Optimal Transport in Imaging Sciences

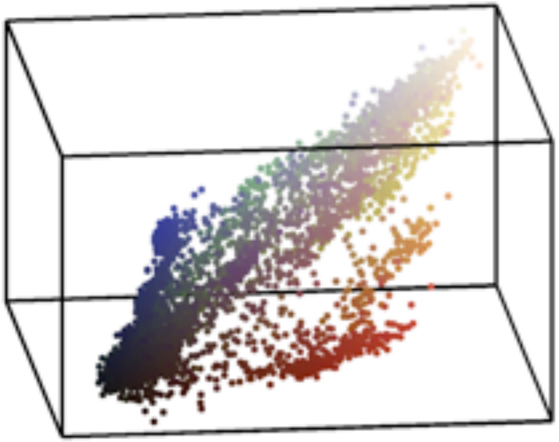
Gabriel Peyré

www.numerical-tours.com

CNRS
CEREMADE
DAUPHINE UNIVERSITÉ PARIS

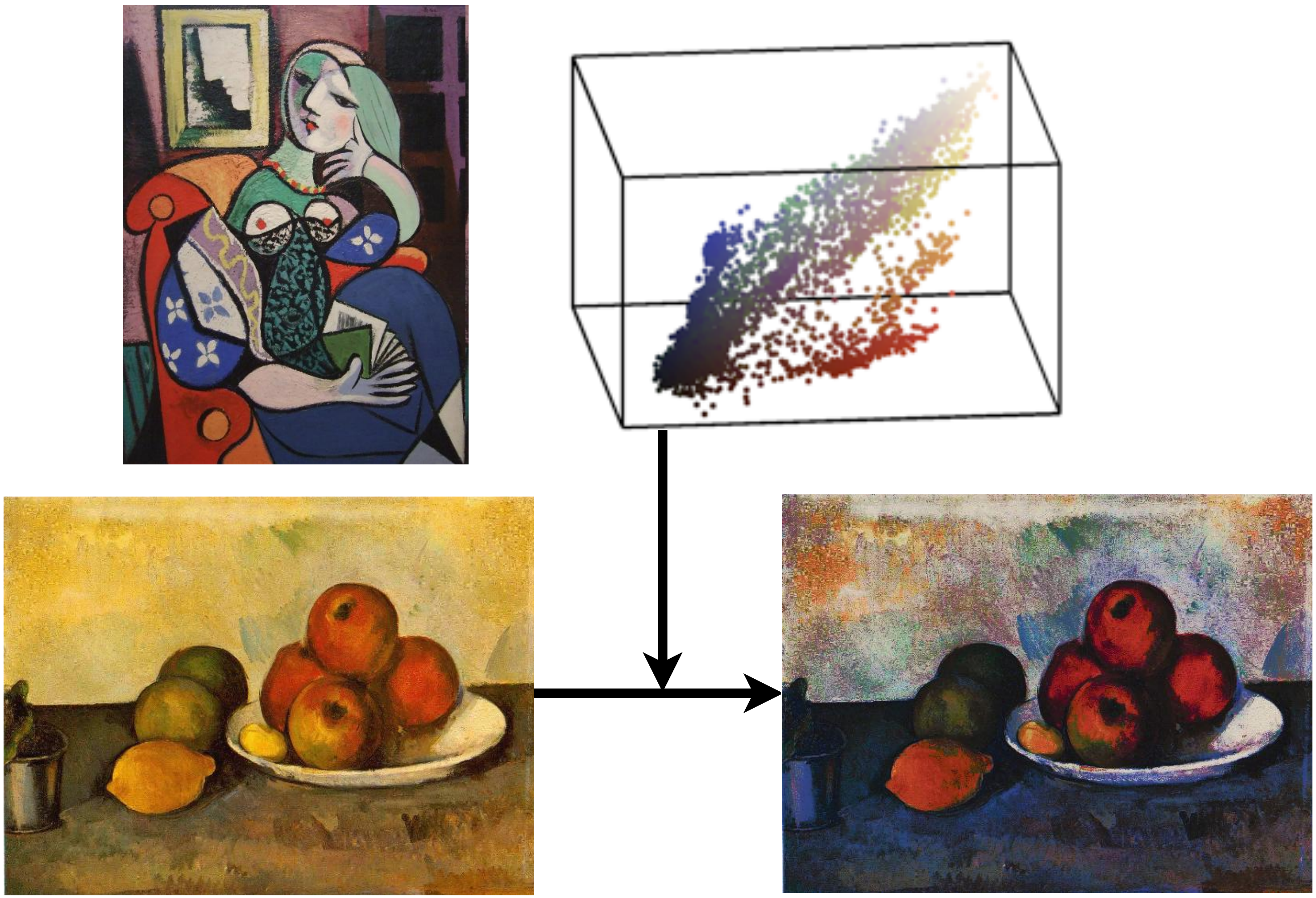
Statistical Image Models

Colors distribution: each pixel $\Leftrightarrow$ point in $\mathbb{R}^3$



Statistical Image Models

Colors distribution: each pixel $\Leftrightarrow$ point in $\mathbb{R}^3$

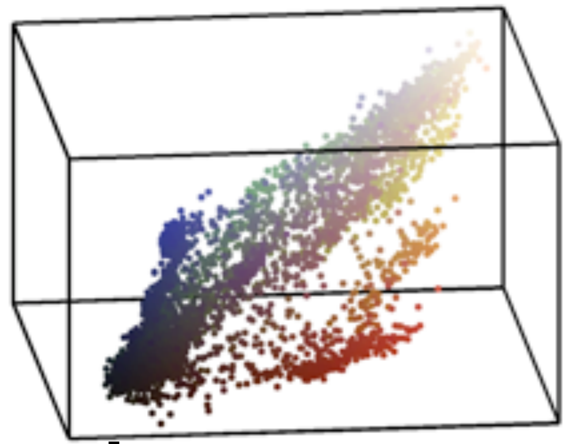


Input image

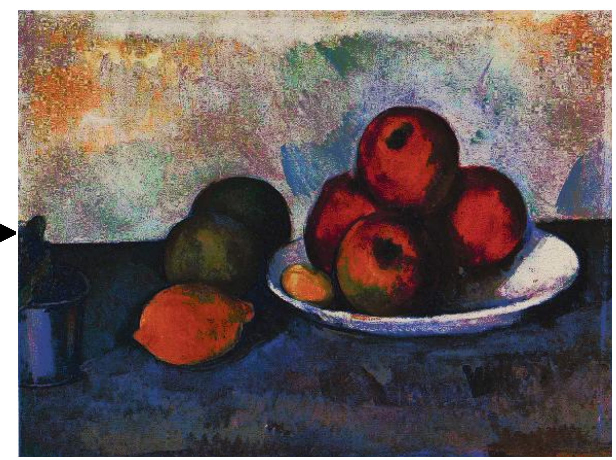
Modified image

Statistical Image Models

Colors distribution: each pixel $\Leftrightarrow$ point in $\mathbb{R}^3$



Input image

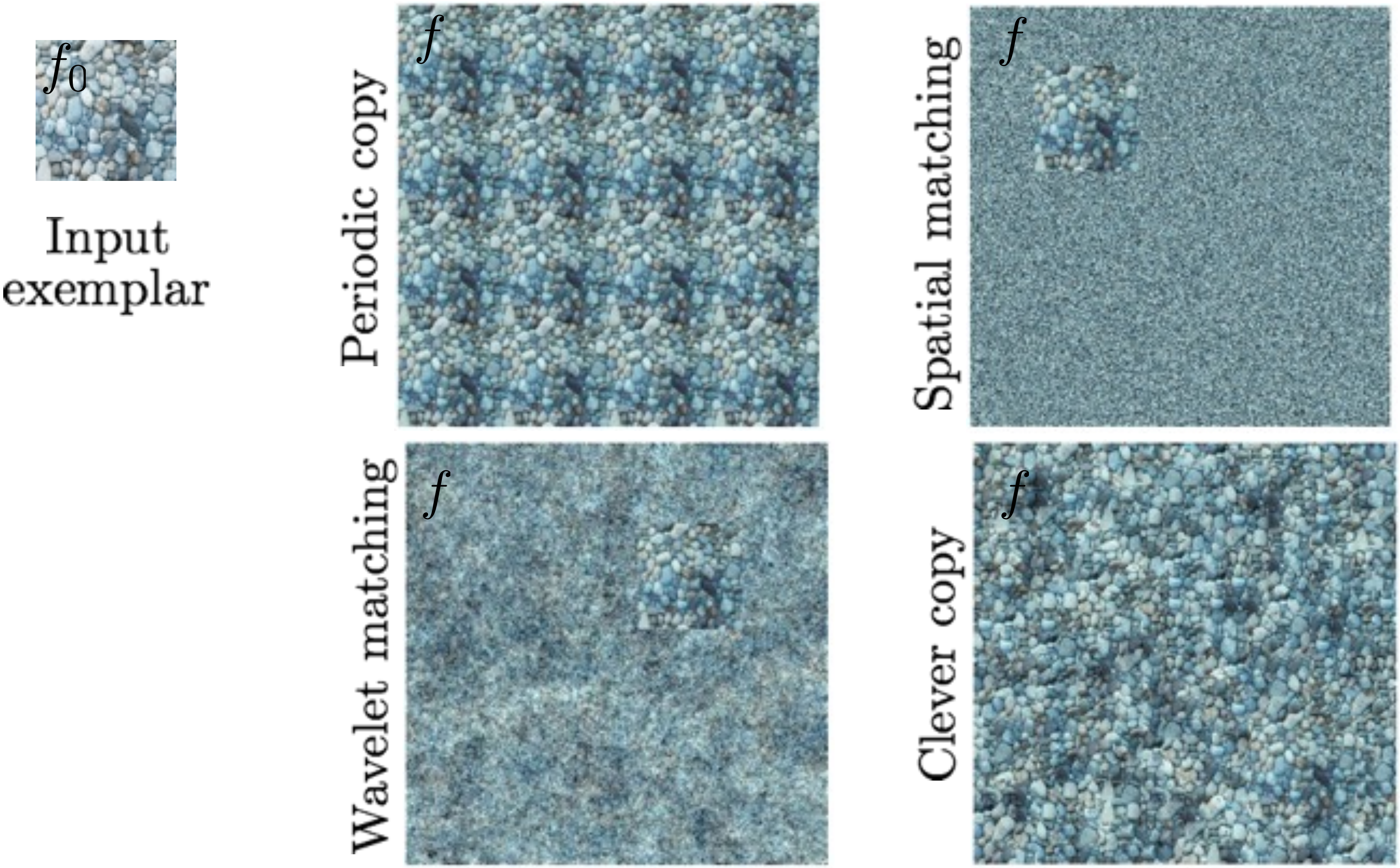


Modified image

Other applications:
- Texture synthesis
- Texture segmentation

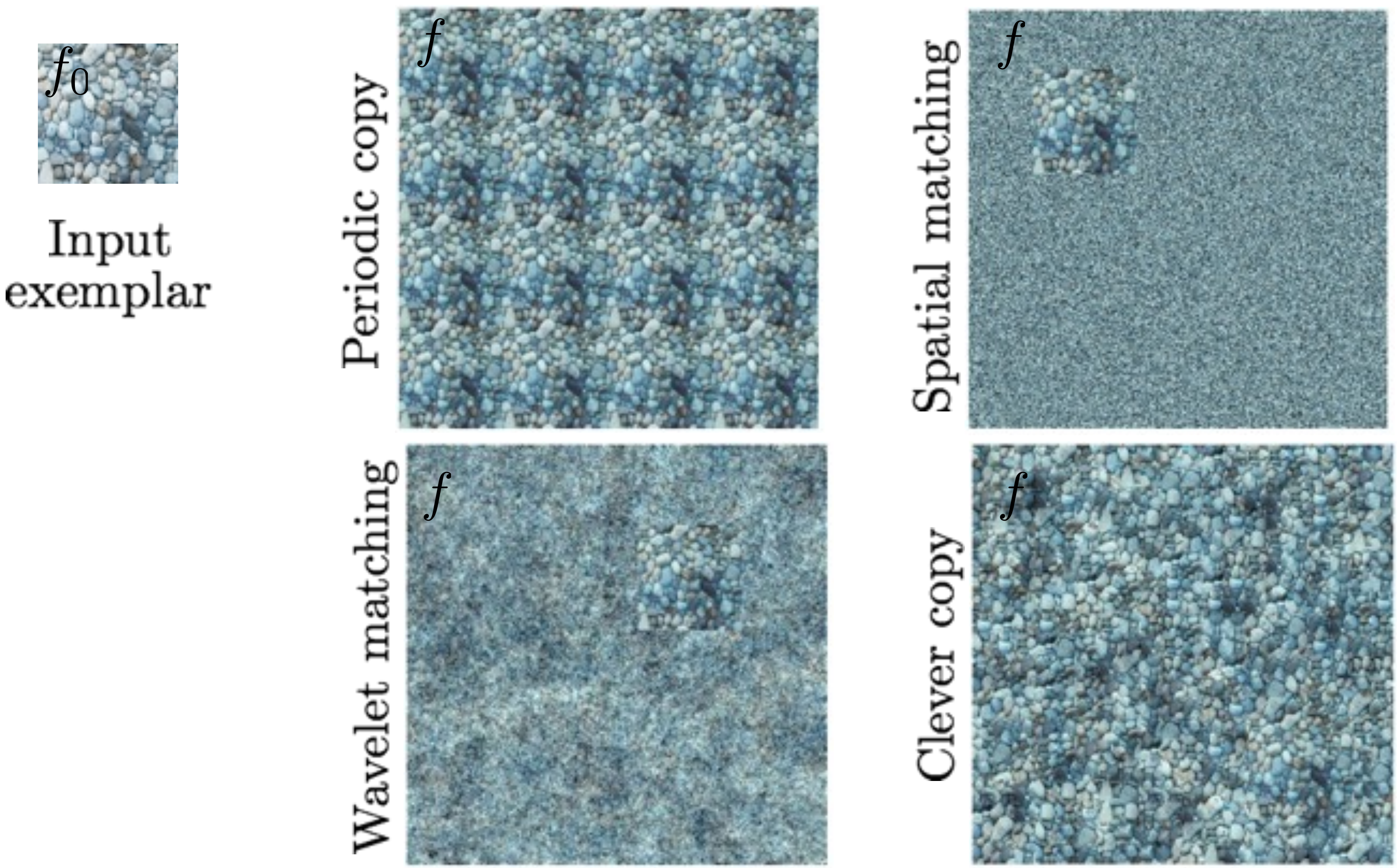
Texture Synthesis

Generate f perceptually similar to some input f_0



Texture Synthesis

Generate f perceptually similar to some input f_0



$\longrightarrow$ Design and manipulate statistical constraints.

$\longrightarrow$ Use statistical constraints for other imaging problems.

Overview

- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- Gaussian Texture Models

Discrete Distributions

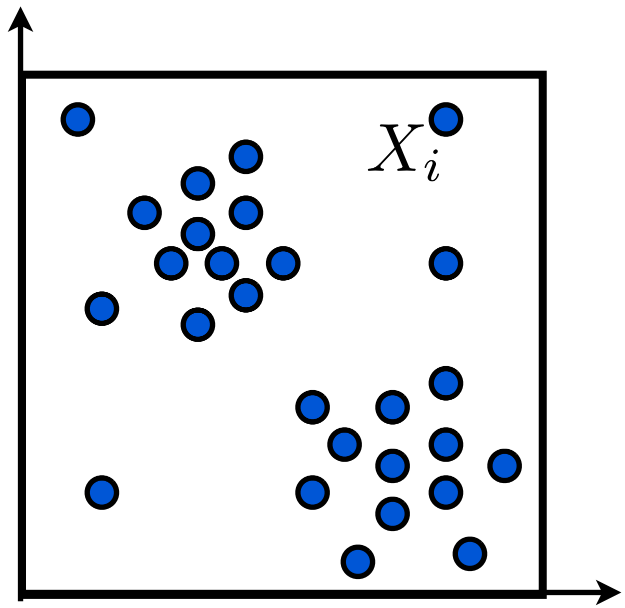
Discrete measure: $\mu = \sum_{i=0}^{N-1} p_i \delta_{X_i} \qquad X_i \in \mathbb{R}^d \qquad \sum_i p_i = 1$

Discrete Distributions

Discrete measure: $\mu = \sum_{i=0}^{N-1} p_i \delta_{X_i} \qquad X_i \in \mathbb{R}^d \qquad \sum_i p_i = 1$

Point cloud

Constant weights: $p_i = 1/N$.



Quotient space:

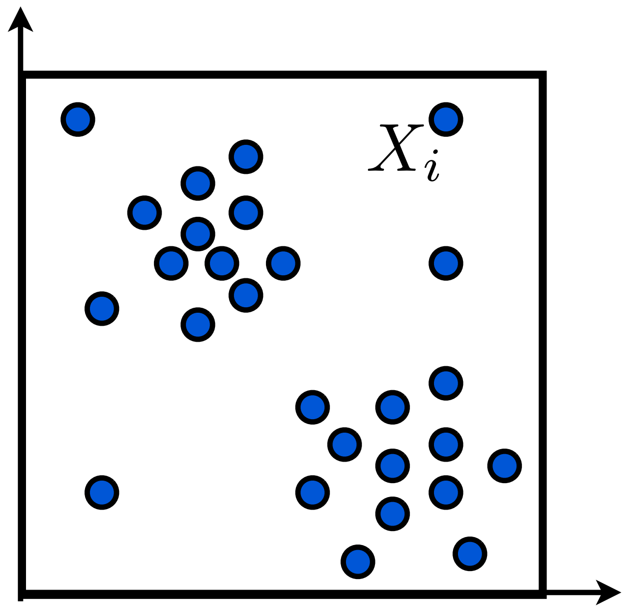
$$\mathbb{R}^{N \times d} / \Sigma_N$$

Discrete Distributions

Discrete measure: $\mu = \sum_{i=0}^{N-1} p_i \delta_{X_i} \qquad X_i \in \mathbb{R}^d \qquad \sum_i p_i = 1$

Point cloud

Constant weights: $p_i = 1/N$.

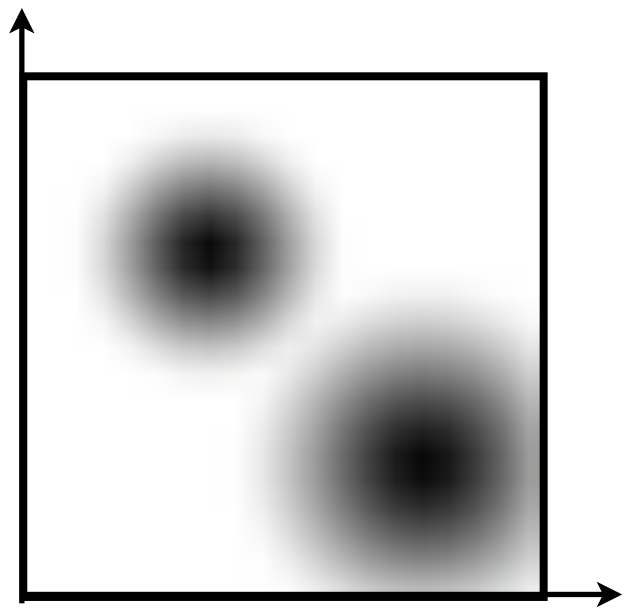


Quotient space:

$\mathbb{R}^{N \times d} / \Sigma_N$

Histogram

Fixed positions X_i (e.g. grid)



Affine space:

$\{(p_i)_i \setminus \sum_i p_i = 1\}$

From Images to Statistics

Discretized image $f \in \mathbb{R}^{N \times d}$

$$N = \#\text{pixels}, \; d = \#\text{colors}.$$

$f_i \in \mathbb{R}^d = \mathbb{R}^3$

From Images to Statistics

Discretized image $f \in \mathbb{R}^{N \times d}$

$N = $ #pixels, $d = $ #colors.

$f_i \in \mathbb{R}^d = \mathbb{R}^3$

Disclamers: images are **not** distributions.

$\rightarrow$ Needs an estimator: $f \longrightarrow \mu_f$

$\rightarrow$ Modify f by controlling μ_f.

From Images to Statistics

Discretized image $f \in \mathbb{R}^{N \times d}$

$N = \#\text{pixels},\ d = \#\text{colors}.$

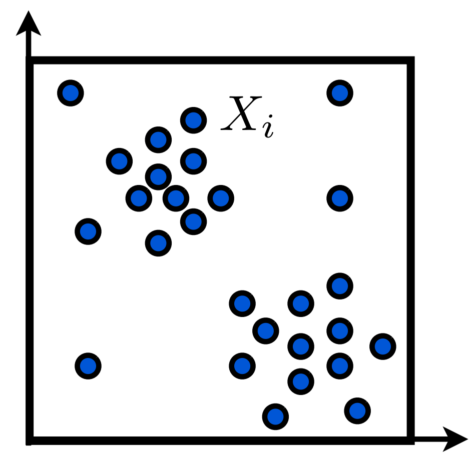
$f_i \in \mathbb{R}^d = \mathbb{R}^3$

Disclamers: images are **not** distributions.

$\rightarrow$ Needs an estimator: $\quad f \longrightarrow \mu_f$

$\rightarrow$ Modify f by controlling μ_f.

Point cloud discretization: $\quad \mu_f = \sum_i \delta_{f_i}$



From Images to Statistics

Discretized image $f \in \mathbb{R}^{N \times d}$

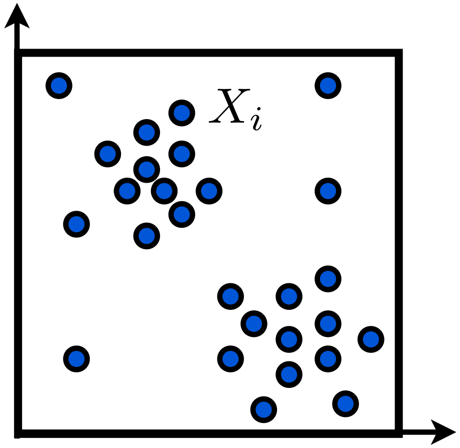
$N = \#\text{pixels},\ d = \#\text{colors}.$

Disclamers: images are **not** distributions.

