A Wasserstein-type distance in the space of Gaussian mixture models

Julie Delon, Agnès Desolneux

Séminaire du LJLL, 16/10/2020

Mixture models

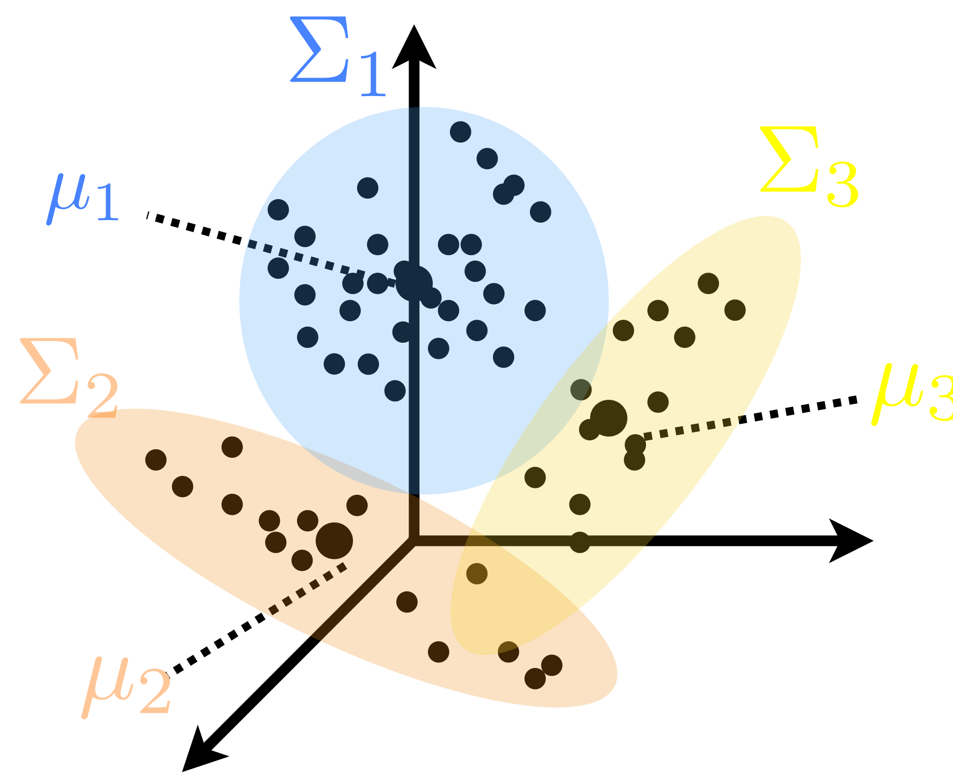
Mixture models remain popular tools in applications (image, vision, machine learning, etc.)

Gaussian mixture models (GMM) among the most used

$$X \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k)$$

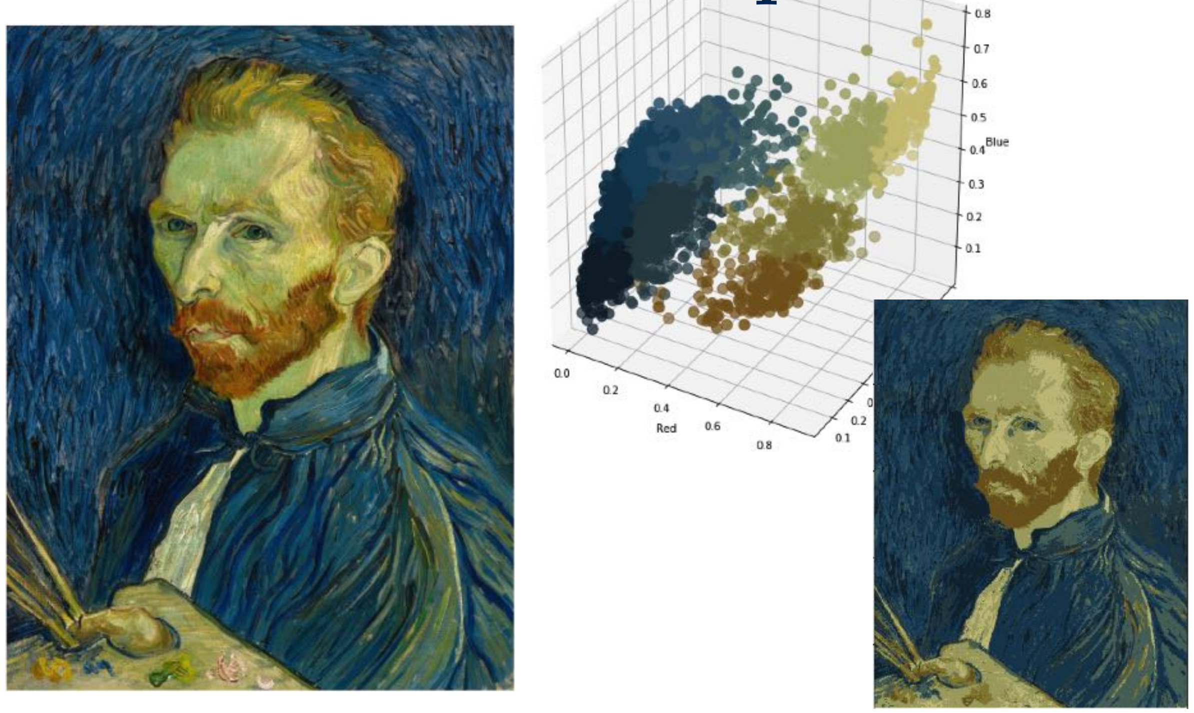
$x_1, \ldots, x_n$ assumed independent realisations of X on $\mathbb{R}^d$

Inference with EM algorithm.

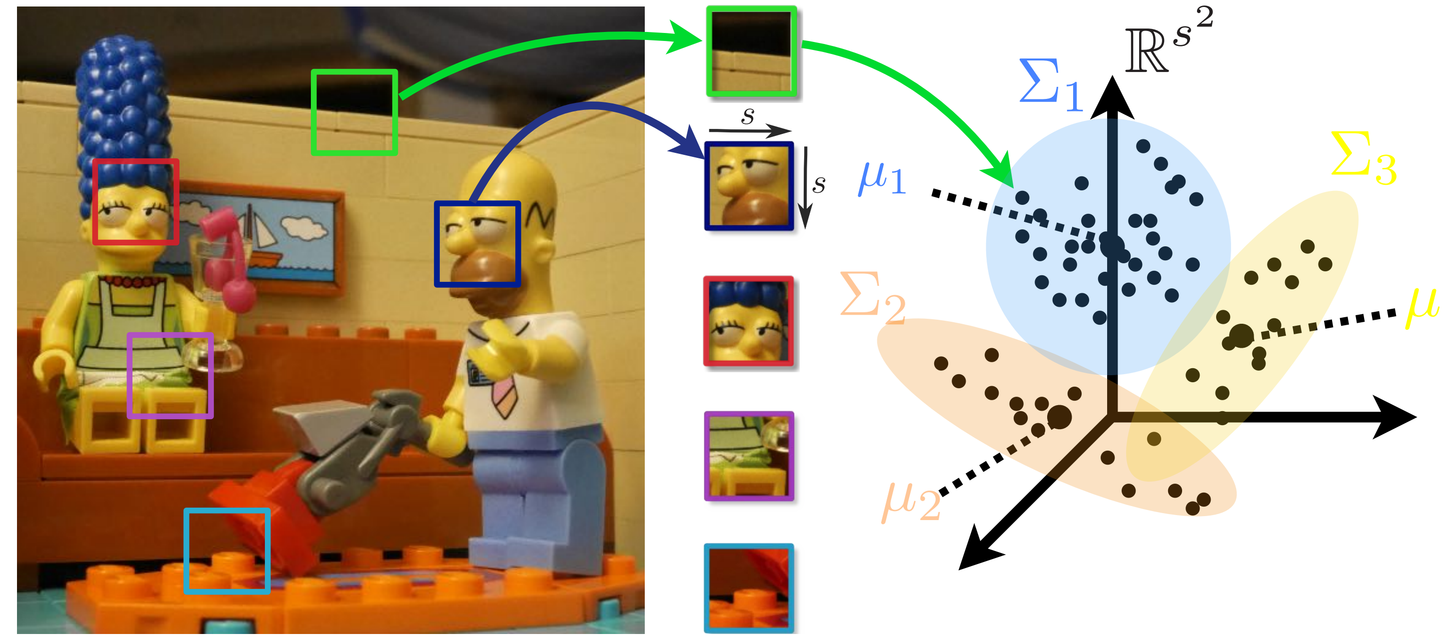


GMMs in image processing

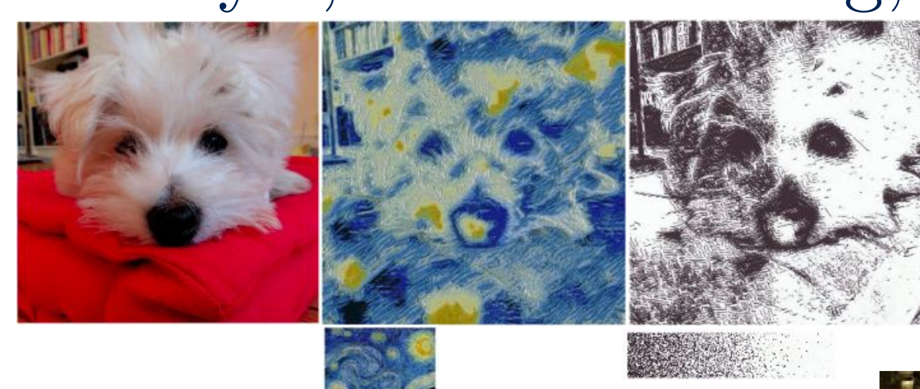
GMM on pixels



GMM on patches



Style, texture editing, restoration



[Leclaire, Rabin 2019]

[EPLL 2011, SPLE 2013, HDMI 2018]

GMM on images

[Richardson, Weiss 2018]

Optimal transport

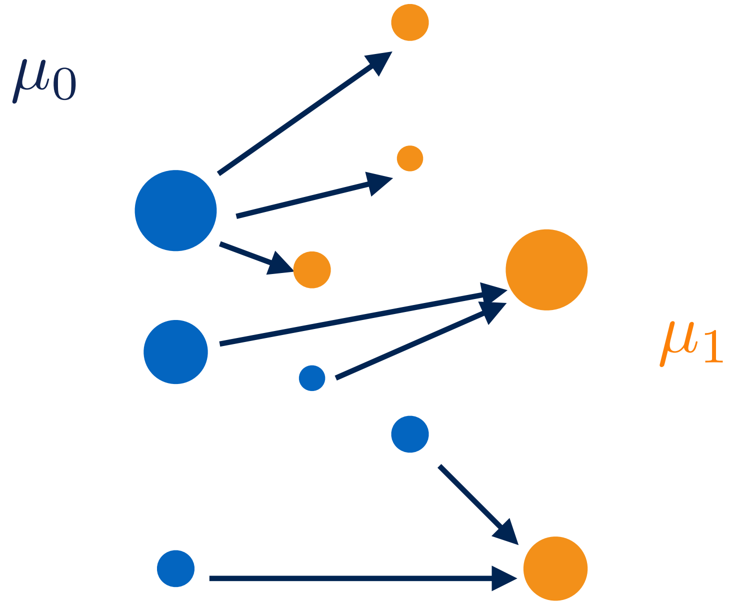
<u>Couplings:</u>

$\Pi(\mu_0, \mu_1)$ = probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0 and μ_1

[Kantorovich, *On the transfer of masses*, 1942]

$$\inf_{Y_0 \sim \mu_0; Y_1 \sim \mu_1} \mathbb{E}(c(Y_0, Y_1))$$

$$= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(y_0, y_1) d\gamma(y_0, y_1).$$



Discrete case: Linear programming

$$\mu_0 = \sum_i \mu_0[i] \delta_{y_0^i}, \quad \mu_1 = \sum_j \mu_1[i] \delta_{y_1^i}$$

$$\inf_{\gamma \in \Pi(\mu_0, \mu_1)} \sum_{i,j} c(y_0^i, y_1^j) d\gamma_{ij}.$$

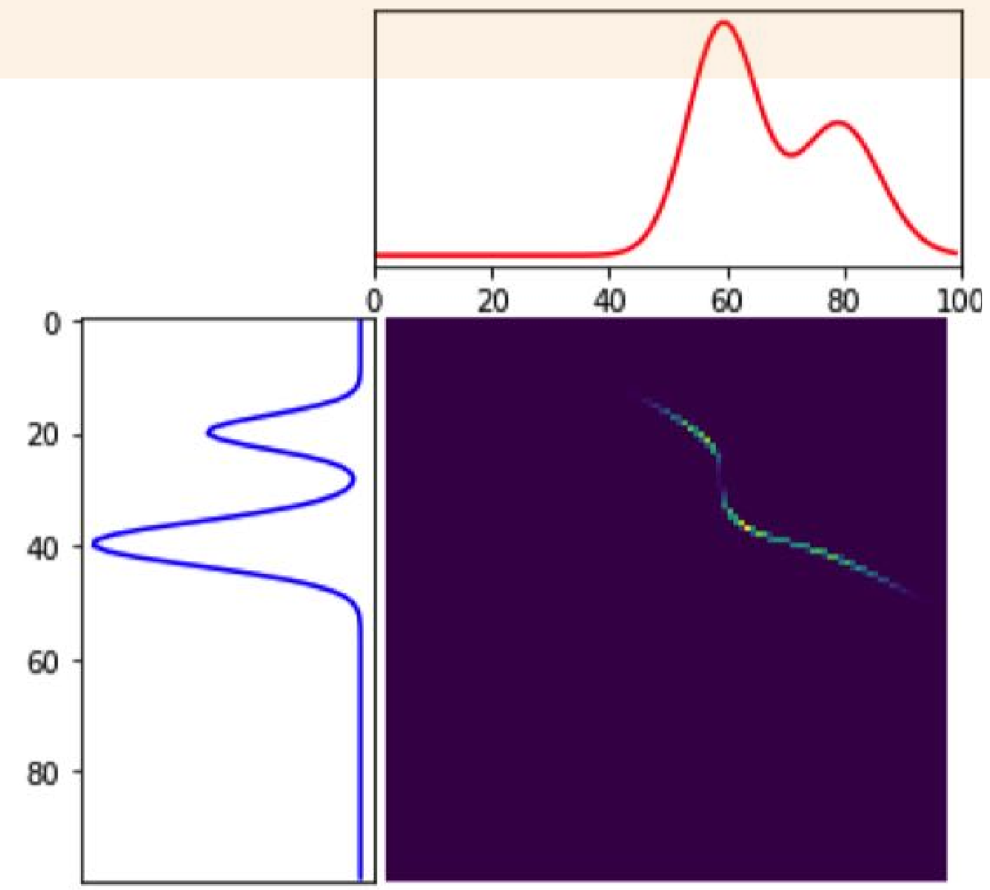
$c(y_0, y_1) = \|y_0 - y_1\|^2$

Wasserstein distance

$$W_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1).$$

Proposition: W_2 is a distance on pr. distr. with moments of order 2.

