1]$$

with $\alpha < \frac{\Delta}{2N}$ has the same optimal assignment as the original $C(k,l)$ cost

- The randomized original problem is non degenerate.

- If all entries of $C(k,l)$ are scaled to be integer, then $\Delta \geq 1$ and it is sufficient to have

- $$\alpha < \frac{1}{2N}$$

When to stop the annealing?

- If the problem is non-degenerate, at low enough temperature, the matrix $X(k,l)$ becomes row-dominant: each row contains one and only one element $> 1/2$ (since $\sum_{k,l} X(k,l) = 1$)

- Theorem: At that point, one can stop and replace in each row the dominant $X(k,l)$ by 1 and all others by 0. This is the ground state assignment $G^*(k,l)$

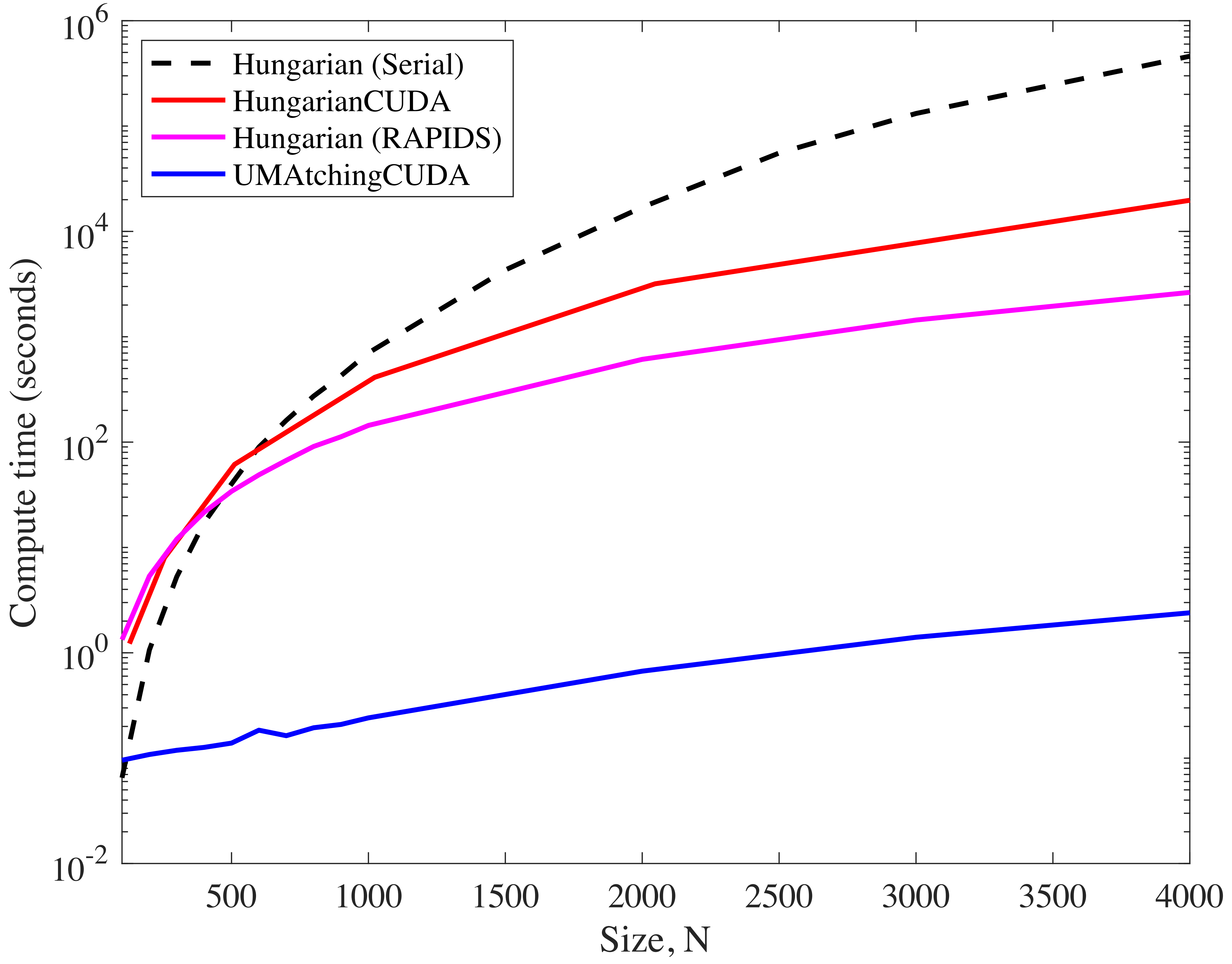
- If the problem is degenerate, make it first non-degenerate by adding the proper random noise then use above theorem.

Implementation

- Apparent computational complexity is $O(N^2)$ compared to $O(N^3)$ for Hungarian algorithm.

- Can be parallelised and run on GPU.

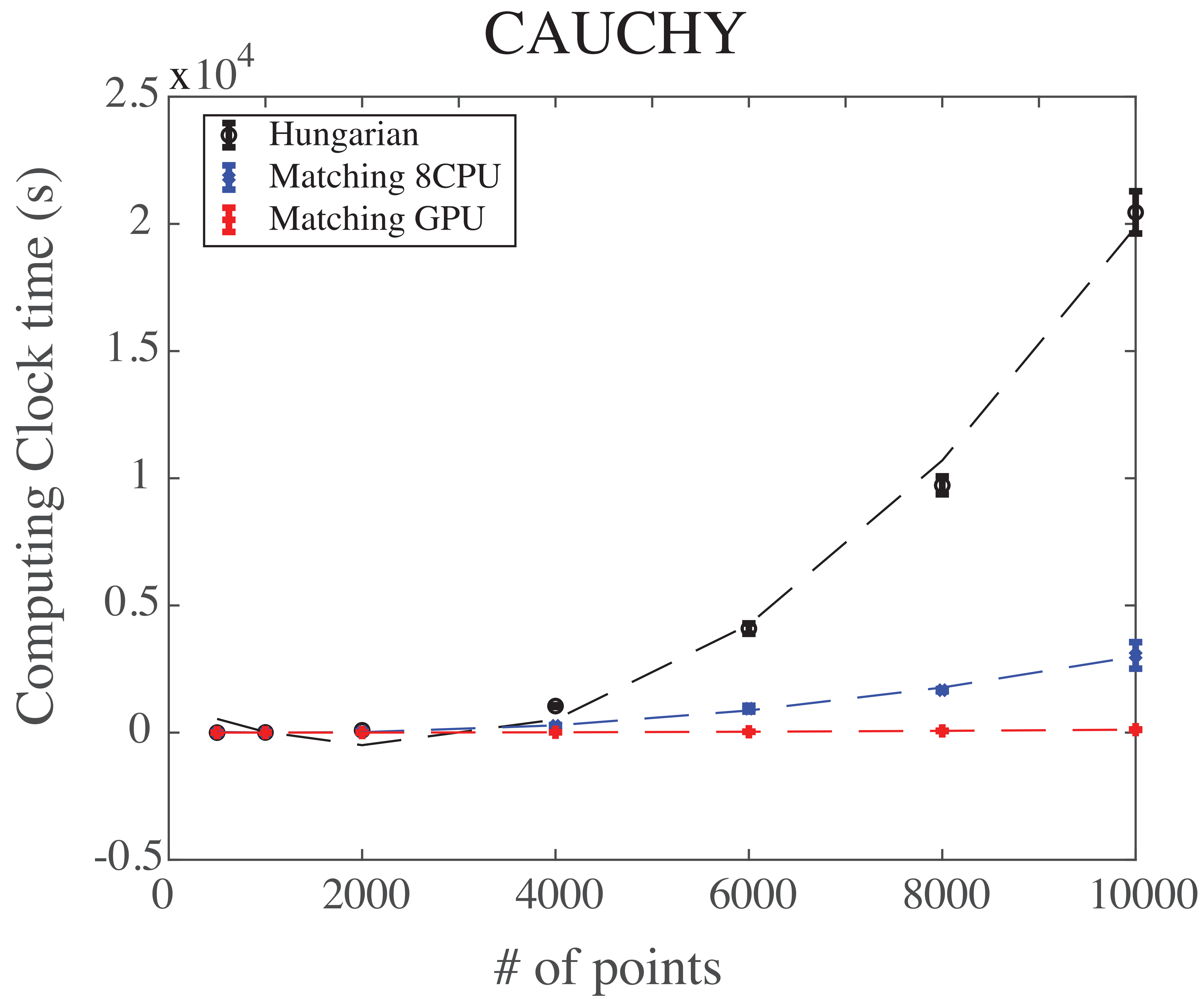
- Storage space required $O(N^2)$. Limit of N=30000 on GPU.

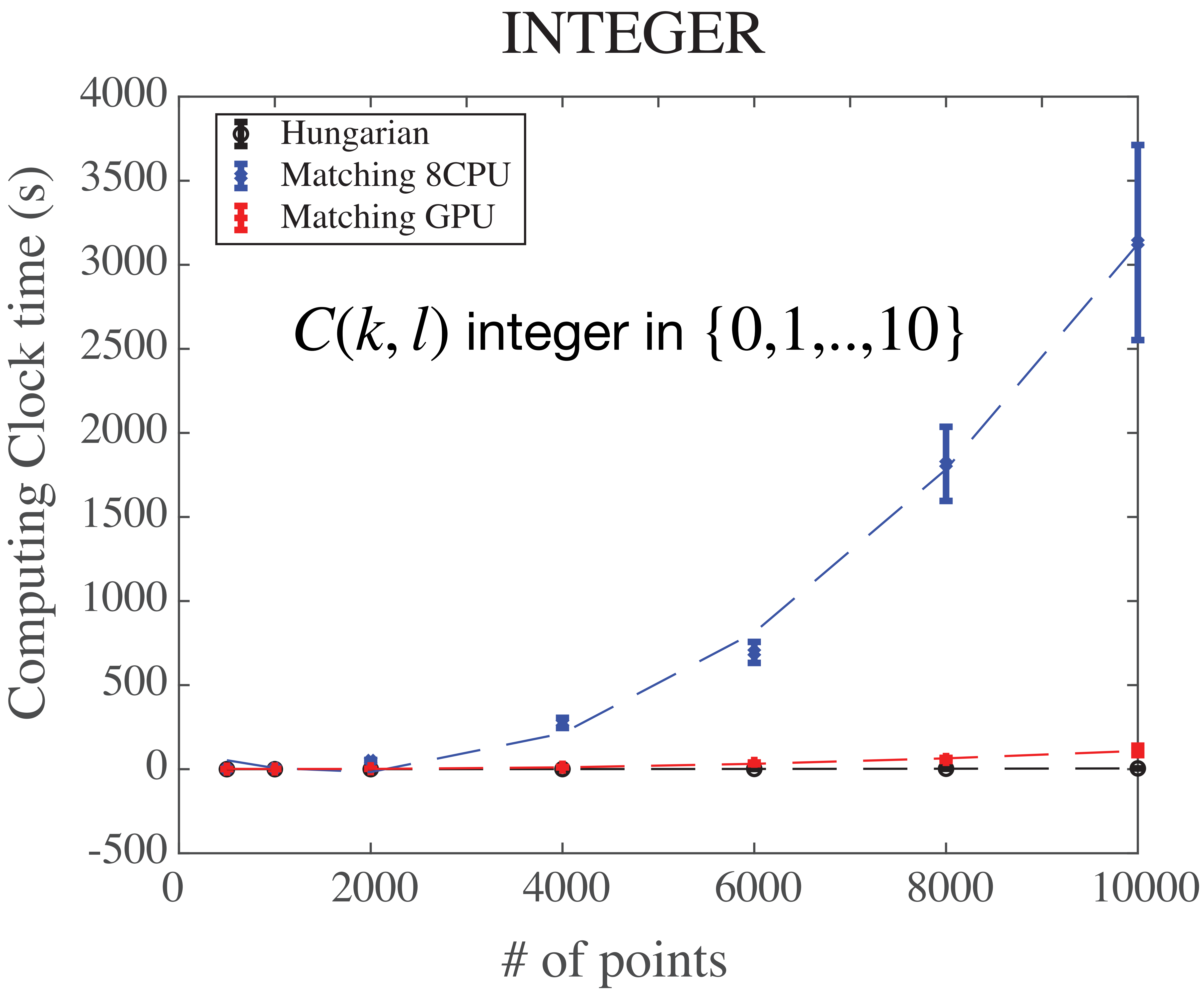


For N = 4000, computational times (in seconds):

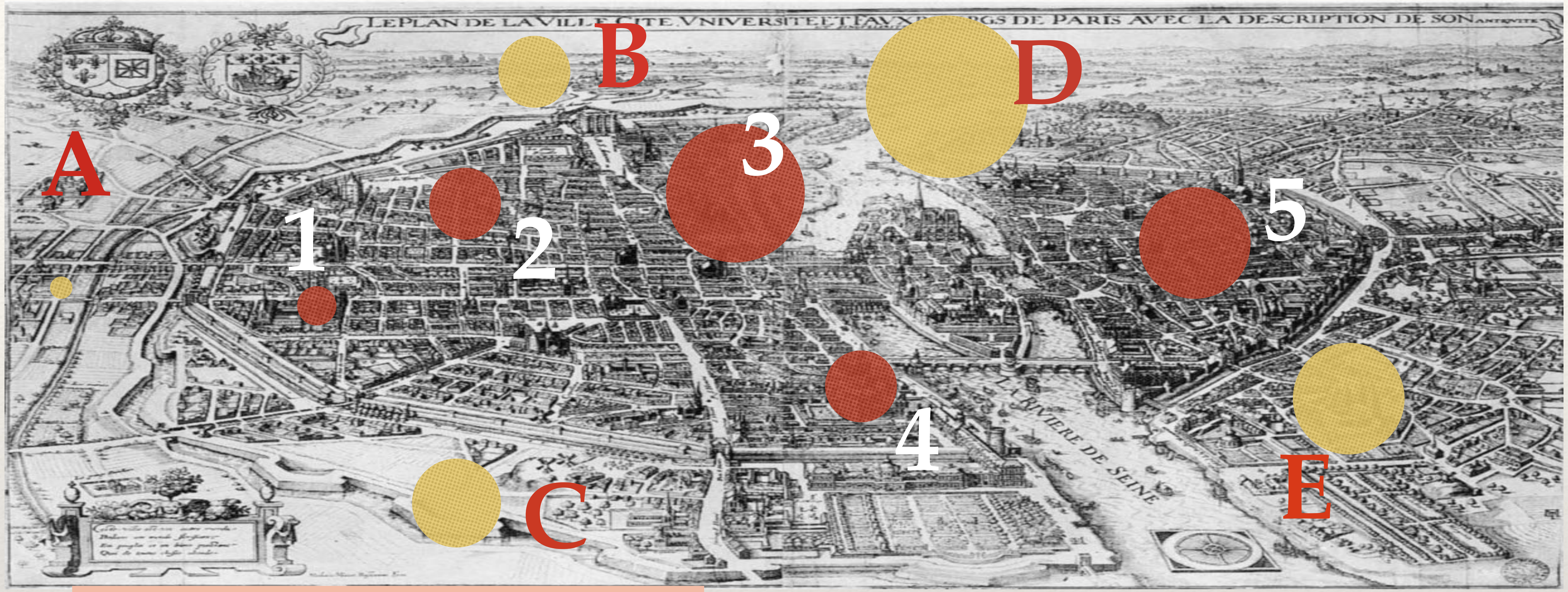
Hungarian sequential	Hungarian CUDA	Hungarian NVIDIA	UMAtching
460000	21590	2638	2.40