$\rightarrow$ Needs an estimator: $f \longrightarrow \mu_f$

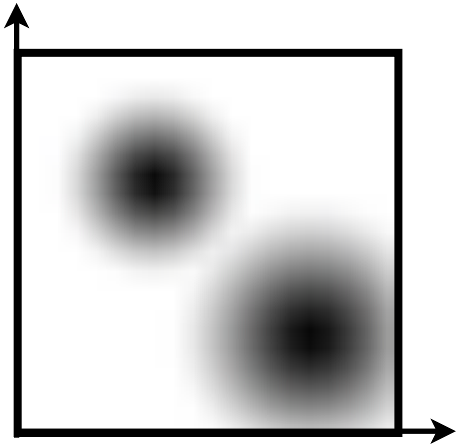
$\rightarrow$ Modify f by controlling μ_f.

Point cloud discretization: $\mu_f = \sum_i \delta_{f_i}$



Histogram discretization: $\mu_f = \sum_i p_i \delta_{X_i}$

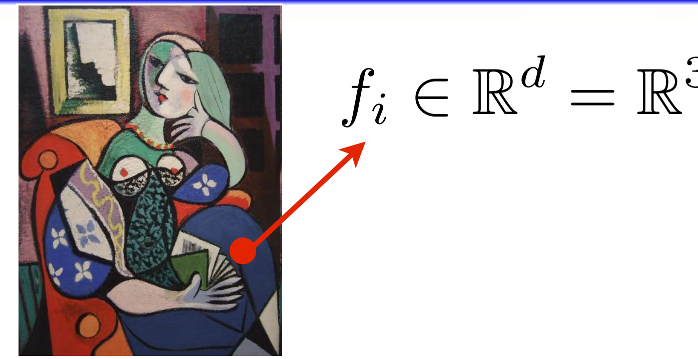
Parzen windows: $p_i = \frac{1}{Z_f} \sum_j \psi(x_i - f_j)$



From Images to Statistics

Discretized image $f \in \mathbb{R}^{N \times d}$

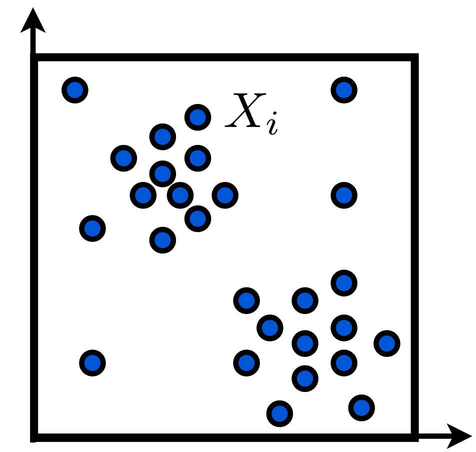
$N = \#\text{pixels}, \; d = \#\text{colors}.$



Disclamers: images are **not** distributions.

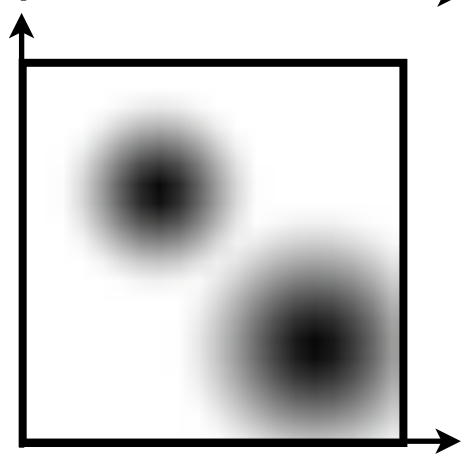
$\rightarrow$ Needs an estimator: $\quad f \longrightarrow \mu_f$

$\rightarrow$ Modify f by controlling μ_f.



Point cloud discretization: $\quad \mu_f = \sum_i \delta_{f_i}$

Today's focus



Histogram discretization: $\quad \mu_f = \sum_i p_i \delta_{X_i}$

Parzen windows: $\; p_i = \dfrac{1}{Z_f} \sum_j \psi(x_i - f_j)$

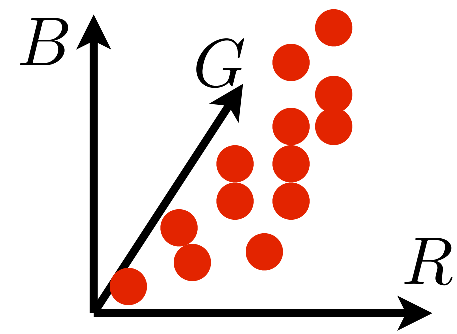
Optimal Transport Distances

Vector $X \in \mathbb{R}^{N \times d}$ $\longrightarrow$ $\mu_X = \sum_i \delta_{X_i}$

(image, coefficients, $\ldots$)

Grayscale: 1-D

Colors: 3-D

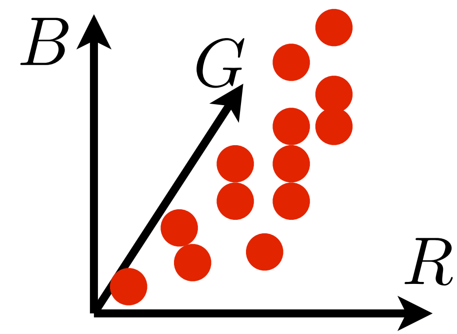


Optimal Transport Distances

Vector $X \in \mathbb{R}^{N \times d}$ $\longrightarrow$ $\mu_X = \sum_i \delta_{X_i}$
(image, coefficients, ...)

Grayscale: 1-D

Colors: 3-D



Optimal assignment:

$$\sigma^\star \in \underset{\sigma \in \Sigma_N}{\operatorname{argmin}} \sum_i \|X_i - Y_{\sigma(i)}\|^p$$

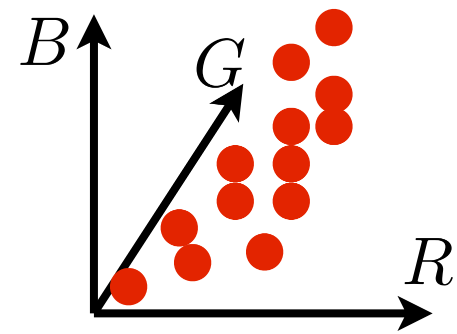
Optimal Transport Distances

Vector $X \in \mathbb{R}^{N \times d} \longrightarrow \mu_X = \sum_i \delta_{X_i}$

(image, coefficients, ...)

Grayscale: 1-D

Colors: 3-D



Optimal assignment:

$$\sigma^\star \in \underset{\sigma \in \Sigma_N}{\operatorname{argmin}} \sum_i \|X_i - Y_{\sigma(i)}\|^p$$

Wasserstein distance: $\quad W_p(\mu_X, \mu_Y)^p = \sum_i \|X_i - Y_{\sigma^\star(i)}\|^p$

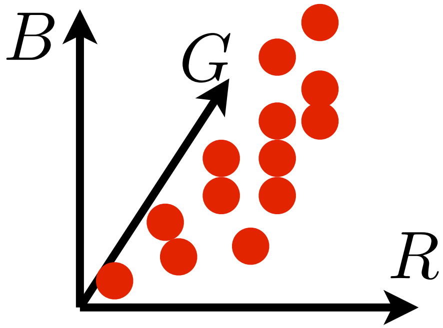
$\longrightarrow$ Metric on the space of distributions.

Optimal Transport Distances

Vector $X \in \mathbb{R}^{N \times d}$ $\longrightarrow$ $\mu_X = \sum_i \delta_{X_i}$
(image, coefficients, ...)

Grayscale: 1-D

Colors: 3-D



Optimal assignment:

$$\sigma^\star \in \underset{\sigma \in \Sigma_N}{\operatorname{argmin}} \sum_i \|X_i - Y_{\sigma(i)}\|^p$$

Wasserstein distance: $W_p(\mu_X, \mu_Y)^p = \sum_i \|X_i - Y_{\sigma^\star(i)}\|^p$

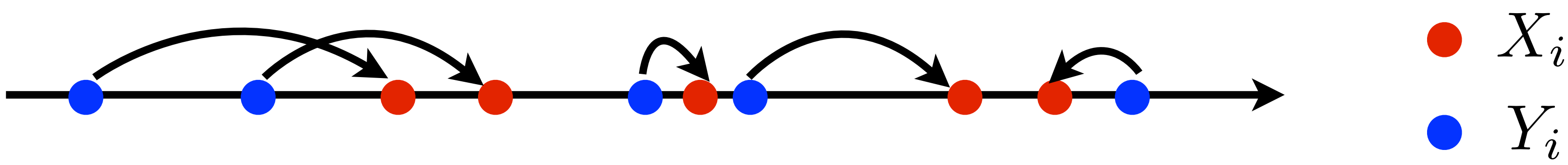
$\longrightarrow$ Metric on the space of distributions.

Projection on statistical constraints: $\mathcal{C} = \{f \setminus \mu_f = \mu_Y\}$

$$\mathrm{Proj}_{\mathcal{C}}(f) = Y \circ \sigma^\star$$

Computing Transport Distances

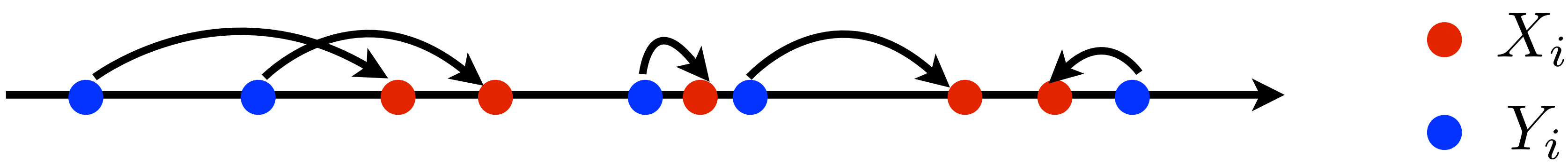
Explicit solution for 1D distribution (e.g. grayscale images):



$\longrightarrow$ sorting the values, $O(N\log(N))$ operations.

Computing Transport Distances

Explicit solution for 1D distribution (e.g. grayscale images):



$\longrightarrow$ sorting the values, $O(N \log(N))$ operations.

Higher dimensions: combinatorial optimization methods

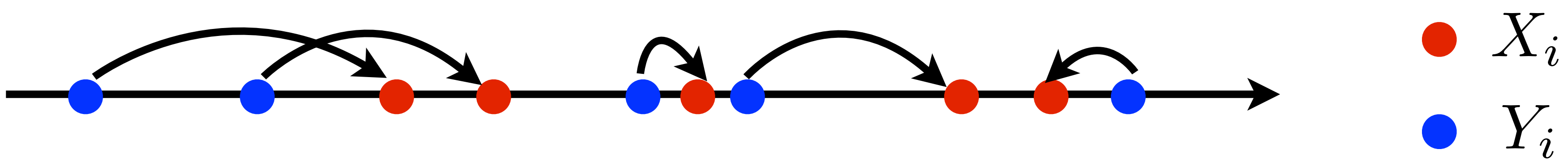
Hungarian algorithm, auctions algorithm, etc.

$\longrightarrow$ $O(N^{5/2} \log(N))$ operations.

$\longrightarrow$ intractable for imaging problems.

Computing Transport Distances

Explicit solution for 1D distribution (e.g. grayscale images):



$\longrightarrow$ sorting the values, $O(N\log(N))$ operations.

Higher dimensions: combinatorial optimization methods

Hungarian algorithm, auctions algorithm, etc.

$\longrightarrow$ $O(N^{5/2}\log(N))$ operations.

$\longrightarrow$ intractable for imaging problems.

Arbitrary distributions: $\mu = \sum_i p_i \delta_{X_i}$ $\quad \nu = \sum_i q_i \delta_{Y_i}$

$\longrightarrow$ $W_p(\mu,\nu)^p$ solution of a linear program.

Convex Formulation

Probabilistic coupling: $\mu = \sum_i p_i \delta_{X_i} \quad \nu = \sum_i q_i \delta_{Y_i}$

$$\Pi_{\mu,\nu} = \left\{ P \in \mathbb{R}^{N \times N} \setminus P \geqslant 0,\; P1 = p,\; P'1 = q \right\}$$

Linear programming (Kantorovitch):

$$P^\star \in \underset{P \in \Pi_{\mu,\nu}}{\operatorname{argmin}} \langle P,\, C \rangle = \sum_{i,j} C_{i,j} P_{i,j}$$

$$C_{i,j} = \|X_i - Y_j\|^p$$

$$W(\mu,\nu)^p = \langle P^\star,\, C \rangle$$

Convex Formulation

Probabilistic coupling: $\mu = \sum_i p_i \delta_{X_i} \qquad \nu = \sum_i q_i \delta_{Y_i}$

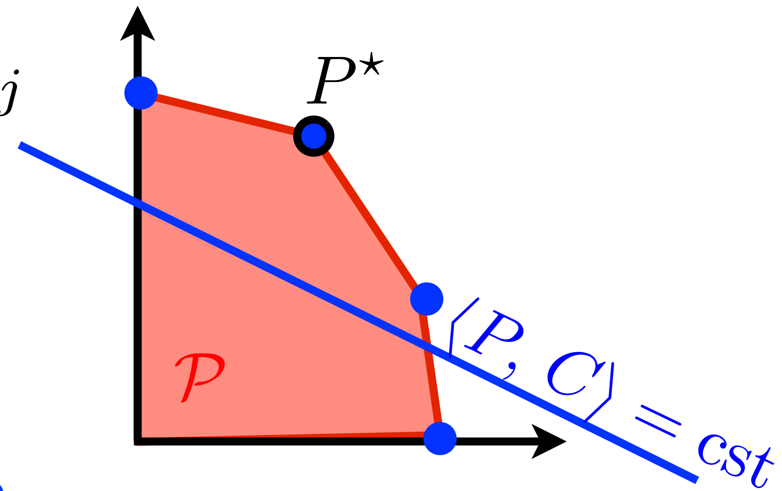
$$\Pi_{\mu,\nu} = \left\{ P \in \mathbb{R}^{N\times N} \setminus P \geqslant 0,\ P\mathbb{1} = p,\ P'\mathbb{1} = q \right\}$$

Linear programming (Kantorovitch):

$$P^\star \in \underset{P\in\Pi_{\mu,\nu}}{\operatorname{argmin}} \langle P, C\rangle = \sum_{i,j} C_{i,j} P_{i,j}$$

$$C_{i,j} = \|X_i - Y_j\|^p$$

$$W(\mu,\nu)^p = \langle P^\star, C\rangle$$



If $p_i = q_i = 1/N$, extremal points:

Permutation matrices: $P = P_\sigma = (\delta_{i-\sigma(j)})_{i,j}$

$$\boxed{\textit{Theorem: } P^\star = P_{\sigma^\star}}$$

Convex Formulation

Probabilistic coupling: $\mu = \sum_i p_i \delta_{X_i} \qquad \nu = \sum_i q_i \delta_{Y_i}$

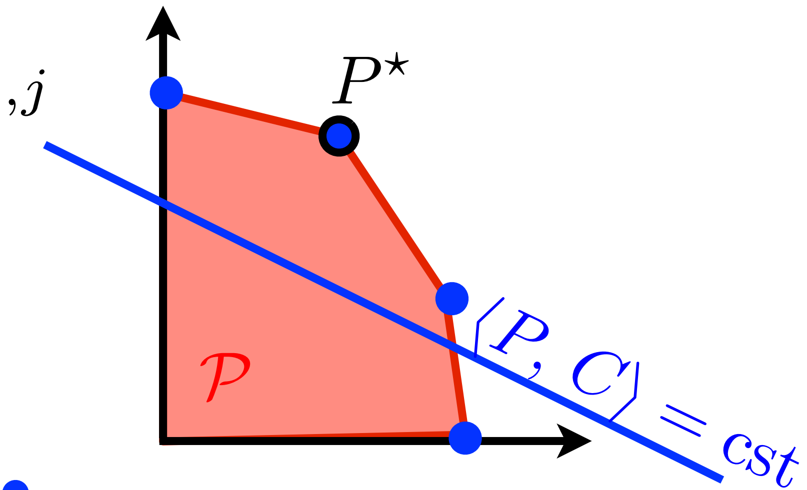
$$\Pi_{\mu,\nu} = \{P \in \mathbb{R}^{N\times N} \setminus P \geqslant 0,\ P1 = p,\ P'1 = q\}$$

Linear programming (Kantorovitch):

$$P^\star \in \underset{P\in\Pi_{\mu,\nu}}{\operatorname{argmin}} \langle P,\, C\rangle = \sum_{i,j} C_{i,j} P_{i,j}$$

$$C_{i,j} = \|X_i - Y_j\|^p$$

$$W(\mu,\nu)^p = \langle P^\star,\, C\rangle$$



If $p_i = q_i = 1/N$, extremal points:

Permutation matrices: $P = P_\sigma = (\delta_{i-\sigma(j)})_{i,j}$

$$\boxed{\textit{Theorem: } P^\star = P_{\sigma^\star}}$$

Faster methods: Hungarian algorithm, auctions algorithm, etc.

$\longrightarrow O(N^{5/2}\log(N))$ operations. $\longrightarrow$ intractable for imaging.

Optimization Codes

Discrete optimal transport:

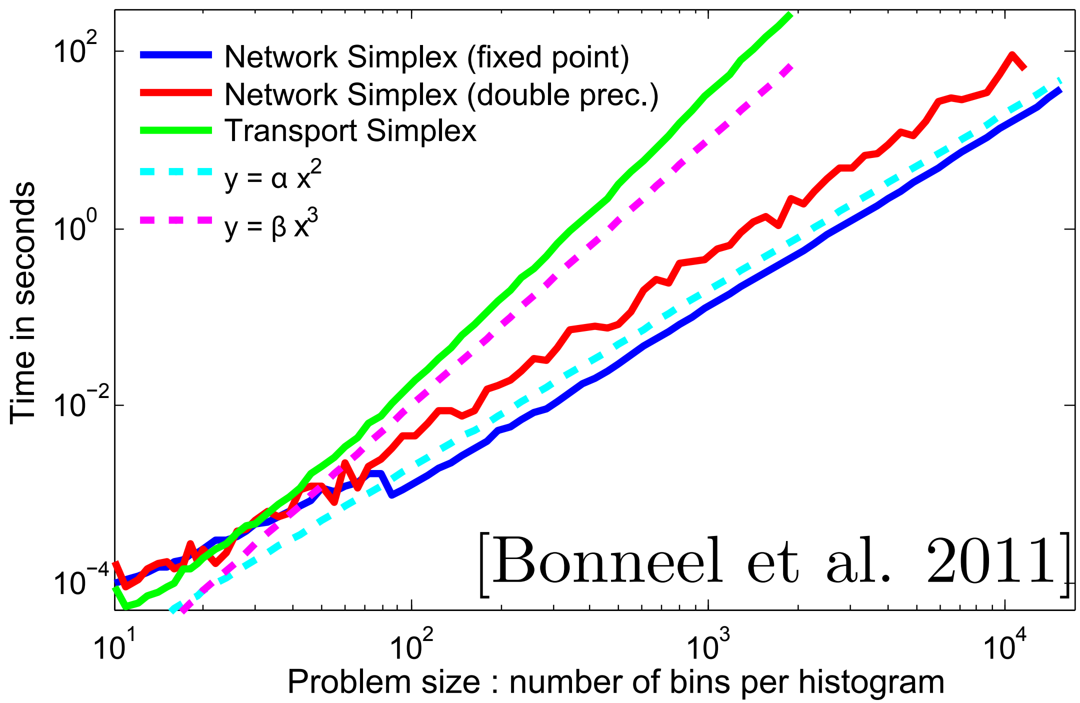
$$P^\star \in \underset{P \in \Pi_{\mu,\nu}}{\operatorname{argmin}} \langle P, C \rangle = \sum_{i,j} C_{i,j} P_{i,j}$$

Linear program:

$\rightarrow$ Interior points: slow.

$\rightarrow$ Network simplex.

$\rightarrow$ Transportation simplex.



[Bonneel et al. 2011]

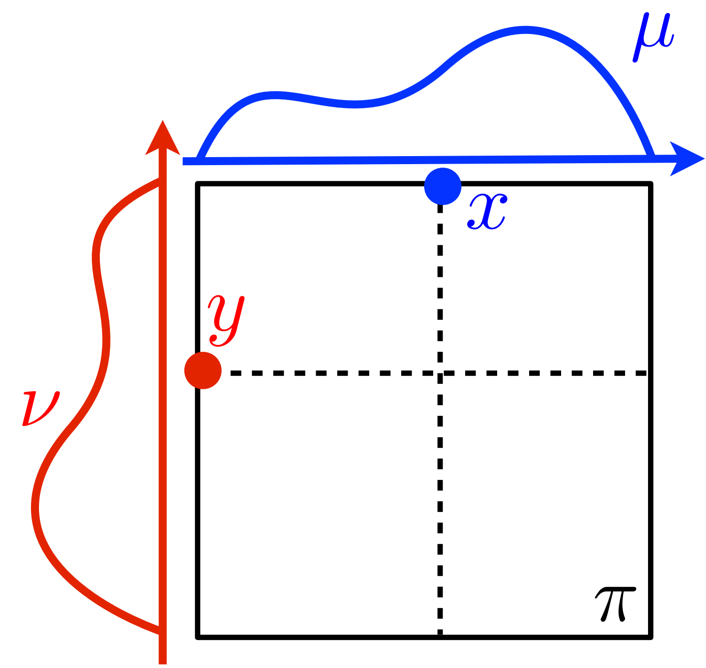
Block search pivoting strategy [Kelly and O'Neill 1991]

Continuous Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.

Couplings: $\pi \in \Pi_{\mu,\nu}$

$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$



Continuous Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.