If μ_0 is abs. continuous, then $\gamma = (Id, T)\#\mu_0$ with $T\#\mu_0 = \mu_1$.



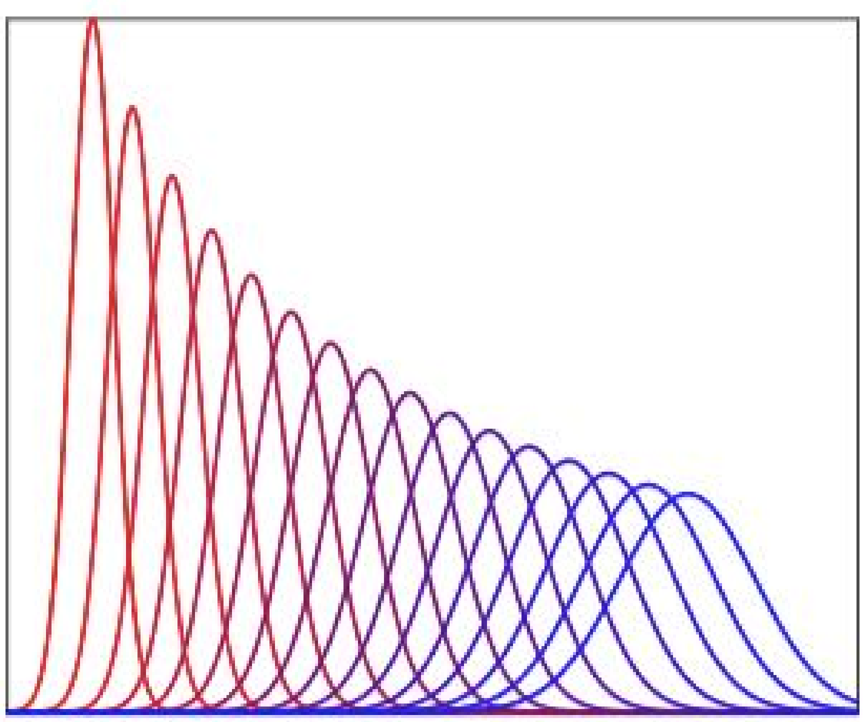
Optimal transport between Gaussians

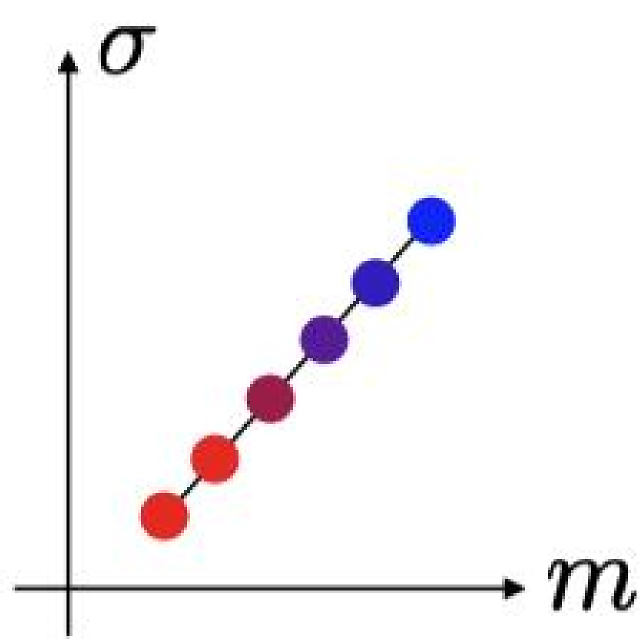
$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0,1\}$ two Gaussian distributions on $\mathbb{R}^d$

$$W_2^2(\mu_0,\mu_1) = \|m_0 - m_1\|^2 + \underbrace{\mathrm{tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)}_{B^2(\Sigma_0,\Sigma_1)}$$

If Σ_0 non-singular, affine optimal map

$$T(x) = m_1 + \Sigma_0^{-\frac{1}{2}}\left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\Sigma_0^{-\frac{1}{2}}(x - m_0)$$





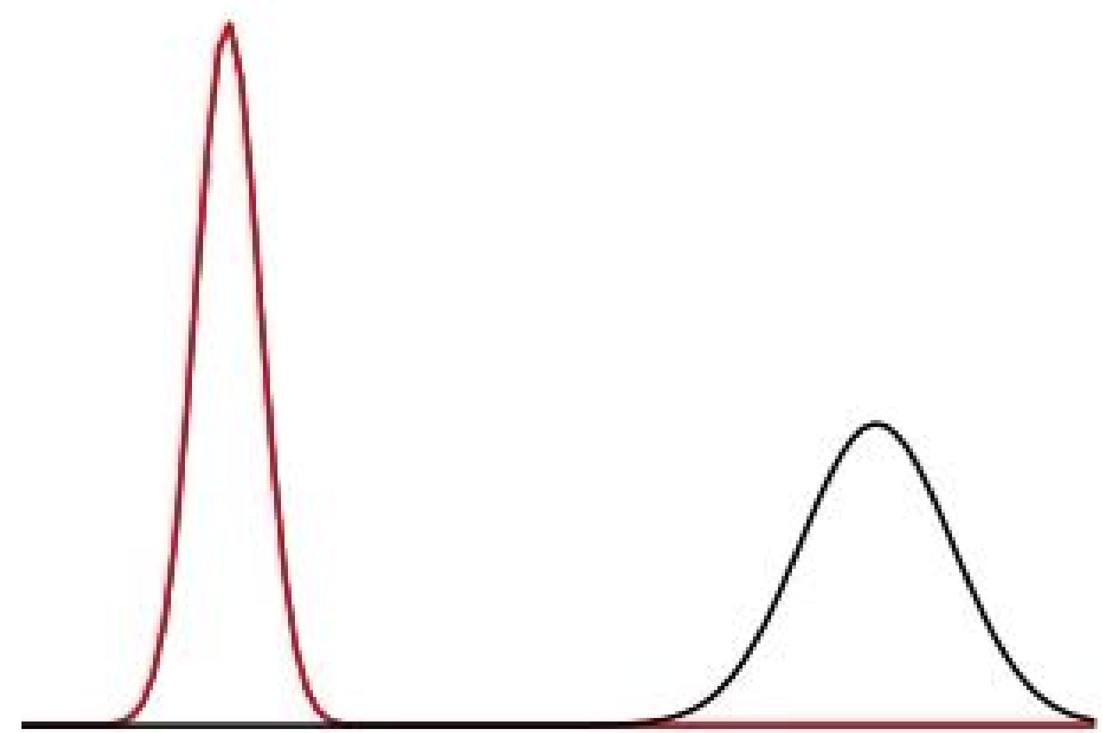


illustration: Cuturi, Peyré, OT book

Barycenters between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, \dots, I-1\}$ Gaussian distributions on $\mathbb{R}^d$

Barycenter [Agueh, Carlier 2011]:

$$\mathrm{argmin}_\mu \sum_{i=0}^{I-1} \lambda_i W_2^2(\mu_i, \mu) = \mathcal{N}(m^*, \Sigma^*)$$

$$m^* = \sum \lambda_i m_i \qquad \Sigma^* = \min_\Sigma \sum_i \lambda_i \mathrm{B}(\Sigma, \Sigma_i)^2$$

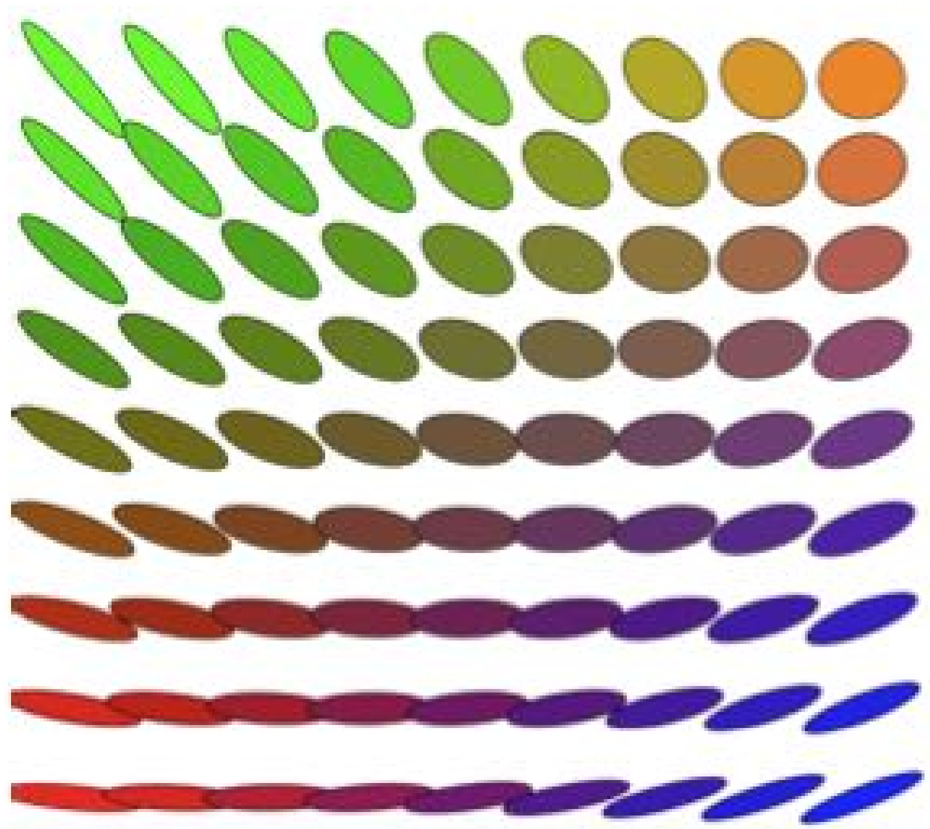
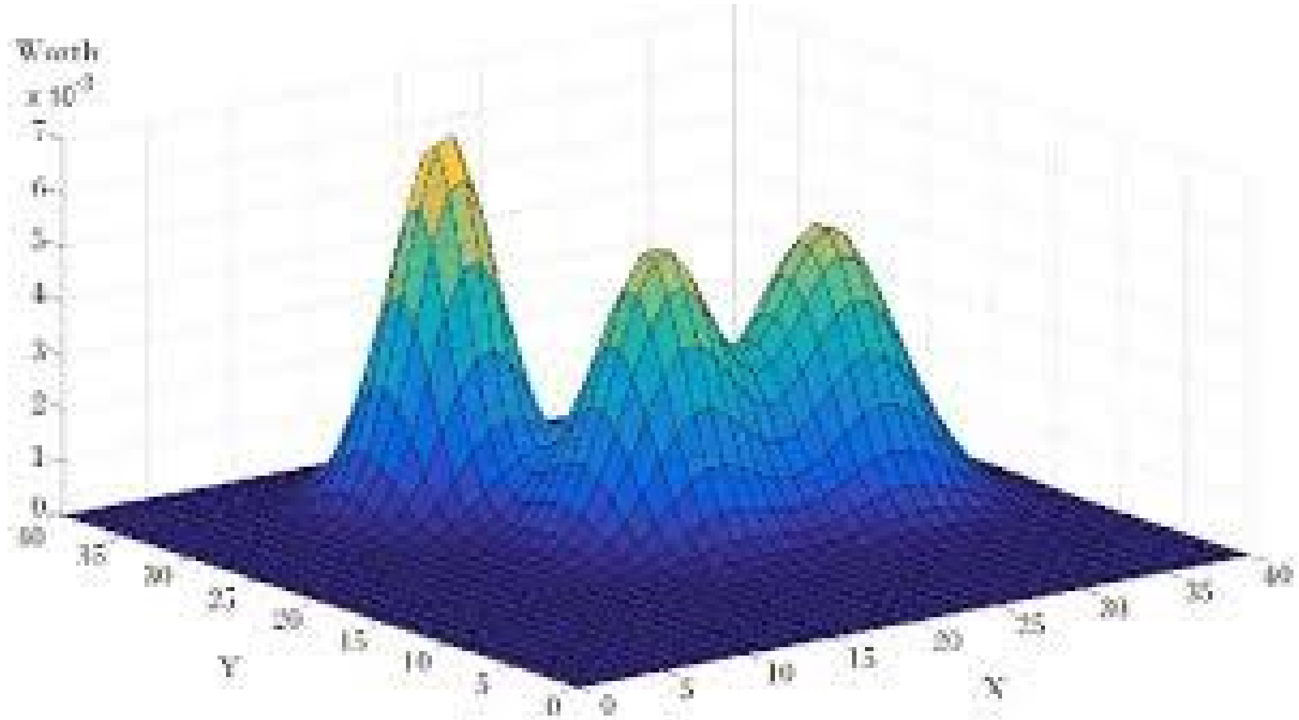


illustration: Cuturi, Peyré, OT book

Texture mixing [Xia et al, 2014]