Test with i.i.d. random costs



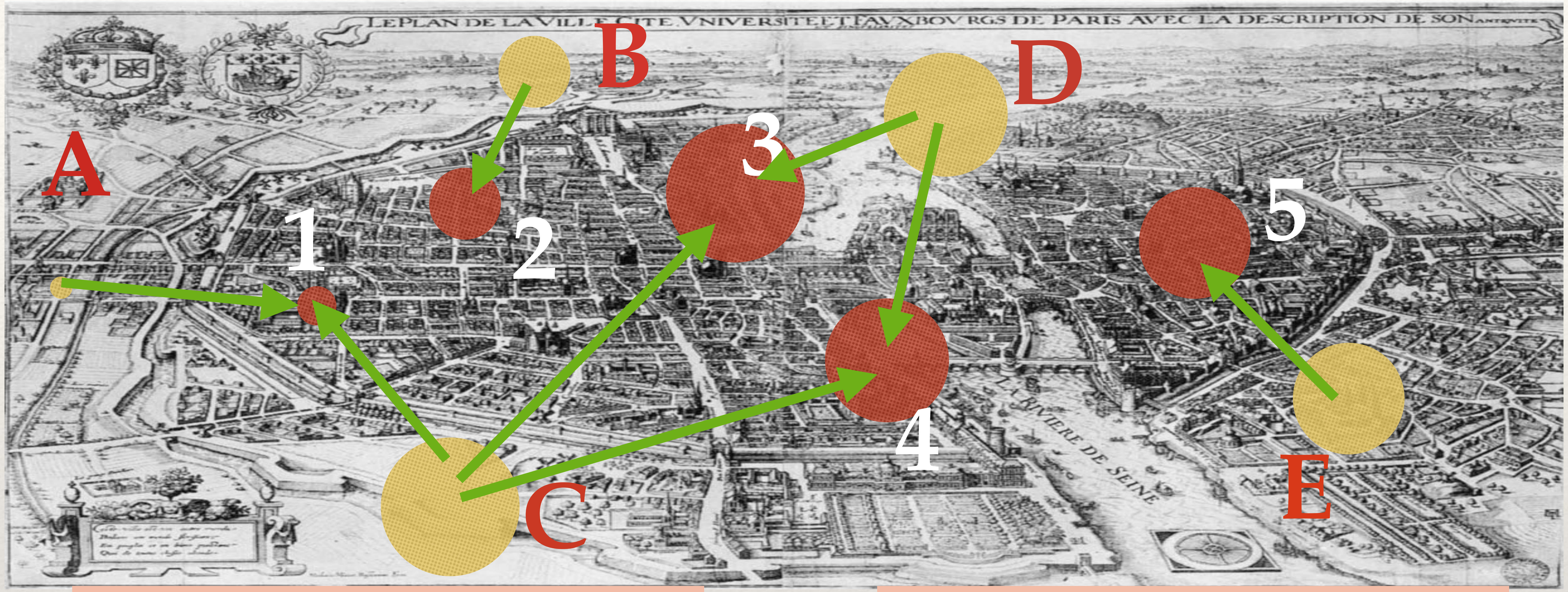


Optimal Transport: Transport Plan

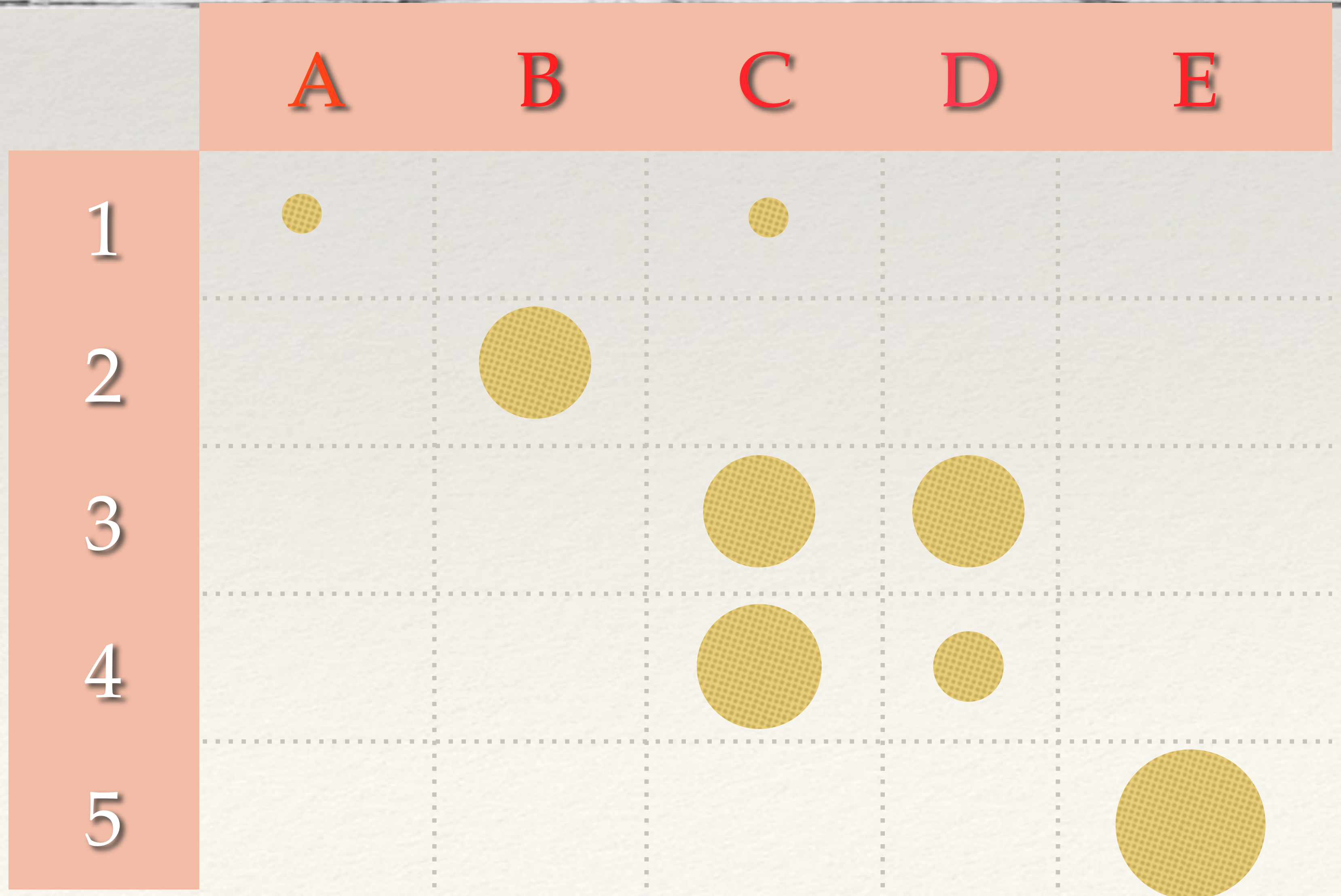


	A	B	C	D	E
1	2	3	2,5	5	8
2	3	1,5	3	4	7,5
3	4	1,5	3,5	2	4,5
4	6	4	3	3	4
5	9	6,5	7	3	2

Optimal Transport: Transport Plan



	A	B	C	D	E
1	2	3	2,5	5	8
2	3	1,5	3	4	7,5
3	4	1,5	3,5	2	4,5
4	6	4	3	3	4
5	9	6,5	7	3	2



Kantorovitch formulation: breaking the stones

$P(x, y)$ is a transport plan from ρ_0 to ρ_1 iff

$$\begin{cases} \int dy P(x,y) = \rho_0(x) \\ \int dx P(x,y) = \rho_1(y) \end{cases}$$

Cost function $C(x, y) \geq 0$

$$d(\rho_0, \rho_1) = \inf_{P \in \text{transport}} \int dx \int dy C(x,y) P(x,y)$$

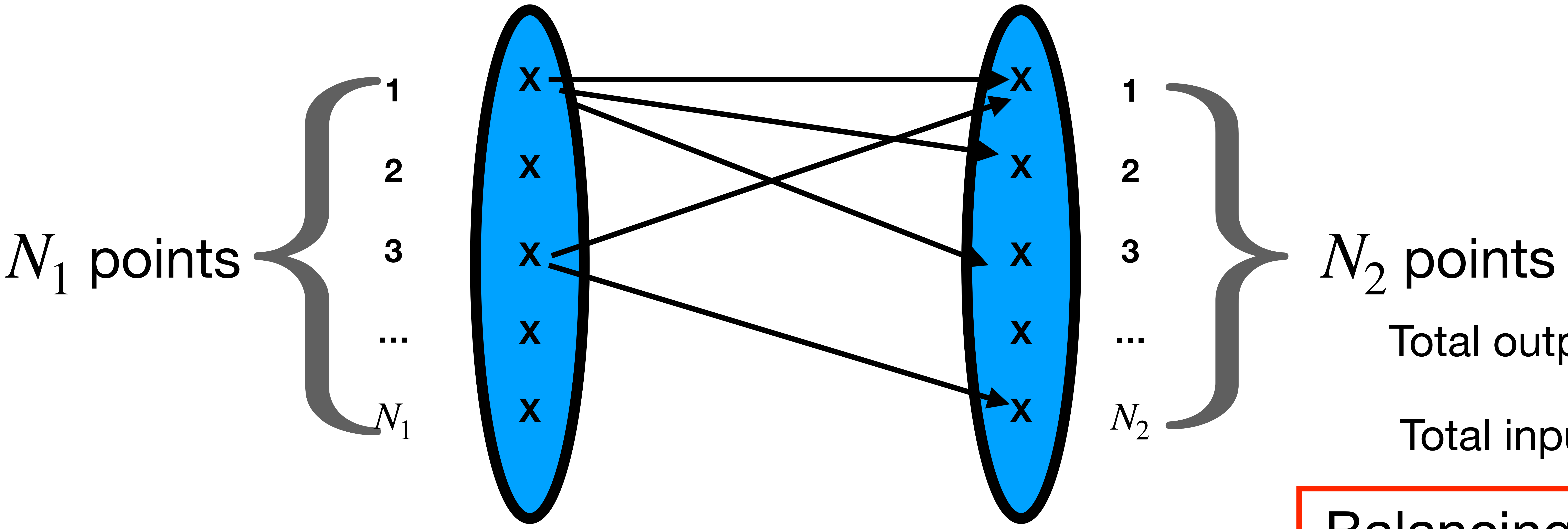
Identical to Monge distance

Monge-Kantorovitch distance

$$d_p(\rho_0, \rho_1) = \inf_{P \in \text{transport}} \int dx \int dy \, |x - y|^p P(x,y)$$

Wasserstein distance

Optimal Transport



Total output of $i = m_1(i)$

Total input of $j = m_2(j)$

$$C_{ij} = \text{transport cost from } i \text{ to } j$$

$$G_{ij} \in [0,1] = \text{transport plan}$$

$N_1 \times N_2$ matrices

Find the best transport plan to minimise $\mathscr{C} = \sum_{ij} C_{ij} G_{ij}$

Balancing input and output

$$\forall i, \sum_j G_{ij} = m_1(i)$$

$$\forall j, \sum_i G_{ij} = m_2(j)$$

$$\sum_i m_1(i) = \sum_j m_2(j)$$

$N \times N$ Cost Matrix C_{ij}

Transport plan G_{ij} with

$$\begin{cases} G_{ij} \in [0,1] \\ \forall i \in \{1,...,N_1\},\ \sum_{j=1}^{N_2} G_{ij} = m_1(i) \\ \forall j \in \{1,...,N_2\},\ \sum_{i=1}^{N_1} G_{ij} = m_2(j) \end{cases}$$

Cost Function $E = \sum_{i,j} C_{ij} G_{ij}$

Partition function at temperature

$$Z(\beta) = \int_0^1 \prod_{ij} dG_{ij} \prod_{i=1}^{N_1} \delta\left(\sum_{j=1}^{N_2} G_{ij} - m_1(i)\right) \prod_{j=1}^{N_2} \delta\left(\sum_{i=1}^{N_1} G_{ij} - m_2(j)\right) e^{-\beta \sum_{ij} C_{ij} G_{ij}}$$

P Koehl, M Delarue, H Orland
Physical review letters 123 (4), 040603 , 2019

P Koehl, M Delarue, H Orland
Physical Review E 100 (1), 013310 , 2019

Optimal Transport at Finite Temperature

$$Z = \int \prod_k d\lambda_k \int \prod_l d\mu_l e^{-\beta F_{\text{eff}}(\beta, i\lambda_k, i\mu_l)}$$

$$F_{\text{eff}}(\beta, \lambda, \mu) = -\left(\sum_k \lambda_k m_1(k) + \sum_l \mu_l m_2(l)\right) - \frac{1}{\beta} \sum_{kl} \ln\left(\frac{1 - e^{-\beta(C_{kl}+\lambda_k+\mu_l)}}{\beta(C_{kl} + \lambda_k + \mu_l)}\right).$$

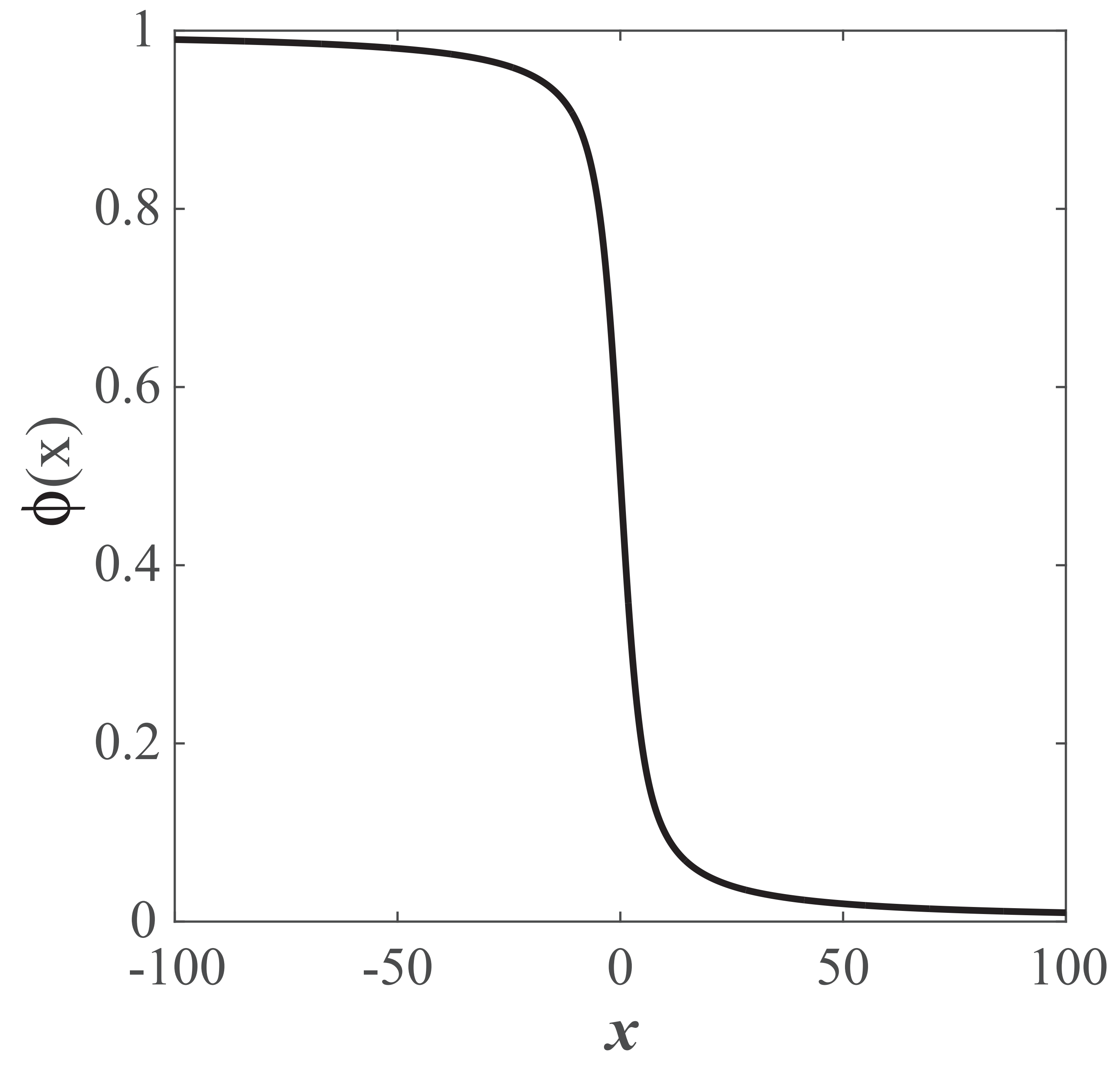
Saddle-Point Approximation

$$\frac{\partial \mathcal{F}_{\text{eff}}(\beta, i\lambda, i\mu)}{\partial \lambda_k} = 0 \quad \text{and} \quad \frac{\partial \mathcal{F}_{\text{eff}}(\beta, i\lambda, i\mu)}{\partial \mu_l} = 0.$$

$$\bar{G}_{kl} = \phi[\beta(C_{kl} + i\lambda_k + i\mu_l)]$$

$$\forall\, k, \qquad \sum_l \bar{G}_{kl} = m_1(k),$$

$$\forall\, l, \qquad \sum_k \bar{G}_{kl} = m_2(l),$$



$$\phi(x) = \frac{e^{-x}}{e^{-x} - 1} + \frac{1}{x}.$$

Implementation

- Hessian of $\mathscr{F}_{\text{eff}}$ is positive, with a trivial 0-mode (set one λ or μ to 0) so effective free energy is concave

- Solve by Newton-Raphson using a temperature annealing scheme $T \searrow 0$

- Free energy and internal energy are decreasing functions of β and converge to the optimal plan at 0 temperature

- Triangular inequality for the internal energy: 3 ensembles N_1, N_2, N_3 points, with $\forall (k,l) \forall j,\ C_{13}(k,l) \leq C_{12}(k,j) + C_{23}(j,l)$

→ $$U_{13} \leq U_{12} + U_{23}$$ at any temperature

Application - Computer Vision

Flip a coin - correct 50% of the time
Software fifteen years ago - not much better
Today - 99%

Application - Computer Vision

Dog or Muffin? Still a challenge

Application - Computer Vision

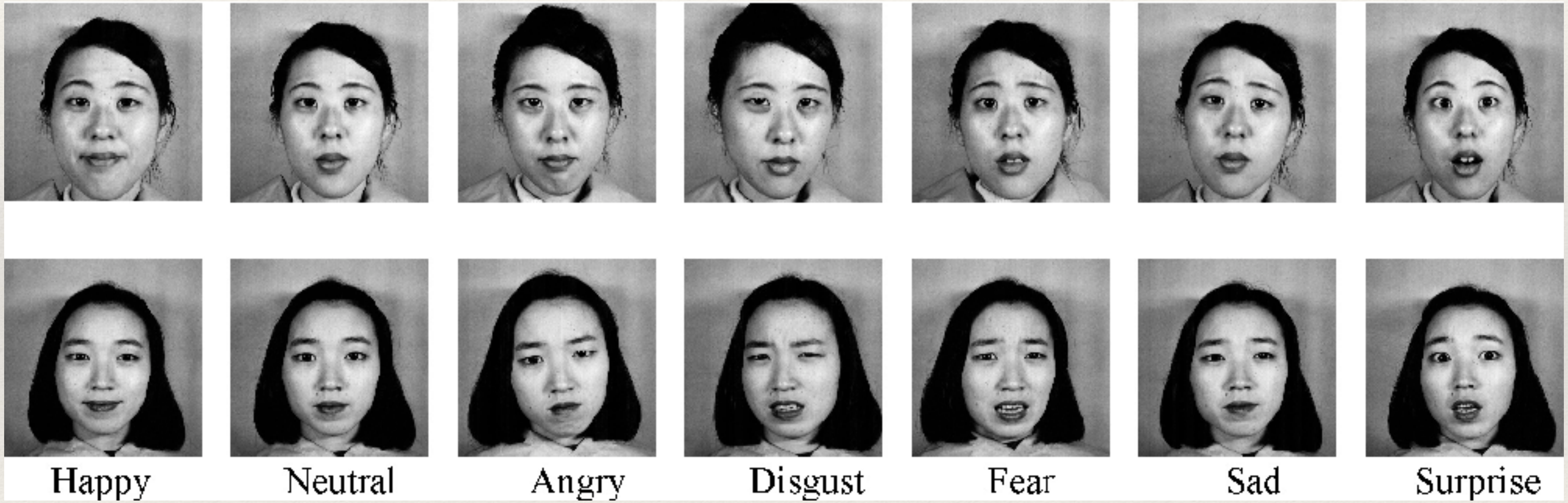
Puppy or Bagel?