Couplings: $\quad \pi \in \Pi_{\mu,\nu}$

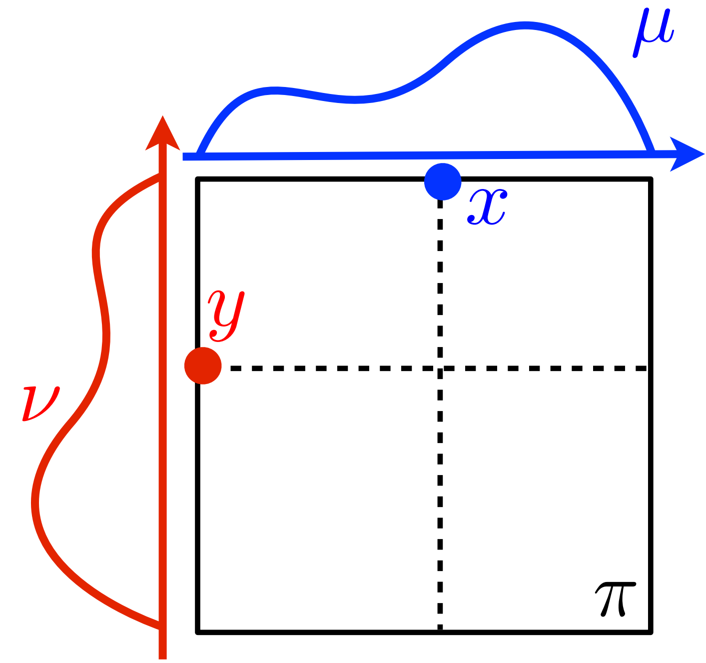
$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$

Transportation cost:

$$c(x,y) = \|x-y\|^p$$

L^p Wasserstein distance:

$$W_p(\mu,\nu)^p = \min_{\pi \in \Pi_{\mu,\nu}} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\mathrm{d}\pi(x,y)$$

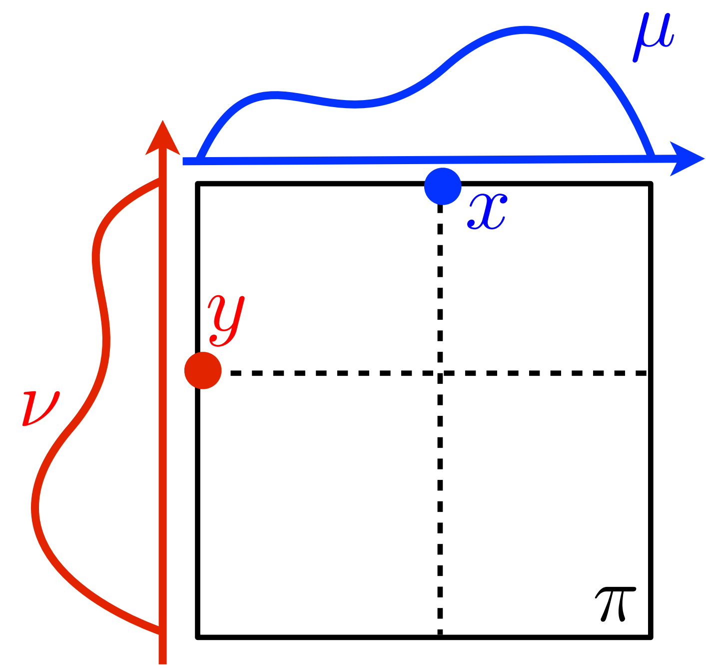


Continuous Wasserstein Distance

Input measures μ, ν on $\mathbb{R}^d$.

Couplings: $\pi \in \Pi_{\mu,\nu}$

$$\forall\, A \subset \mathbb{R}^d, \pi(A \times \mathbb{R}^d) = \mu(A)$$
$$\forall\, B \subset \mathbb{R}^d, \pi(\mathbb{R}^d \times B) = \nu(B)$$



Transportation cost:

$$c(x, y) = \|x - y\|^p$$

L^p Wasserstein distance:

$$W_p(\mu, \nu)^p = \min_{\pi \in \Pi_{\mu,\nu}} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \mathrm{d}\pi(x, y)$$

Uniqueness: (μ does not vanish on small sets)

If $p > 1$, $\exists! \, \pi$ optimal.

Continuous Optimal Transport

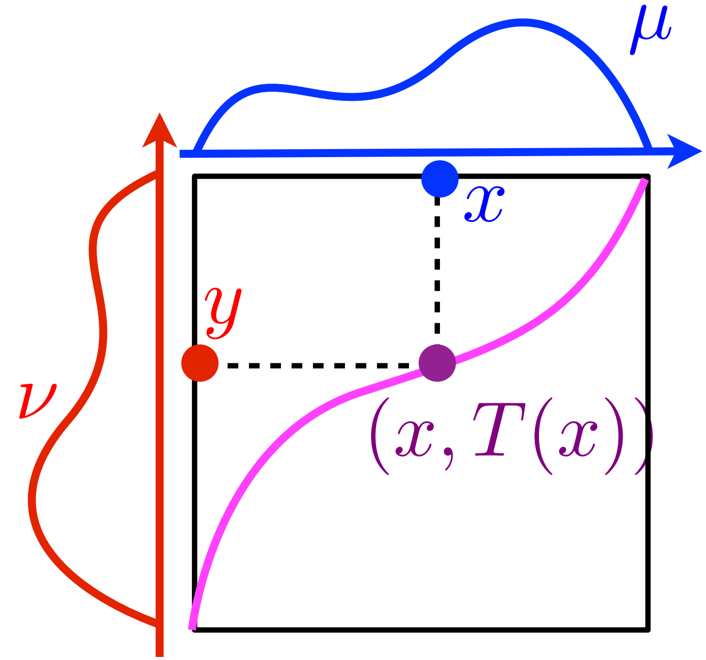
Let $p > 1$ and μ does not vanish on small sets.

$$\text{Unique } \pi \in \Pi_{\mu,\nu} \text{ s.t. } \quad W_p(\mu,\nu)^p = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \mathrm{d}\pi(x,y)$$

Optimal transport $T : \mathbb{R}^d \to \mathbb{R}^d$:

π is supported on the graph of $x \mapsto y = T(x)$.

$$\mathrm{d}\pi(x,y) = \mathrm{d}\mu(x)\delta(y = T(x))$$



Continuous Optimal Transport

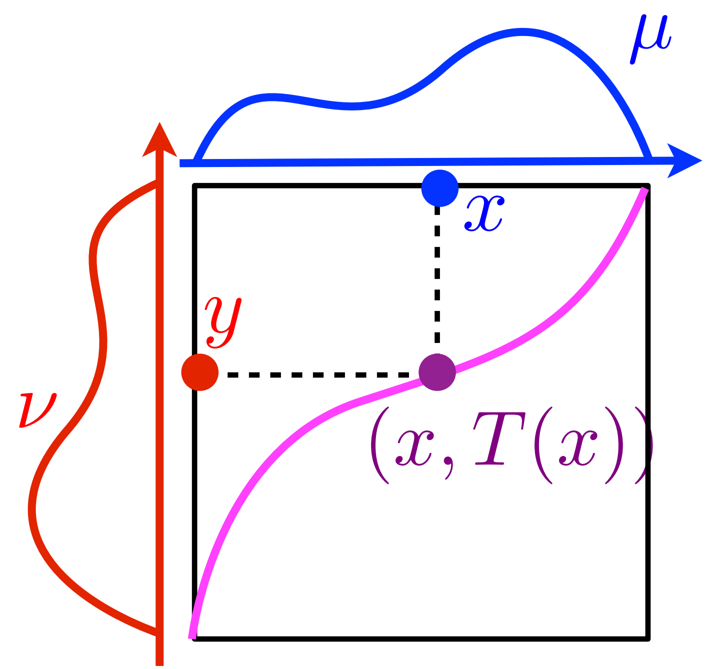
Let $p > 1$ and μ does not vanish on small sets.

$$\text{Unique } \pi \in \Pi_{\mu,\nu} \text{ s.t.} \quad W_p(\mu,\nu)^p = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \mathrm{d}\pi(x,y)$$

Optimal transport $T : \mathbb{R}^d \to \mathbb{R}^d$:

π is supported on the graph of $x \mapsto y = T(x)$.

$$\mathrm{d}\pi(x,y) = \mathrm{d}\mu(x)\delta(y = T(x))$$



$p = 2$: $T = \nabla\varphi$ unique solution of

$$\begin{cases} \varphi \text{ is convex l.s.c.} \\ (\nabla\varphi)\sharp\mu = \nu \end{cases}$$

1-D Continuous Wasserstein

Distributions μ, ν on $\mathbb{R}$.

Cumulative functions: $$C_\mu(t) = \int_{-\infty}^{t} \mathrm{d}\mu(x)$$

For all $p > 1$: $$T = C_\nu^{-1} \circ C_\mu$$

T is non-decreasing ("change of contrast")

1-D Continuous Wasserstein

Distributions μ, ν on $\mathbb{R}$.

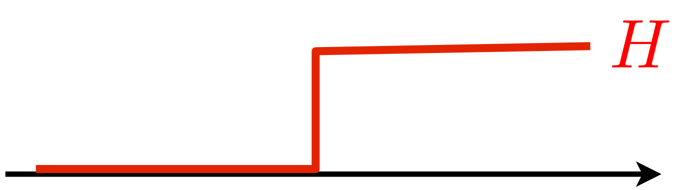
Cumulative functions: $$C_\mu(t) = \int_{-\infty}^{t} \mathrm{d}\mu(x)$$

For all $p > 1$: $$T = C_\nu^{-1} \circ C_\mu$$

T is non-decreasing ("change of contrast")

Explicit formulas:

$$W_p(\mu,\nu)^p = \int_0^1 |C_\mu^{-1} - C_\nu^{-1}|^p$$



$$W_1(\mu,\nu) = \int_{\mathbb{R}} |C_\mu - C_\nu| = \|(C_\mu - C_\nu) \star H\|_1$$

Continuous Histogram Transfer

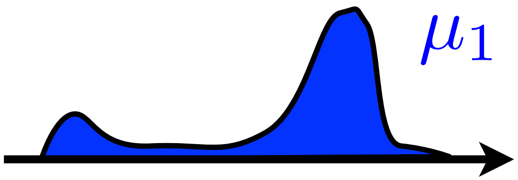
Input images: $f_i : [0,1]^2 \to [0,1], i = 0, 1.$

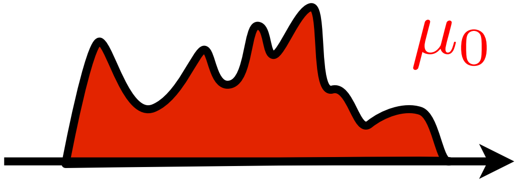
Continuous Histogram Transfer

Input images: $f_i : [0,1]^2 \rightarrow [0,1], i = 0, 1.$

Gray-value distributions: μ_i defined on $[0,1]$.

$$\mu_i([a,b]) = \int_{[0,1]^2} 1_{\{a \leqslant f \leqslant b\}}(x)\mathrm{d}x$$





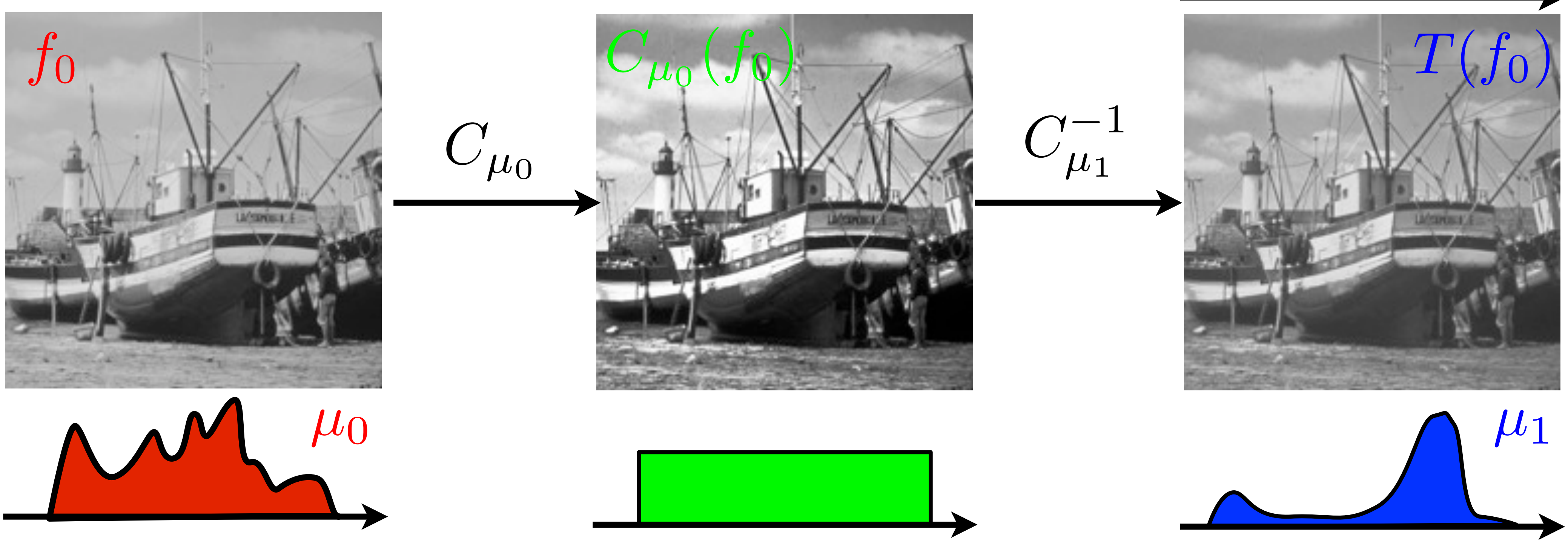
Continuous Histogram Transfer

Input images: $f_i : [0,1]^2 \to [0,1], i = 0, 1.$

Gray-value distributions: μ_i defined on $[0,1]$.

$$\mu_i([a,b]) = \int_{[0,1]^2} 1_{\{a \leqslant f \leqslant b\}}(x) \mathrm{d}x$$

Optimal transport: $T = C_{\mu_1}^{-1} \circ C_{\mu_0}$.



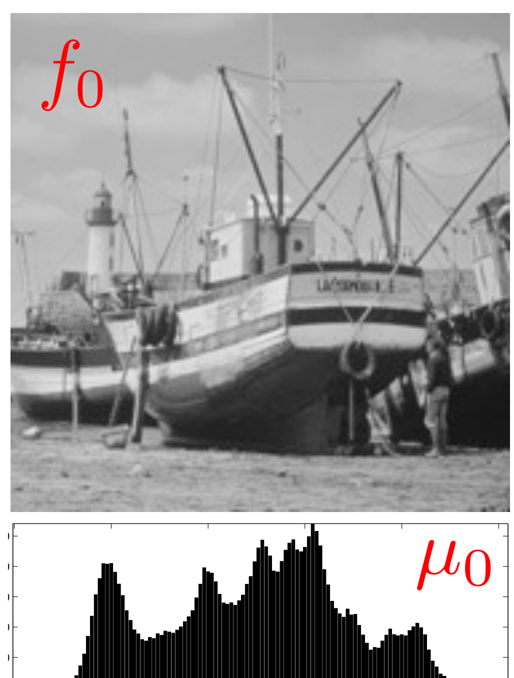
Discrete Histogram Transfer

Discretized grayscale images $f_0, f_1 \in \mathbb{R}^N$.

Discrete Histogram Transfer

Discretized grayscale images $f_0, f_1 \in \mathbb{R}^N$.

Discrete distributions $\mu_i = \mu_{f_i} = N^{-1} \sum_k \delta_{f_i(k)}$.



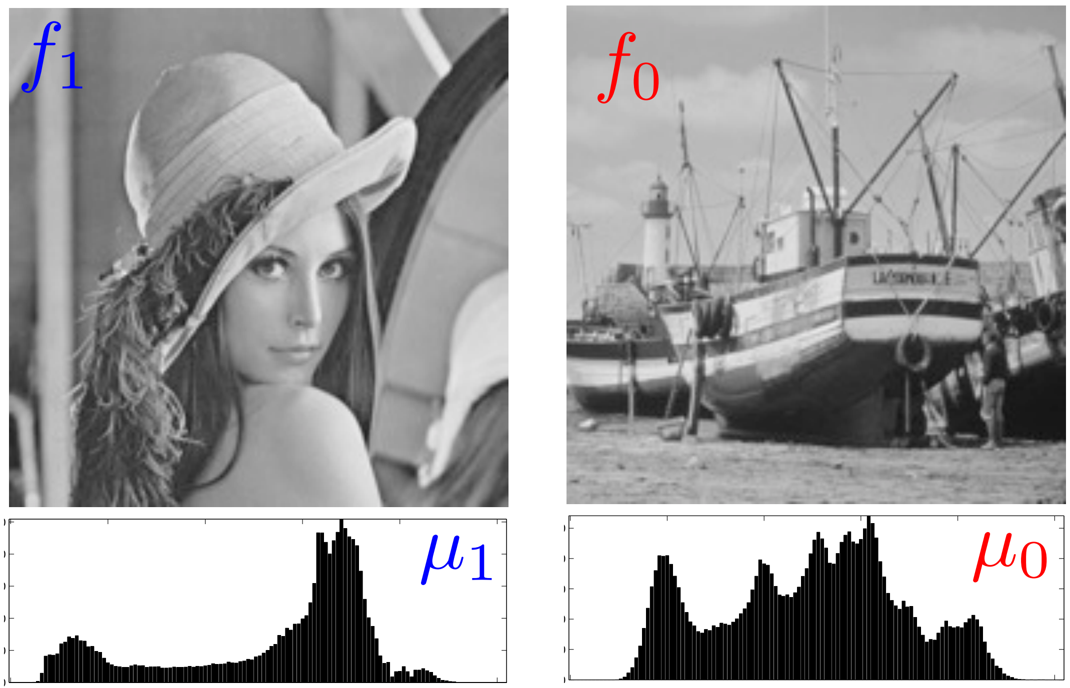
Discrete Histogram Transfer

Discretized grayscale images $f_0, f_1 \in \mathbb{R}^N$.

Discrete distributions $\mu_i = \mu_{f_i} = N^{-1} \sum_k \delta_{f_i(k)}$.

Sorting the values : $\sigma_i \in \Sigma_N$ s.t. $f_i(\sigma_i(k)) \leqslant f_i(\sigma_i(k+1))$.

Optimal transport: $T : f_0(\sigma_0(k)) \mapsto f_1(\sigma_1(k))$



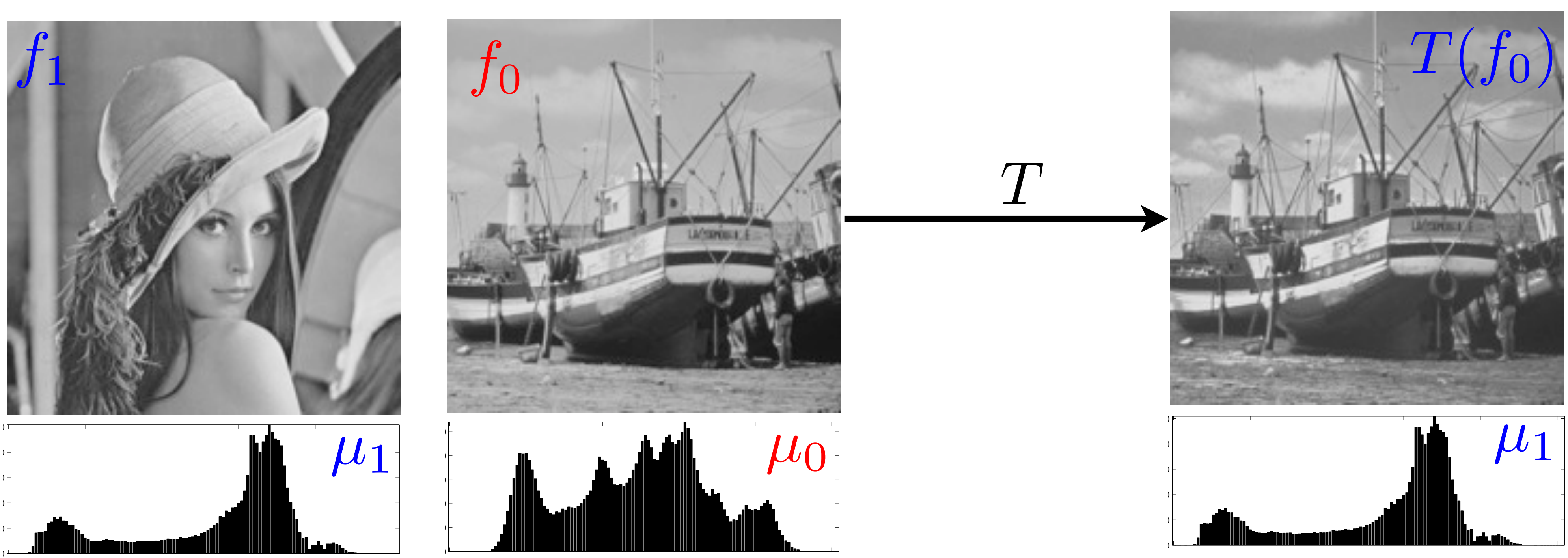
Discrete Histogram Transfer

Discretized grayscale images $f_0, f_1 \in \mathbb{R}^N$.

Discrete distributions $\mu_i = \mu_{f_i} = N^{-1} \sum_k \delta_{f_i(k)}$.

Sorting the values : $\sigma_i \in \Sigma_N$ s.t. $f_i(\sigma_i(k)) \leqslant f_i(\sigma_i(k+1))$.

Optimal transport: $T : f_0(\sigma_0(k)) \mapsto f_1(\sigma_1(k))$



Discrete Histogram Transfer

Discretized grayscale images $f_0, f_1 \in \mathbb{R}^N$.

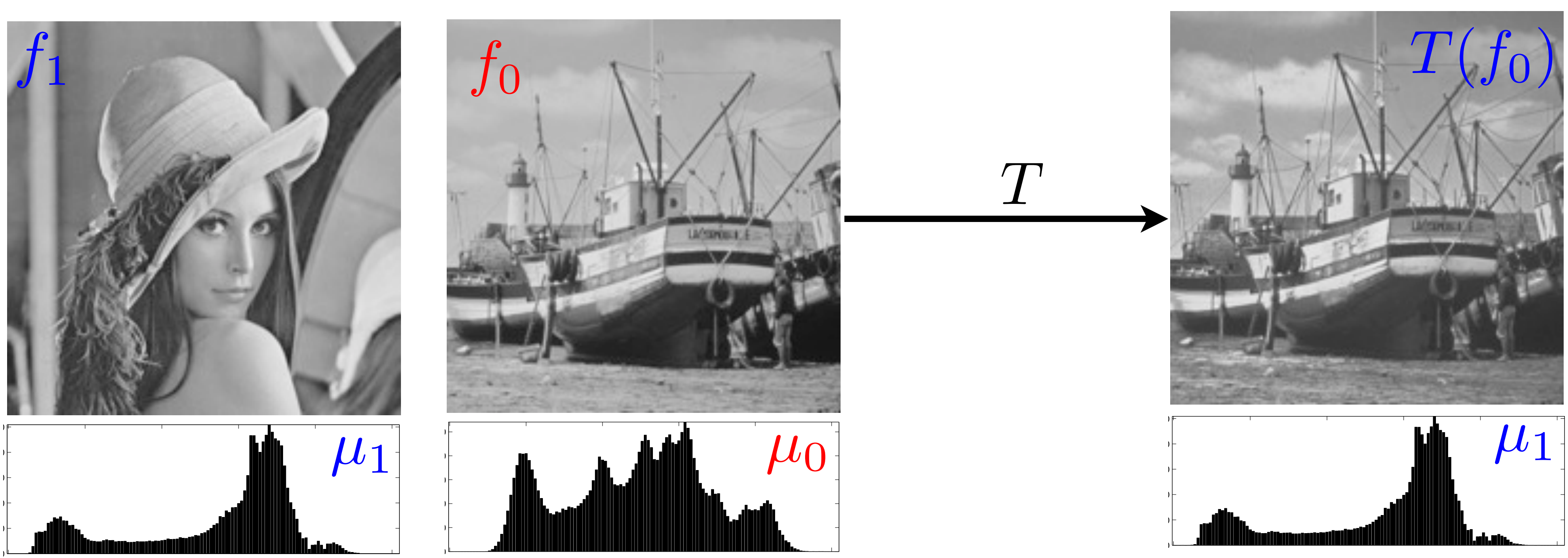
Discrete distributions $\mu_i = \mu_{f_i} = N^{-1} \sum_k \delta_{f_i(k)}$.