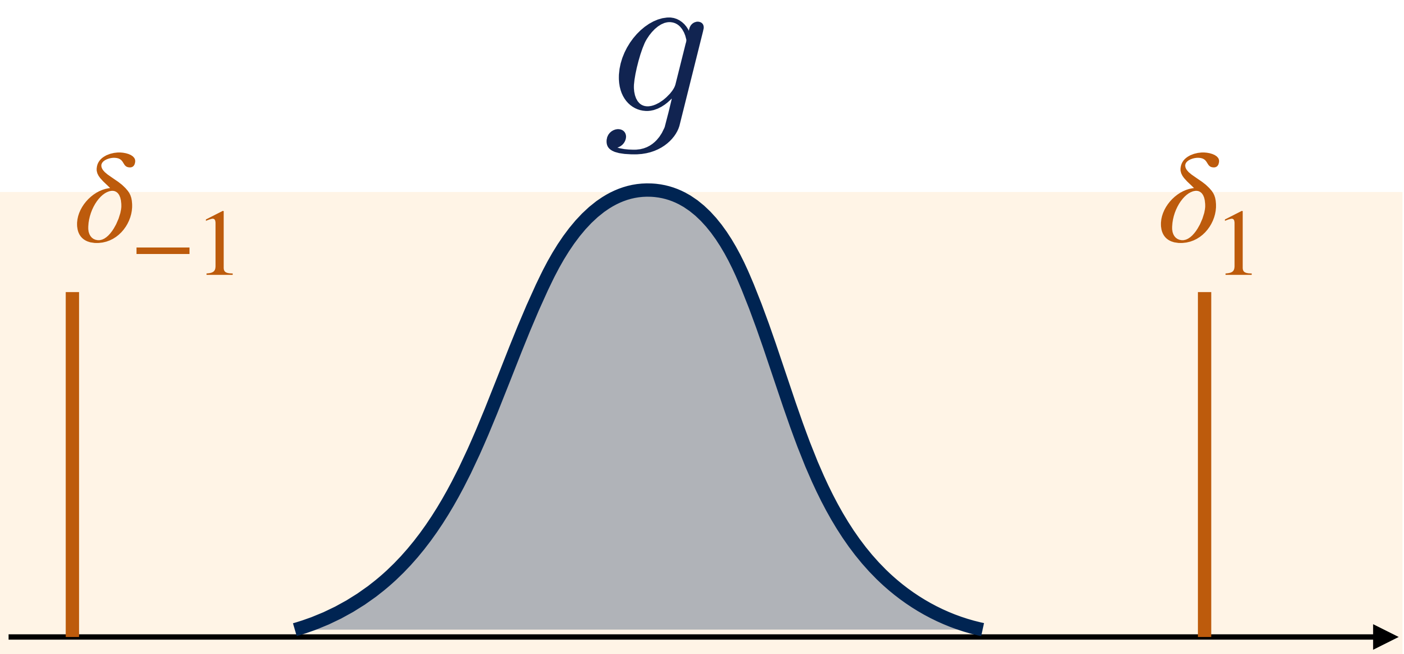
What about GMMs ?



Optimal transport between GMM

OT plans between GMM: usually not GMM themselves.

Same remark for barycenters.



- $\mu_0 = \mathcal{N}(0,1)$
- $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$
- density of μ_t :

$$f_t(x) = \frac{1}{1-t}\left(g\left(\frac{x+t}{1-t}\right)\mathbf{1}_{x<-t} + g\left(\frac{x-t}{1-t}\right)\mathbf{1}_{x>t}\right)$$

Restricting the set of couplings to GMMs

Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

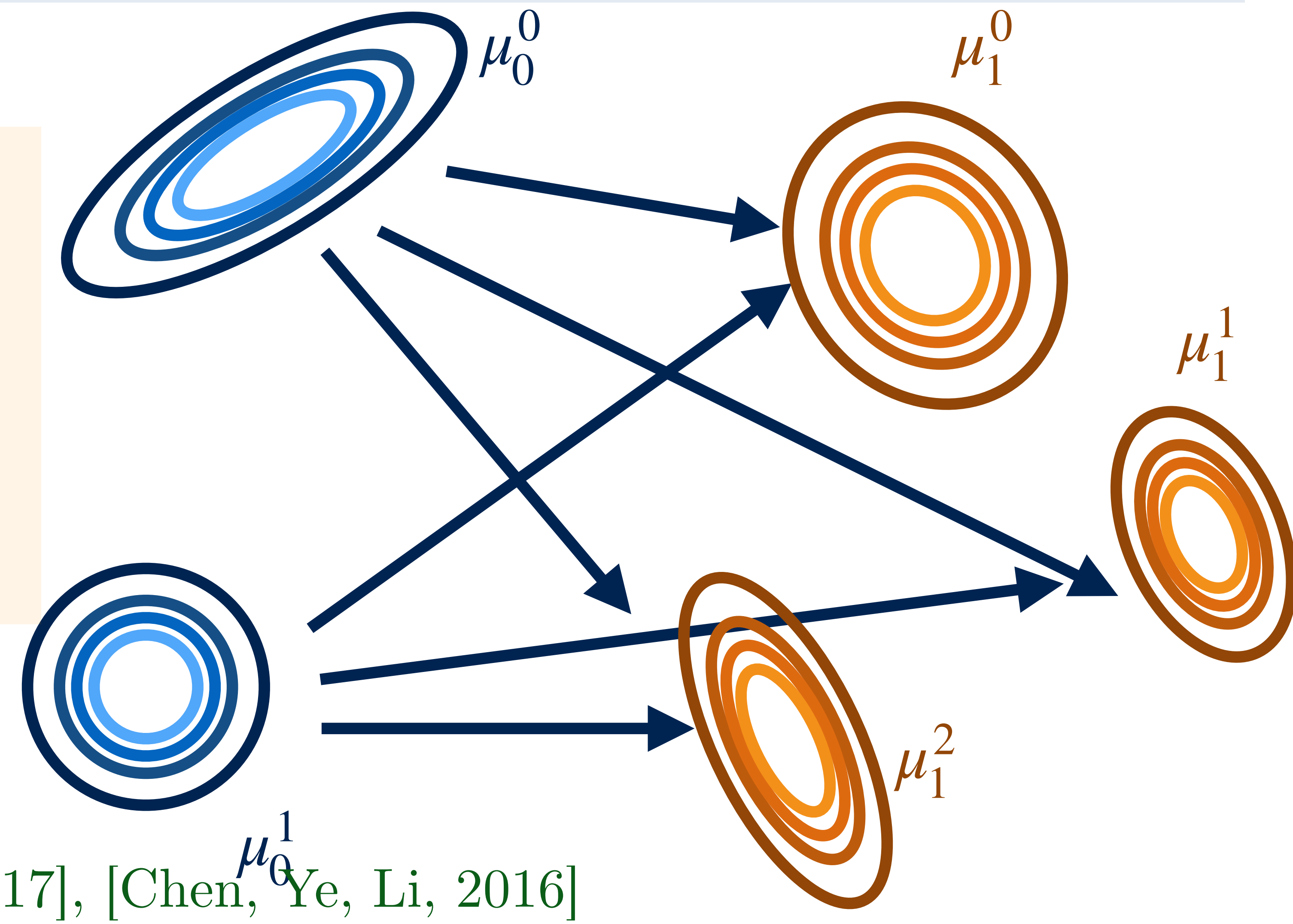
Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0,\mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$

Proposition [D.D.,2019]:

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0,\pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l).$$

$K_0 \times K_1$ OT problem



See also [Chen, Georgiu, Tannenbaum, 2017], [Chen, Ye, Li, 2016]

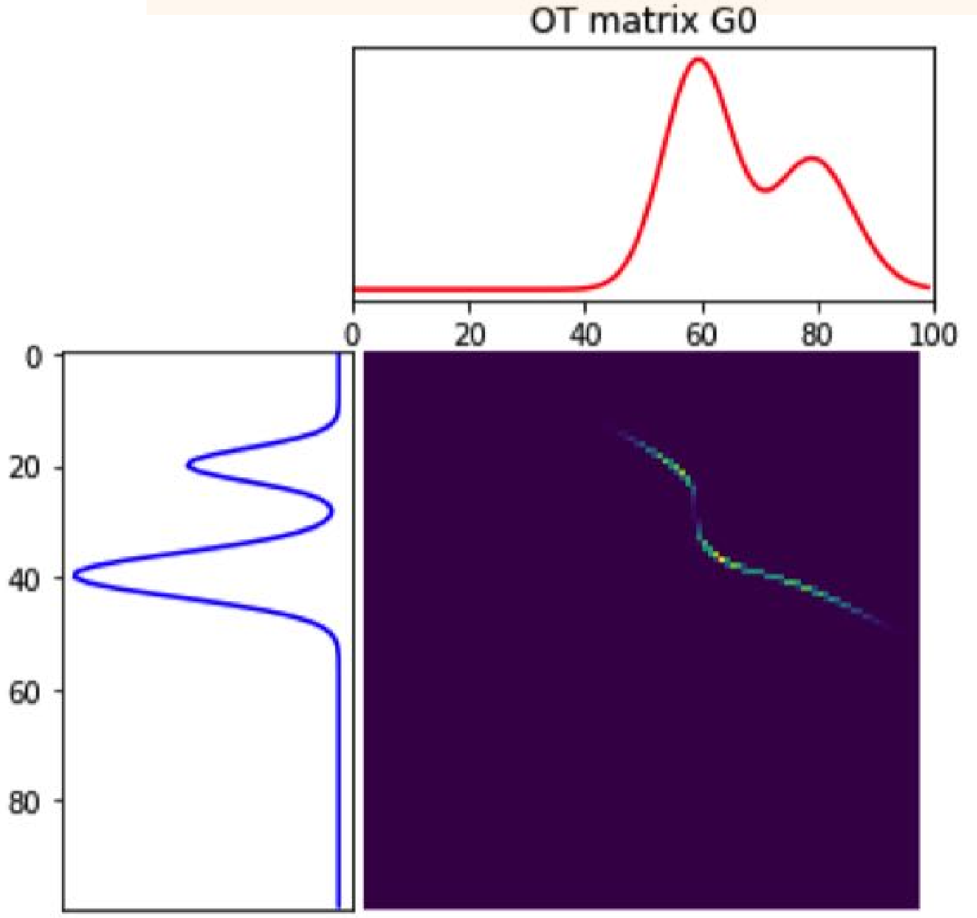
Optimal plan

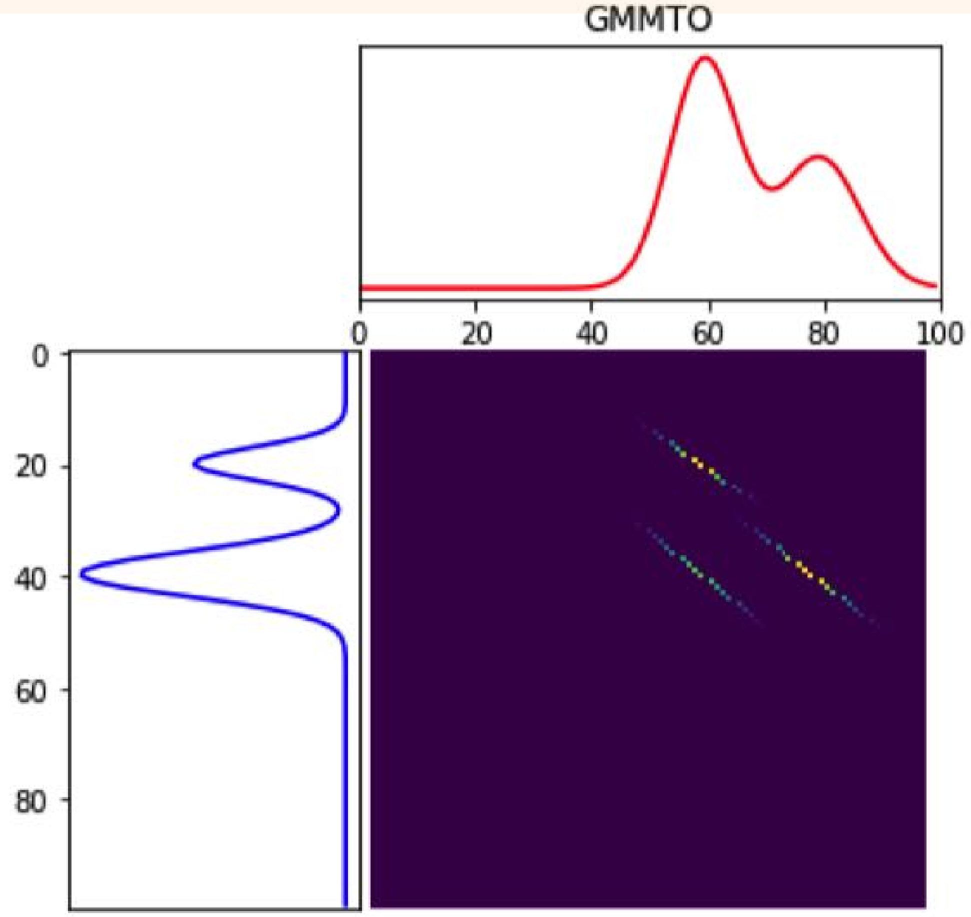
Proposition: an optimal plan γ^* for MW_2 can be written

$$\gamma^*(x,y) = \sum_{k,l} w^*_{k,l}\, g_{m_0^k,\Sigma_0^k}(x)\, \delta_{y=T_{k,l}(x)}, \text{ where}$$

$w^* =$ solution of the discrete OT problem of size $K_0 \times K_1$

$T_{kl} =$ optimal W_2 - map between the Gaussians μ_0^k and μ_1^l





At least one solution has less than $K_0 + K_1 - 1$ components!