Application - Computer Vision

Pain au chocolat or Sloth ?

Comparing images: the Jaffe Dataset

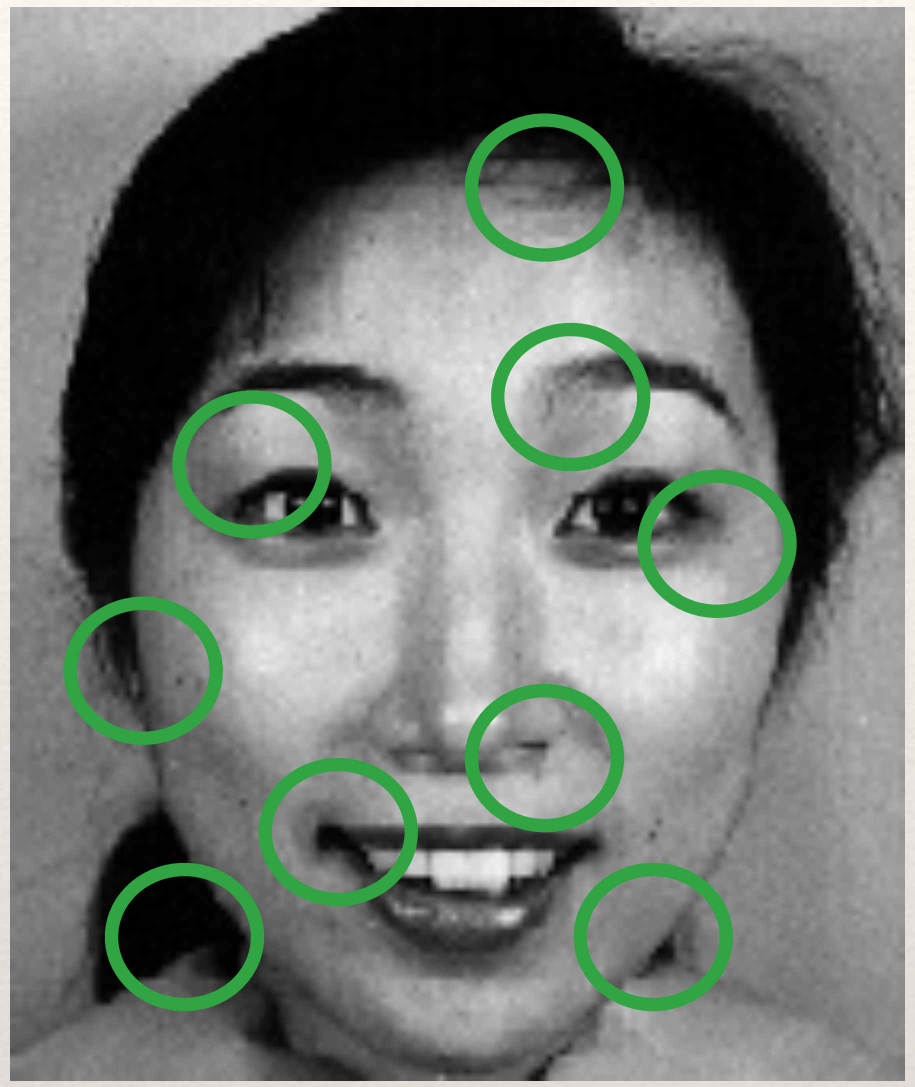
The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models.

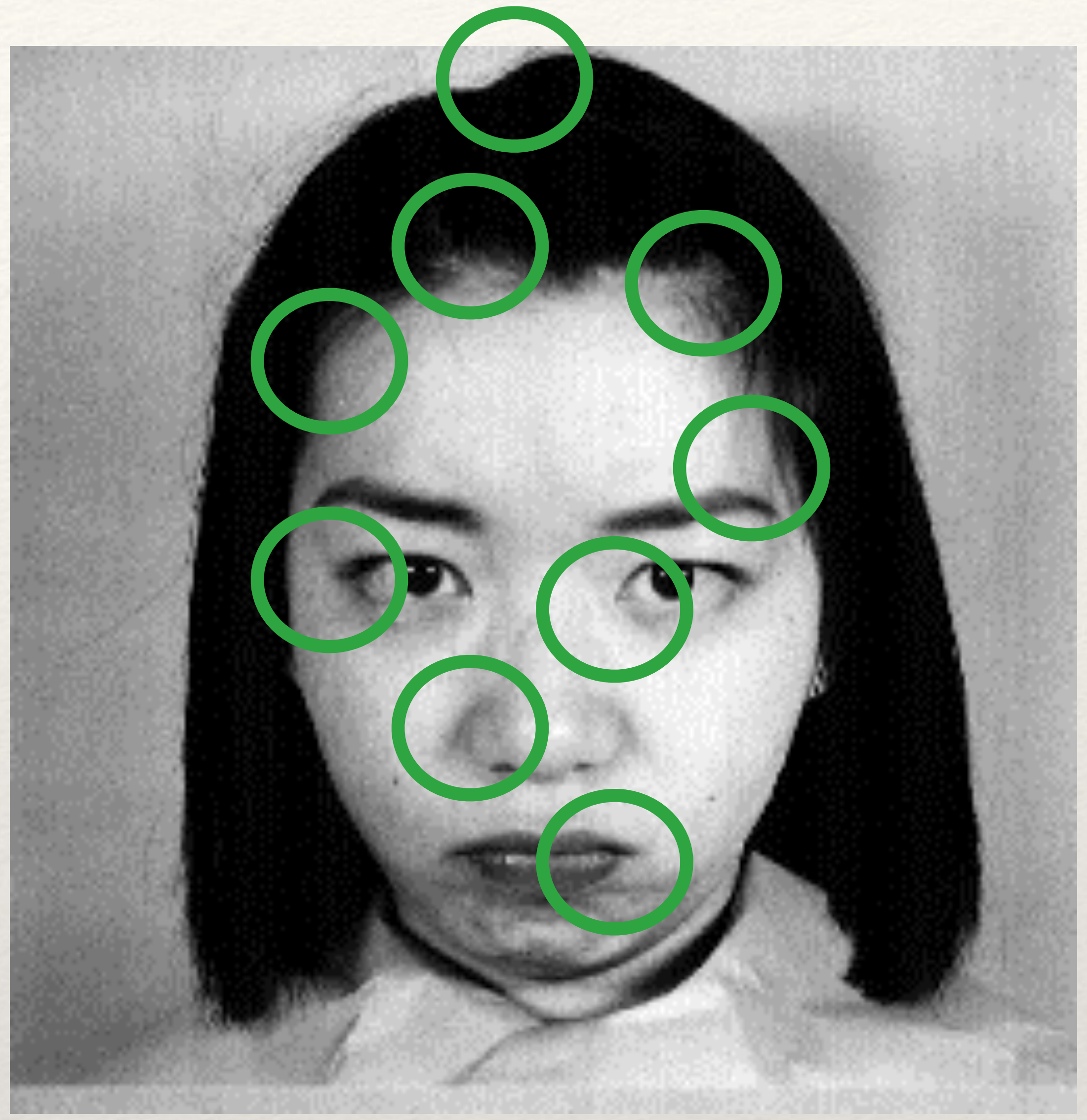


Comparing images: the Jaffe Dataset

Comparing two images I_1 and I_2:

a) Detect keypoints:





b) Assign mass: 1/N

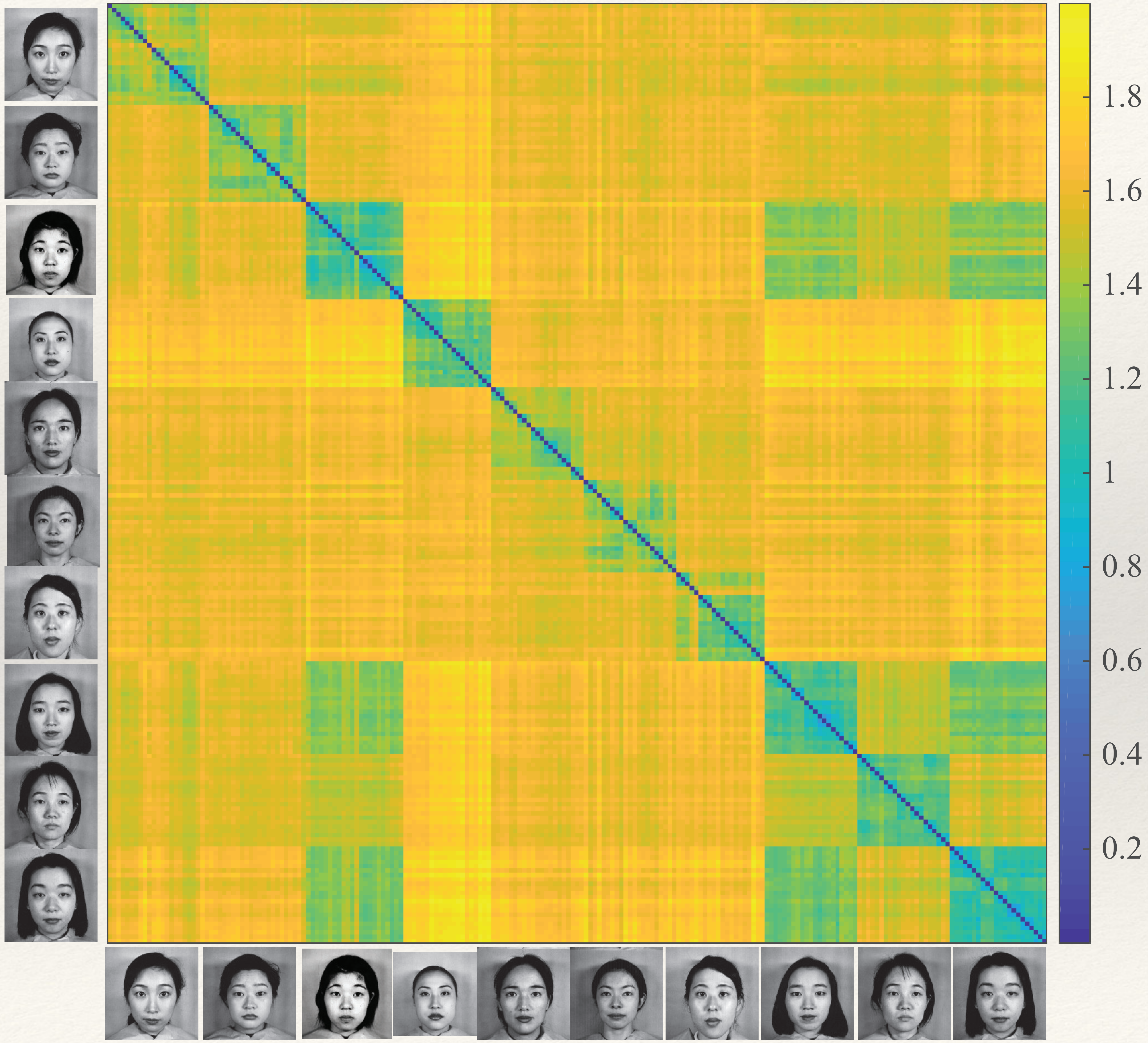
c) Assign cost matrix

- Each keypoint is characterized by a vector F of 64 "features": SURF
- For keypoint i of image 1 and j of image 2, $C(i, j) = ||F(i) - F(j)||$

d) Compute distance:

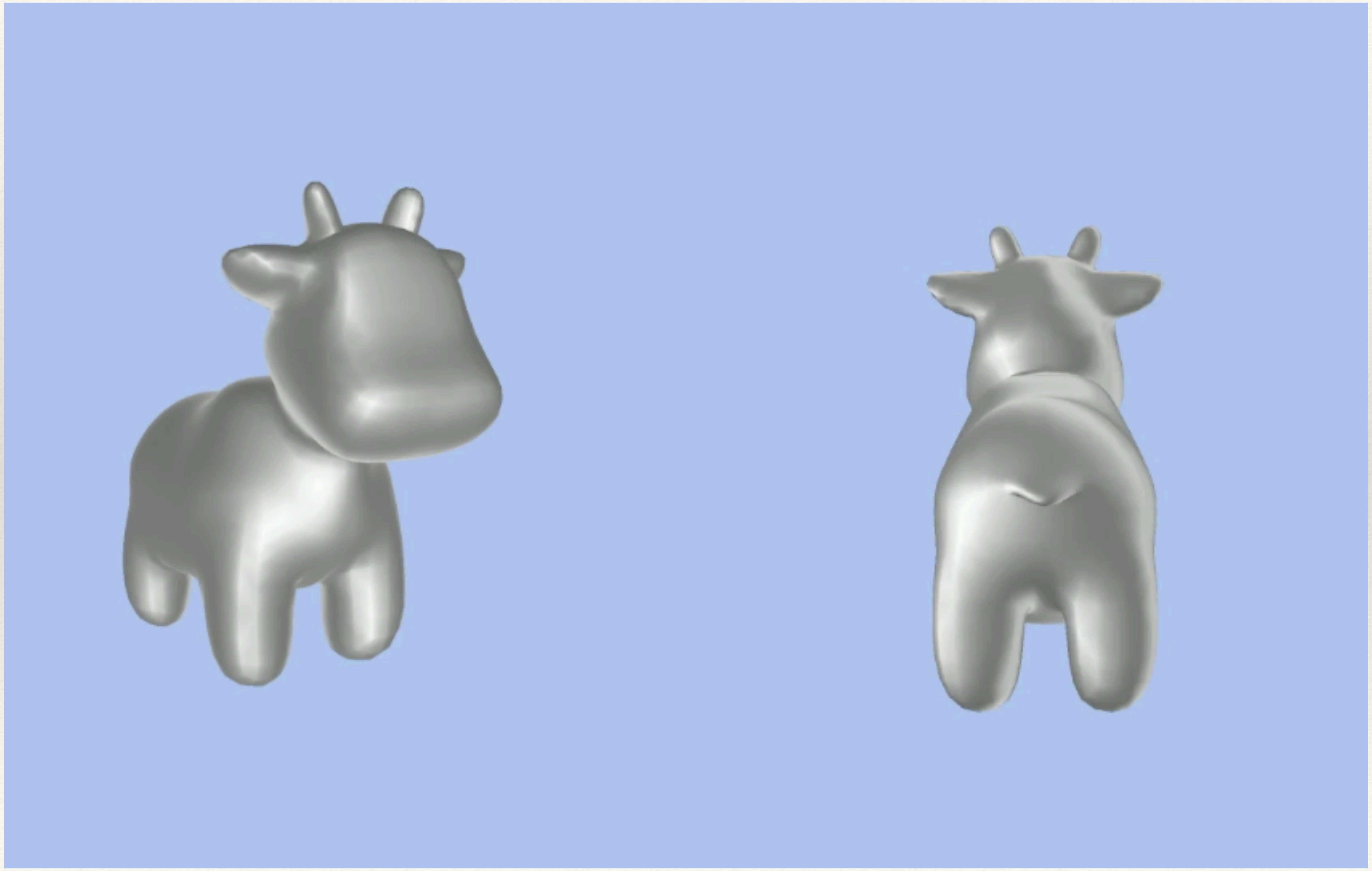
$$d(I_1, I_2) = U_{\beta}^{MF}(I_1, I_2)$$