Sorting the values : $\sigma_i \in \Sigma_N$ s.t. $f_i(\sigma_i(k)) \leqslant f_i(\sigma_i(k+1))$.

Optimal transport: $T : f_0(\sigma_0(k)) \mapsto f_1(\sigma_1(k))$

Matlab code:

```
[a,I] = sort(f0(:));
f0(I) = sort(f1(:));
```



PDE Formulations

Smooth distributions: $\mu_i = \rho_i(x)\mathrm{d}x$

$$T\sharp\mu_0 = \mu_1 \quad \Longleftrightarrow \quad \rho_1(T(x))|\det \partial T(x)| = \rho_0(x)$$

PDE Formulations

Smooth distributions: $\mu_i = \rho_i(x)\mathrm{d}x$

$$T\sharp\mu_0 = \mu_1 \quad \Longleftrightarrow \quad \rho_1(T(x))|\det \partial T(x)| = \rho_0(x)$$

L^2 optimal transport map $T = \nabla\psi$:

$$\rho_1(\nabla\psi(x))\det(H\psi) = \rho_0(x) \qquad \text{(Monge-Ampère)}$$

PDE Formulations

Smooth distributions: $\mu_i = \rho_i(x)\mathrm{d}x$

$$T\sharp\mu_0 = \mu_1 \quad \Longleftrightarrow \quad \rho_1(T(x))|\det \partial T(x)| = \rho_0(x)$$

L^2 optimal transport map $T = \nabla\psi$:

$$\rho_1(\nabla\psi(x))\det(H\psi) = \rho_0(x) \qquad \text{(Monge-Ampère)}$$

Fluid dynamic formulation: find $\rho(x,t) \geqslant 0, m(x,t) \in \mathbb{R}^d$

$$W(\mu_0,\mu_1)^2 = \min_{\rho,m} \int_{\mathbb{R}^d}\int_0^1 \frac{\|m\|^2}{\rho} \quad \text{s.t.} \quad \begin{array}{l} \partial_t\rho + \nabla\cdot m = 0 \\ \rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1 \end{array}$$

$\rightarrow$ Finite element discretization [Benamou-Brenier]

PDE Formulations

Smooth distributions: $\mu_i = \rho_i(x)\mathrm{d}x$

$$T\sharp\mu_0 = \mu_1 \quad \Longleftrightarrow \quad \rho_1(T(x))|\det \partial T(x)| = \rho_0(x)$$

L^2 optimal transport map $T = \nabla\psi$:

$$\rho_1(\nabla\psi(x))\det(H\psi) = \rho_0(x) \qquad \text{(Monge-Ampère)}$$

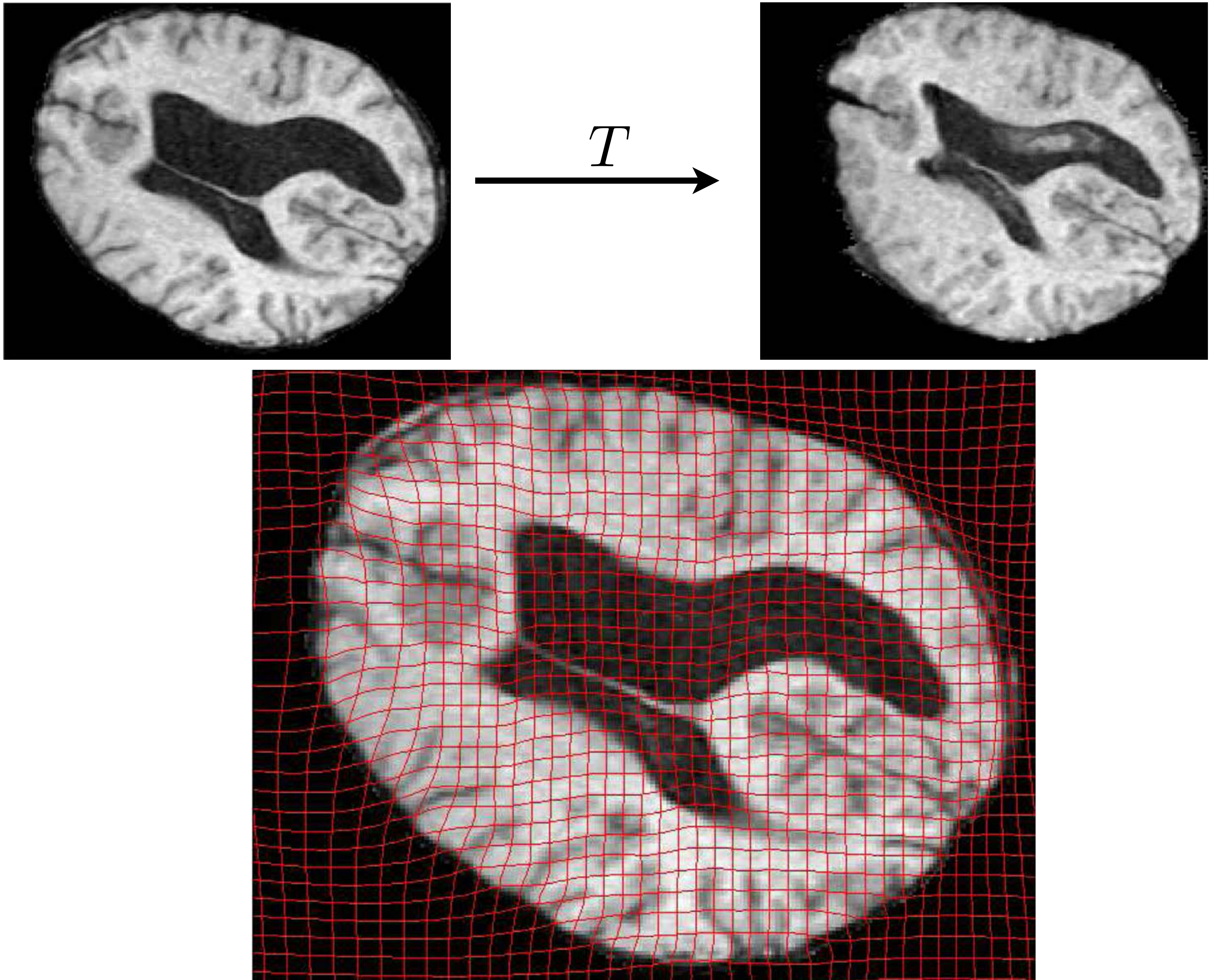
Fluid dynamic formulation: find $\rho(x,t) \geqslant 0, m(x,t) \in \mathbb{R}^d$

$$W(\mu_0,\mu_1)^2 = \min_{\rho,m} \int_{\mathbb{R}^d}\int_0^1 \frac{\|m\|^2}{\rho} \quad \text{s.t.} \quad \begin{array}{l} \partial_t\rho + \nabla\cdot m = 0 \\ \rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1 \end{array}$$

$\rightarrow$ Finite element discretization [Benamou-Brenier]

Related works of [Tannenbaum et al.].

Image Registration



[ur Rehman et al, 2009]

Overview

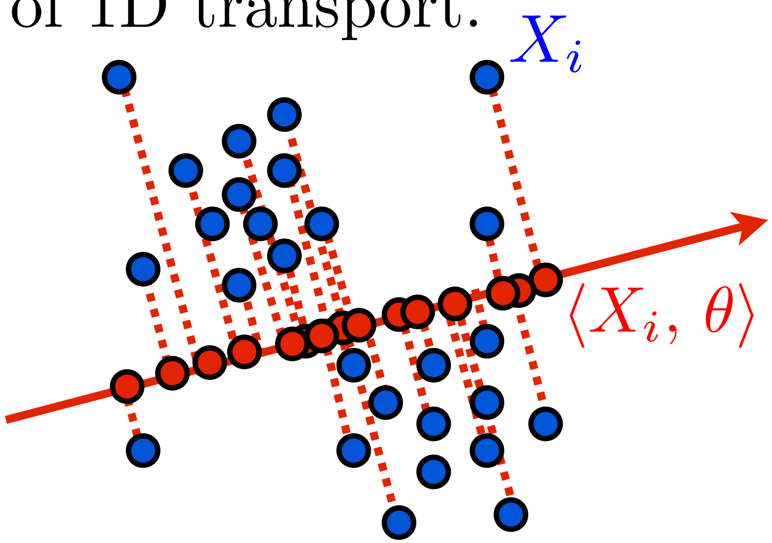
- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- Gaussian Texture Models

Approximate Sliced Distance

Key idea: replace transport in $\mathbb{R}^d$ by series of 1D transport.

[Rabin, Peyré, Delon & Bernot 2010]

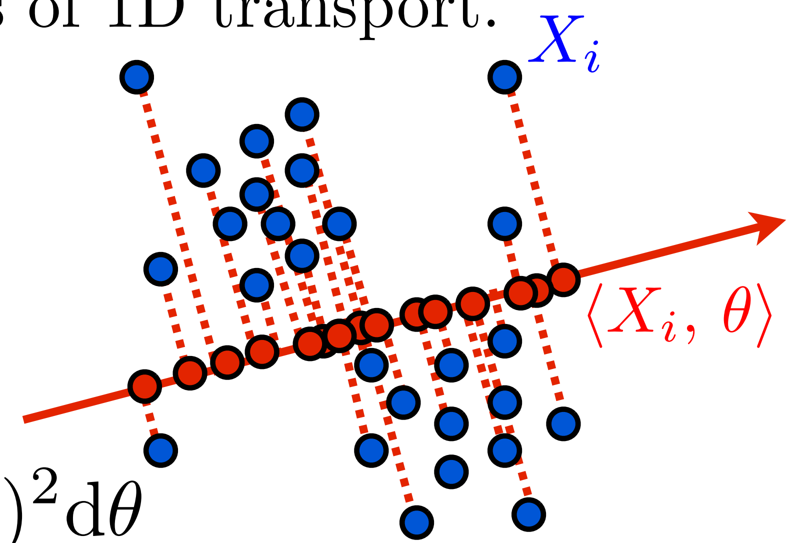
Projected point cloud: $X_\theta = \{\langle X_i, \theta \rangle\}_i$.



Approximate Sliced Distance

Key idea: replace transport in $\mathbb{R}^d$ by series of 1D transport.

[Rabin, Peyré, Delon & Bernot 2010]



Projected point cloud: $X_\theta = \{\langle X_i,\, \theta\rangle\}_i$.

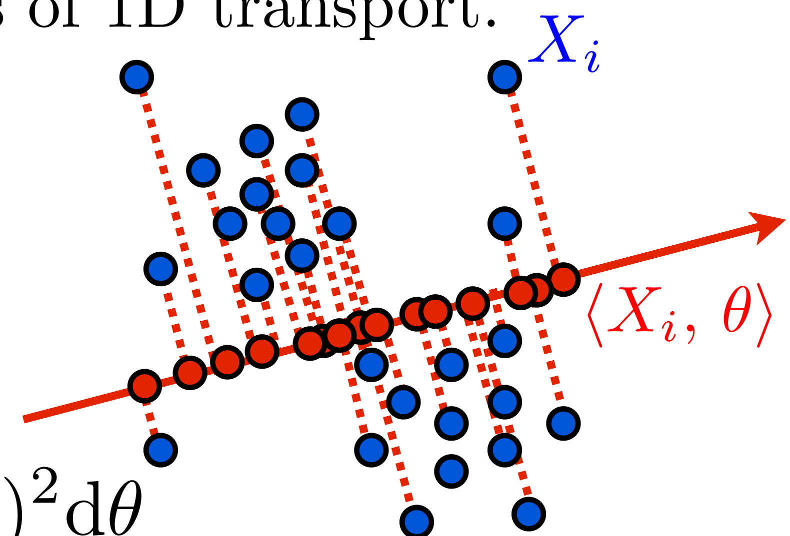
Sliced Wasserstein distance: $(p = 2)$

$$SW(\mu_X, \mu_Y)^2 = \int_{\|\theta\|=1} W(\mu_{X_\theta}, \mu_{Y_\theta})^2 \mathrm{d}\theta$$

Approximate Sliced Distance

Key idea: replace transport in $\mathbb{R}^d$ by series of 1D transport.

[Rabin, Peyré, Delon & Bernot 2010]



Projected point cloud: $X_\theta = \{\langle X_i,\, \theta\rangle\}_i$.

Sliced Wasserstein distance: $(p = 2)$

$$SW(\mu_X, \mu_Y)^2 = \int_{\|\theta\|=1} W(\mu_{X_\theta}, \mu_{Y_\theta})^2 \mathrm{d}\theta$$

Theorem: $E(X) = SW(\mu_X, \mu_Y)^2$ is of class C^1 and

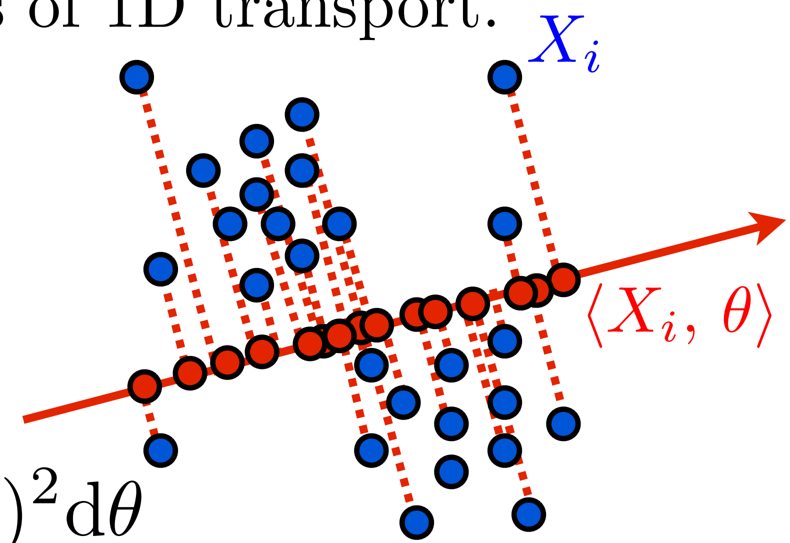
$$\nabla E(X) = \int_\theta \langle X_i - Y_{\sigma_\theta(i)},\, \theta\rangle\, \theta\, d\theta.$$

where $\sigma_\theta \in \Sigma_N$ are 1-D optimal assignents of X_θ and Y_θ.

Approximate Sliced Distance

Key idea: replace transport in $\mathbb{R}^d$ by series of 1D transport.

[Rabin, Peyré, Delon & Bernot 2010]



Projected point cloud: $X_\theta = \{\langle X_i,\, \theta\rangle\}_i$.

Sliced Wasserstein distance: $(p = 2)$

$$SW(\mu_X, \mu_Y)^2 = \int_{\|\theta\|=1} W(\mu_{X_\theta}, \mu_{Y_\theta})^2 \mathrm{d}\theta$$

Theorem: $E(X) = SW(\mu_X, \mu_Y)^2$ is of class C^1 and

$$\nabla E(X) = \int_\theta \langle X_i - Y_{\sigma_\theta(i)},\, \theta\rangle\, \theta\, d\theta.$$

where $\sigma_\theta \in \Sigma_N$ are 1-D optimal assignents of X_θ and Y_θ.

$\longrightarrow$ Possible to use SW in variational imaging problems.

$\longrightarrow$ Fast numerical scheme : use a few random θ.

Sliced Assignment

Theorem: X is a local minima of $E(X) = SW(\mu_X, \mu_Y)^2$

$$\iff \exists \sigma \in \Sigma_N,\ X = Y \circ \sigma$$

Sliced Assignment

Theorem: X is a local minima of $E(X) = SW(\mu_X, \mu_Y)^2$

$$\iff \exists \sigma \in \Sigma_N,\ X = Y \circ \sigma$$

Stochastic gradient descent of $E(X)$:

Step 1: choose Θ at random. $\quad E_\Theta(X) = \sum_{\theta \in \Theta} W(X_\theta, Y_\theta)^2$

Step 2: $X^{(\ell+1)} = X^{(\ell)} - \tau \nabla E_\Theta(X^{(\ell)})$

Sliced Assignment

Theorem: X is a local minima of $E(X) = SW(\mu_X, \mu_Y)^2$

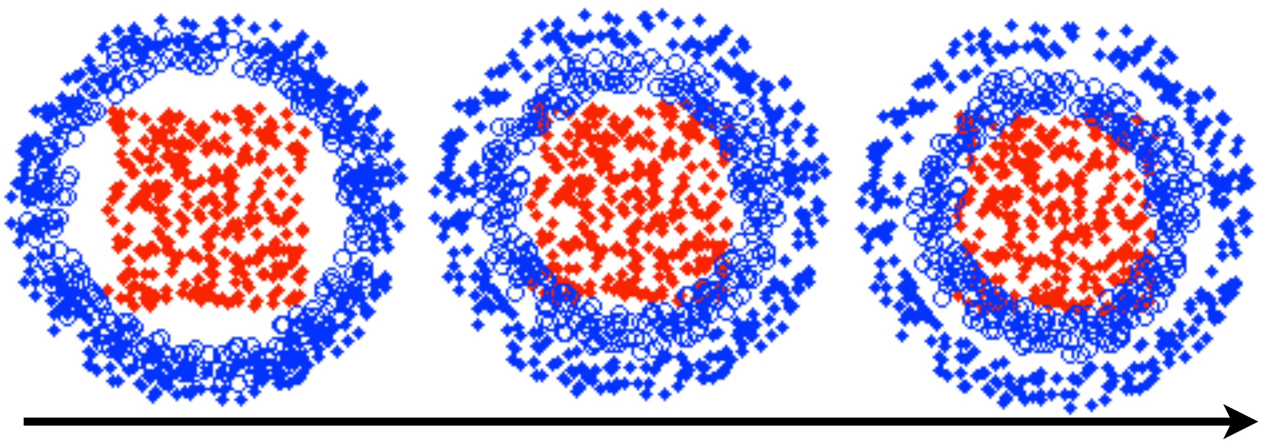
$$\iff \exists \sigma \in \Sigma_N,\ X = Y \circ \sigma$$

Stochastic gradient descent of $E(X)$:

Step 1: choose Θ at random. $\quad E_\Theta(X) = \sum_{\theta \in \Theta} W(X_\theta, Y_\theta)^2$

Step 2: $X^{(\ell+1)} = X^{(\ell)} - \tau \nabla E_\Theta(X^{(\ell)})$

$X^{(\ell)}$ converges to $\mathcal{C} = \{X \setminus \mu_X = \mu_Y\}$.



Sliced Assignment

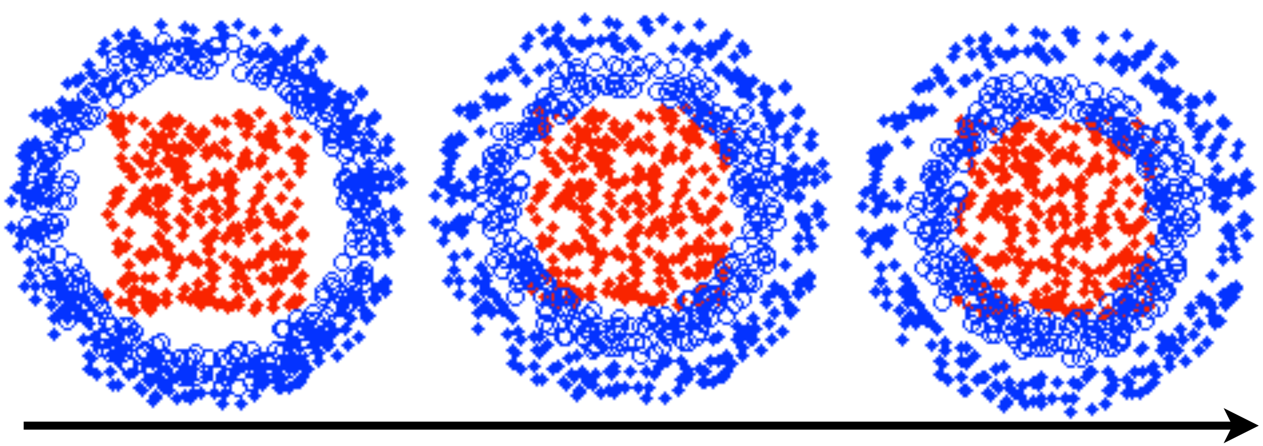
> *Theorem:* X is a local minima of $E(X) = SW(\mu_X, \mu_Y)^2$
> $$\iff \exists \sigma \in \Sigma_N,\ X = Y \circ \sigma$$

Stochastic gradient descent of $E(X)$:

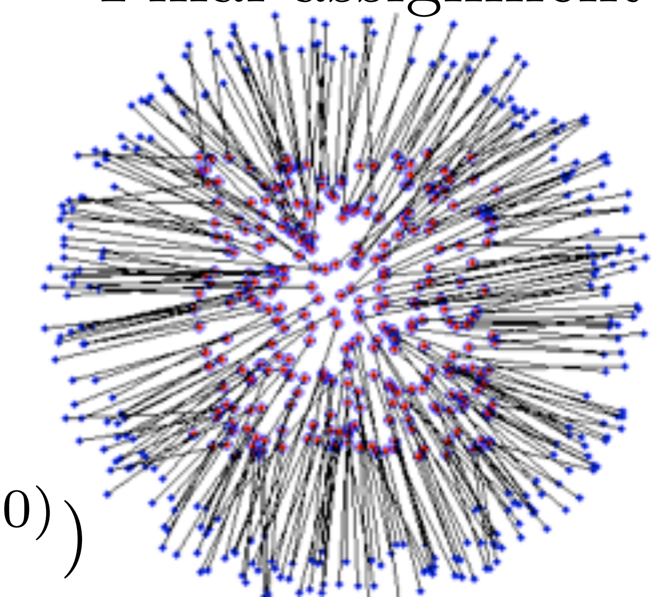
Step 1: choose Θ at random. $\quad E_\Theta(X) = \sum_{\theta \in \Theta} W(X_\theta, Y_\theta)^2$

Step 2: $X^{(\ell+1)} = X^{(\ell)} - \tau \nabla E_\Theta(X^{(\ell)})$

$X^{(\ell)}$ converges to $\mathcal{C} = \{X \setminus \mu_X = \mu_Y\}$.