Metric properties of MW_2

Proposition: MW_2 defines a metric on $GMM_d(\infty)$. This space equipped with the distance MW_2 is a geodesic space.

Corollary: Barycenters between $\mu_0 = \sum_k \pi_0^k \mu_0^k$ and $\mu_1 = \sum_l \pi_1^l \mu_1^l$ belong to $GMM_d(\infty)$ and can be written explicitly as

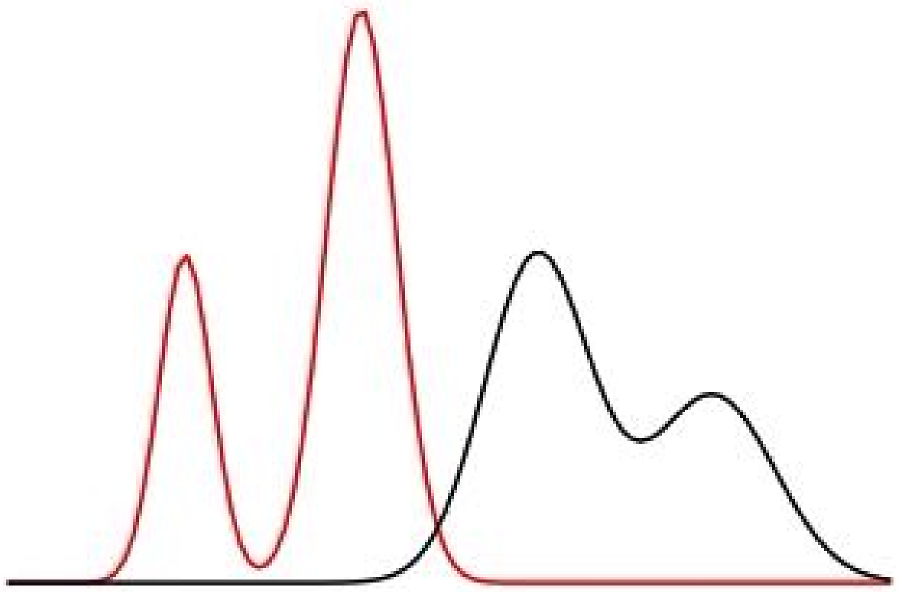
$$\forall t \in [0,1], \quad \mu_t = \sum_{k,l} w^*_{k,l} \mu_t^{k,l},$$

where $\mu_t^{k,l}$ is the displacement interpolation between μ_0^k and μ_1^l.

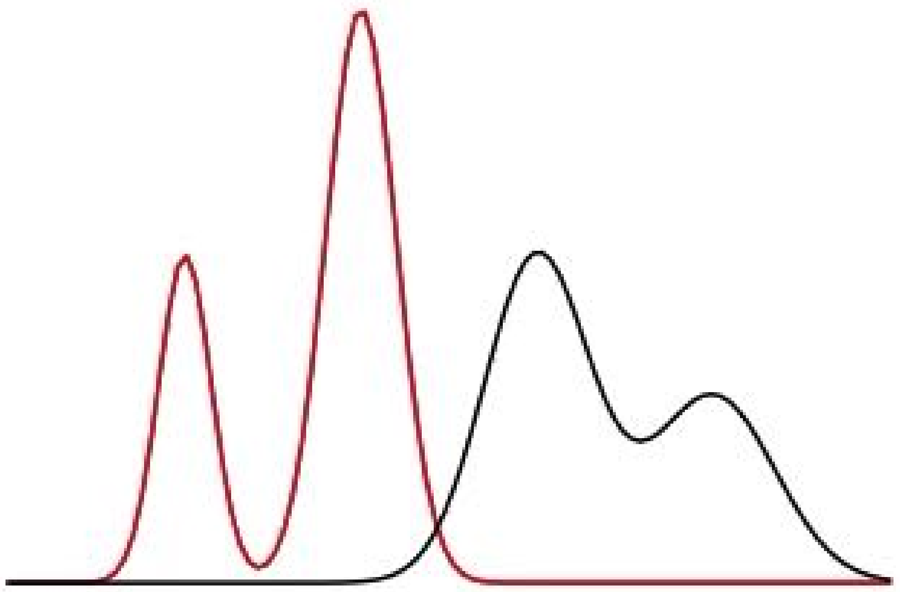
Barycenters can be chosen with less than $K_0 + K_1 - 1$ components.

Displacement interpolation

W_2

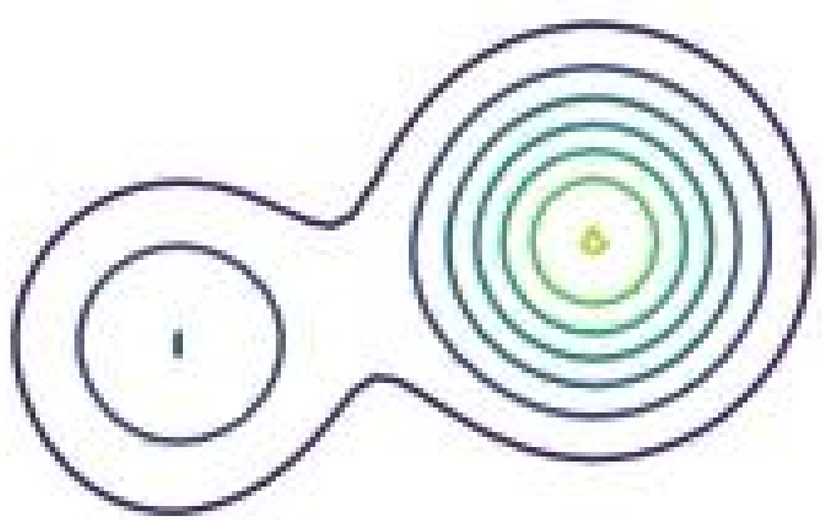


MW_2

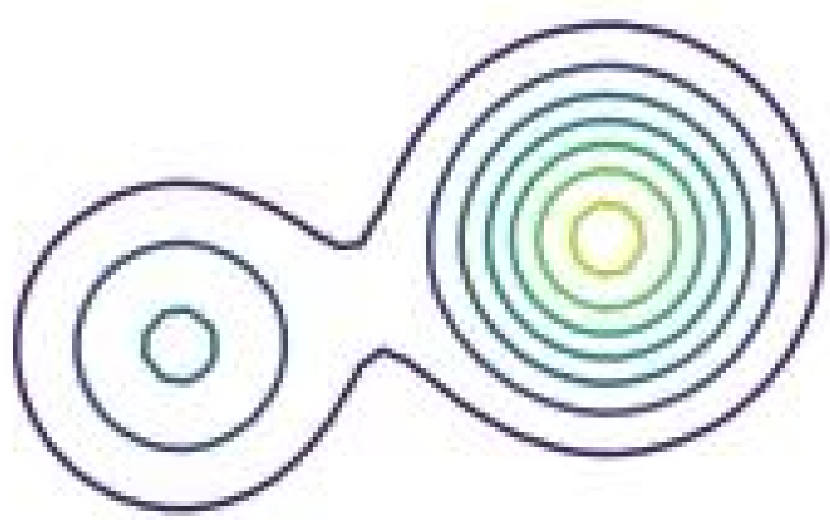


Displacement interpolation

W_2



MW_2



Comparison between W_2 and MW_2

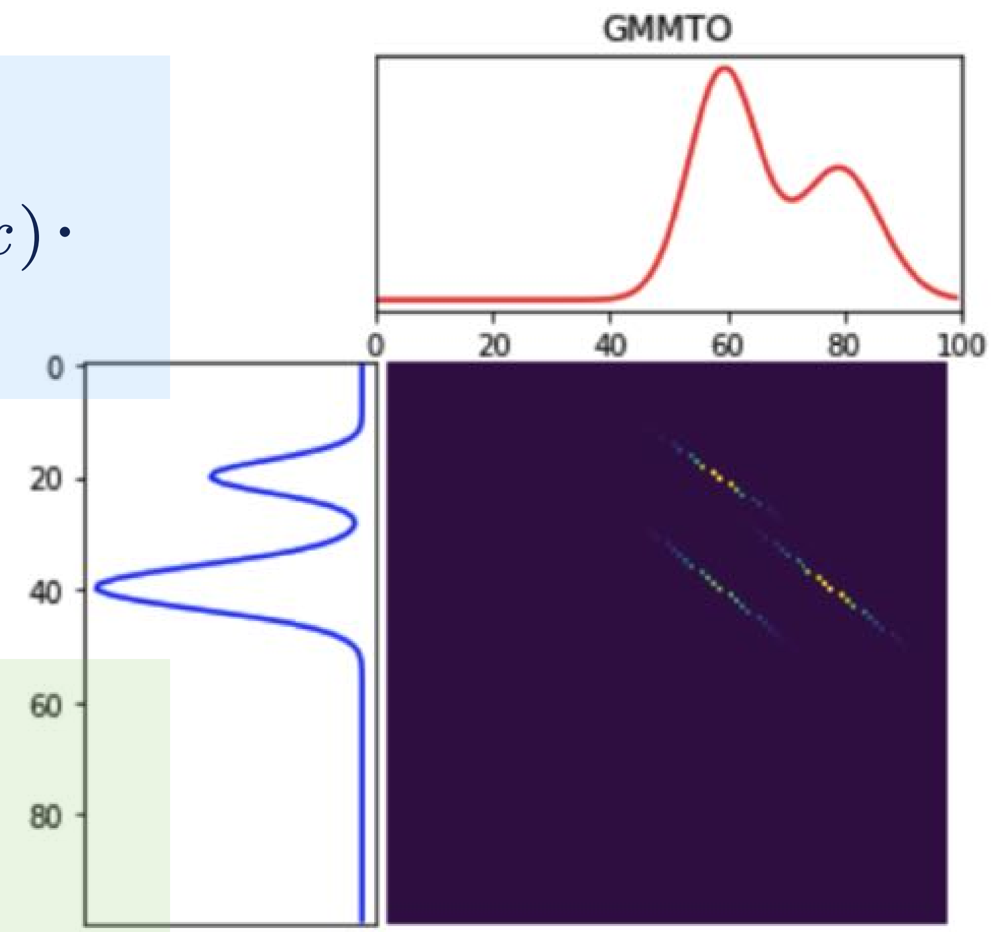
Proposition: Let $\mu_0 \in GMM_d(K_0)$ and $\mu_1 \in GMM_d(K_1)$ be two Gaussian mixtures, then

$$W_2(\mu_0, \mu_1) \leq MW_2(\mu_0, \mu_1) \leq W_2(\mu_0, \mu_1) + \sum_{i=0,1} \left(2 \sum_{k=1}^{K_i} \pi_i^k \mathrm{trace}(\Sigma_i^k) \right)^{\frac{1}{2}}.$$

Using MW_2 on real data

From a transport plan to a map

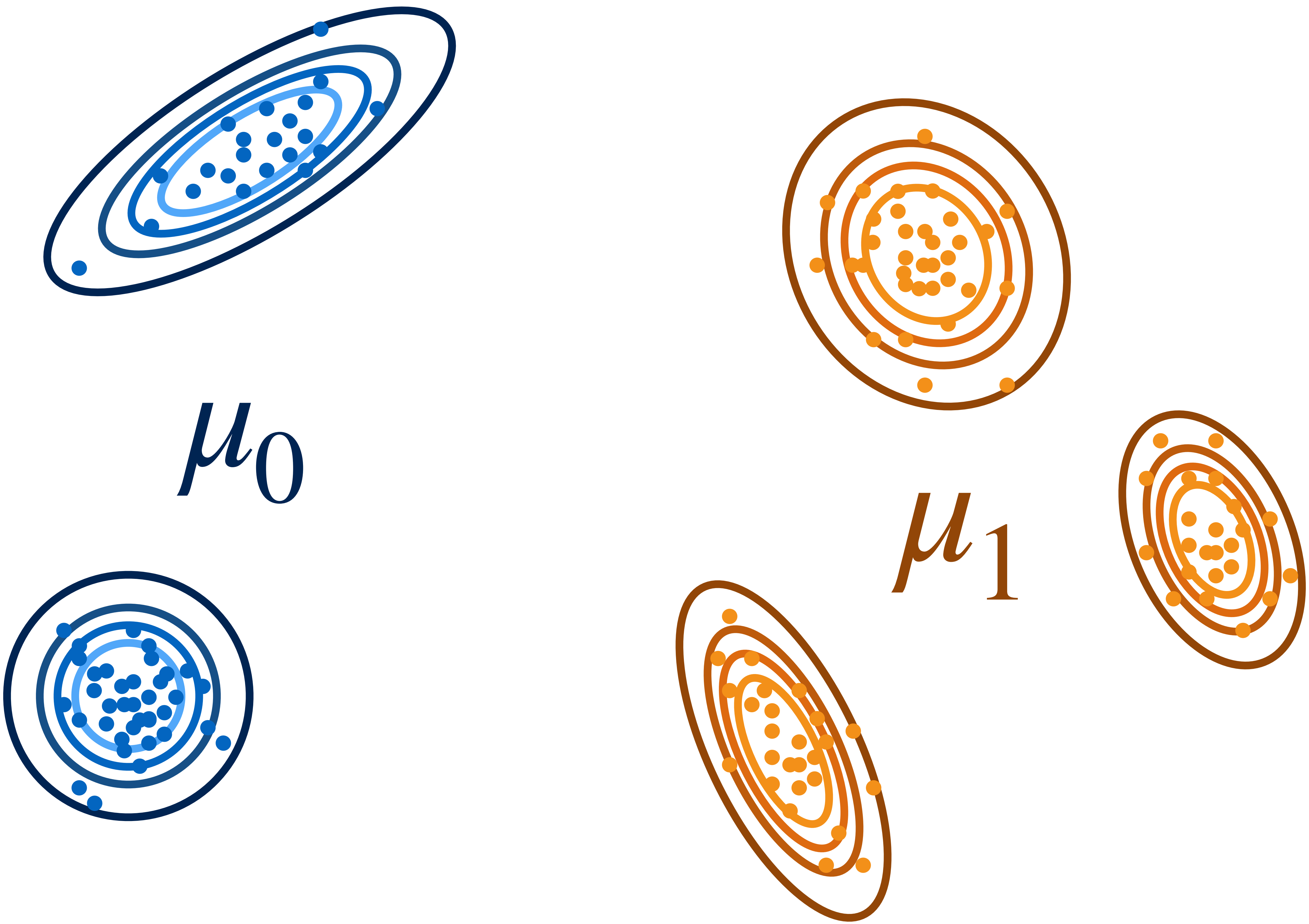
$$\gamma(x,y) = \sum_{k,l} w^*_{k,l}\, g_{m_0^k,\Sigma_0^k}(x)\,\delta_{y=T_{k,l}(x)}.$$