Comparing images: the Jaffe Dataset

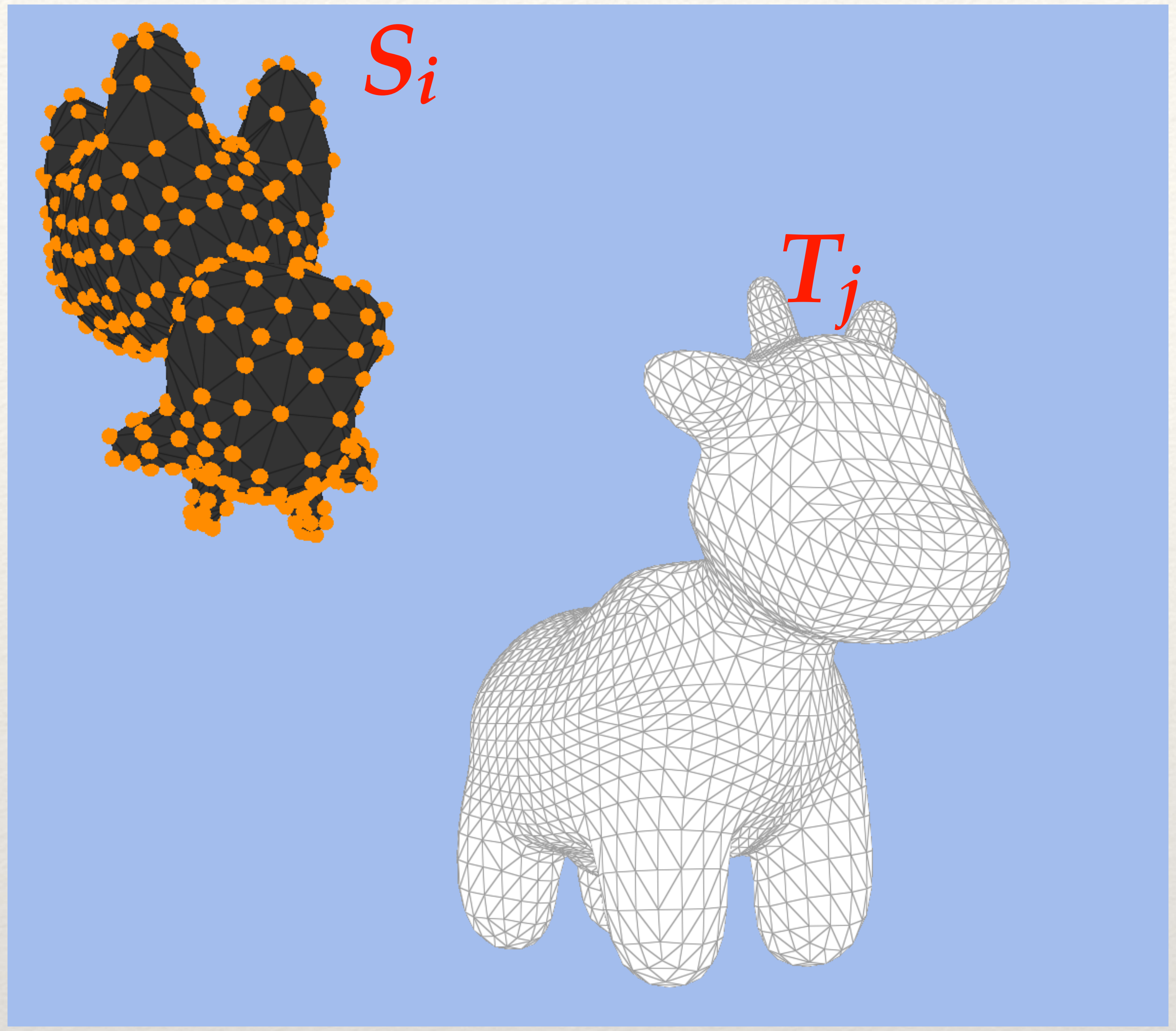


Comparing 3D shapes with OT

SPOT



Comparing 3d shapes



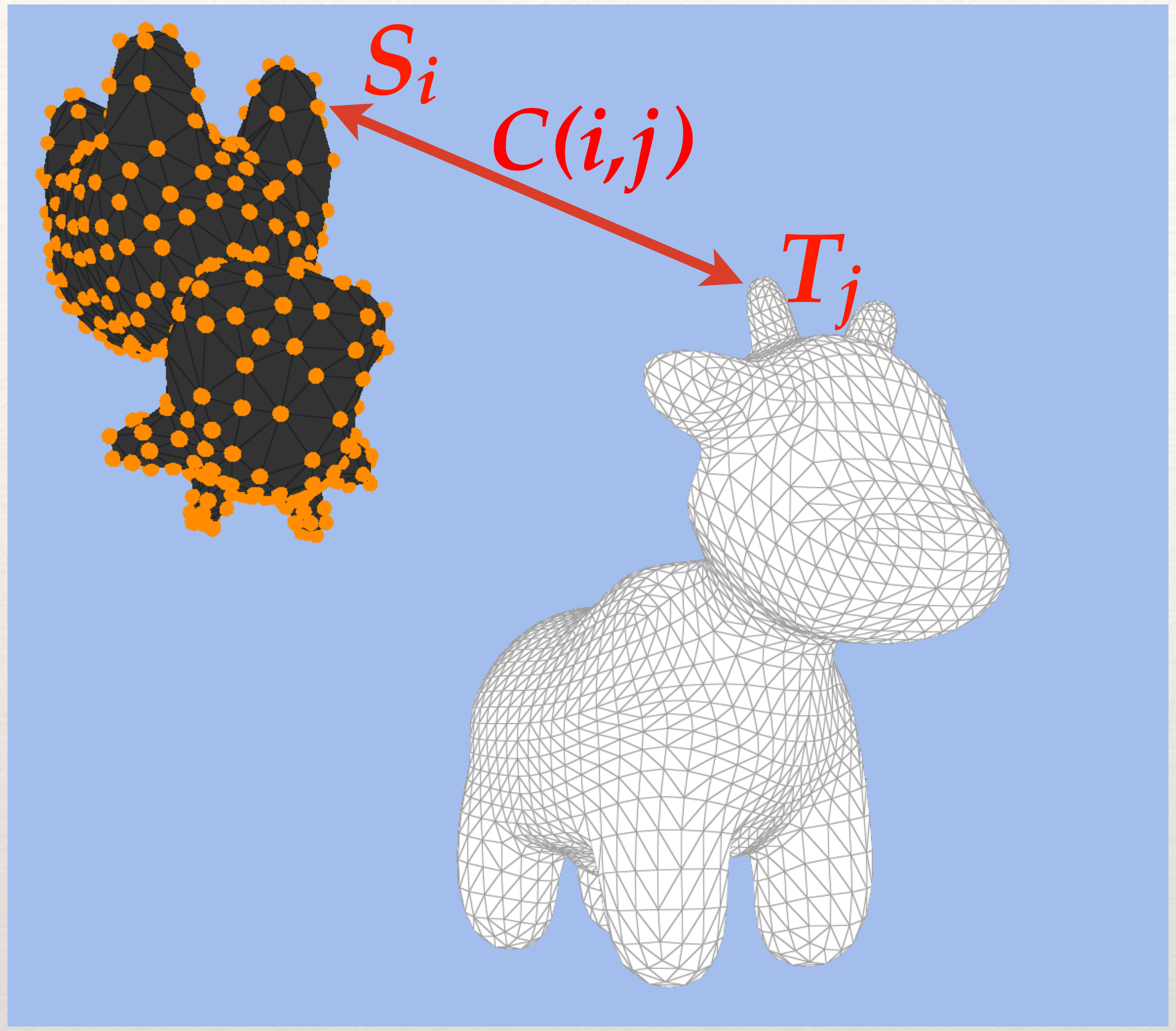
$$S = \{S_i\}_{i \in [1,N]}$$

$$m_1(i) = \frac{1}{N}$$

$$T = \{T_j\}_{j \in [1,M]}$$

$$m_2(j) = \frac{1}{M}$$

Comparing shapes



$$S = \{S_i\}_{i \in [1,N]}$$

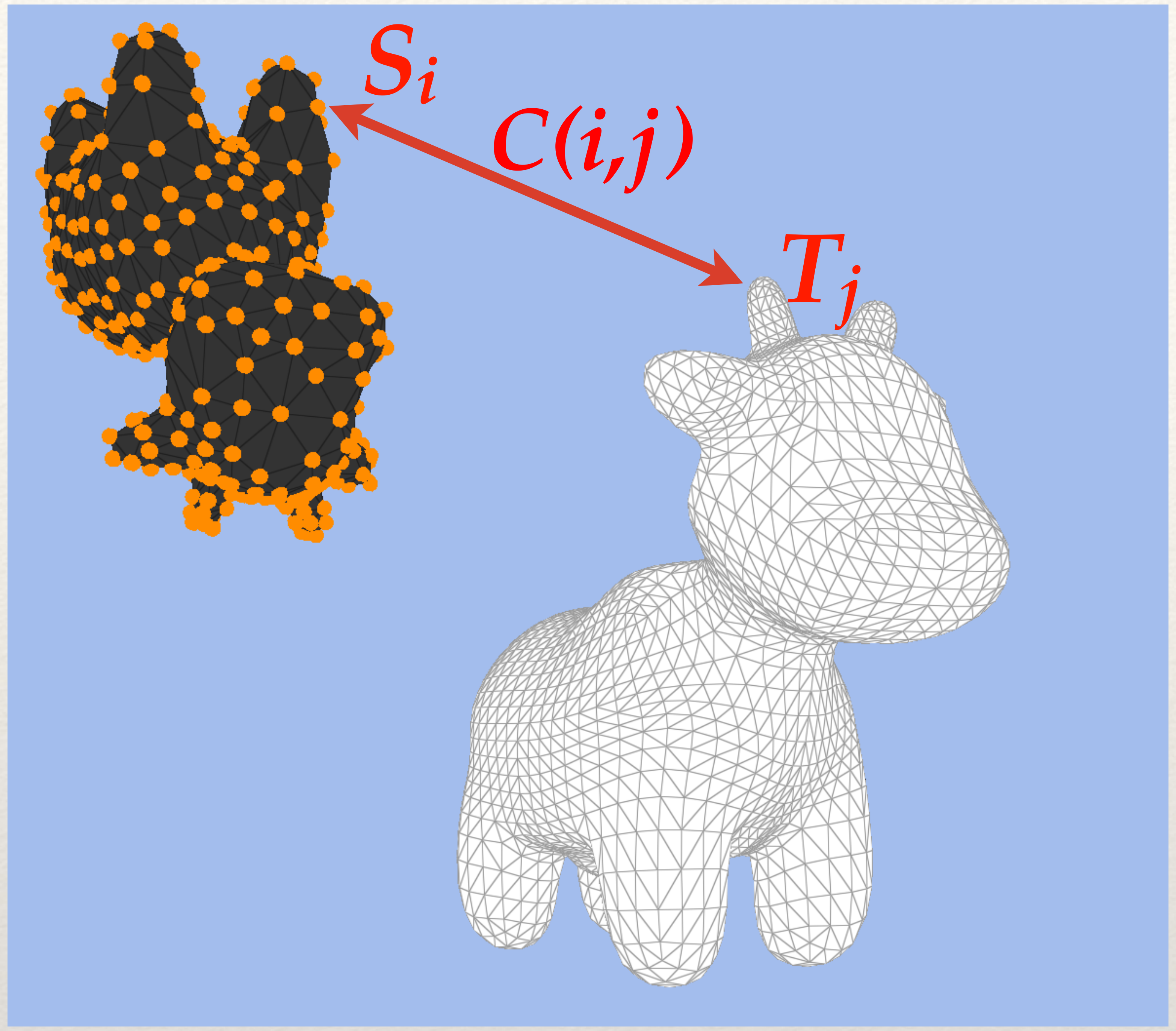
$$m_1(i) = \frac{1}{N}$$

$$T = \{T_j\}_{j \in [1,M]}$$

$$m_2(j) = \frac{1}{M}$$

$$C(i,j) = dist(S_i, T_j)$$

Comparing shapes



$$S = \{S_i\}_{i \in [1,N]}$$

$$m_1(i) = \frac{1}{N}$$

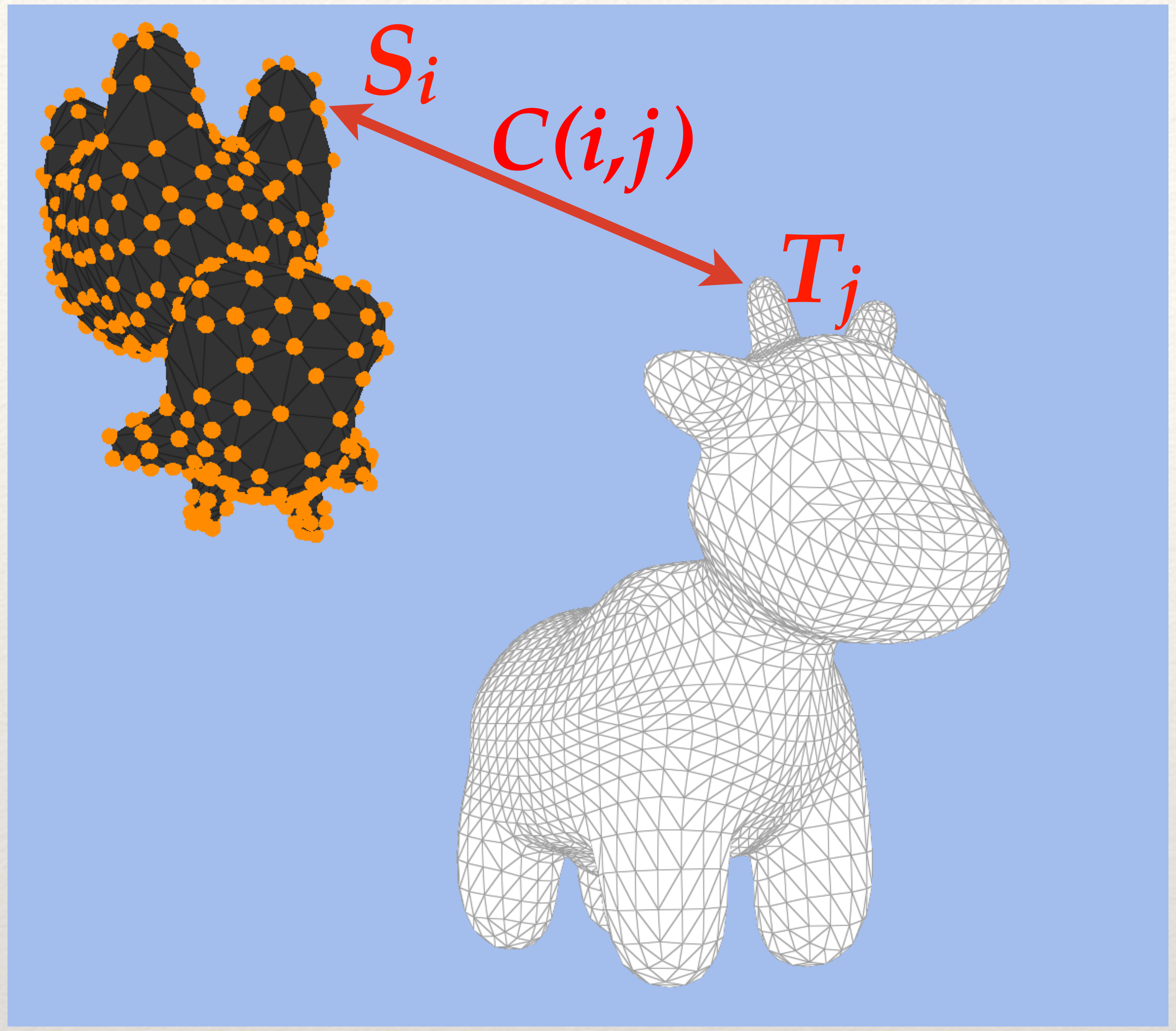
$$T = \{T_j\}_{j \in [1,M]}$$

$$m_2(j) = \frac{1}{M}$$

$$C(i,j) = dist(S_i, T_j)$$

$$d_{OT}(S,T) = \min_G \sum_i \sum_j C(i,j)G(i,j)$$

Comparing shapes



$$S = \{S_i\}_{i \in [1,N]}$$

$$m_1(i) = \frac{1}{N}$$

$$T = \{T_j\}_{j \in [1,M]}$$

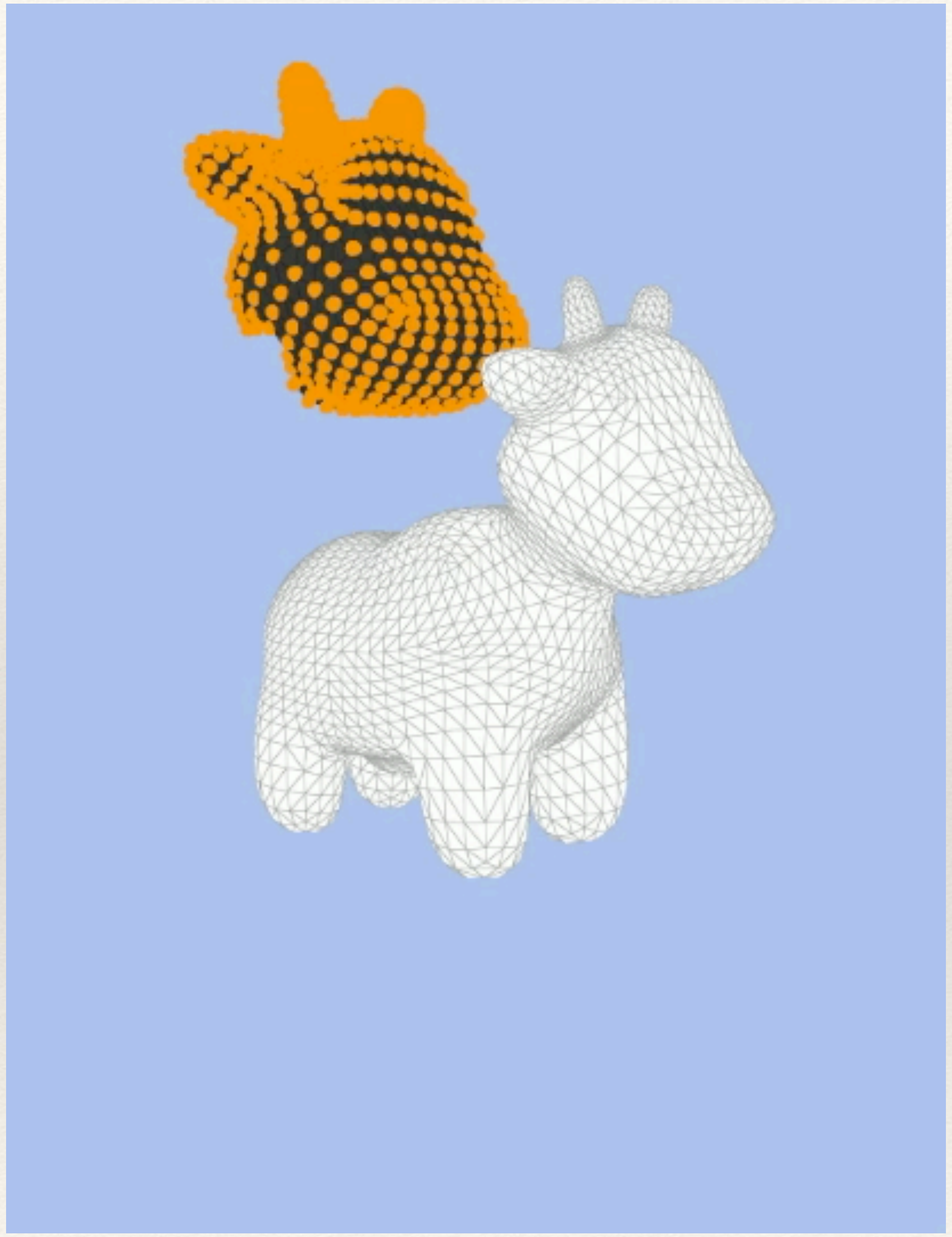
$$m_2(j) = \frac{1}{M}$$

$$C(i,j) = dist(S_i, T_j)$$

$$d_{OT}(S,T) = \min_G \sum_i \sum_j C(i,j)G(i,j)$$

$$d(S,T) = \min_R d_{OT}(R(S),T) = \min_R \left(\min_G \sum_i \sum_j C_R(i,j)G(i,j) \right)$$

Comparing shapes: Problem!



OT with variable masses

Minimize

$$U(G, m_1, m_2) = \sum_i \sum_j C(i,j)G(i,j)$$

under the constraints:

$$Gij \geq 0$$

$$\sum_j G(i,j) = m_1(i) \quad \forall i$$

$$\sum_i G(i,j) = m_2(j) \quad \forall j$$

P Koehl, M Delarue, H Orland
Physical Review E 103 (1), 012113 (2021)

OT with variable masses

Minimize

$$U(G, m_1, m_2) = \sum_i \sum_j C(i,j)G(i,j)$$

under the constraints:

$$Gij \geq 0$$

$$\sum_j G(i,j) = m_1(i) \quad \forall i \qquad m_1(i) \geq 0$$

$$\sum_i G(i,j) = m_2(j) \quad \forall j \qquad m_2(i) \geq 0$$

$$\sum_i \sum_j G(i,j) = 1$$

P Koehl, M Delarue, H Orland
Physical Review E 103 (1), 012113 (2021)

OT with variable masses

Minimize