Final assignment



Numerical observation: $X^{(\infty)} \approx \mathrm{Proj}_\mathcal{C}(X^{(0)})$

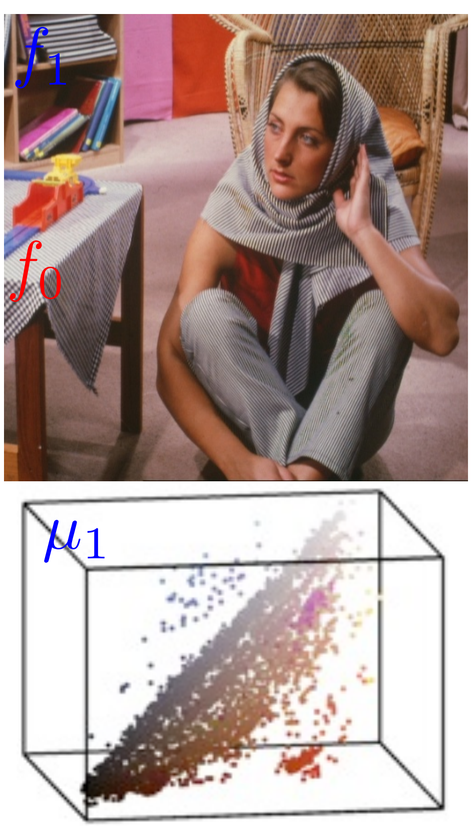
Overview

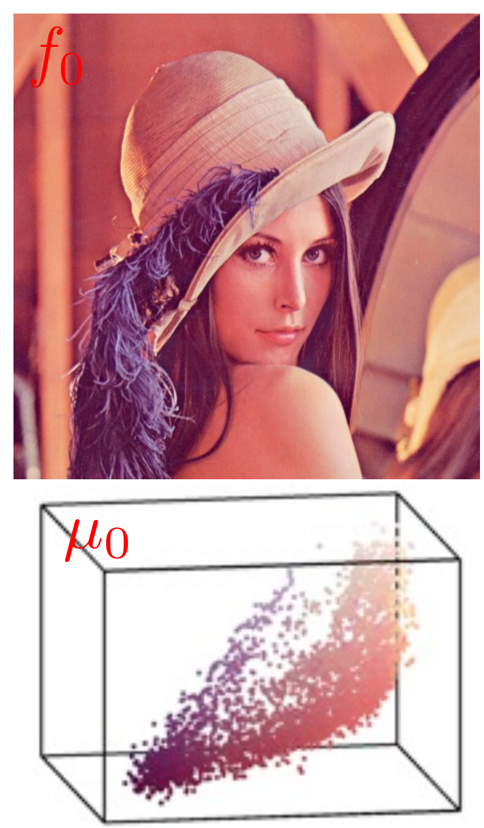
- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- Gaussian Texture Models

Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$.

$$\nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$$

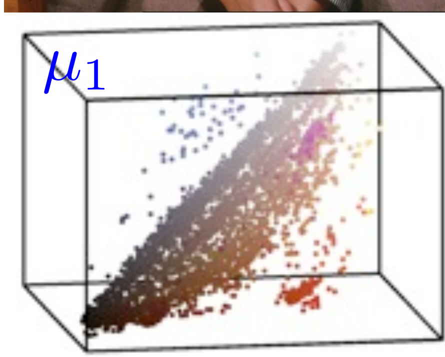


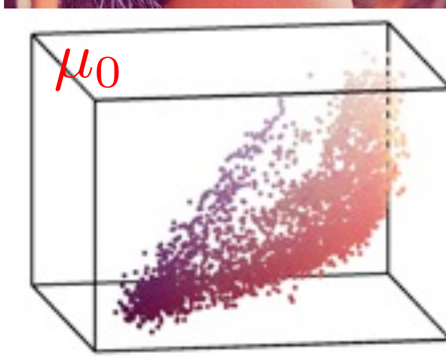


Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$. $\qquad \nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$

Optimal assignement: $\min_{\sigma \in \Sigma_N} \|f_0 - f_1 \circ \sigma\|$



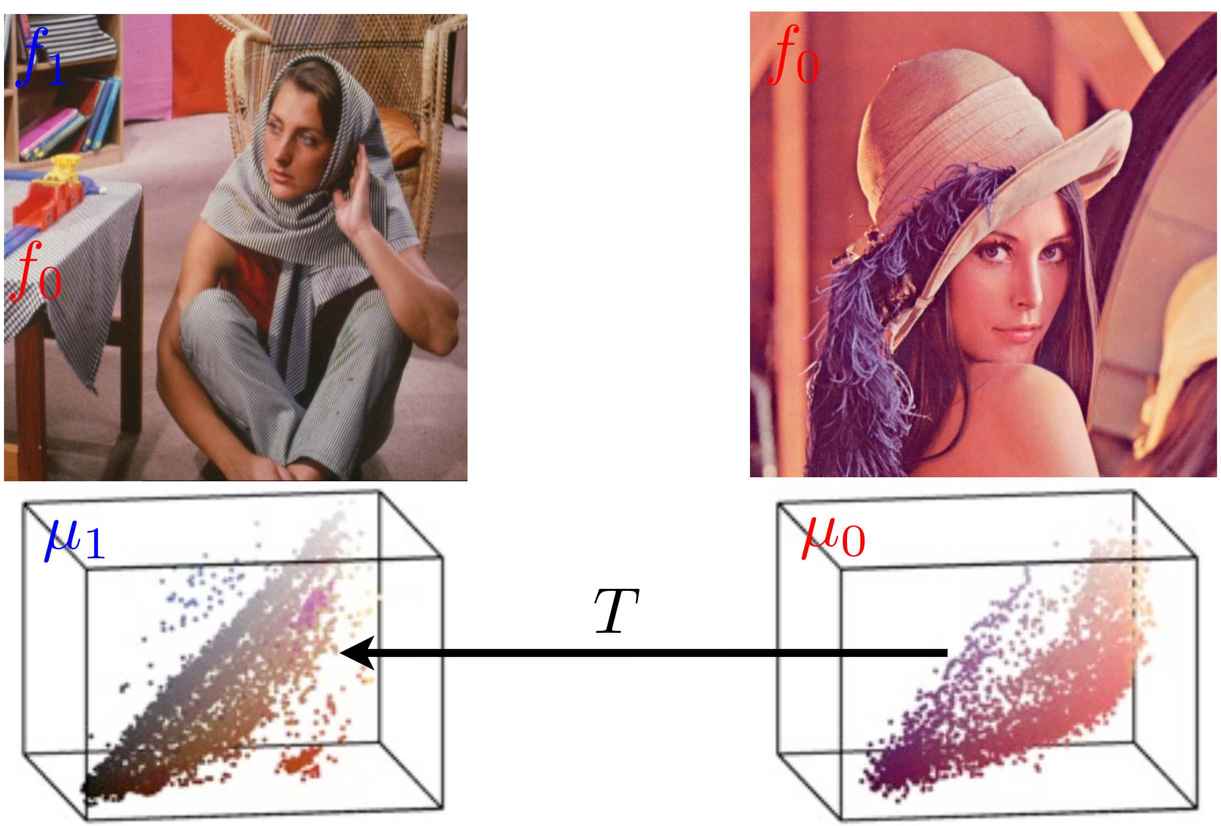


Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N\times 3}$. $\qquad \nu_i = \frac{1}{N}\sum_x \delta_{f_i(x)}$

Optimal assignement: $\min_{\sigma\in\Sigma_N} \|f_0 - f_1 \circ \sigma\|$

Transport: $T : f_0(x) \in \mathbb{R}^3 \mapsto f_1(\sigma(i)) \in \mathbb{R}^3$



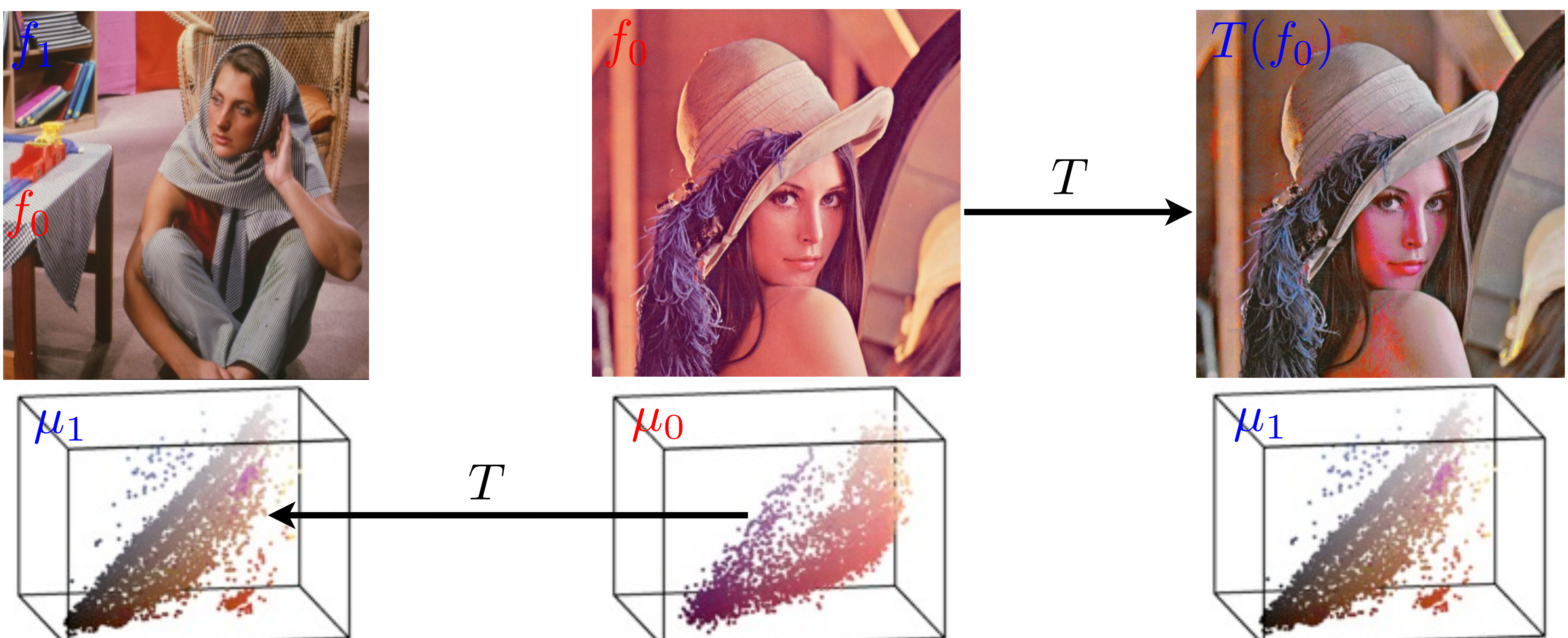
Color Histogram Equalization

Input color images: $f_i \in \mathbb{R}^{N \times 3}$. $\quad \nu_i = \frac{1}{N} \sum_x \delta_{f_i(x)}$

Optimal assignement: $\min_{\sigma \in \Sigma_N} \|f_0 - f_1 \circ \sigma\|$

Transport: $T : f_0(x) \in \mathbb{R}^3 \mapsto f_1(\sigma(i)) \in \mathbb{R}^3$

Equalization: $\tilde{f}_0 = T(f_0) \iff \tilde{f}_0 = f_1 \circ \sigma$



Sliced Wasserstein Transfert

Solving $\min_{\sigma \in \Sigma_N} \|f_0 - f_1 \circ \sigma\|$ is computationally untractable.

Approximate Wasserstein projection:

$$\tilde{f}_0 \text{ solves } \min_f E(f) = SW(\mu_f, \mu_{f_0})$$

and is close to f_0

Sliced Wasserstein Transfert

Solving $\min_{\sigma \in \Sigma_N} \|f_0 - f_1 \circ \sigma\|$ is computationally untractable.

Approximate Wasserstein projection:

$\tilde{f}_0$ solves $\min_f E(f) = SW(\mu_f, \mu_{f_0})$

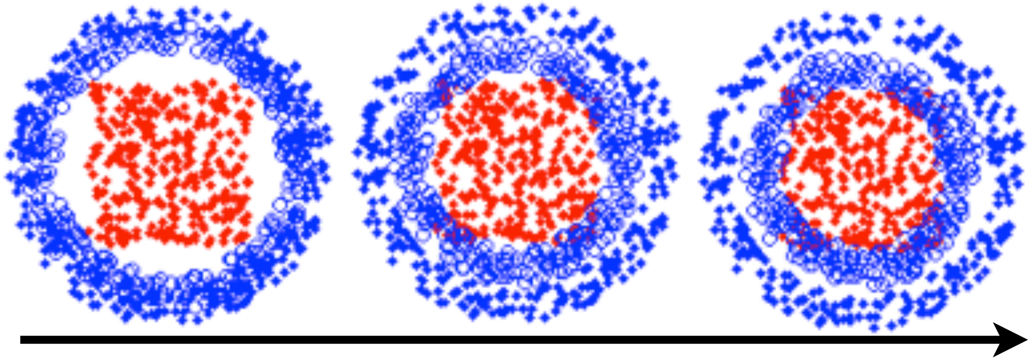
and is close to f_0

(Stochastic) gradient descent:

$$f^{(0)} = f_0$$

$$f^{(\ell+1)} = f^{(\ell)} - \tau_\ell \nabla E(f^{(\ell)})$$

$$f^{(\ell)} \to \tilde{f}_1$$



At convergence: $\mu_{\tilde{f}_0} = \mu_{f_1}$

Color Transfert

Input image f_0

Target image f_1

Transfered image $\tilde{f}_0$

Color Exchange

Input image f_0

Target image f_1

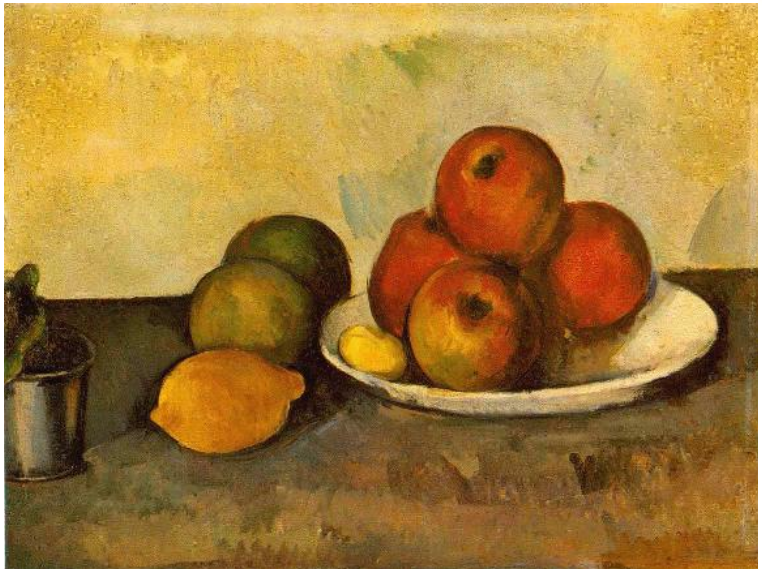
Transfered image $\tilde{f}_0$

Transfered image $\tilde{f}_1$

Overview

- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- Gaussian Texture Models

Color Transfer Artifacts



Input image f_0

Transfert: $\tilde{f}_0 = T(f_0)$.

$T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ not regular.

$\longrightarrow T$ amplifies noise.

Target image f_1

Transfered image $\tilde{f}_0$

Wasserstein Fidelity in Imaging

Variational regularization: $E(f) = SW_2(\mu_f, \nu)^2$

target distribution

$$\min_f \boxed{\frac{1}{2}\|f - f_0\|^2} + \boxed{\lambda R(f)} + \boxed{\mu E(f)} \qquad (\star)$$

Data fidelity $F(f)$ | Regularization | Histogram fidelity

Total variation regularization: $R(f) = \sum_x \|\nabla f(x)\|$

Wasserstein Fidelity in Imaging

Variational regularization: $E(f) = SW_2(\mu_f, \nu)^2$

target distribution

$$\min_f \boxed{\frac{1}{2}\|f - f_0\|^2} + \boxed{\lambda R(f)} + \boxed{\mu E(f)} \qquad (\star)$$

Data fidelity $F(f)$ | Regularization | Histogram fidelity

Smooth | Non-smooth | Non-convex

Total variation regularization: $R(f) = \sum_x \|\nabla f(x)\|$

Forward-backward proximal algorithm:

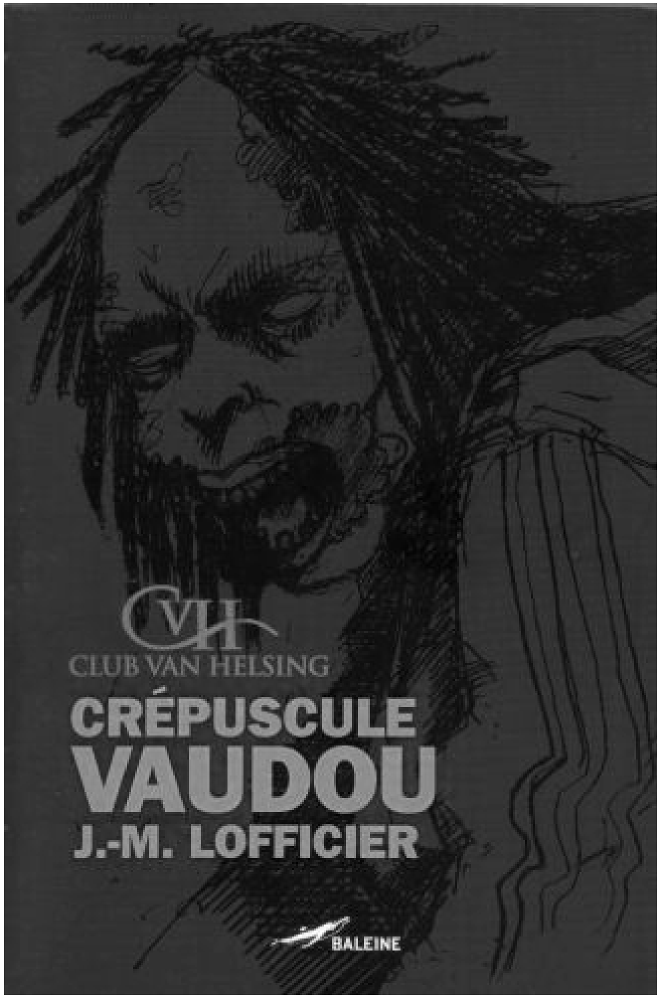
$$f^{(\ell+1)} = \mathrm{Prox}_{\tau\lambda R}\left(f^{(\ell)} - \tau(\nabla F(f) + \mu \nabla E(f^{(\ell)}))\right)$$

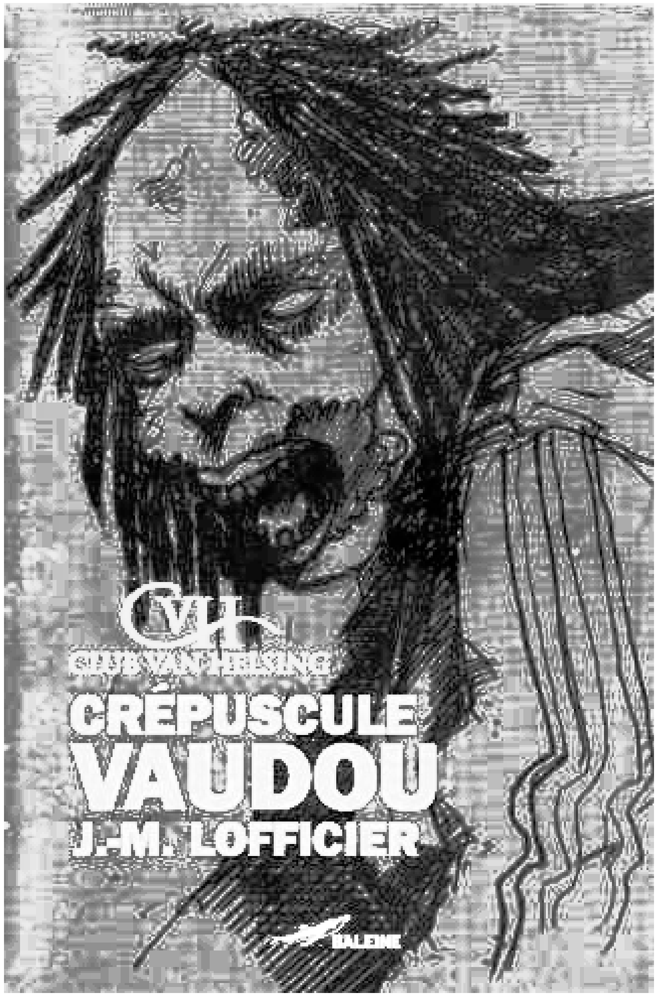
where $\mathrm{Prox}_{\eta R}(f) = \underset{g}{\mathrm{argmin}}\ \frac{1}{2}\|f - g\|^2 + \eta R(g)$

$\longrightarrow$ converges to a local minimum of $(\star)$

Regularized Equalization

Here μ_{f_1} is uniform.