$$T_{mean}(x) = \frac{\sum_{k,l} w^*_{k,l}\, g_{m_0^k,\Sigma_0^k}(x)\, T_{k,l}(x)}{\sum_k \pi_0^k\, g_{m_0^k,\Sigma_0^k}(x)},$$

$$T_{rand}(x) = T_{k,l}(x) \quad \text{with probability } p_{k,l}(x) = \frac{w^*_{k,l}\, g_{m_0^k,\Sigma_0^k}(x)}{\sum_j \pi_0^j\, g_{m_0^j,\Sigma_0^j}(x)}.$$

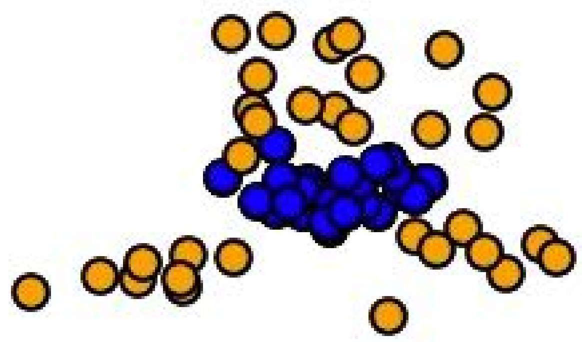
Discrete data : EM + MW_2



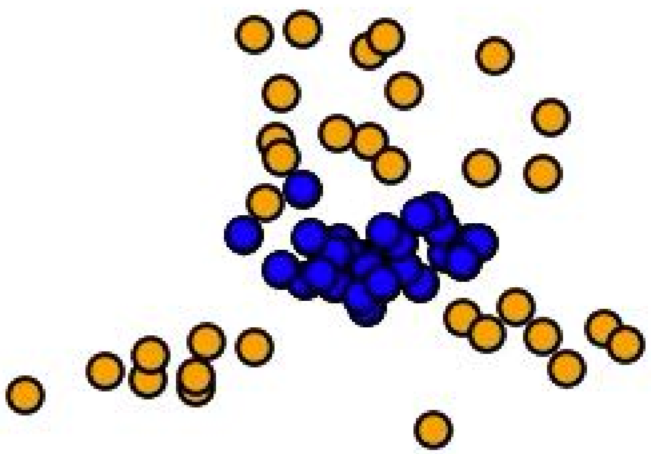
Discrete data : EM $+MW_2$

T_{rand}

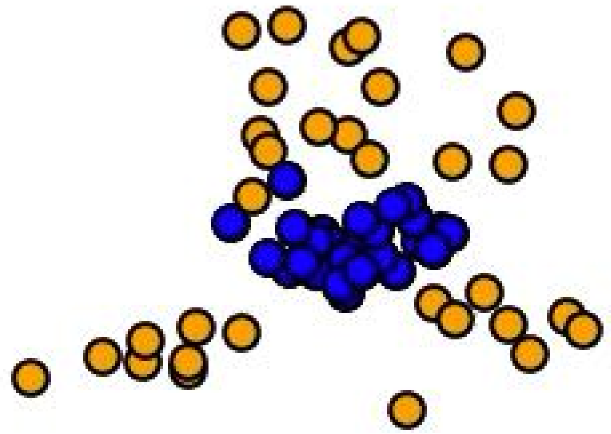
$K=1$



$K=10$

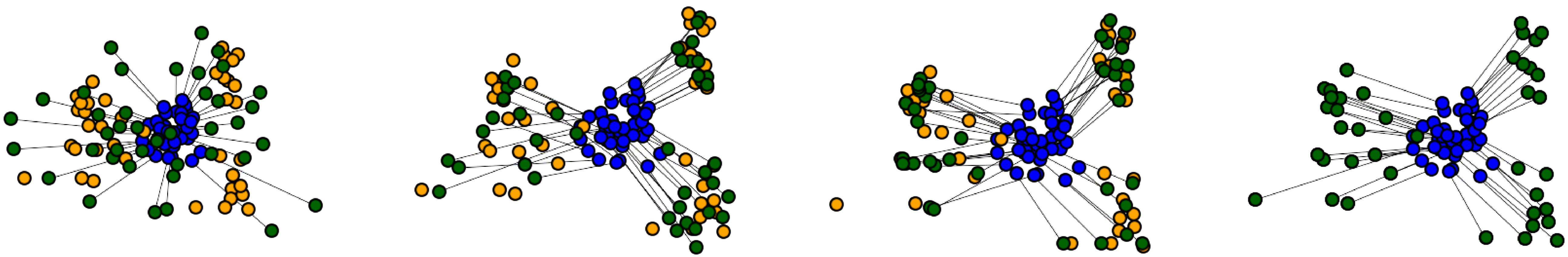


$K=N$

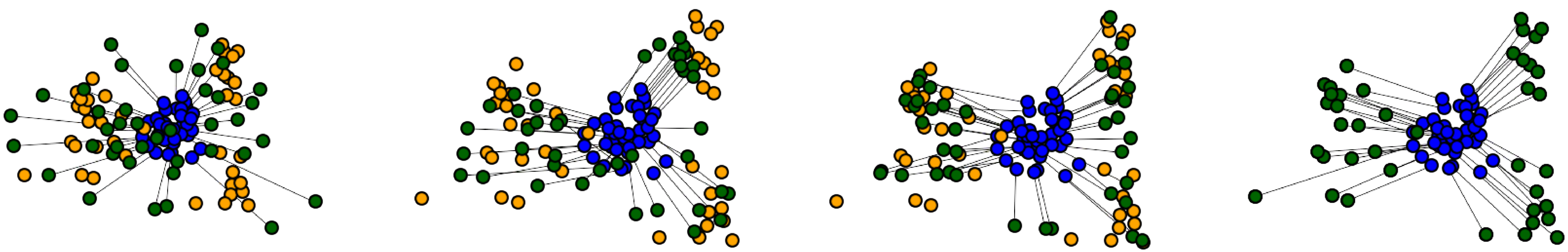


Assignment between cloud points

$K = 1$ $K = 5$ $K = 10$ $K = 40$



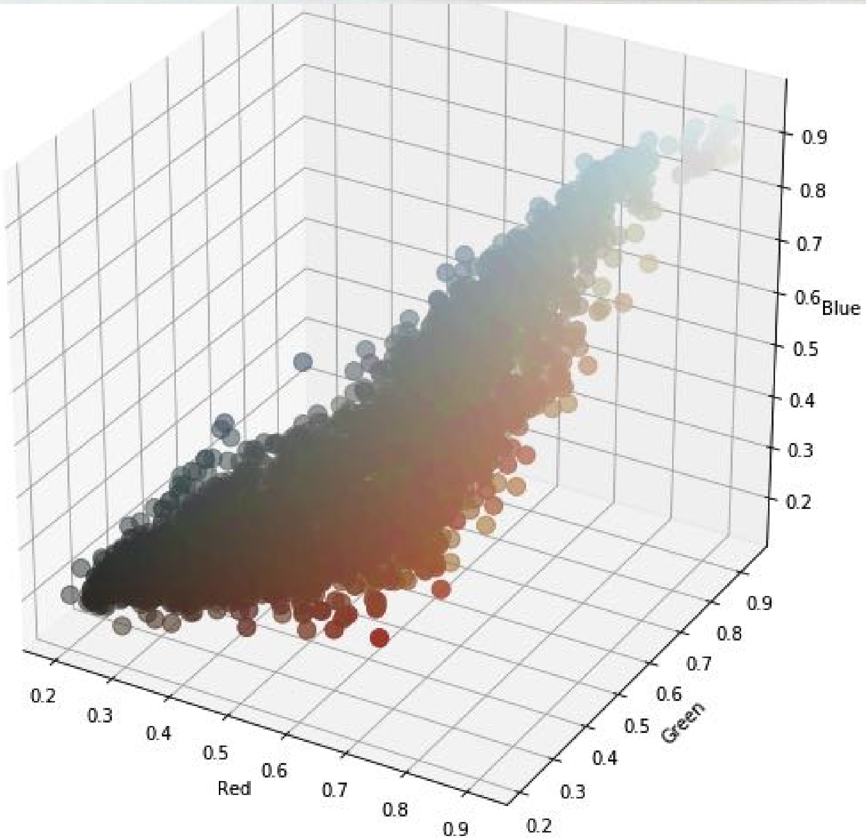