$$U(G, m_1, m_2) = \sum_i \sum_j C(i,j)G(i,j) + \alpha_1 \sum_i m_1(i)^2 + \alpha_2 \sum_j m_2(j)^2$$

under the constraints:

$$Gij \geq 0$$

$$\sum_j G(i,j) = m_1(i) \quad \forall i$$

$$\sum_i G(i,j) = m_2(j) \quad \forall j$$

$$\sum_i \sum_j G(i,j) = 1$$

P Koehl, M Delarue, H Orland
Physical Review E 103 (1), 012113 (2021)

$$Z_\beta(S_1, S_2) = \int_0^1 \prod_{kl} dG(k,l) \int_{-\infty}^{+\infty} \prod_k dm_1(k) \int_{-\infty}^{+\infty} \prod_l dm_2(l) e^{-\beta\left(\sum_{kl} C(k,l)G(k,l) + \sum_k \alpha_1(k) m_1^2(k) + \sum_l \alpha_2(l) m_2^2(l)\right)}$$

$$\times \int_{-\infty}^{+\infty} \prod_k d\lambda(k) e^{-i\beta\lambda(k)\left(\sum_l G(k,l) - m_1(k)\right)} \int_{-\infty}^{+\infty} \prod_l d\mu(l) e^{-i\beta\mu(l)\left(\sum_k G(k,l) - m_2(l)\right)}$$

$$\times \int_{-\infty}^{+\infty} e^{-i\beta x\left(\sum_{kl} G(k,l) - 1\right)} dx$$

P Koehl, M Delarue, H Orland
Physical Review E 103 (1), 012113 (2021)

OT with variable masses

Effective free energy:

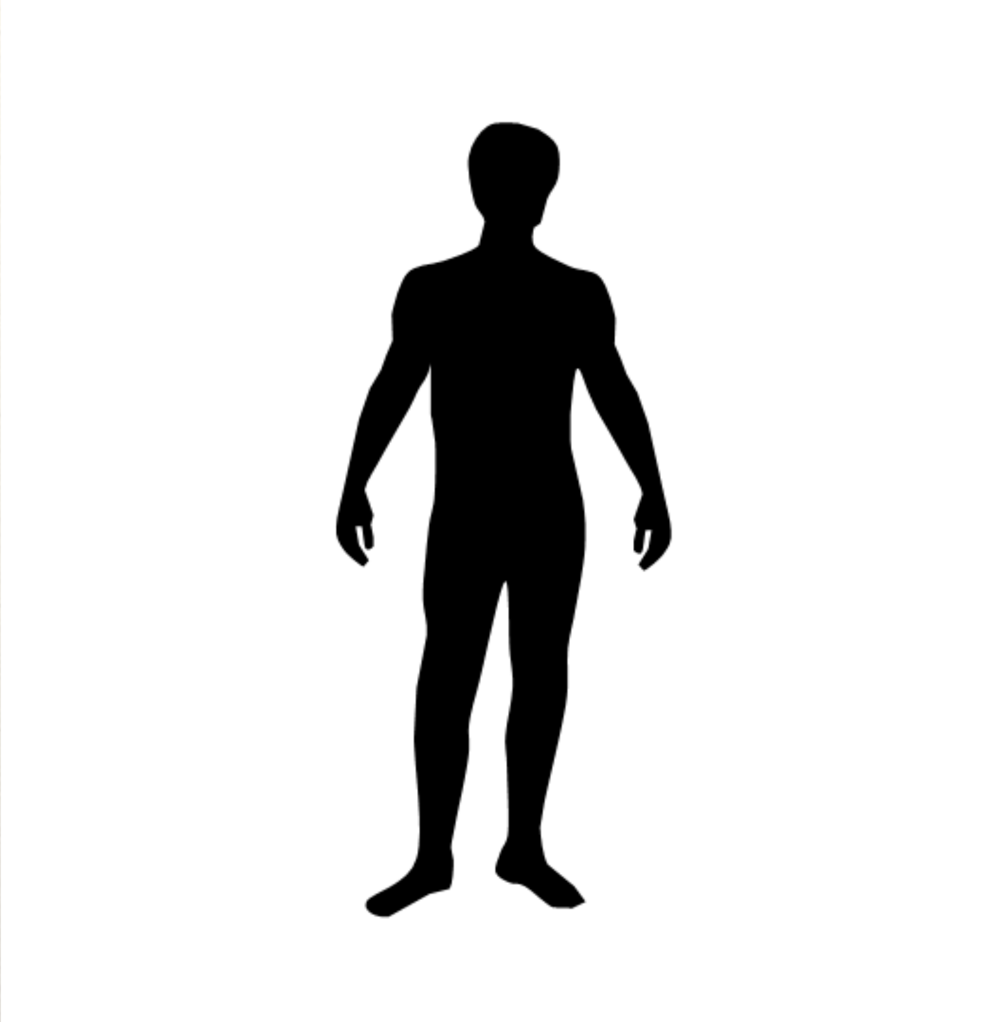
$$F_\beta\left(\boldsymbol{\lambda},\boldsymbol{\mu},x\right) = -x - \frac{1}{4\alpha_1}\sum_k \lambda_k^2 - \frac{1}{4\alpha_2}\sum_l \mu_l^2 - \frac{1}{\beta}\sum_{kl} \ln\left[\frac{1-e^{-\beta\left(C(k,l)+\lambda(k)+\mu(l)\right)}}{\beta\left(C(k,l)+\lambda(k)+\mu(l)\right)}\right]$$

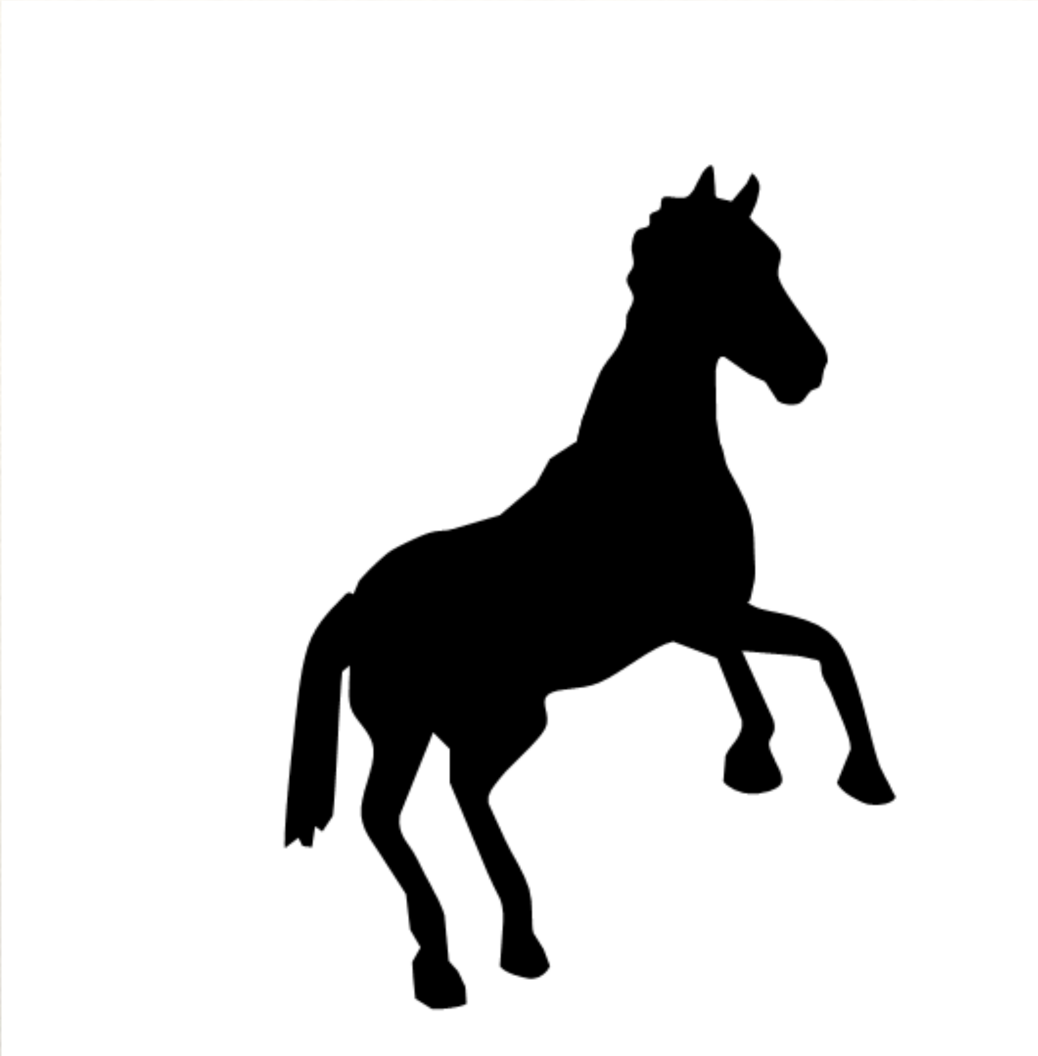
Saddle Point Approximation:

$$\begin{cases} \bar{G}(k,l) = \phi(\beta(C(k,l)+\lambda(k)+\mu(l))) \\ \sum_l \bar{G}(k,l) = \frac{\lambda_k}{2\alpha_1} \quad \forall k \\ \sum_k \bar{G}(k,l) = \frac{\mu_l}{2\alpha_2} \quad \forall l \\ \sum_{k,l} G(k,l) = 1 \end{cases}$$

where $\phi(x) = \dfrac{e^{-x}}{e^{-x}-1} + \dfrac{1}{x}$

OT with variable masses





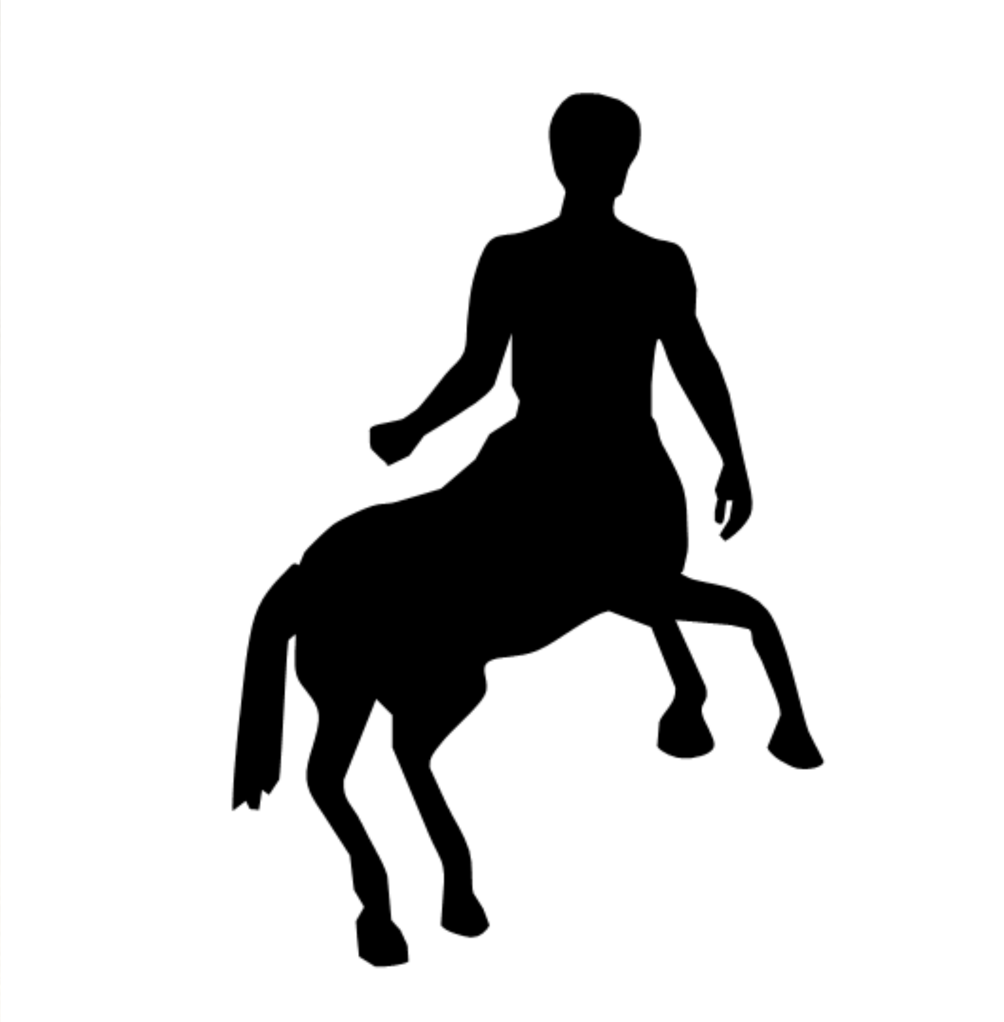
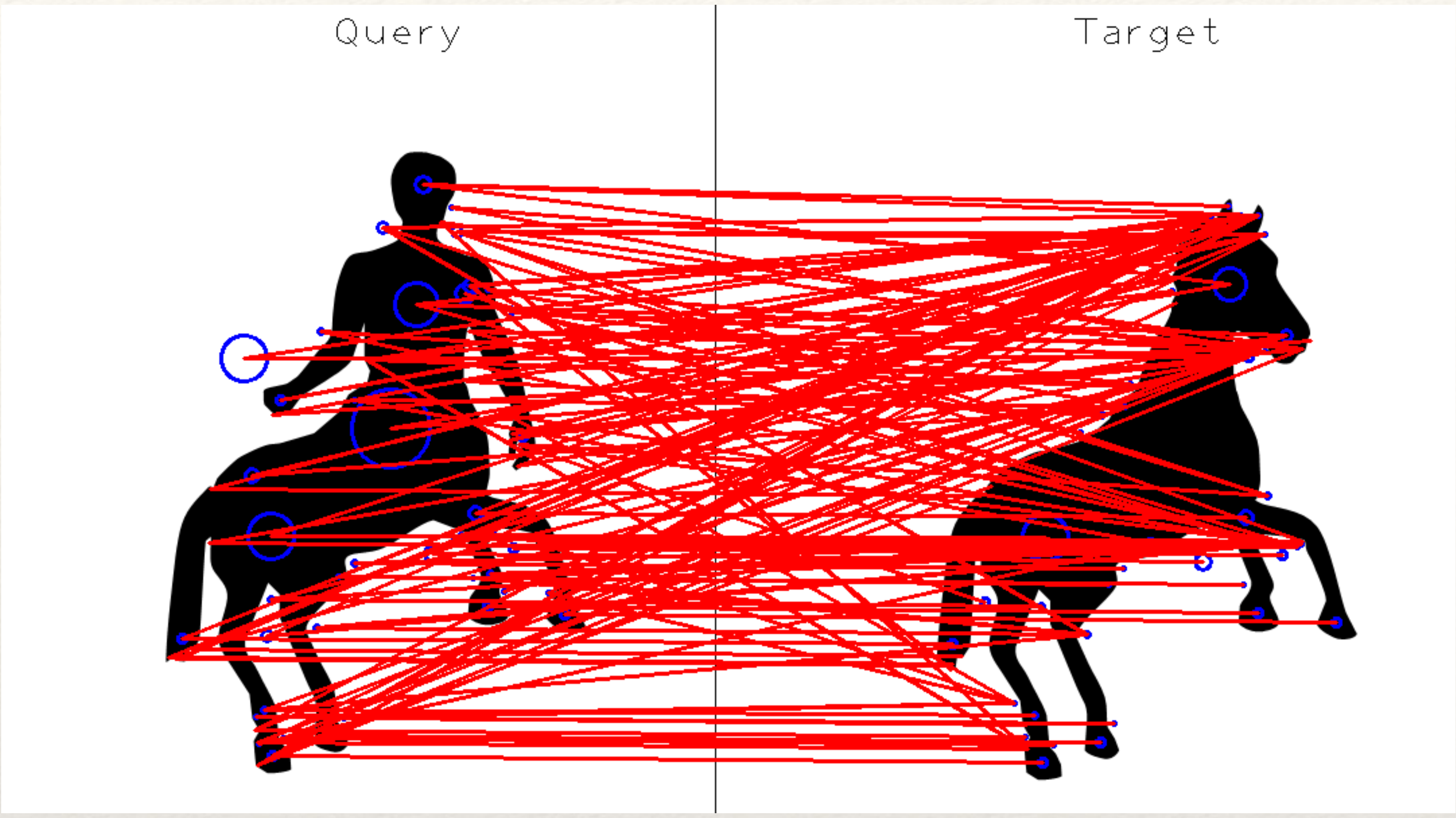