No regularization

With regularization

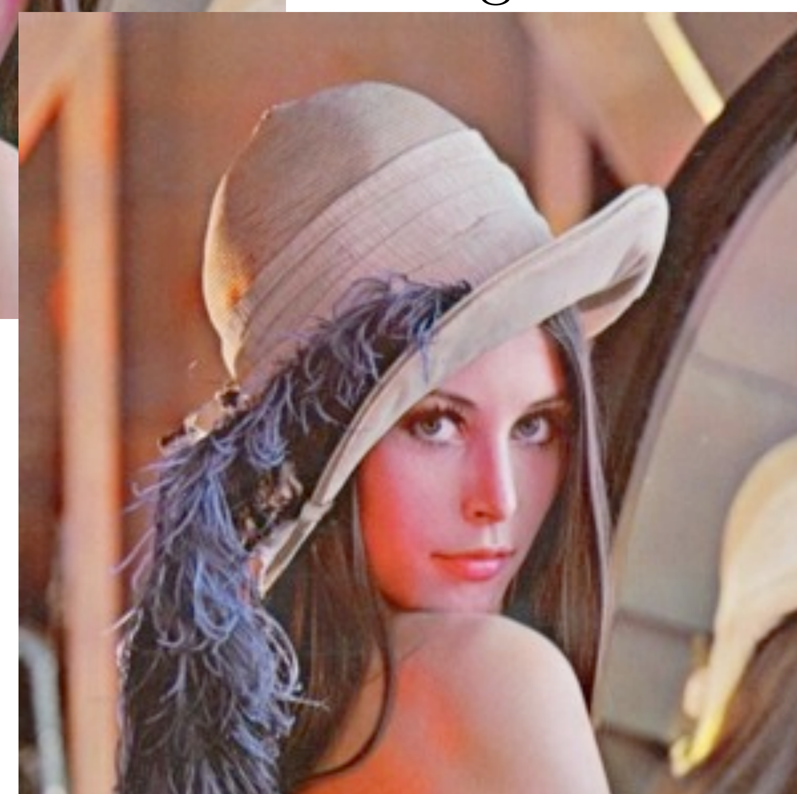
Regularized Color Transfer

Here $\nu = \mu_{f_1}$

Original images

No regularization

With regularization



Regularized Color Transfer

Here $\nu = \mu_{f_1}$

Original images

Overview

- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- Gaussian Texture Models

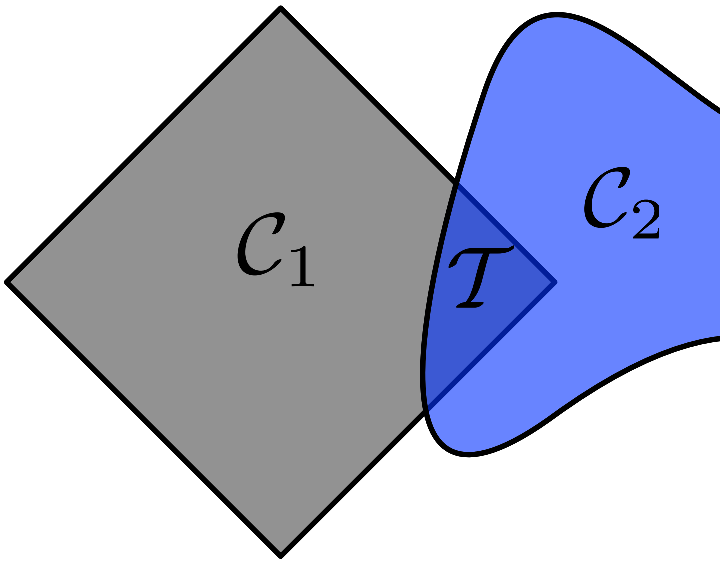
Textures Ensembles

Sets of constraints $\{\mathcal{C}_i\}_i$. Texture ensemble: $\mathcal{T} = \bigcap_i \mathcal{C}_i$

– Spacial constraints.

Fixed energy: $\mathcal{C}_i = \{f \setminus \|f\| = 1\}$.

Fixed histograms: $\mathcal{C}_i = \{f \setminus \mu_f = \nu\}$



Textures Ensembles

Sets of constraints $\{\mathcal{C}_i\}_i$. Texture ensemble: $\mathcal{T} = \bigcap_i \mathcal{C}_i$

– Spacial constraints.

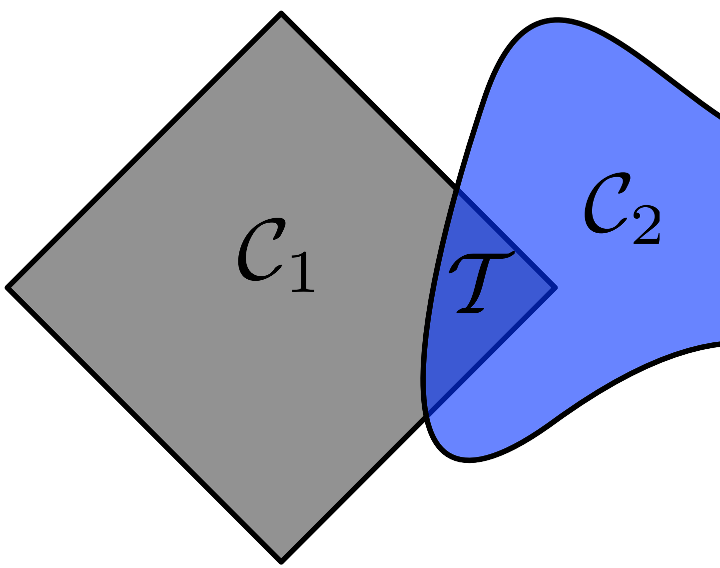
Fixed energy: $\mathcal{C}_i = \{f \setminus \|f\| = 1\}$.

Fixed histograms: $\mathcal{C}_i = \{f \setminus \mu_f = \nu\}$

– Smoothness constraints.

Sobolev: $\mathcal{C}_i = \left\{f \setminus \int \|\nabla f\|^2 \leqslant \tau\right\}$

TV: $\mathcal{C}_i = \left\{f \setminus \int \|\nabla f\| \leqslant \tau\right\}$



Textures Ensembles

Sets of constraints $\{\mathcal{C}_i\}_i$. Texture ensemble: $\mathcal{T} = \bigcap_i \mathcal{C}_i$

– Spacial constraints.

- Fixed energy: $\mathcal{C}_i = \{f \setminus \|f\| = 1\}$.
- Fixed histograms: $\mathcal{C}_i = \{f \setminus \mu_f = \nu\}$

– Smoothness constraints.

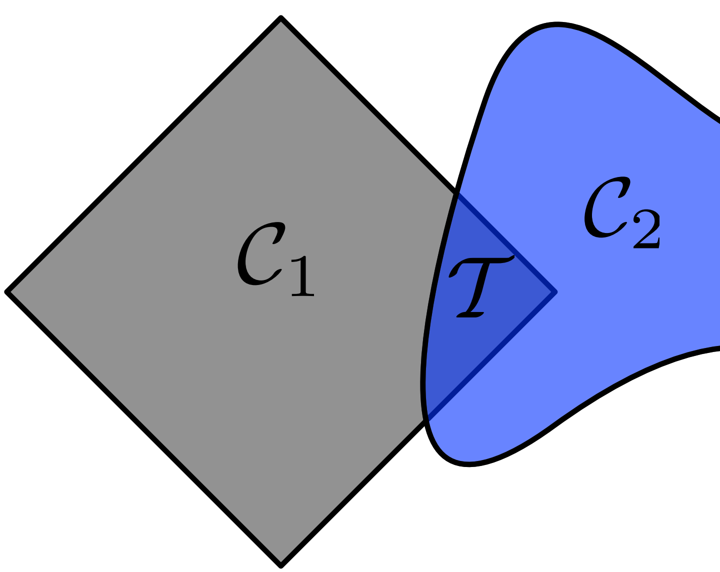
- Sobolev: $\mathcal{C}_i = \{f \setminus \int \|\nabla f\|^2 \leqslant \tau\}$
- TV: $\mathcal{C}_i = \{f \setminus \int \|\nabla f\| \leqslant \tau\}$

– Transformed domain constraints on $\Psi f = \{\langle f, \psi_m \rangle\}_m$

$$\mathcal{C}_i = \left\{ f \setminus \forall \omega,\ |\hat{f}(\omega)| = c_\omega \right\}$$ (see the work of Bruno Galerne)

$$\mathcal{C}_i = \{f \setminus \sum |\langle f, \psi_m \rangle| \leqslant \tau\}$$ (Besov)

$$\mathcal{C}_i = \{f \setminus \mu_{\Psi f} = \nu\}$$



Textures Ensembles

Sets of constraints $\{\mathcal{C}_i\}_i$. Texture ensemble: $\mathcal{T} = \bigcap_i \mathcal{C}_i$

– Spacial constraints.

Fixed energy: $\mathcal{C}_i = \{f \setminus \|f\| = 1\}$.

Fixed histograms: $\mathcal{C}_i = \{f \setminus \mu_f = \nu\}$

– Smoothness constraints.

Sobolev: $\mathcal{C}_i = \left\{f \setminus \int \|\nabla f\|^2 \leqslant \tau\right\}$

TV: $\mathcal{C}_i = \left\{f \setminus \int \|\nabla f\| \leqslant \tau\right\}$

– Transformed domain constraints on $\Psi f = \{\langle f, \psi_m \rangle\}_m$

$$\mathcal{C}_i = \left\{f \setminus \forall \omega,\ |\hat{f}(\omega)| = c_\omega\right\} \quad \text{(see the work of Bruno Galerne)}$$

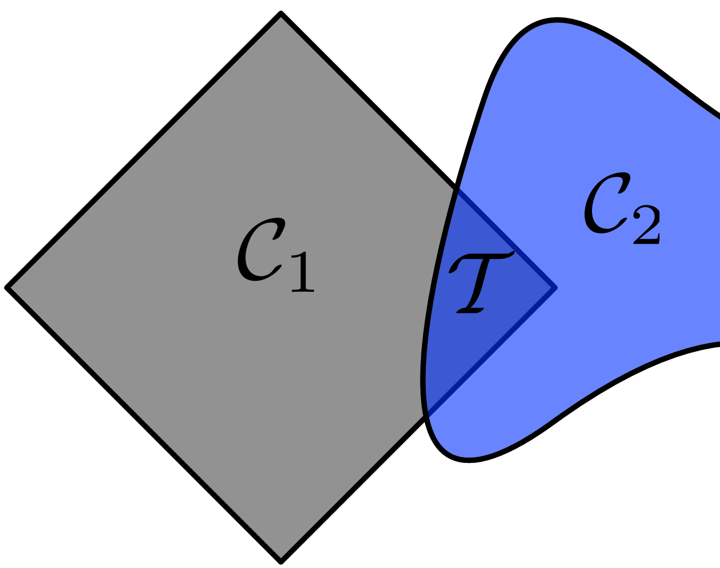
$$\mathcal{C}_i = \{f \setminus \textstyle\sum |\langle f, \psi_m \rangle| \leqslant \tau\} \quad \text{(Besov)}$$

$$\mathcal{C}_i = \{f \setminus \mu_{\Psi f} = \nu\}$$

– Higher order statistical constraints on Ψf

$$\mathcal{C}_i = \{f \setminus \mathrm{Curtosis}(\Psi f) \leqslant \tau\}.$$

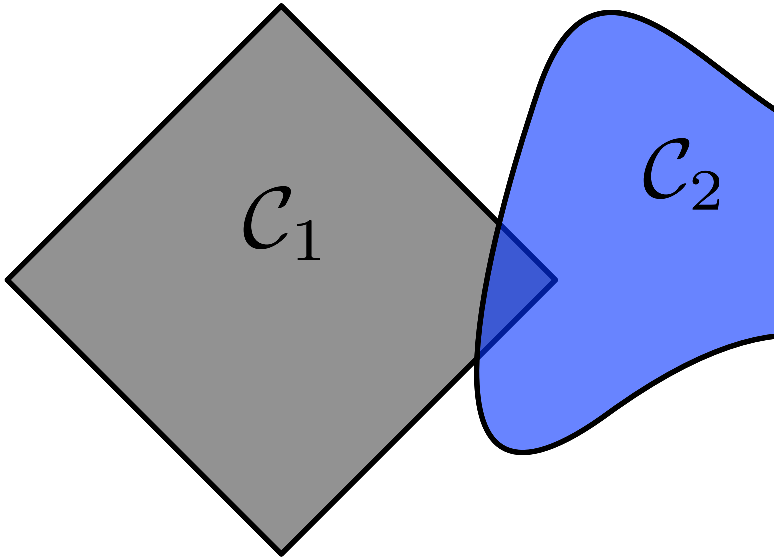
Correlations, pairwise histograms, etc.



Synthesis Using Iterative Projections

Distribution in $\mathcal{T}$ with maximal entropy: uniform distribution.

Synthesis: draw $f \in \mathcal{T}$ uniformly at random. [Zhu,Mumford]
$\rightarrow$ computationaly untractable, needs Gibbs sampler.

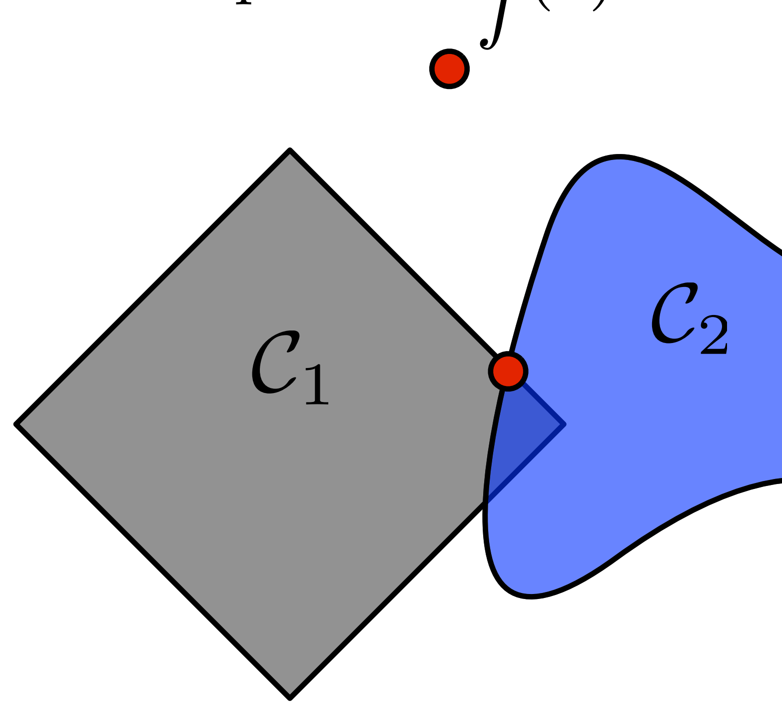


Synthesis Using Iterative Projections

Distribution in $\mathcal{T}$ with maximal entropy: uniform distribution.

Synthesis: draw $f \in \mathcal{T}$ uniformly at random. [Zhu,Mumford]

$\rightarrow$ computationaly untractable, needs Gibbs sampler.



Approximation #1: project noise on $\mathcal{T}$.

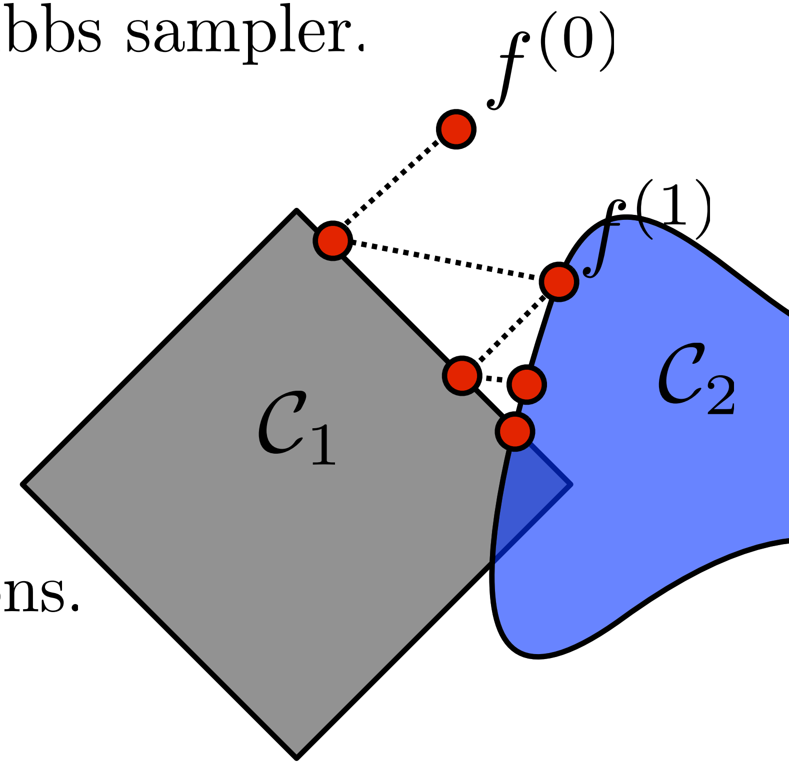
[Portilla, Simoncelli]

Synthesis Using Iterative Projections

Distribution in $\mathcal{T}$ with maximal entropy: uniform distribution.

Synthesis: draw $f \in \mathcal{T}$ uniformly at random. [Zhu,Mumford]

$\rightarrow$ computationaly untractable, needs Gibbs sampler.



Approximation #1: project noise on $\mathcal{T}$.

[Portilla, Simoncelli]

Approximation #2: use iterative projections.

$$f^{(k+1)} = \mathrm{Proj}_{\mathcal{C}_{i_k}}(f^{(k)})$$

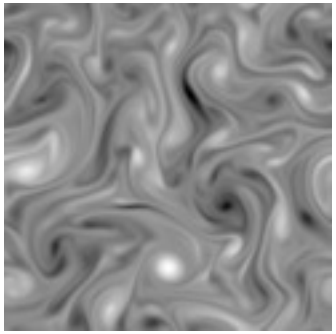
Local convergence under conditions on $\mathcal{C}_i$.

[Lewis, Malick, Luke]

Multiscale Statistical Constraints

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$. (gray $d = 1$, color $d = 3$)

Exemplar f_0

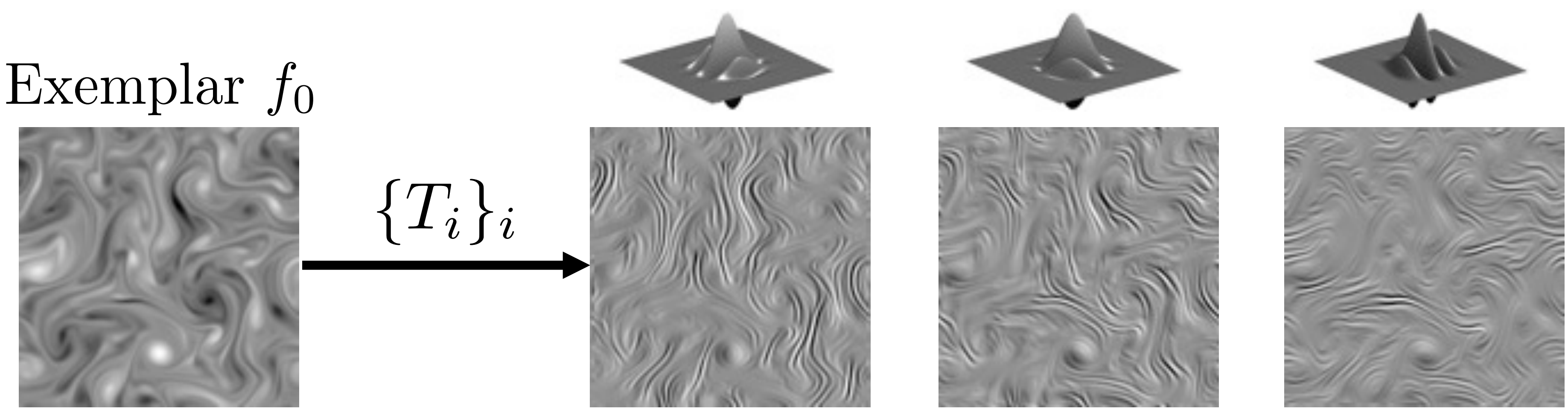


Multiscale Statistical Constraints

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$. (gray $d = 1$, color $d = 3$)

Oriented multiscale transform: $T_0(f) = f$

$$\forall\, i = 1, \ldots, I, \quad T_i(f) = f \star \psi_i$$



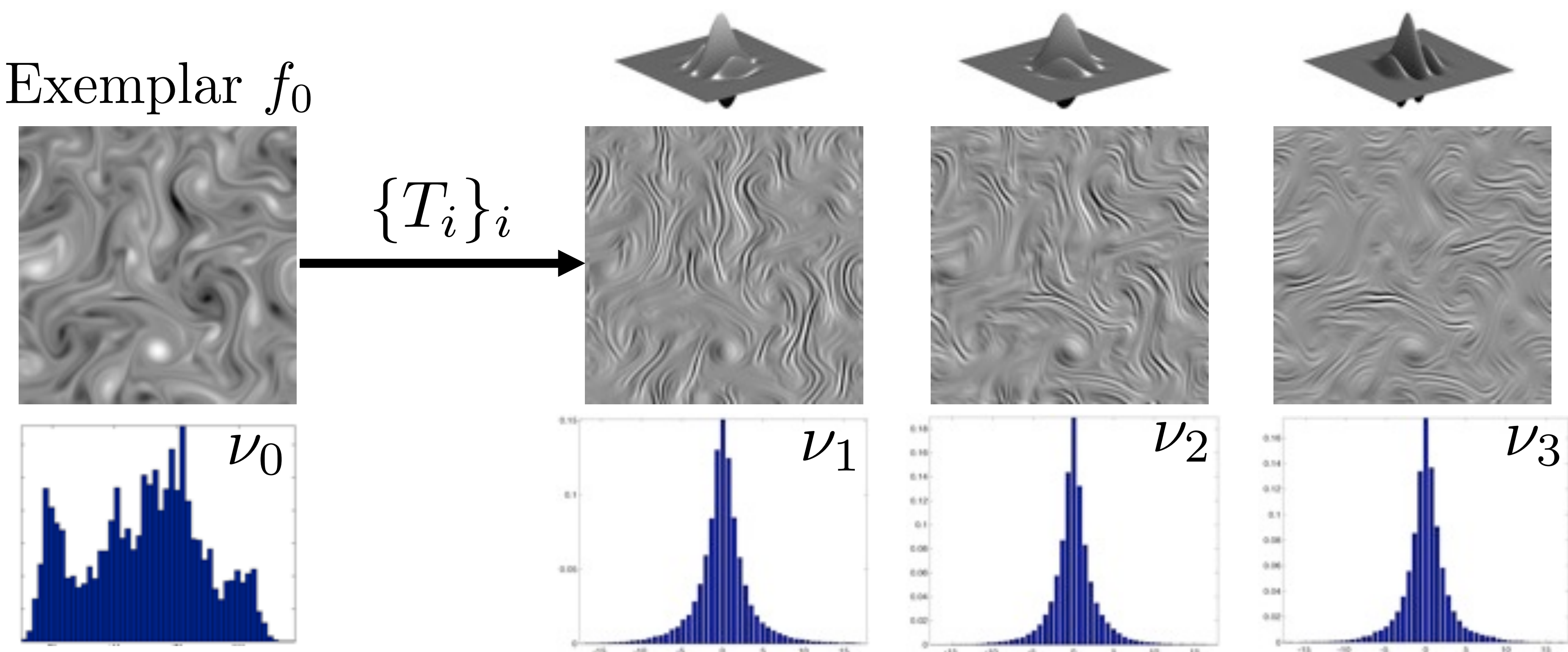
Multiscale Statistical Constraints

Input exemplar: $f_0 \in \mathbb{R}^{N\times d}$. (gray $d=1$, color $d=3$)