T_{rand}

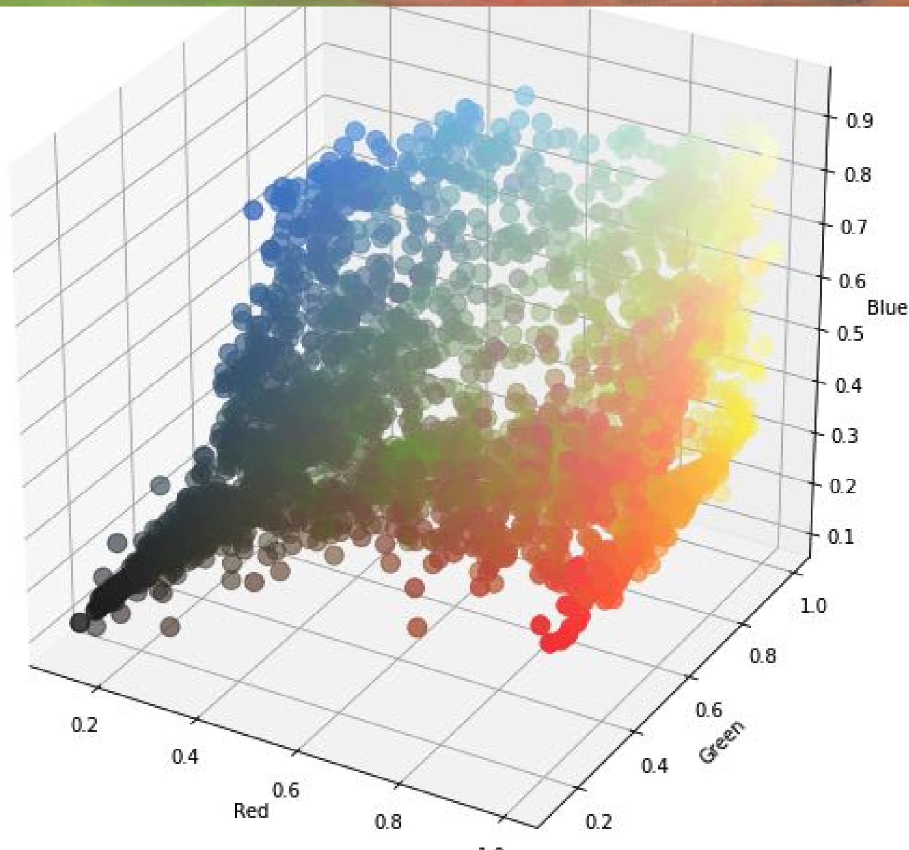


T_{mean}

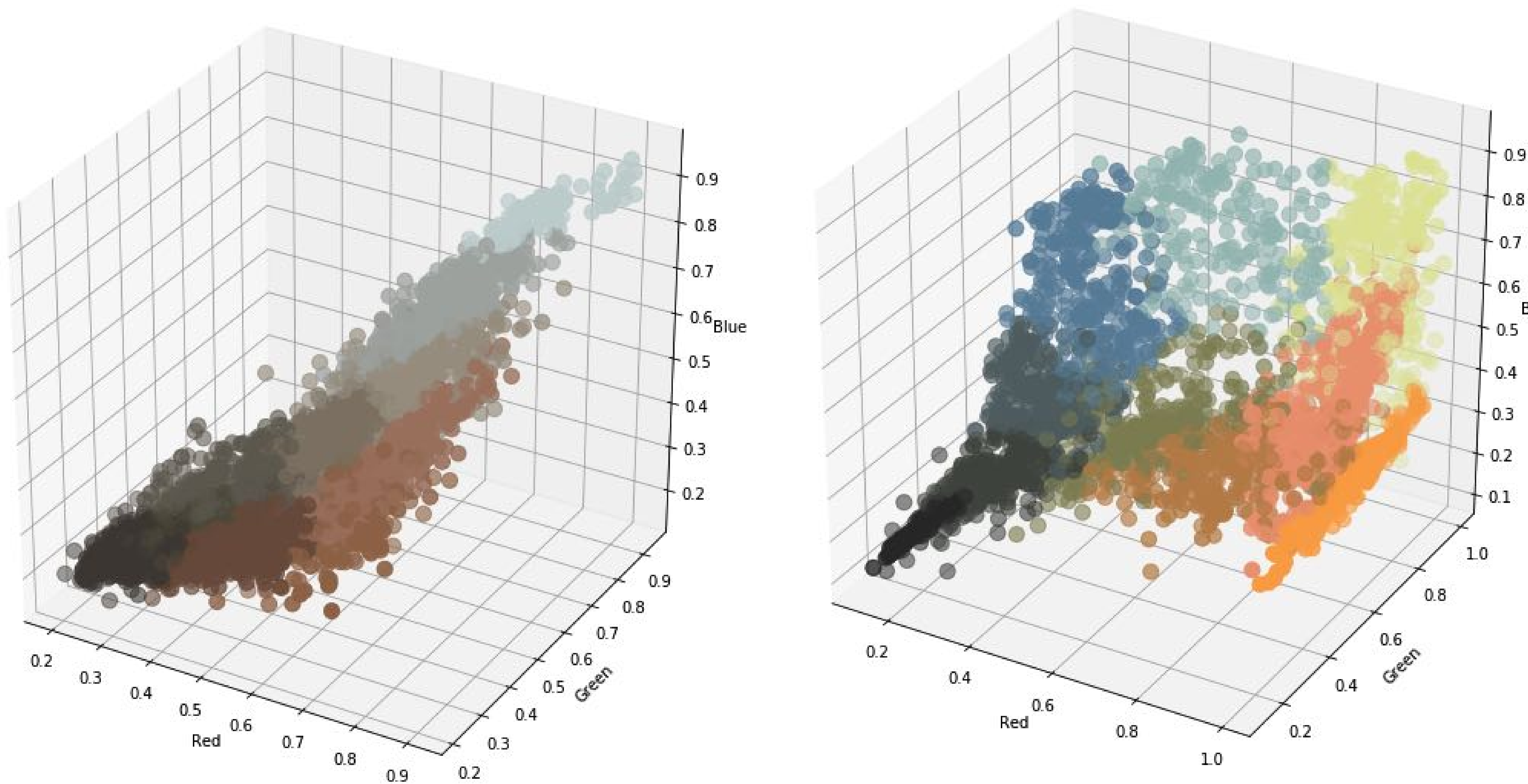
Two applications in image processing

Color transfer





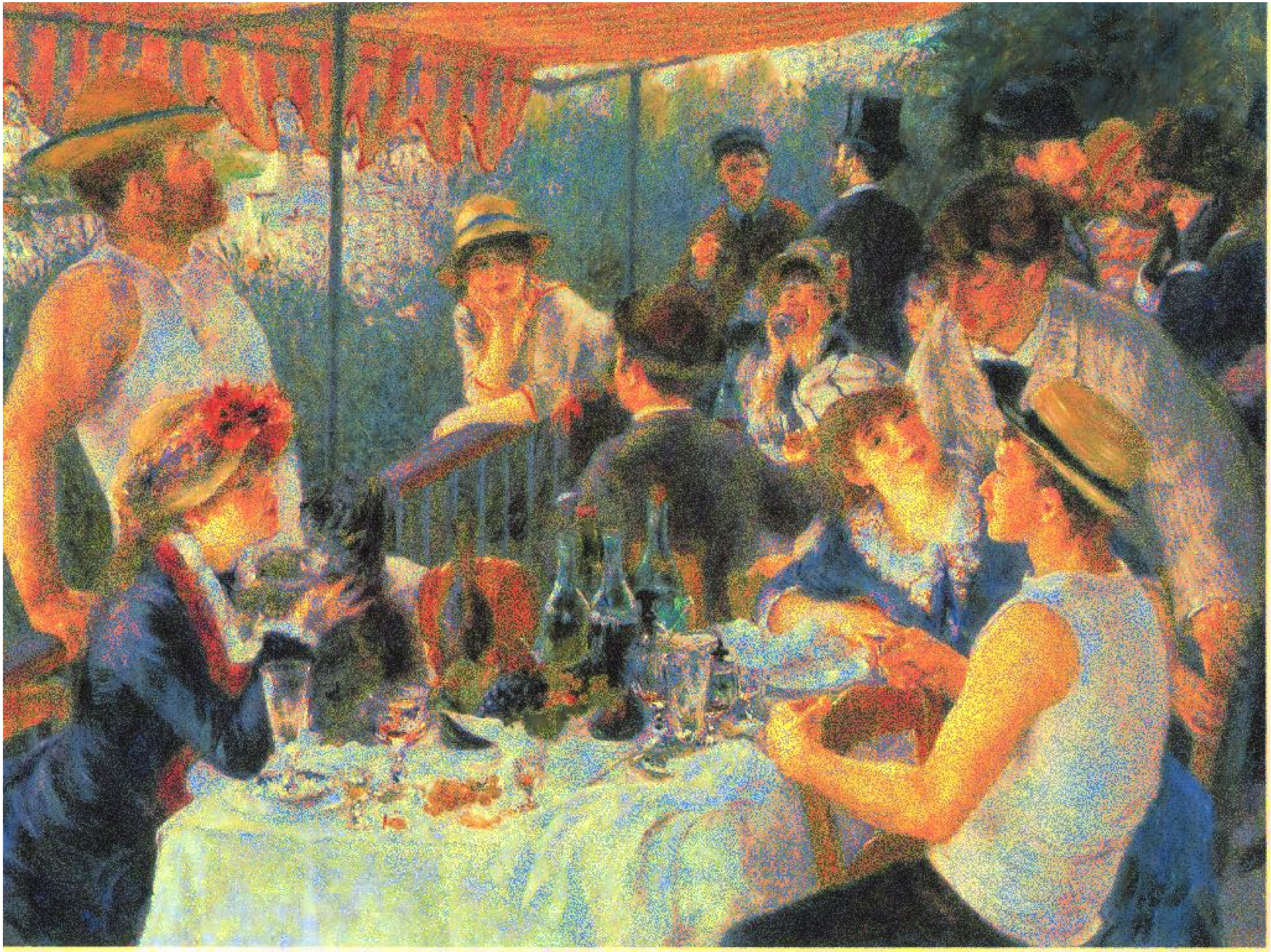
10 classes GMM



Color transfer

Color transfer, T_{mean}

Color transfer, T_{rand}

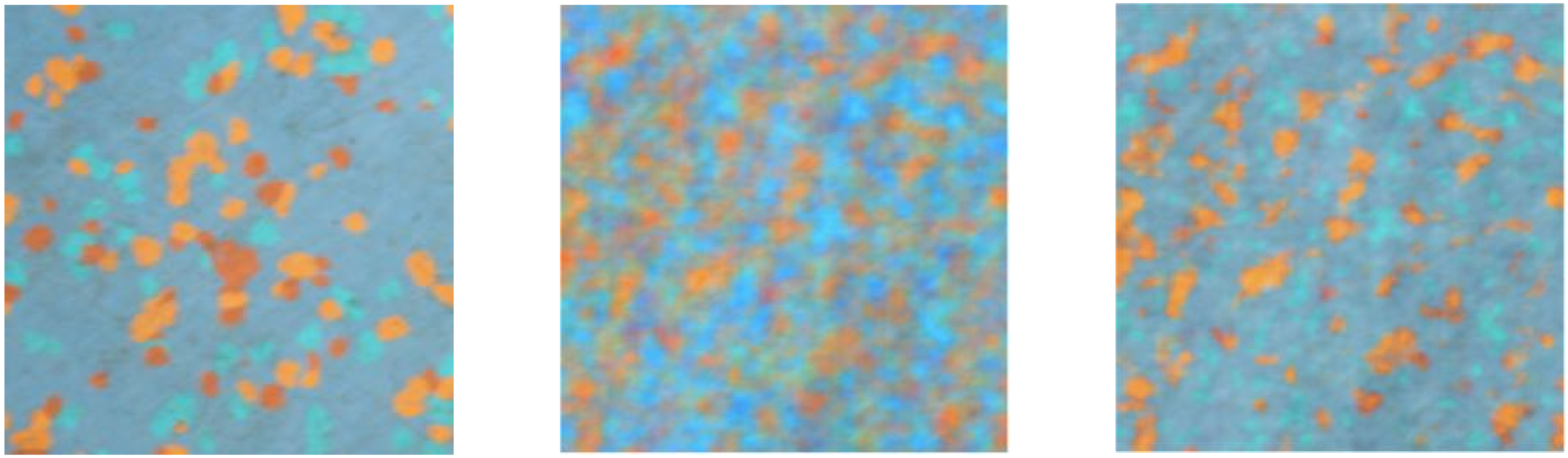


Color transfer (sliced OT)

Color transfer, T_{mean}, iterated

Texture synthesis

Original texture u ADSN(u) Synthetized texture



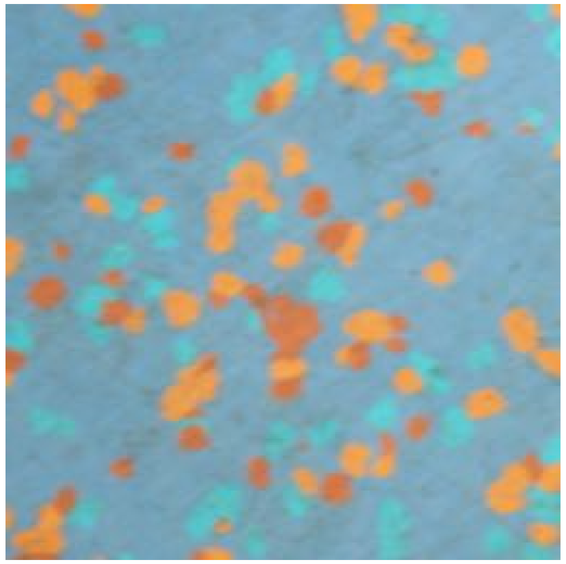
ADSN(u): Stationary Gaussian random field $U : \mathbb{Z}^2 \to \mathbb{R}^3$ with same mean and covariance as u.

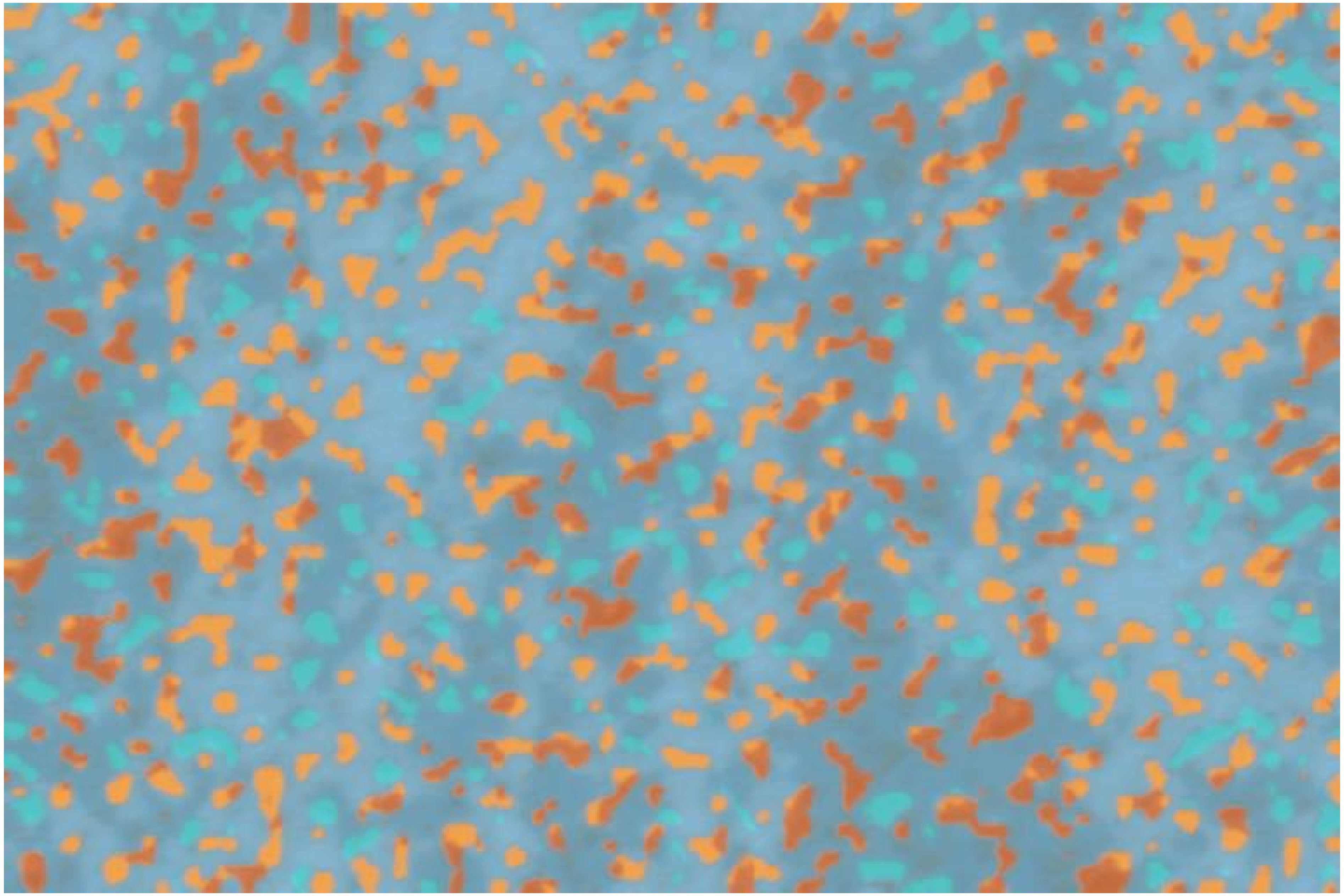
Texture synthesis for one scale:

- decompose u and $ADSN(u)$ into two set of patches
- compute the optimal plan (for MW_2) between corresponding GMMs
- replace patches from $ADSN(u)$ with matching patches in u

[Ongoing work with A. Leclaire] Inspired by [Leclaire, Galerne, Rabin, 2018]

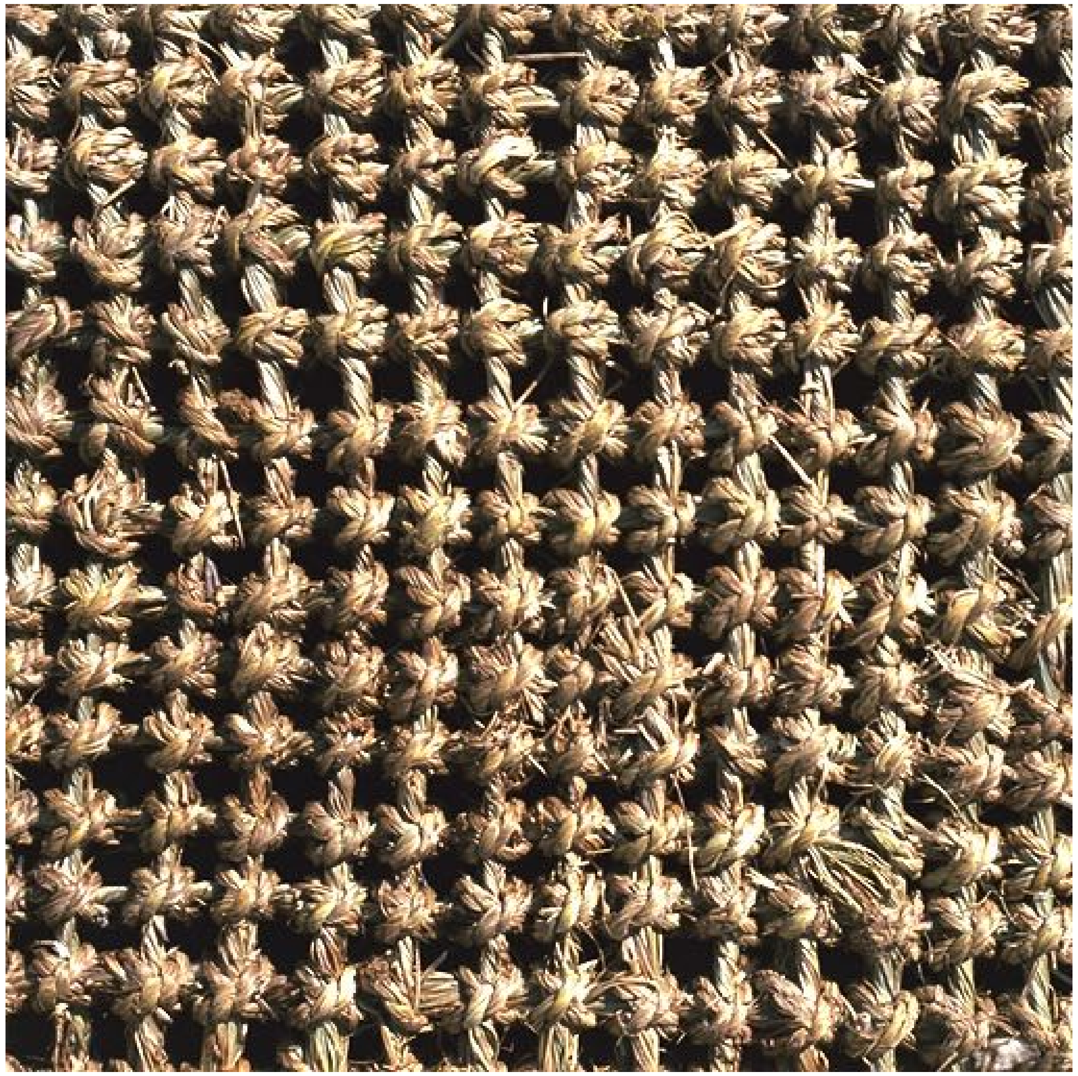
Multiscale Texture synthesis

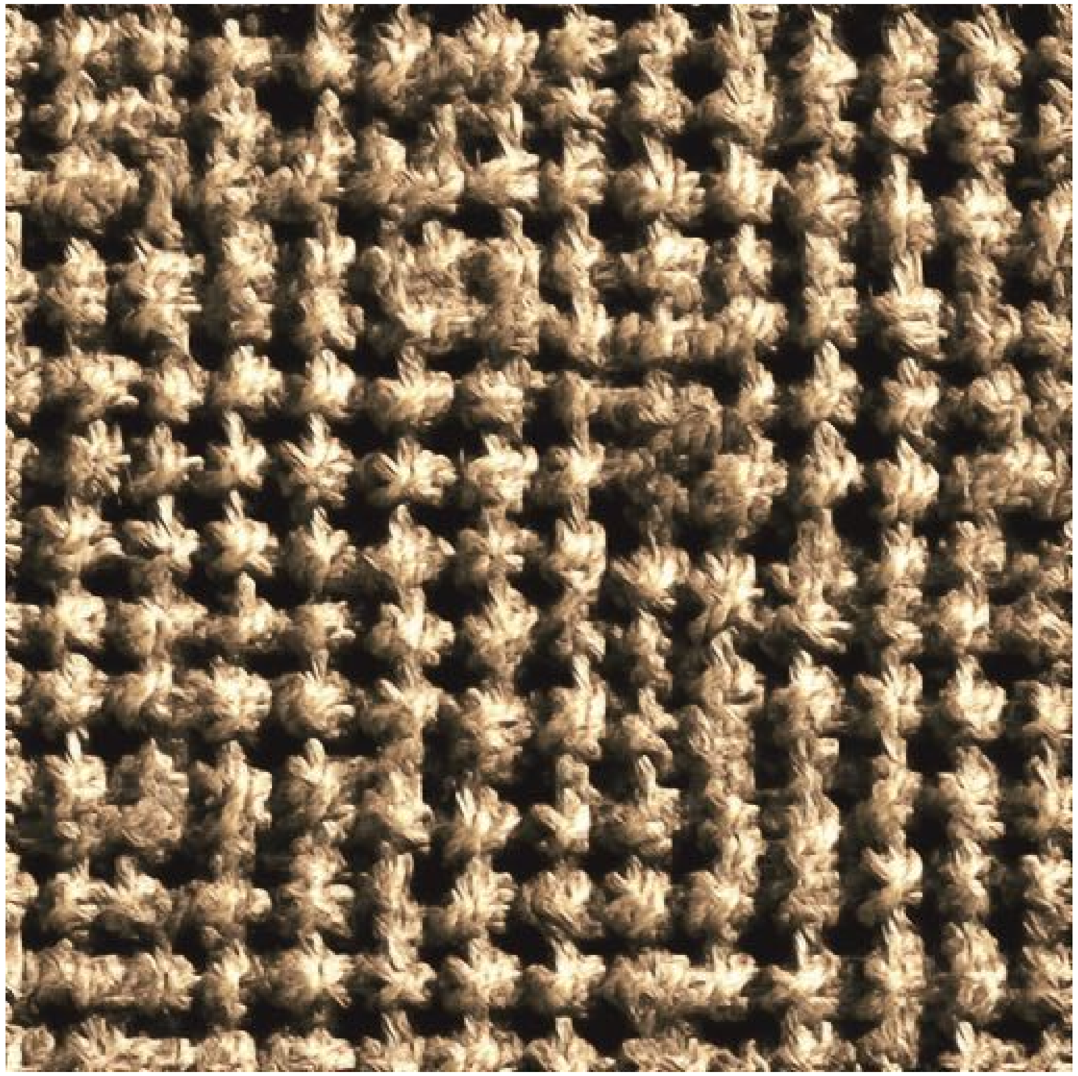




[Ongoing work with A. Leclaire]

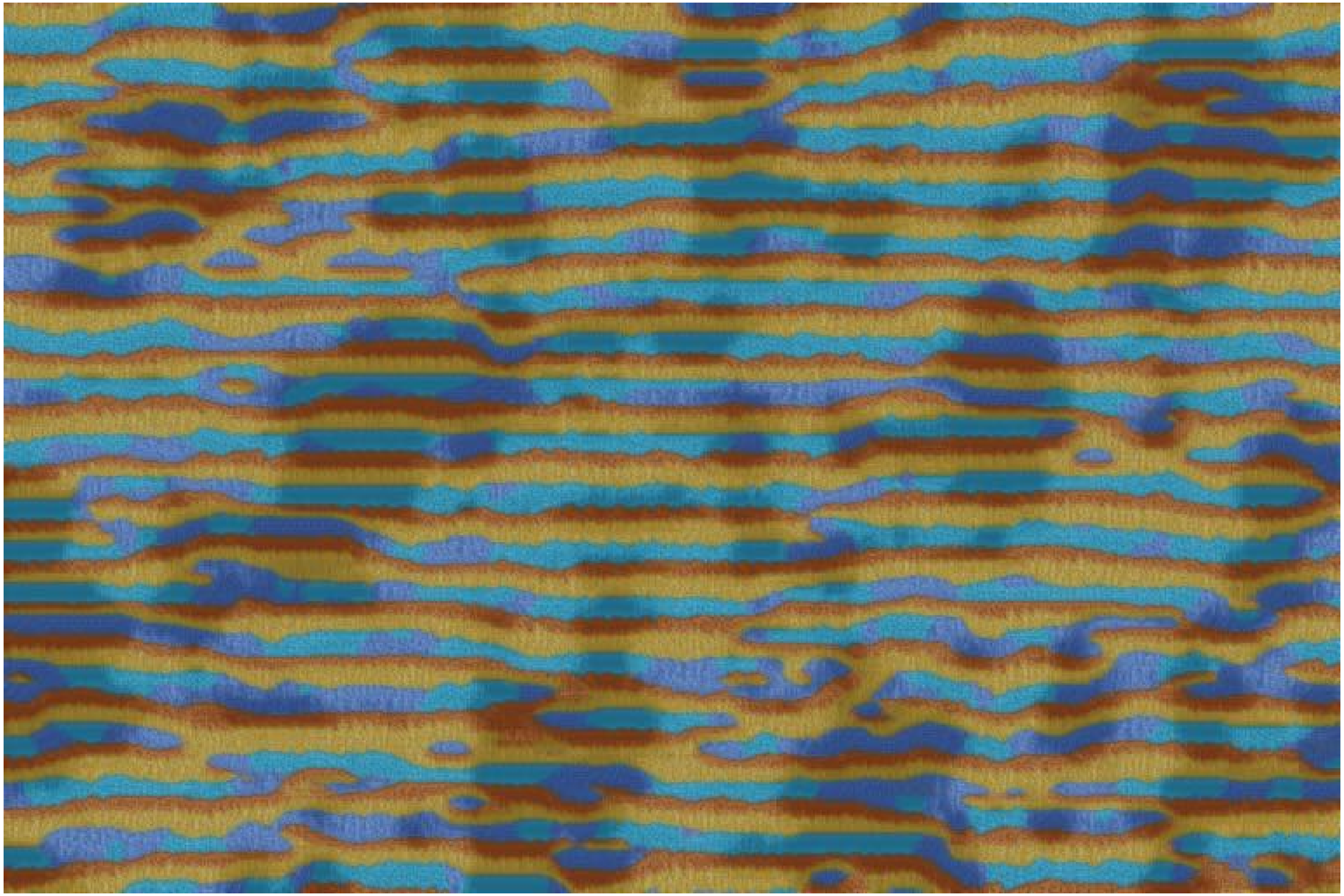
Multiscale Texture synthesis





[Ongoing work with A. Leclaire]

Multiscale Texture synthesis



[Ongoing work with A. Leclaire]

Multiscale Texture synthesis

[Ongoing work with A. Leclaire]