Image comparisons

Fixed masses



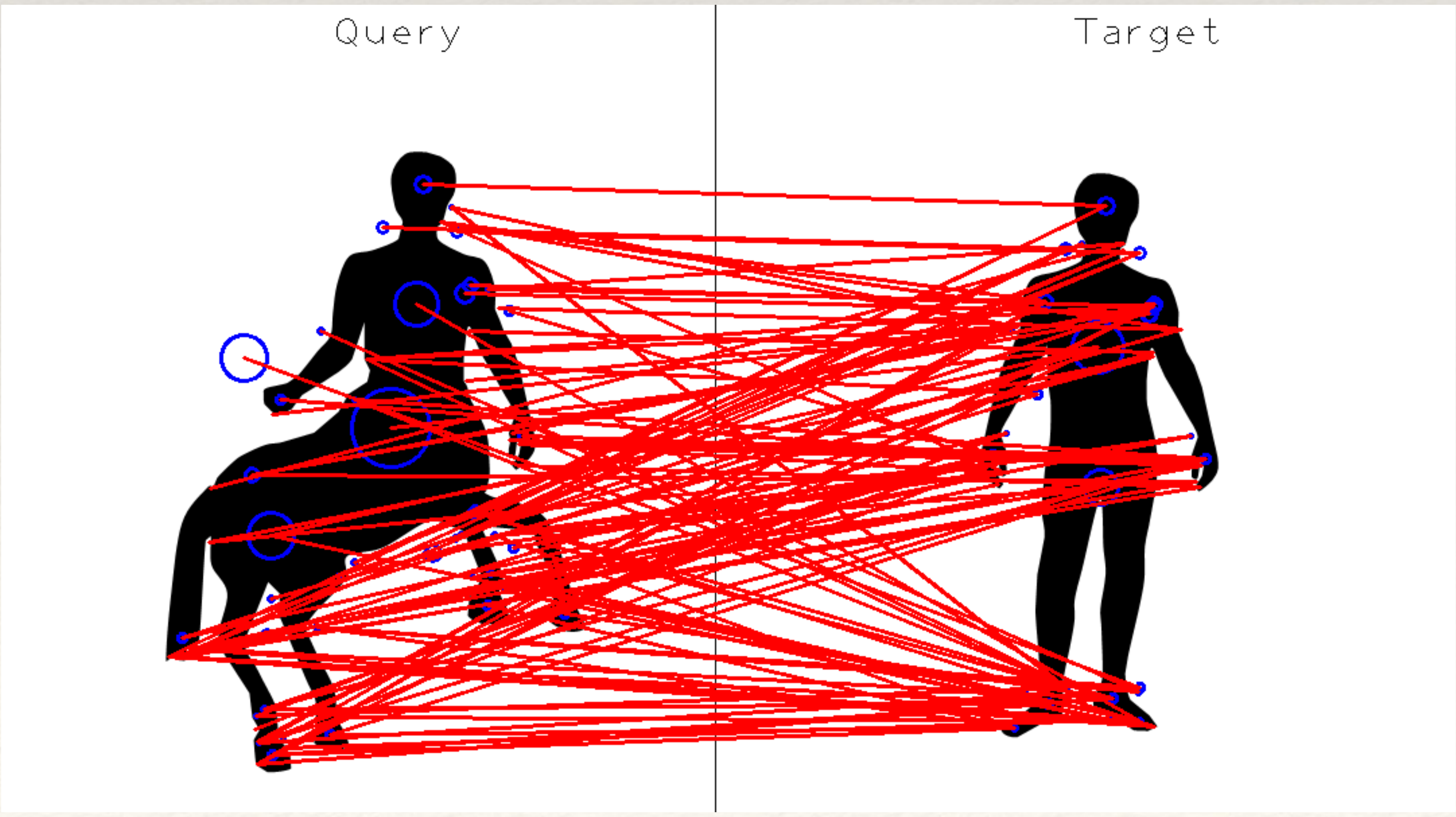
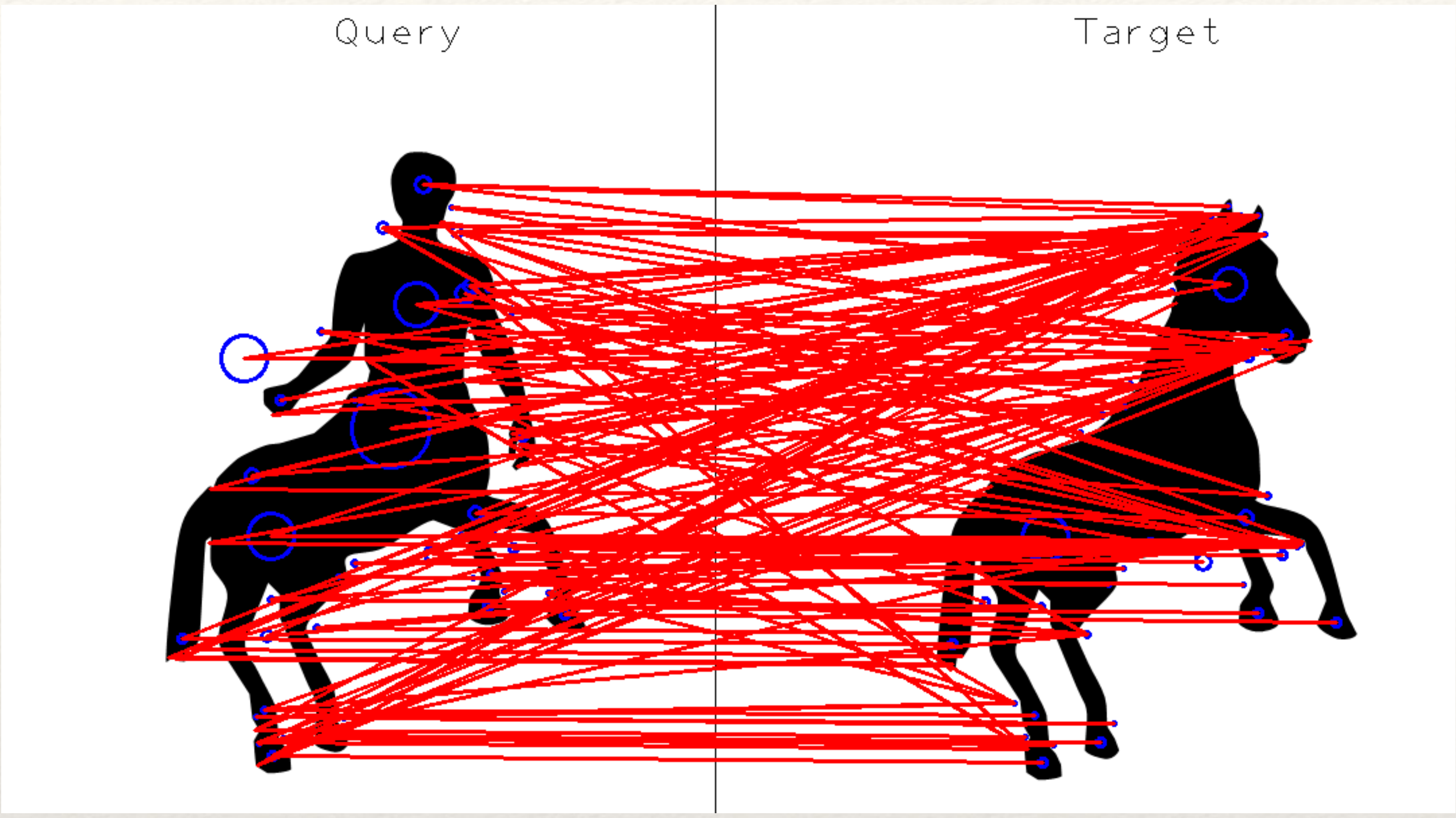
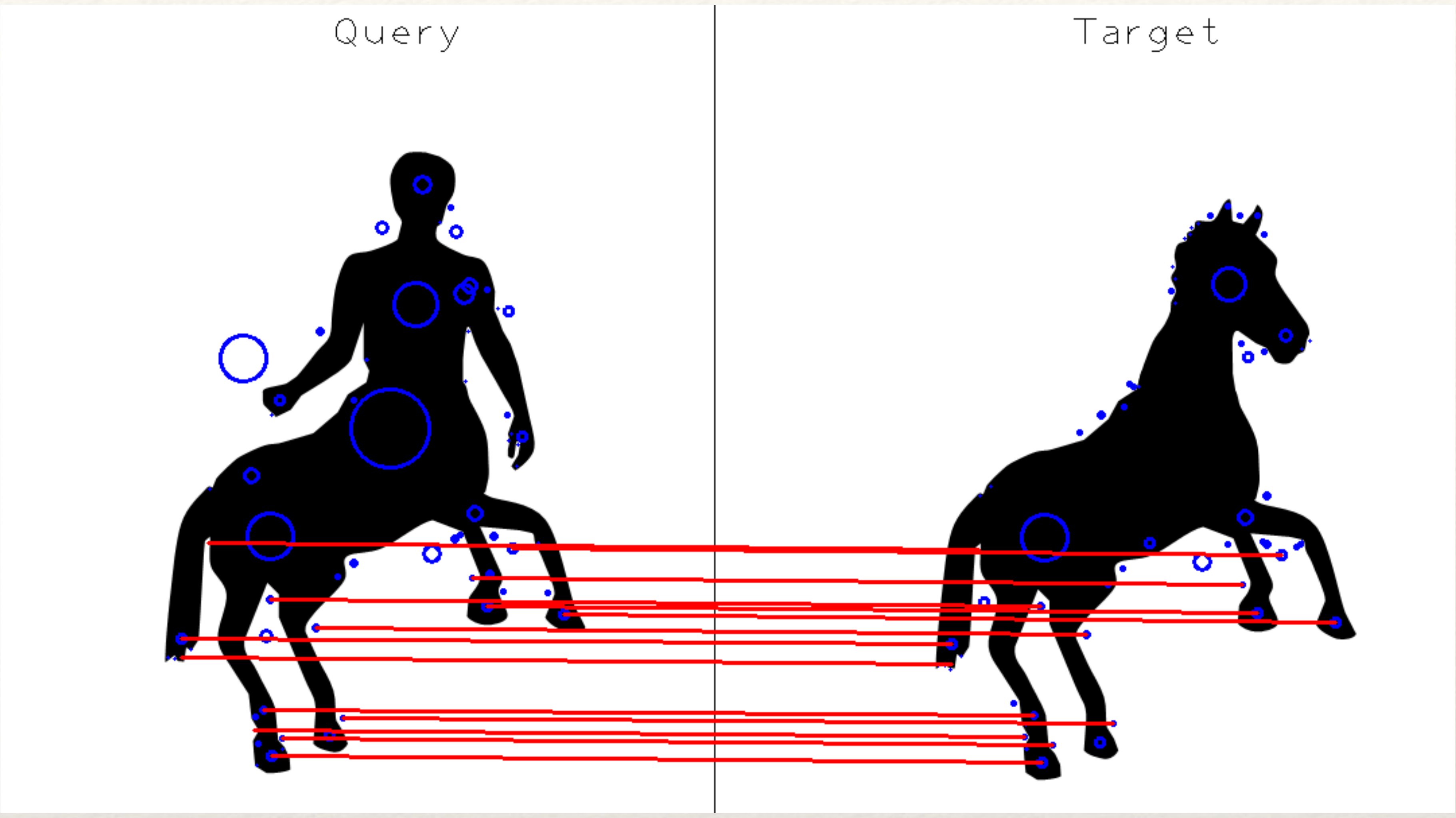