Oriented multiscale transform: $T_0(f) = f$

$$\forall\, i = 1, \ldots, I, \quad T_i(f) = f \star \psi_i$$

Learning the model: $\nu_i = \mu_{T_i(f_0)}$



Multiscale Statistical Constraints

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$. (gray $d = 1$, color $d = 3$)

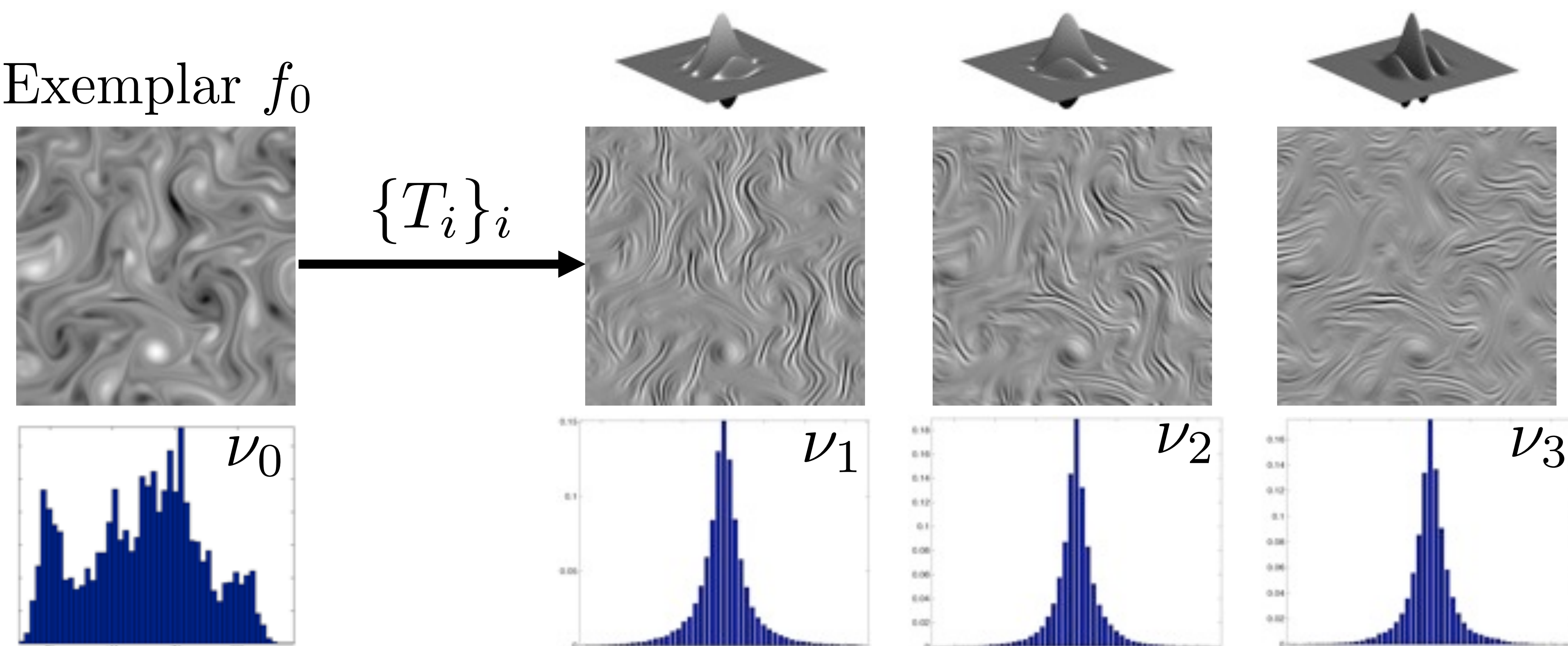
Oriented multiscale transform: $T_0(f) = f$

$$\forall\, i = 1, \ldots, I, \quad T_i(f) = f \star \psi_i$$

Learning the model: $\nu_i = \mu_{T_i(f_0)}$

Statistical constraints: $\mathcal{C}_i = \{ f \setminus \mu_{T_i(f)} = \nu_i \}$

Exemplar f_0



Synthesis Method

Synthesized texture: stationary point of

$$\mathcal{E}(f) = \sum_{i=0}^{K} W_2(\mu_{T_i(f)}, \nu_i)^2$$

Synthesis Method

Synthesized texture: stationary point of

$$\mathcal{E}(f) = \sum_{i=0}^{K} W_2(\mu_{T_i(f)}, \nu_i)^2$$

Randomized sampling of stationary point:

$$f^{(0)} \sim \mathcal{N}(0, \mathrm{Id}) \qquad f^{(\ell+1)} = f^{(\ell)} - \tau_\ell \nabla\mathcal{E}(f^{(\ell)})$$

$$\nabla\mathcal{E}(f) = \sum_{i=0}^{K} T_i^* \circ \nabla E_i \circ T_i(f)$$

where $\quad E_i(u) = W_2(\mu_u, \nu_i)^2$

Synthesis Method

Synthesized texture: stationary point of

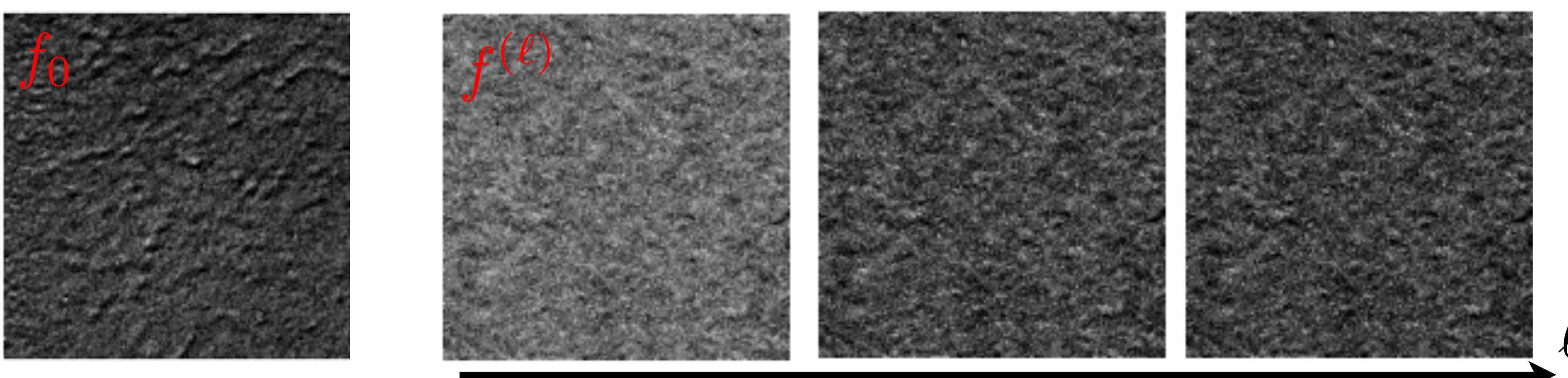
$$\mathcal{E}(f) = \sum_{i=0}^{K} W_2(\mu_{T_i(f)}, \nu_i)^2$$

Randomized sampling of stationary point:

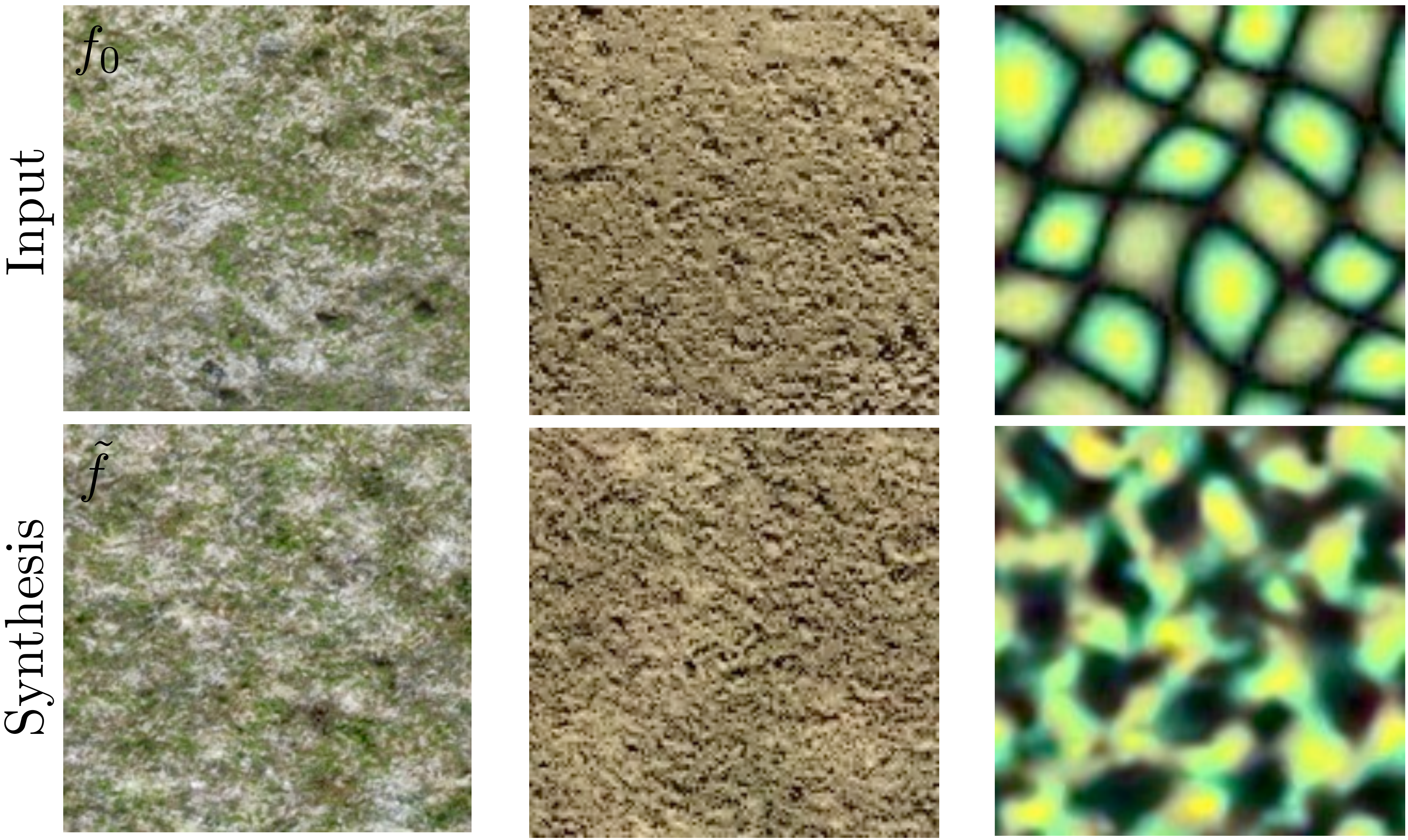
$$f^{(0)} \sim \mathcal{N}(0, \mathrm{Id}) \qquad\qquad f^{(\ell+1)} = f^{(\ell)} - \tau_\ell \nabla \mathcal{E}(f^{(\ell)})$$

$$\nabla \mathcal{E}(f) = \sum_{i=0}^{K} T_i^* \circ \nabla E_i \circ T_i(f)$$

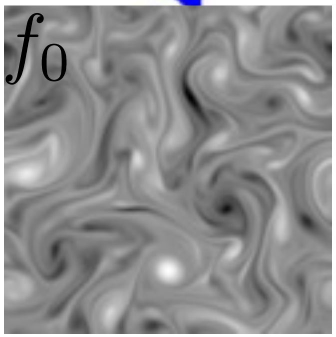
where $\quad E_i(u) = W_2(\mu_u, \nu_i)^2$



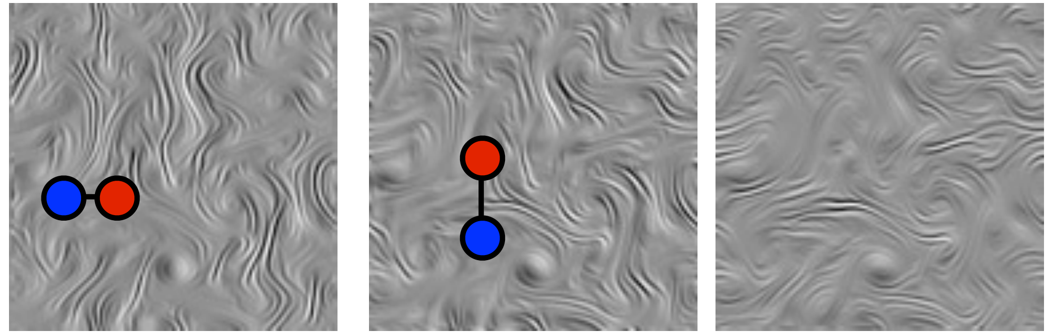
Color Texture Synthesis



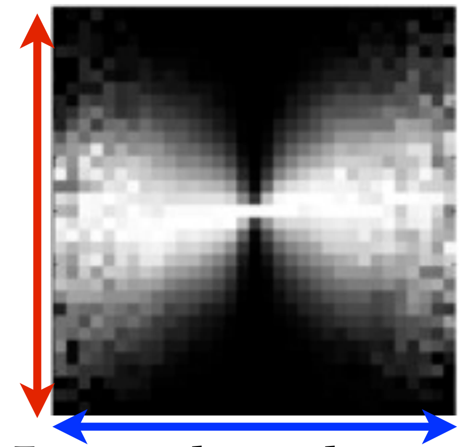
Pairwise Statistics



$\xrightarrow{\{T_i\}_i}$

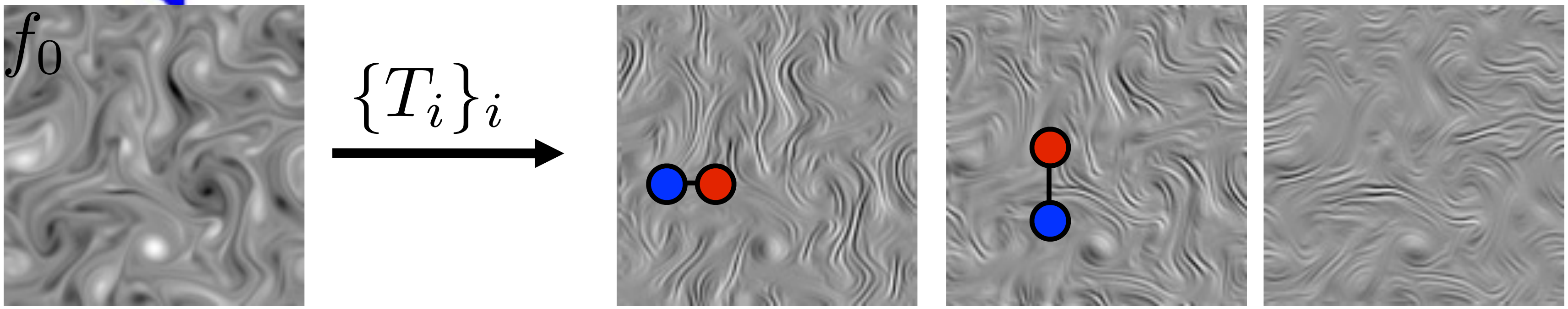


Local neighborood $\mathcal{N}$:



Joint distribution

Pairwise Statistics



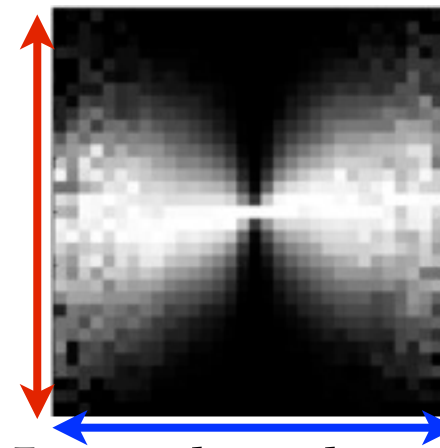
Local neighborood $\mathcal{N}$: 

Higher-dimensional transforms:

$$\tilde{T}_i : \mathbb{R}^{N\times d} \to \mathbb{R}^{N\times d|\mathcal{N}|}$$

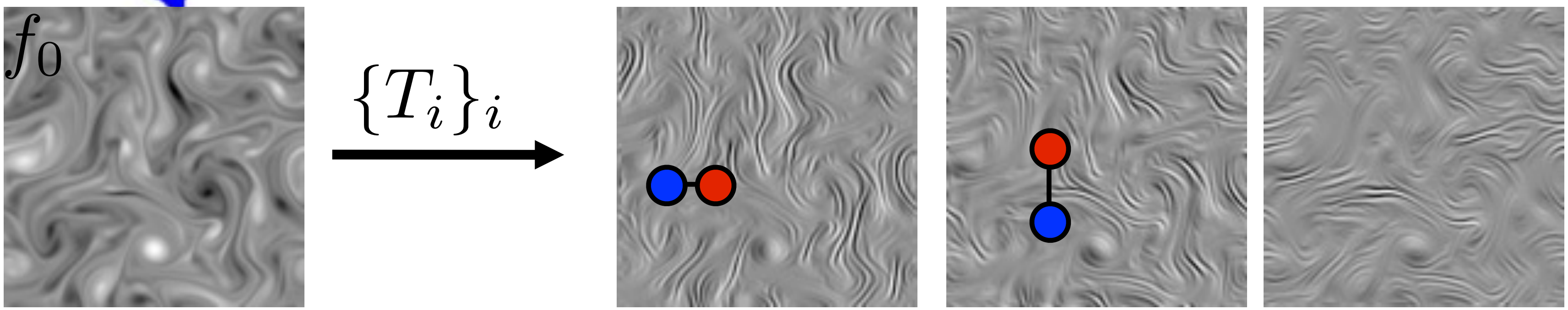
where $\quad y = T_i(f) \qquad \tilde{y} = \tilde{T}_i(f)$

$$\tilde{y}(x) = (y(x+k))_{k\in\mathcal{N}}$$



Joint distribution

Pairwise Statistics



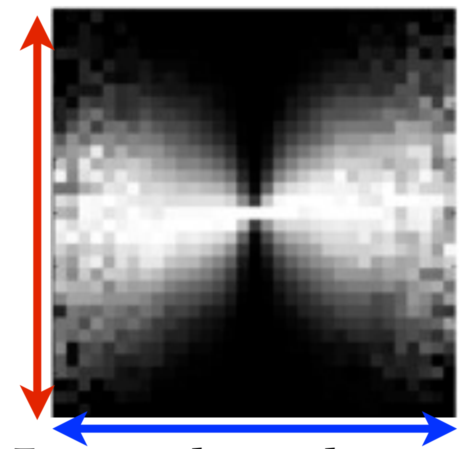
Local neighborood $\mathcal{N}$: 

Higher-dimensional transforms:

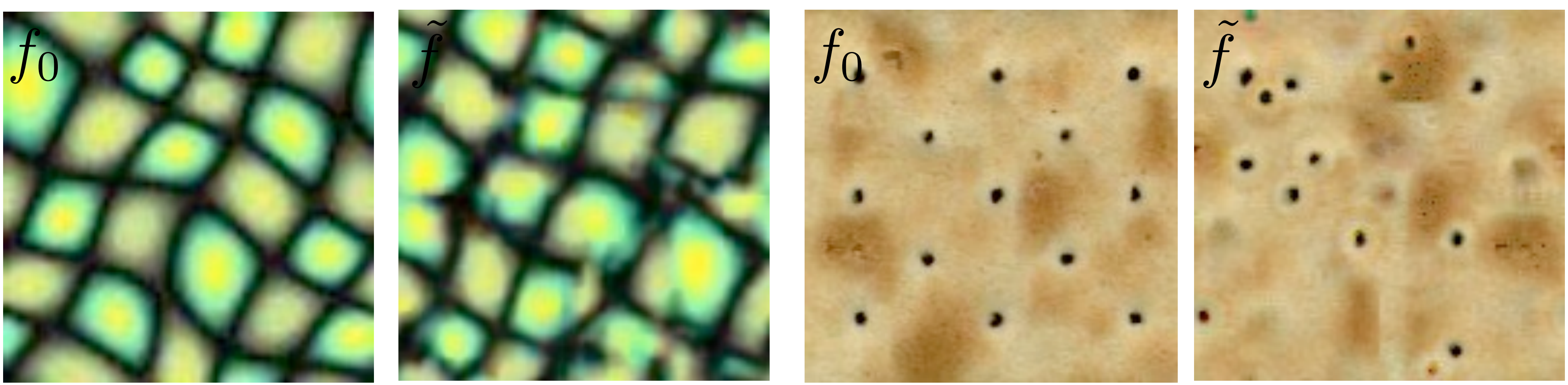
$$\tilde{T}_i : \mathbb{R}^{N\times d} \rightarrow \mathbb{R}^{N\times d|\mathcal{N}|}$$

where $\quad y = T_i(f) \qquad \tilde{y} = \tilde{T}_i(f)$

$$\tilde{y}(x) = (y(x+k))_{k\in\mathcal{N}}$$



Joint distribution



Overview

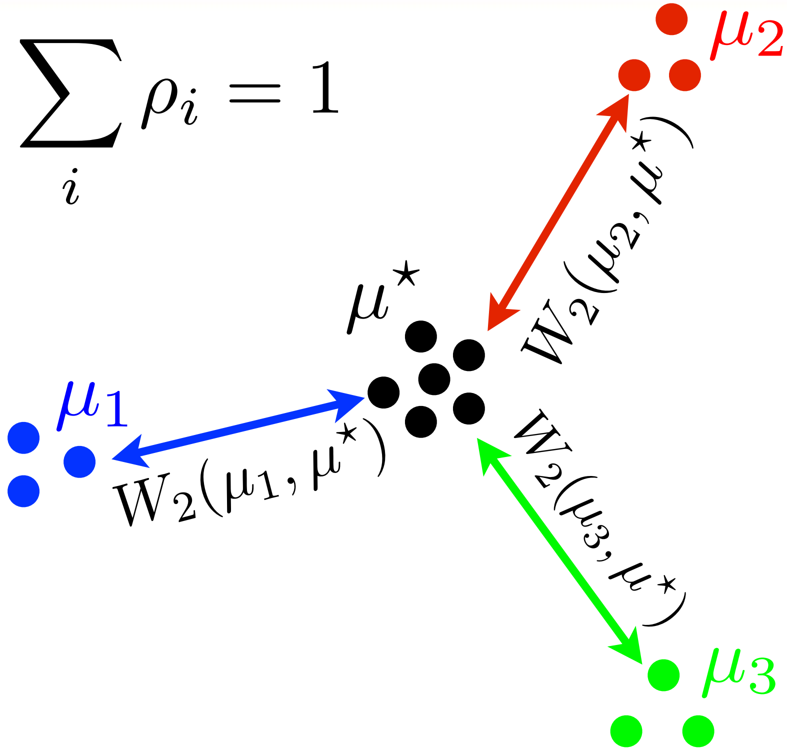
- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- **Wasserstein Barycenter for Texture Mixing**
- Gaussian Texture Models