Barycenters for more than 2 GMMs

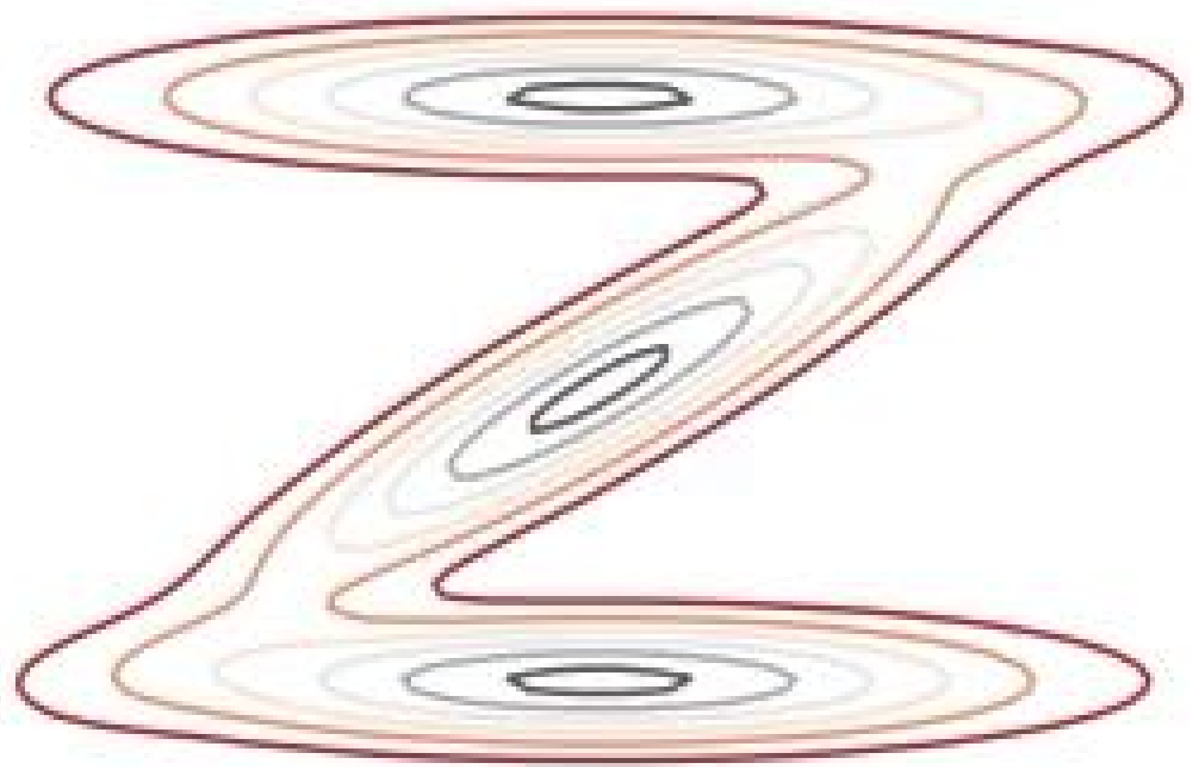
Barycenters for MW_2

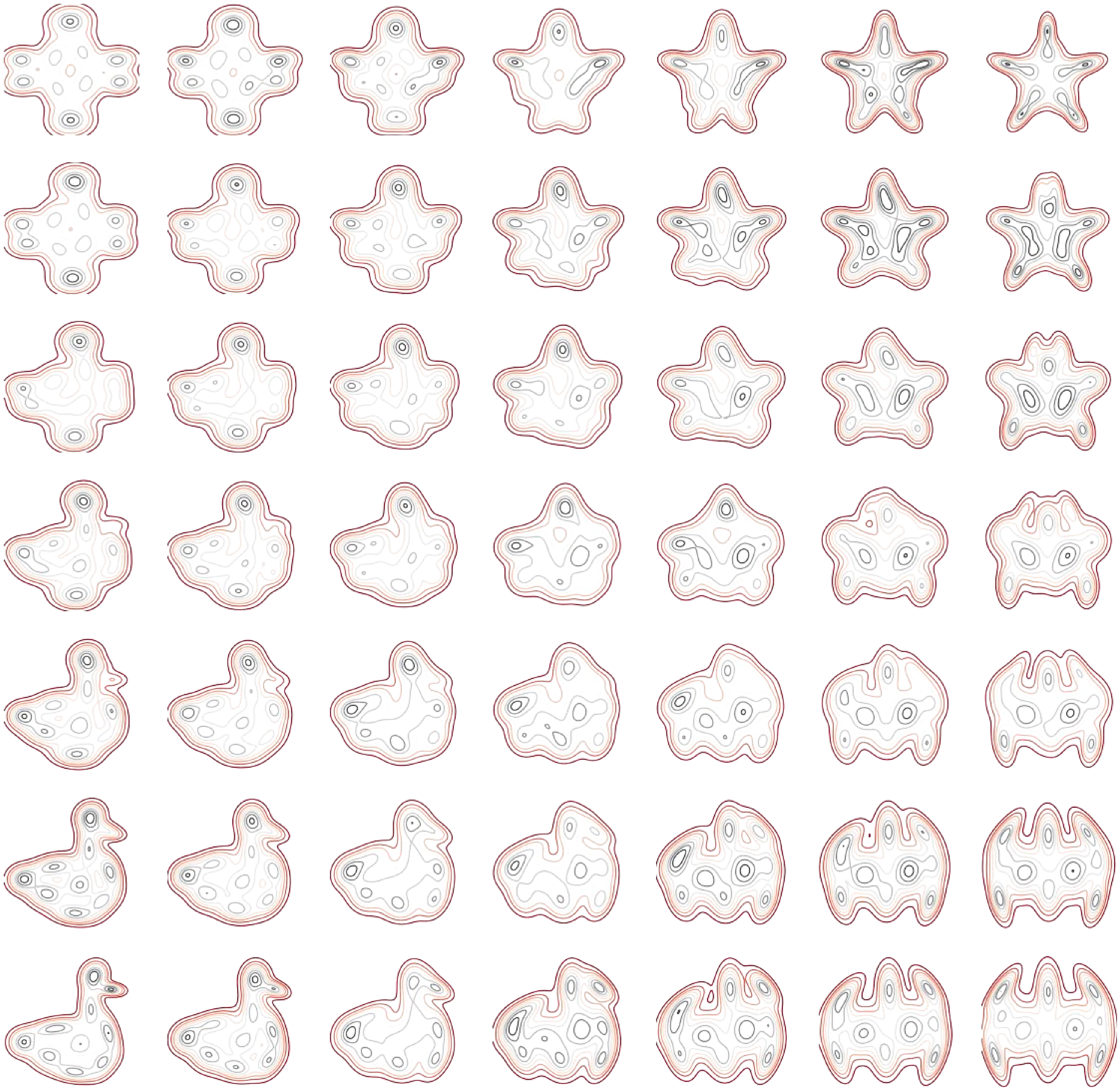
Problem: $\inf_{\nu \in GMM_d(\infty)} \sum_{j=0}^{J-1} \lambda_j MW_2^2(\mu_j, \nu)$, for $\mu_0, \ldots, \mu_{J-1}$ GMMs

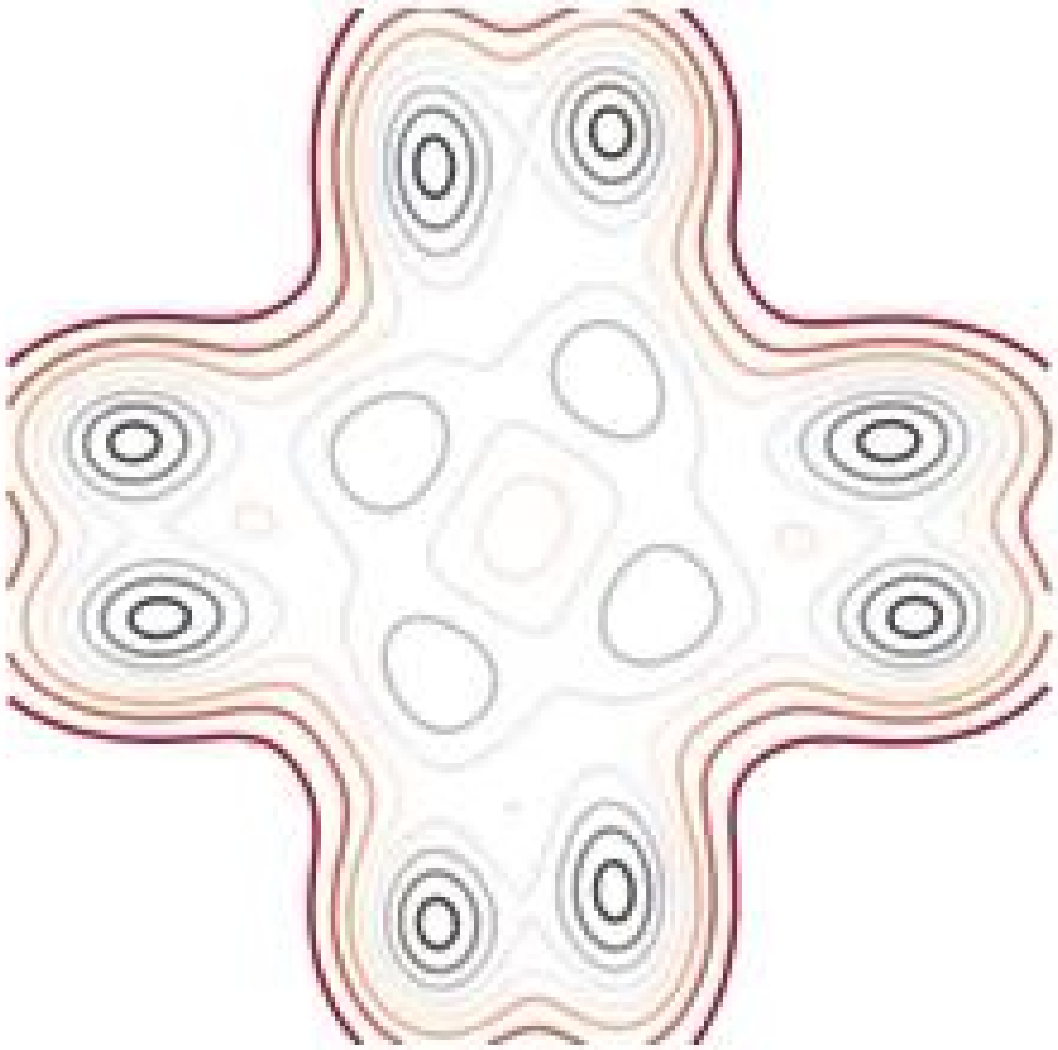
Proposition: solution $\nu = \sum_{k_0, \ldots, k_{J-1}} w^*_{k_0 \ldots k_{J-1}} \nu_{k_0 \ldots k_{J-1}}$, where

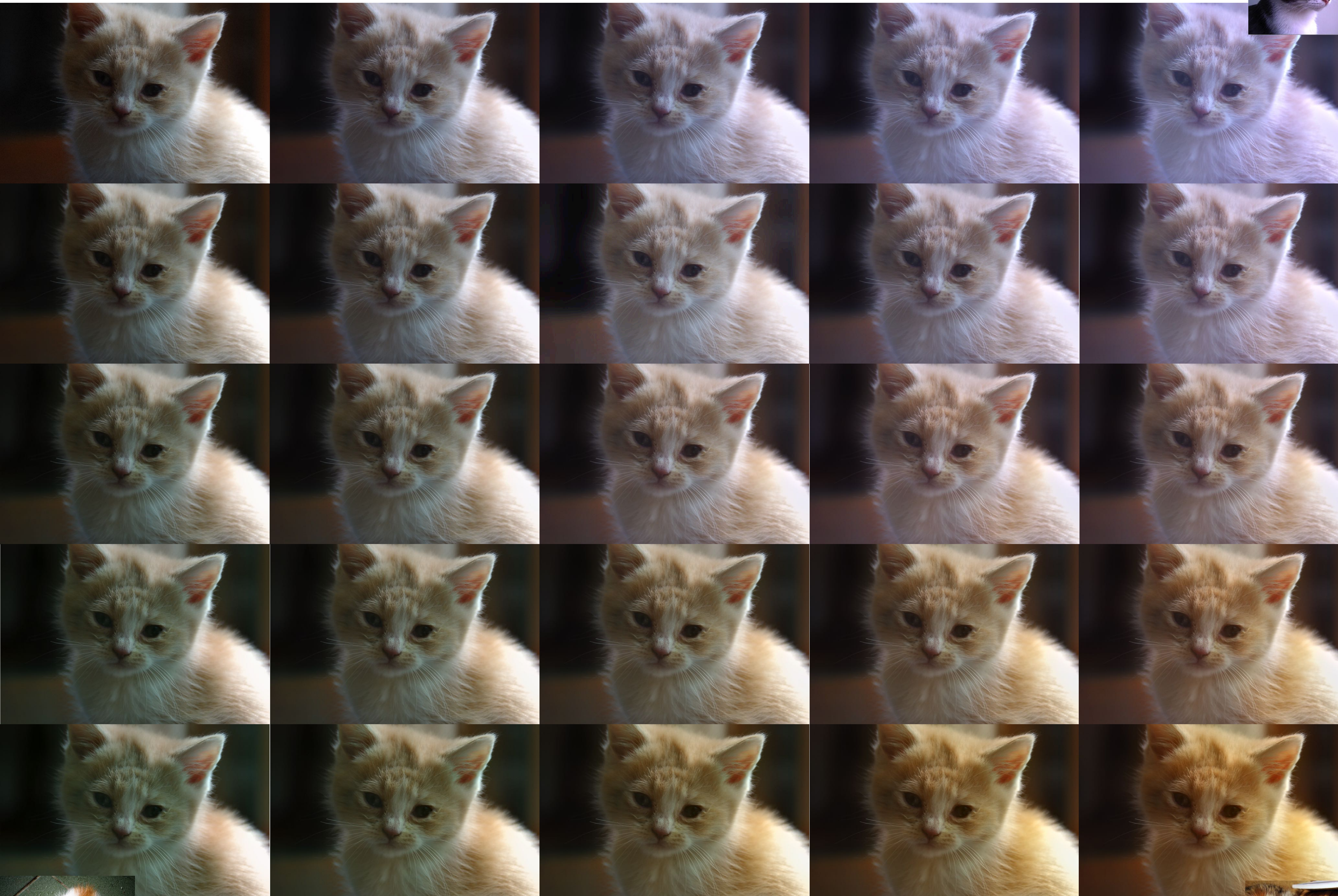
$\nu_{k_0 \ldots k_{J-1}}$ Gaussian barycenter between components $\mu_0^{k_0}, \ldots, \mu_{J-1}^{k_{J-1}}$ and w^* optimal solution of the discrete multi-marginal associated problem.

A solution can be chosen s.t. the barycenters have less than $K_0 + \ldots + K_{J-1} - J + 1$ components.









Perspectives

A distance on GMMs suited for high dimensional data:

- Reduced complexity: OT on a $K_0 \times K_1$ problem
- Relevant for data structured in classes

Limitations:

- use of EM $\rightarrow$ could be replaced by GD or SGD for large data + differentiability.

https://hal.archives-ouvertes.fr/hal-02178204

https://github.com/judelo/gmmot