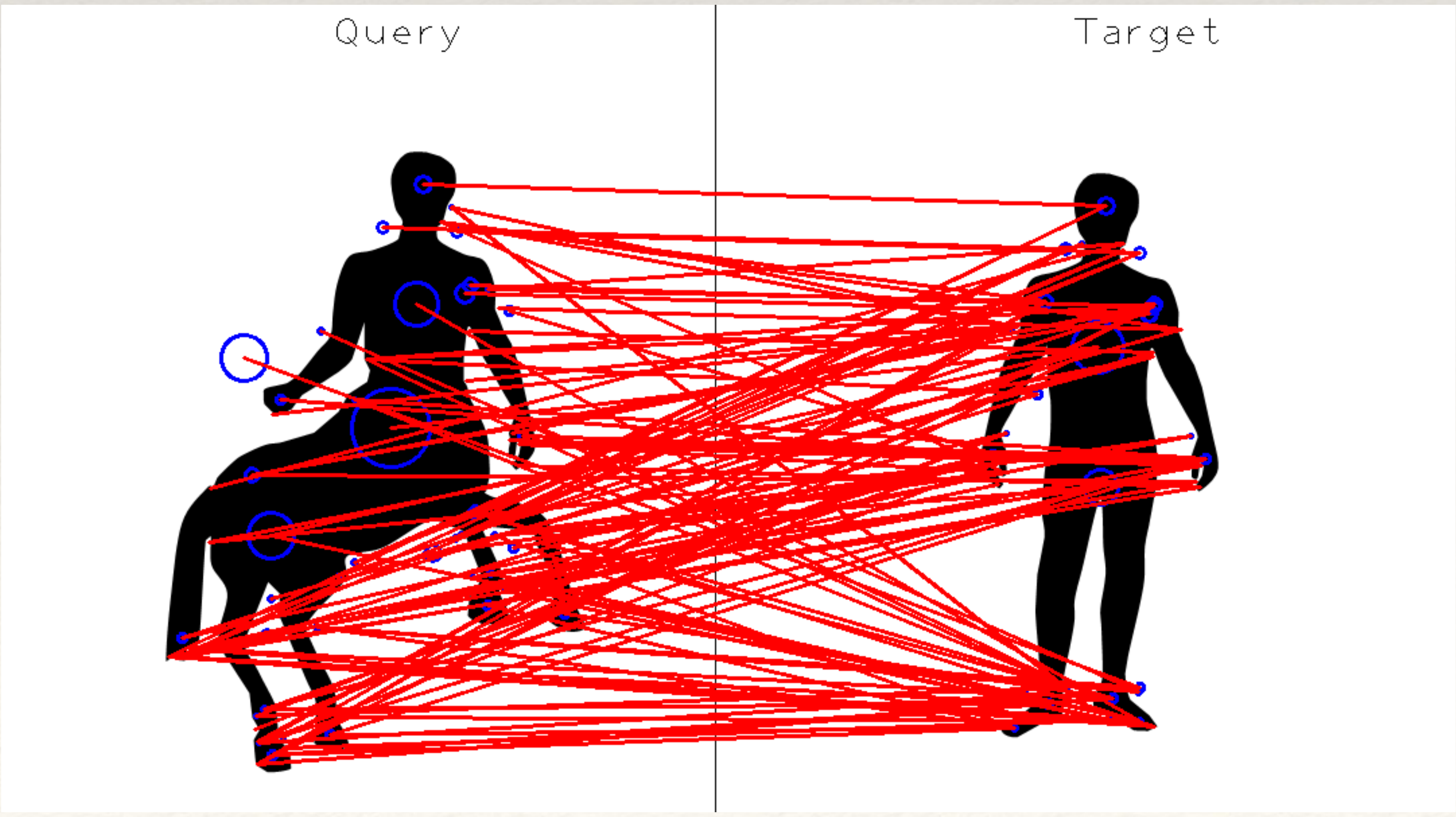
Image comparisons

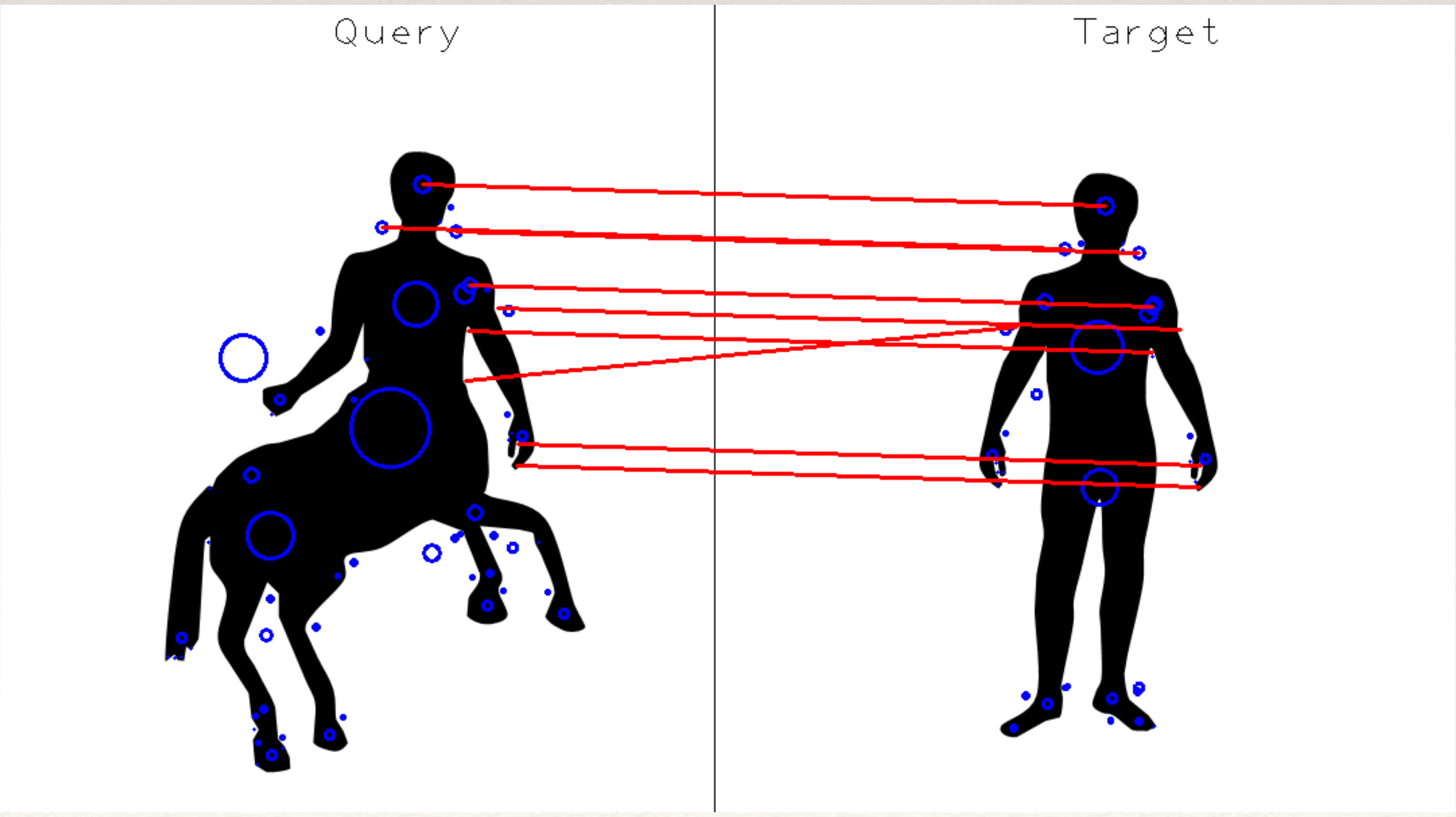
Fixed masses

Variable masses

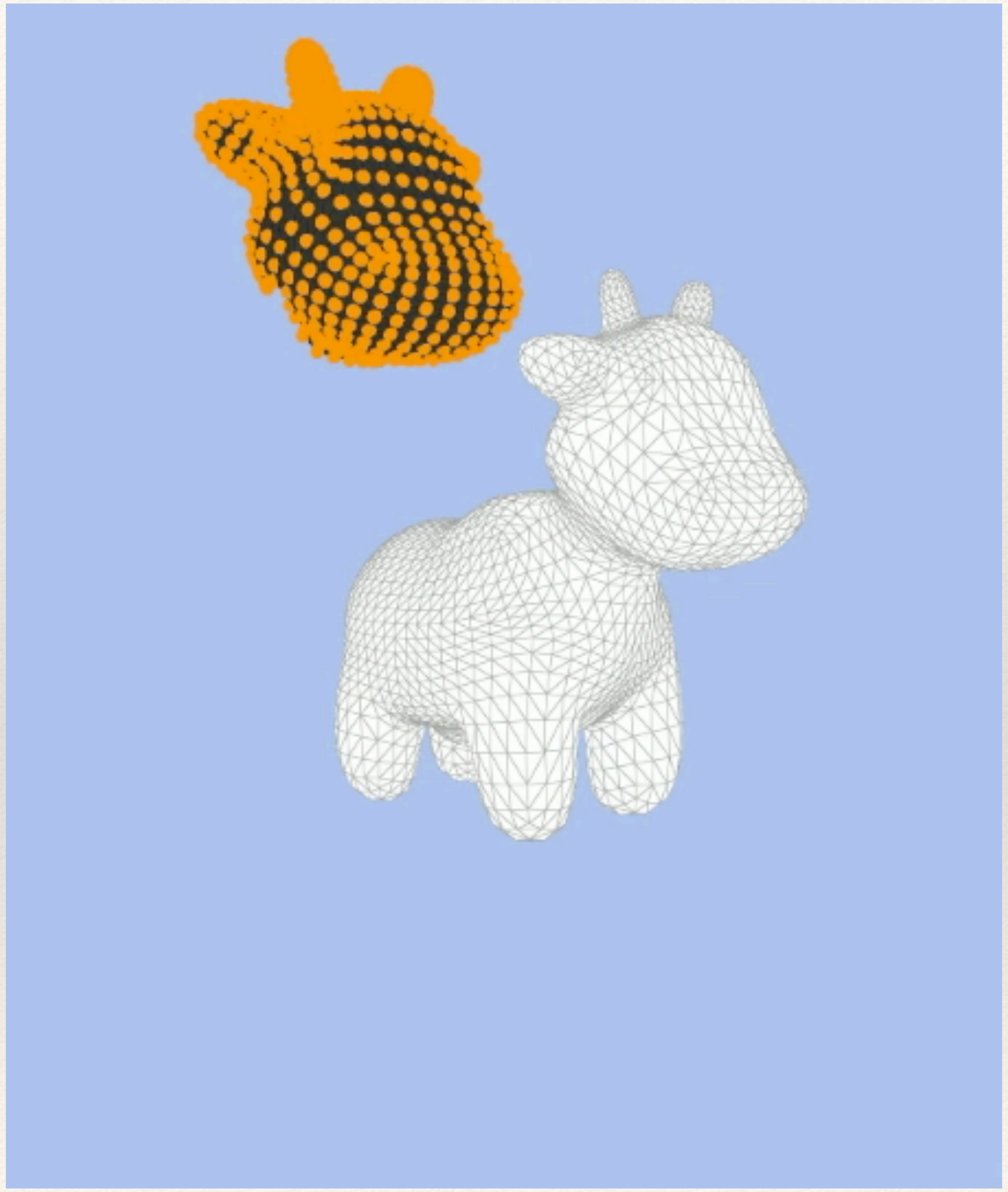




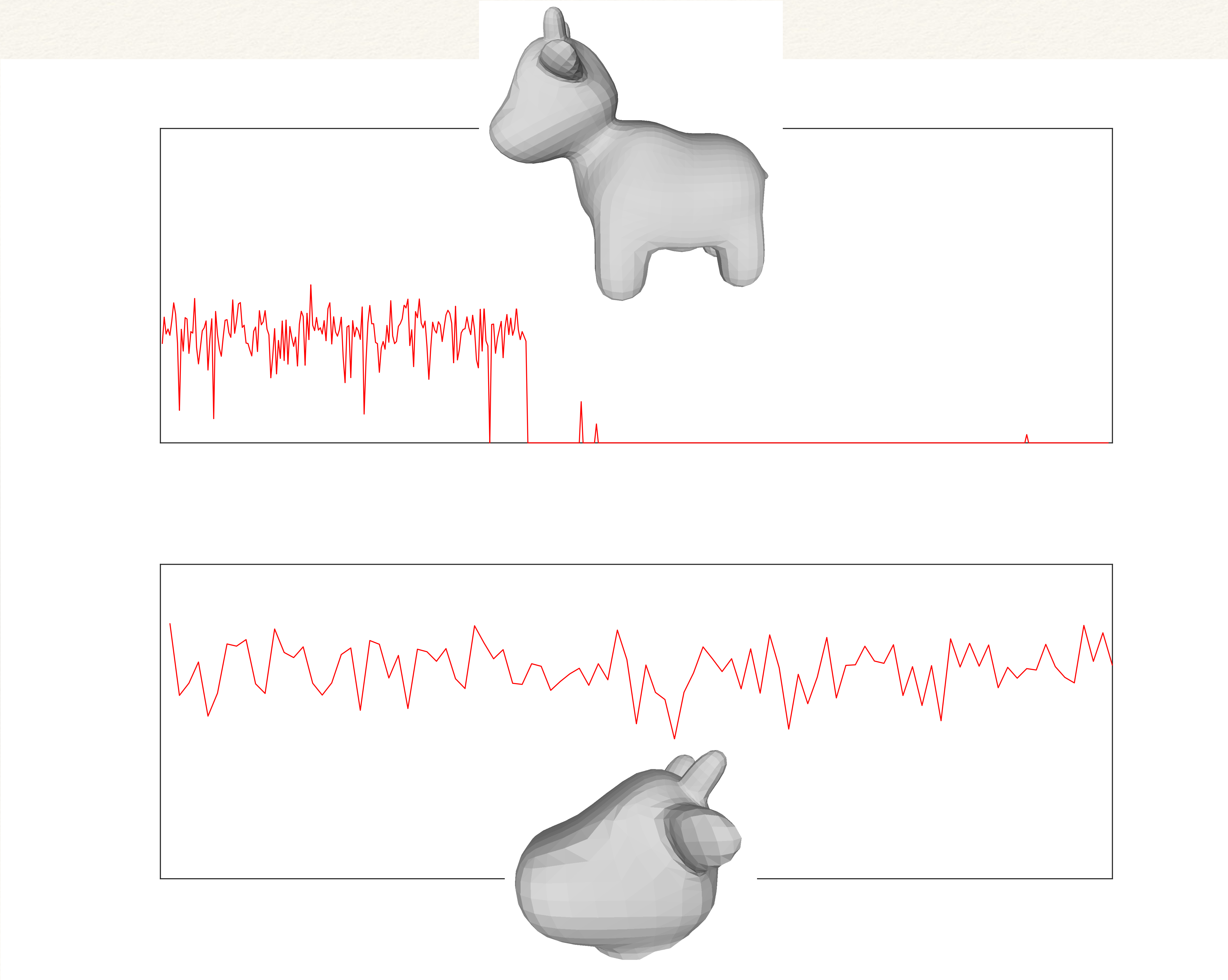




OT with variable masses



OT with variable masses

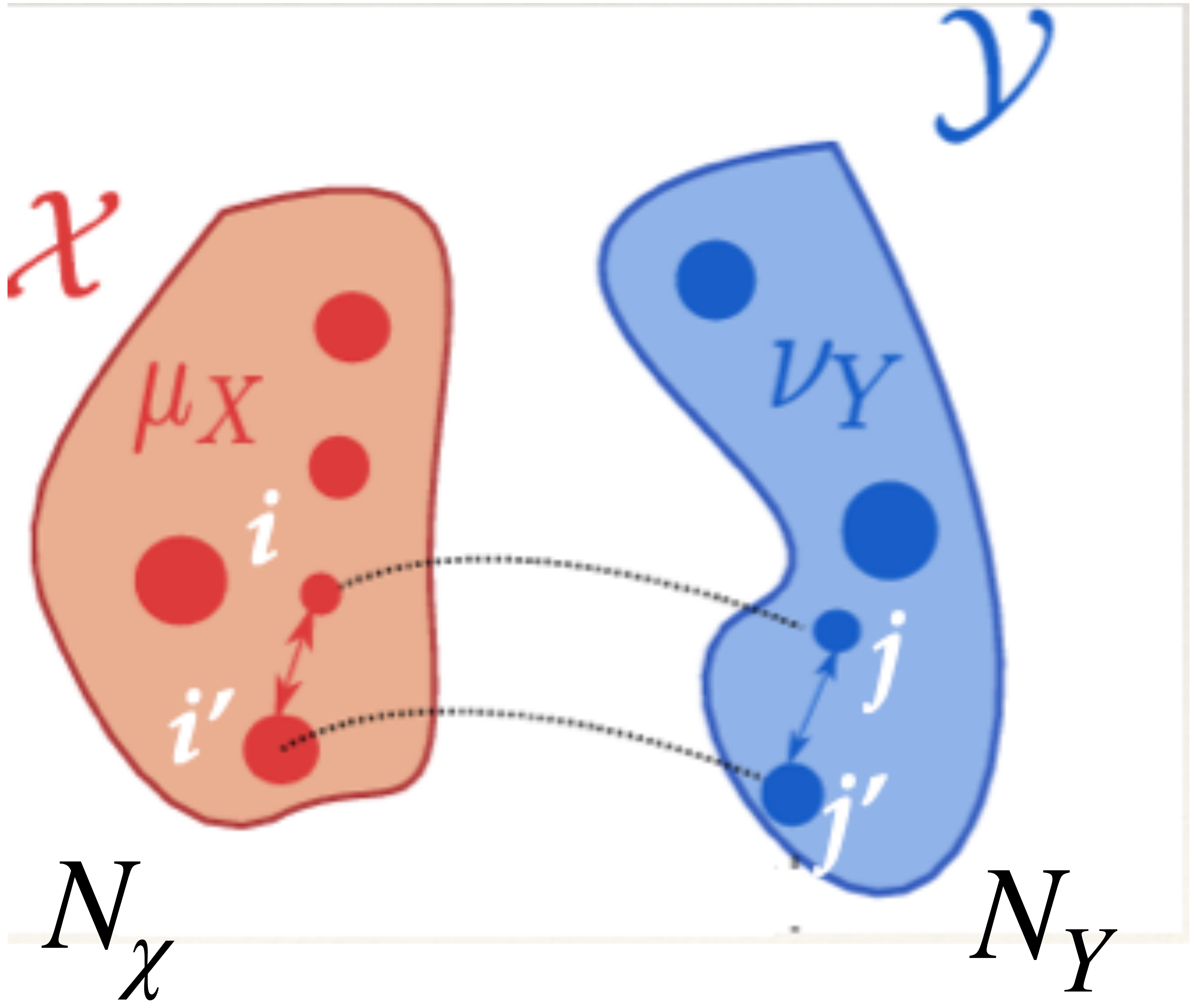


Gromov-Wasserstein OT

$$U_{GW} = \sum_{i,i'} \sum_{j,j'} G(i,j)(d_X(i,i') - d_Y(j,j'))^p G(i'j')$$

$$U_{GW} = \sum_{i,j} G(i,j) C_{GW}(i,j)$$

$$C_{GW}(i,j) = \sum_{i',j'} (d_X(i,i') - d_Y(j,j'))^p G(i'j')$$



$$G_{ij} \in [0,1]$$

Constraints

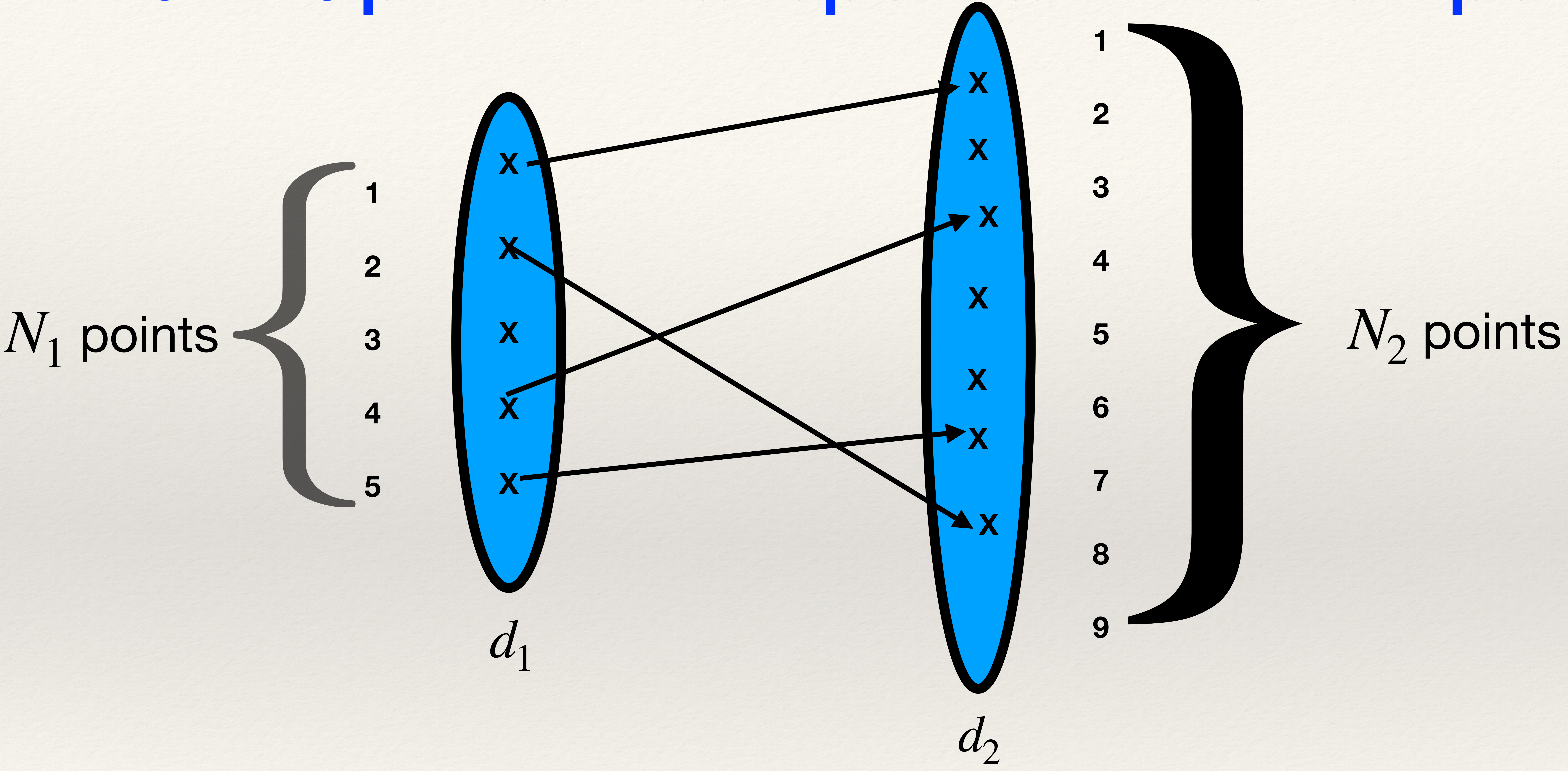
$$\forall i, \sum_j G_{ij} = m_1(i)$$

$$\forall j, \sum_i G_{ij} = m_2(j)$$

Balancing

$$\sum_i m_1(i) = \sum_j m_2(j)$$

GW Optimal Transport at finite temperature



Two different spaces with their own distances d_1 and d_2

GW Optimal Transport at finite temperature

$$E = \sum_{i,j} G(i,j)C(i,j)$$

With

$$C(i,j) = \sum_{i',j'} \left(d_1(i,i') - d_2(j,j')\right)^p G(i',j')$$

Invariance with respect to translations and rotations.
To account for the possibility of different scales, introduce a scale factor and minimize the energy w.r.t. this scale

$$C(i,j) = \sum_{i',j'} \left(d_1(i,i') - s \, . \, d_2(j,j')\right)^p G(i',j')$$

Partition Function

$$Z_\beta = \int_0^1 \prod_{ij} dG_{ij} \prod_i \delta\Big(\sum_j G_{ij} - m_1(i)\Big) \prod_j \delta\Big(\sum_i G_{ij} - m_2(j)\Big) e^{-\beta \sum_{i,j} \sum_{i',j'} G(i,j)\big(d_1(i,i') - s\,.\,d_2(j,j')\big)^p G(i',j')}$$

Introduce a cost variable

$$\begin{aligned} Z_\beta(s, S_1, S_2) = \quad & \int_0^1 \prod_{k,l} dG(k,l) \int_{-\infty}^{+\infty} \prod_{k,l} dC(k,l) \\ & e^{-\beta \sum\limits_{k,l} G(k,l) C(k,l)} \times \\ & \prod_k \delta\left(\sum_l G(k,l) - m_1(k)\right) \prod_l \delta\left(\sum_k G(k,l) - m_2(l)\right) \\ & \prod_{k,l} \delta\left(\sum_{k',l'} |d_1(k,k') - s \cdot d_2(l,l')|^p G(k',l') - C(k,l)\right). \end{aligned}$$

Use the Fourier representation of $\delta-$functions

$$Z_\beta(s,S_1,S_2) = \int_{-\infty}^{+\infty} \prod_{k,l} dC(k,l) \int_0^1 \prod_{k,l} dG(k,l) e^{-\beta \sum\limits_{k,l} C(k,l)G(k,l)} \times$$

$$\int_{-\infty}^{+\infty} \prod_k d\lambda(k) e^{-i\beta \sum\limits_{k,l} \lambda(k)G(k,l) + i\beta \sum\limits_k \lambda(k) m_1(k)} \int_{-\infty}^{+\infty} \prod_l d\mu(l) e^{-i\beta \sum\limits_{k,l} \mu(l)G(k,l) + i\beta \sum\limits_l \mu(l) m_2(l)}$$

$$\int_{-\infty}^{+\infty} \prod_{k,l} dD(k,l) e^{i\beta \sum\limits_{k,l} D(k,l)C(k,l) - i\beta \sum\limits_{k,l} D(k,l) \sum\limits_{k',l'} |d_1(k,k') - s \cdot d_2(l,l')|^p G(k',l')} .$$

Do the G integrals

After some manipulations, the partition function can be written as

$$Z_\beta = \int dCdDd\lambda d\mu \; e^{-\beta F_\beta}$$

where

$$\begin{aligned} F_\beta \quad &= -\sum_{k,l} D(k,l)C(k,l) - \left(\sum_k \lambda(k) m_1(k) + \sum_l \mu_l m_2(l) \right) \\ &+ \sum_{k,k'} \sum_{l,l'} D(k,l) \left| d_1(k,k') - s \cdot d_2(l,l') \right|^p D(k',l') \\ &- \frac{1}{\beta} \sum_{k,l} \ln \left(\frac{1 - e^{-\beta(C(k,l)+\lambda(k)+\mu(l))}}{\beta(C(k,l)+\lambda(k)+\mu(l))} \right). \end{aligned}$$

Saddle-Point Equations

Define average G_{ij} by $\bar{G}_{ij} = D_{ij}$

Standard Optimal Transport

$$\begin{cases} \bar{G}_{ij} = \phi\left(\beta(C_{ij} + \lambda_i + \mu_j)\right) \\ \sum_j \bar{G}_{ij} = m_1(i) \\ \sum_i \bar{G}_{ij} = m_2(j) \end{cases}$$

$$\phi(x) = \frac{e^{-x}}{e^{-x} - 1} + \frac{1}{x}.$$

Self-consistency condition

$$C_{ij} = 2 \sum_{i'j'} \left(d_1(i,i') - s\,.\,d_2(j,j')\right)^p \bar{G}_{i'j'}$$

The scaling factor is obtained by minimising the free energy w.r.t. s

$$\sum_{k,k'}\sum_{l,l'} D(k,l)d_2(l,l')\left|d_1(k,k') - s \cdot d_2(l,l')\right|^{p-2}(d_1(k,k') - s \cdot d_2(l,l'))D(k',l') = 0,$$

If $p = 2$

$$s = \frac{\sum_{k,k'}\sum_{l,l'} D(k,l)d_1(k,k')d_2(l,l')D(k',l')}{\sum_{k,k'}\sum_{l,l'} D(k,l)d_2(l,l')^2 D(k',l')}$$

1. Do a temperature annealing from high to low temperature
2. Start with an initial $\bar{G}_{ij}$, for instance $\bar{G}_{ij} = m_1(i)m_2(j)$ and scaling factor $s = 1$
3. Calculate C_{ij}
4. Solve optimal transport with C_{ij} and get new $\bar{G}_{ij}$ and new s
5. Back to 3 till convergence
6. Decrease temperature till low enough temperature.
7. Back to 3 till convergence

TOSCA Dataset: 11 classes, 133 shapes of 3400 vertices and 6600 faces

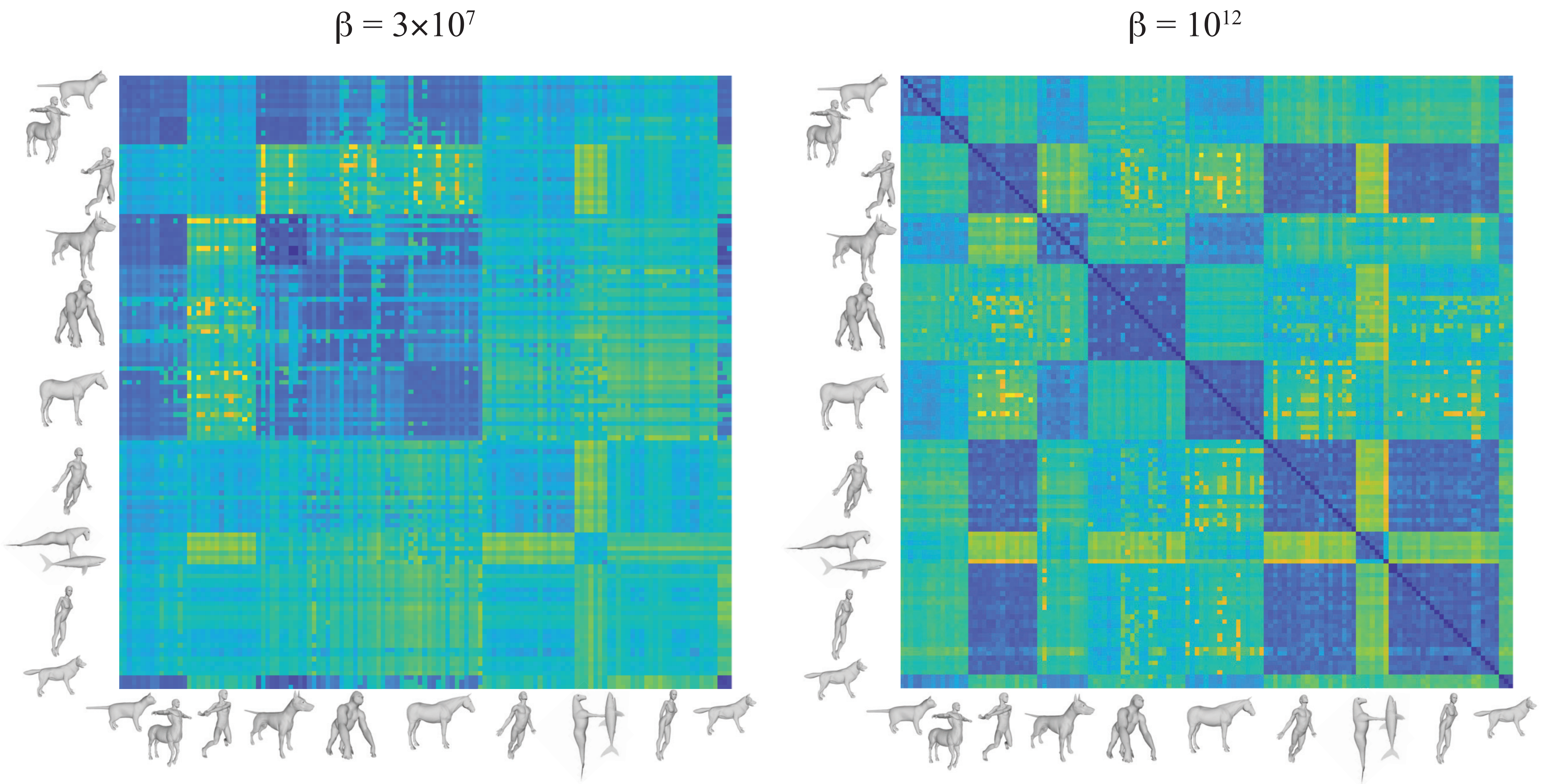
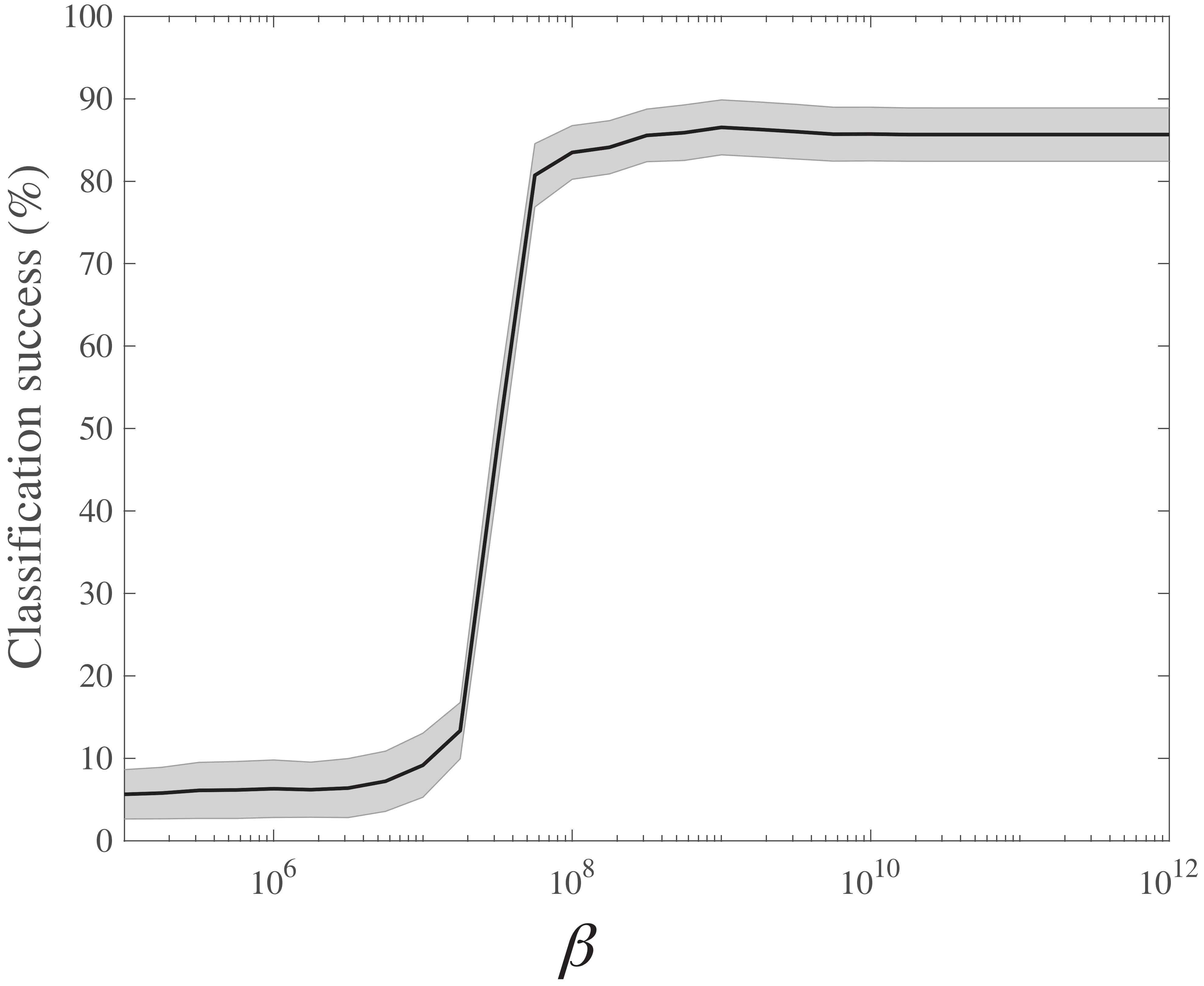


Figure 1. Distance matrices for shape similarity within the TOSCA dataset using the Gromov-Wassertein framework at two different "temperatures", $\beta = 3 \times 10^7$ (**left**) and $\beta = 10^{12}$ (**right**). Blue colors represent small distances (high similarity), while yellow colors represent large distances



Quality of 3d shape recognition (10000 experiments)

Conclusion

- Statistical mechanics is a natural framework to study assignment and optimal transport problems.
- Numerical methods are more stable than standard entropy regularisation
- Possible to do local matching by using variable masses
- GW OT very powerful tool but requires large computational resources