Wasserstein Barycenter

Barycenter of $\{(\mu_i, \rho_i)\}_{i=1}^{L}$:

$$\sum_i \rho_i = 1$$

$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{L} \rho_i W_2(\mu_i, \mu)^2$$

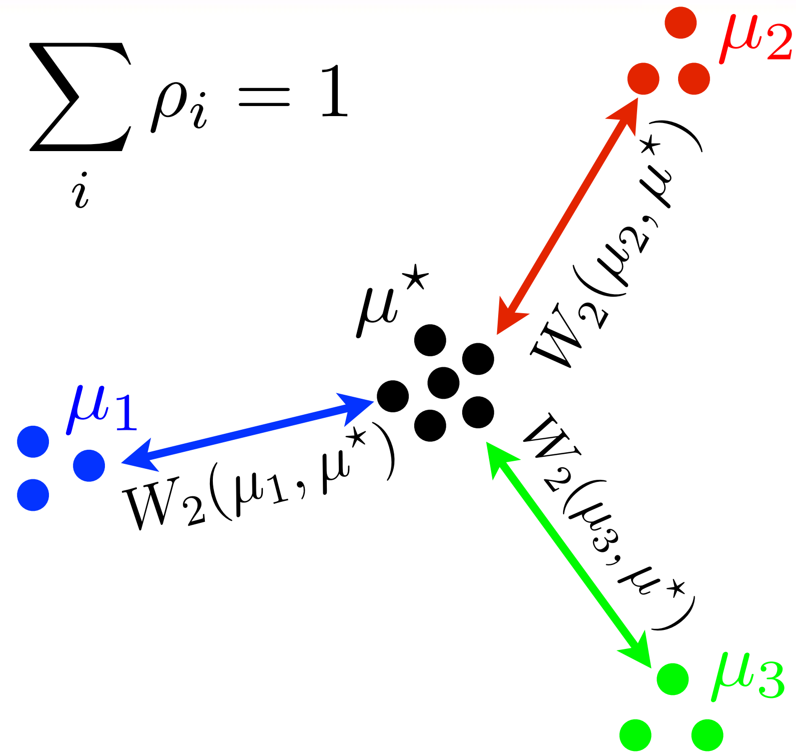


Wasserstein Barycenter

Barycenter of $\{(\mu_i, \rho_i)\}_{i=1}^{L}$:

$$\sum_i \rho_i = 1$$

$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{L} \rho_i W_2(\mu_i, \mu)^2$$



If $\mu_i = \delta_{X_i}$, then $\mu^\star = \delta_{X^\star}$

$$X^\star = \sum_i \rho_i X_i$$

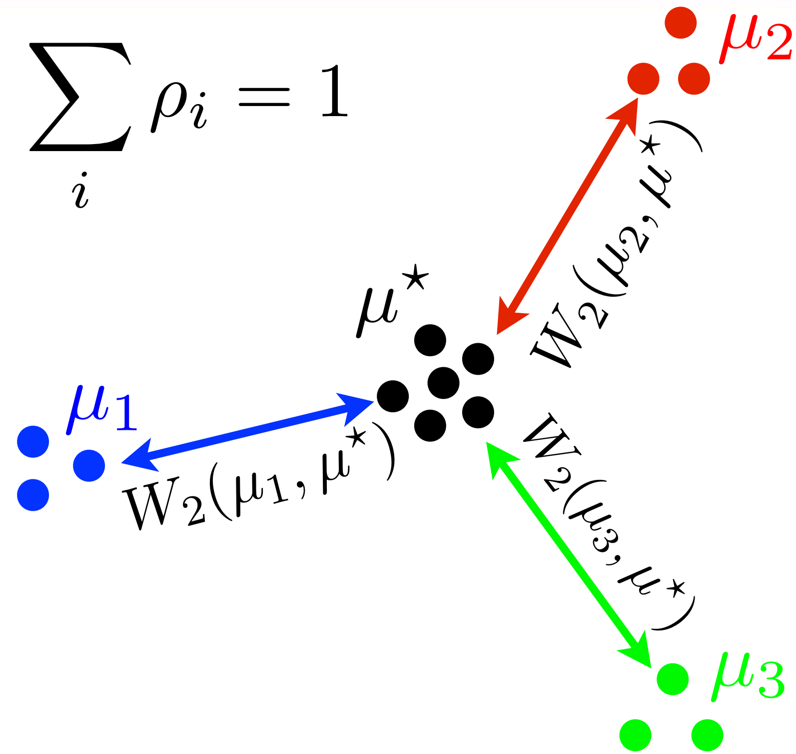
$\longrightarrow$ Generalizes Euclidean barycenter.

Wasserstein Barycenter

Barycenter of $\{(\mu_i, \rho_i)\}_{i=1}^{L}$:

$$\sum_i \rho_i = 1$$

$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{L} \rho_i W_2(\mu_i, \mu)^2$$



If $\mu_i = \delta_{X_i}$, then $\mu^\star = \delta_{X^\star}$

$$X^\star = \sum_i \rho_i X_i$$

$\longrightarrow$ Generalizes Euclidean barycenter.

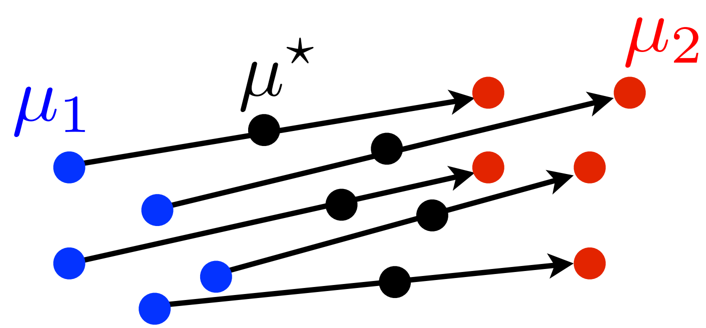
> *Theorem:* [Agueh, Carlier, 2010]
> if μ_0 does not vanish on small sets,
> $\mu^\star$ exists and is unique.

Special Case: 2 Distributions

Case $L = 2$*:*

$$\mu_t \in \underset{\mu}{\operatorname{argmin}}\ (1-t)W_2(\mu_0,\mu)^2 + tW_2(\mu_1,\mu)^2$$

$t \mapsto \mu_t$ is the geodesic path.

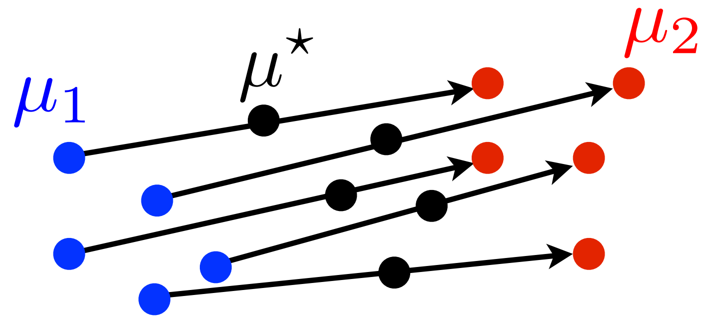


Special Case: 2 Distributions

Case $L = 2$:

$$\mu_t \in \underset{\mu}{\operatorname{argmin}}\ (1-t)W_2(\mu_0,\mu)^2 + tW_2(\mu_1,\mu)^2$$

$t \mapsto \mu_t$ is the geodesic path.



Discrete point clouds: $\mu = \sum_{k=1}^N \delta_{X_i(k)}$,

Assignment: $$\sigma^\star \in \underset{\sigma\in\Sigma_N}{\operatorname{argmin}} \sum_{k=1}^N \|X_1(k) - X_2(\sigma(k))\|^2$$

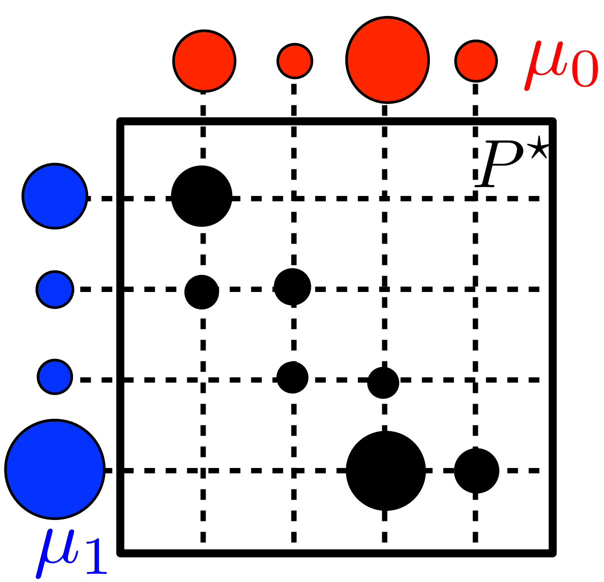
$\mu^\star = \sum_{k=1}^N \delta_{X^\star(k)}$,

where $X^\star = (1-t)X_1(k) + tX_2(\sigma^\star(k))$

Weighted Discrete Case

Linear program: $\mu_i = \sum_{k=1}^{N_i} p_i(k)\delta_{X_i(k)}$

$$P^\star \in \underset{P \in \Pi_{\mu_0,\mu_1}}{\operatorname{argmin}} \sum_{k,\ell} P_{k,\ell} \|X_0(k) - Y_1(\ell)\|^2$$

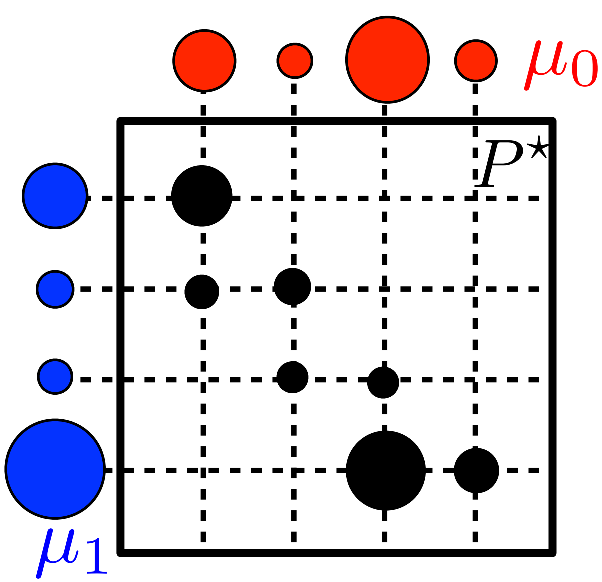


Weighted Discrete Case

Linear program: $\mu_i = \sum_{k=1}^{N_i} p_i(k)\delta_{X_i(k)}$

$$P^\star \in \underset{P \in \Pi_{\mu_0,\mu_1}}{\operatorname{argmin}} \sum_{k,\ell} P_{k,\ell} \|X_0(k) - Y_1(\ell)\|^2$$

Theorem: [Folklore] $|\{(k,\ell) \setminus P^\star_{k,\ell}\} \neq 0| \leqslant N_0 + N_1 - 1.$



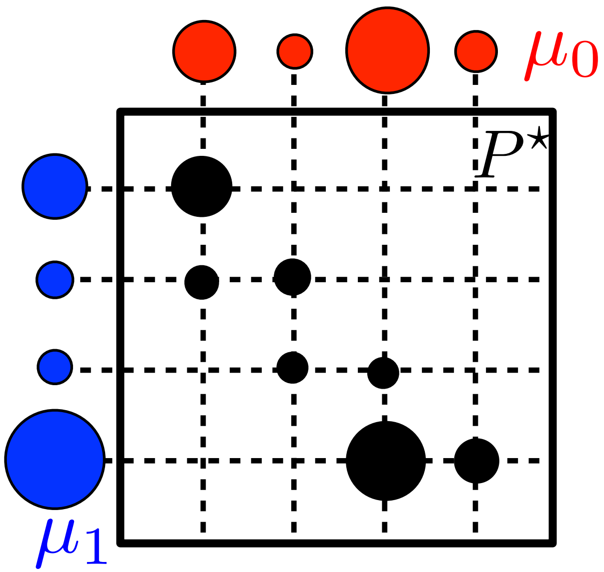
Weighted Discrete Case

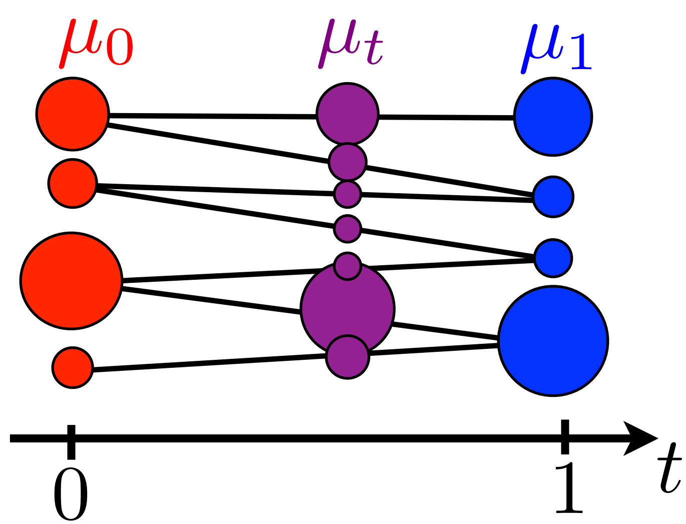
Linear program: $\mu_i = \sum_{k=1}^{N_i} p_i(k)\delta_{X_i(k)}$

$$P^\star \in \underset{P\in\Pi_{\mu_0,\mu_1}}{\operatorname{argmin}} \sum_{k,\ell} P_{k,\ell}\|X_0(k) - Y_1(\ell)\|^2$$

Theorem: [Folklore] $|\{(k,\ell) \setminus P^\star_{k,\ell}\} \neq 0| \leqslant N_0 + N_1 - 1.$

Barycenter: $\mu_t = \sum_{i,j} P^\star_{i,j}\delta_{(1-t)X_i+tY_j}$





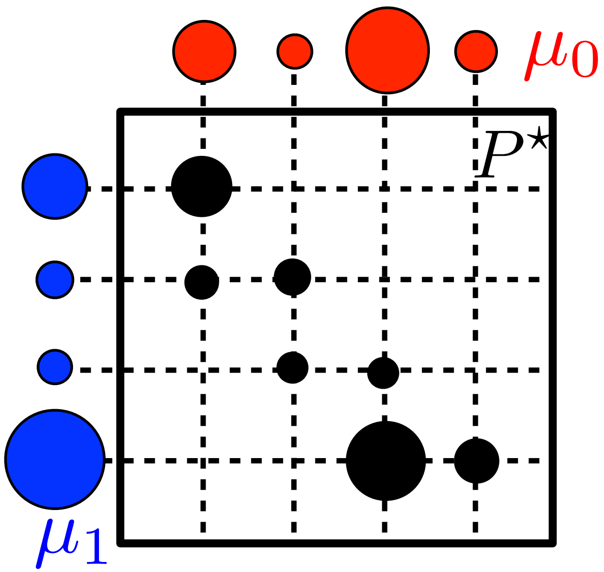
Weighted Discrete Case

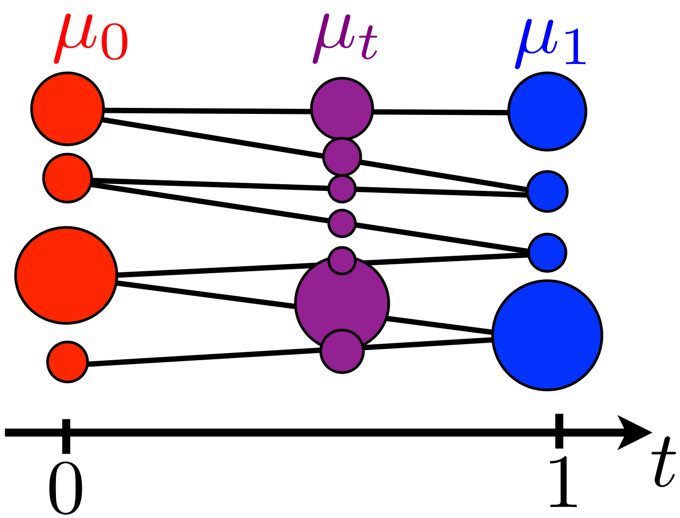
Linear program: $\mu_i = \sum_{k=1}^{N_i} p_i(k)\delta_{X_i(k)}$

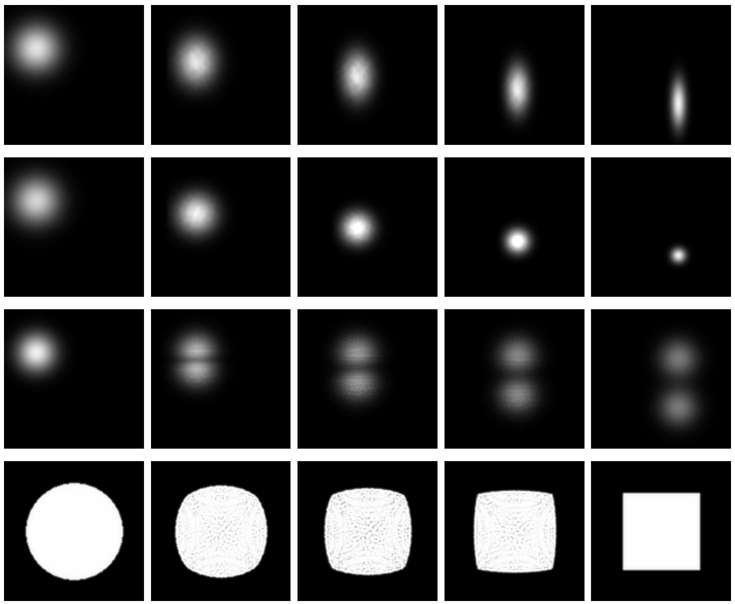
$$P^\star \in \underset{P \in \Pi_{\mu_0,\mu_1}}{\operatorname{argmin}} \sum_{k,\ell} P_{k,\ell} \|X_0(k) - Y_1(\ell)\|^2$$

Theorem: [Folklore] $|\{(k,\ell) \setminus P^\star_{k,\ell}\} \neq 0| \leqslant N_0 + N_1 - 1.$

Barycenter: $\mu_t = \sum_{i,j} P^\star_{i,j} \delta_{(1-t)X_i + tY_j}$







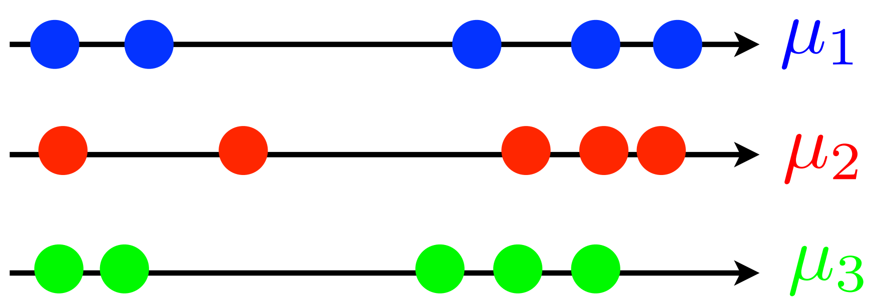
[Bonneel et al. 2011]

Special Case : 1-D Distributions

Discrete 1-D point clouds: $\mu_i = \sum_{k=1}^{N} \delta_{X_i(k)}$,

Ordering the points: $X_i(k) \leqslant X_i(k+1)$

$\rightarrow O(N \log(N))$ operations.



Special Case : 1-D Distributions

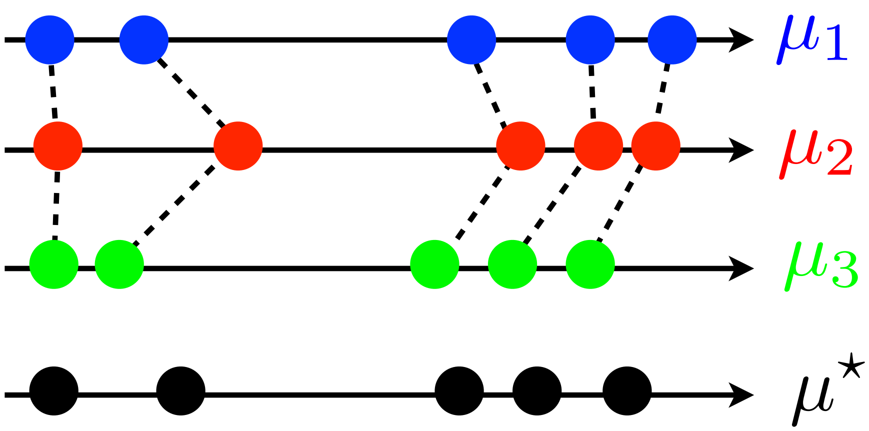
Discrete 1-D point clouds: $\mu_i = \sum_{k=1}^{N} \delta_{X_i(k)}$,

Ordering the points: $X_i(k) \leqslant X_i(k+1)$

$\rightarrow O(N \log(N))$ operations.

Barycenter: $\mu^\star = \sum_k \delta_{X^\star(k)}$

where $X^\star(k) = \sum_i \rho_i X_i(k)$



Special Case : 1-D Distributions

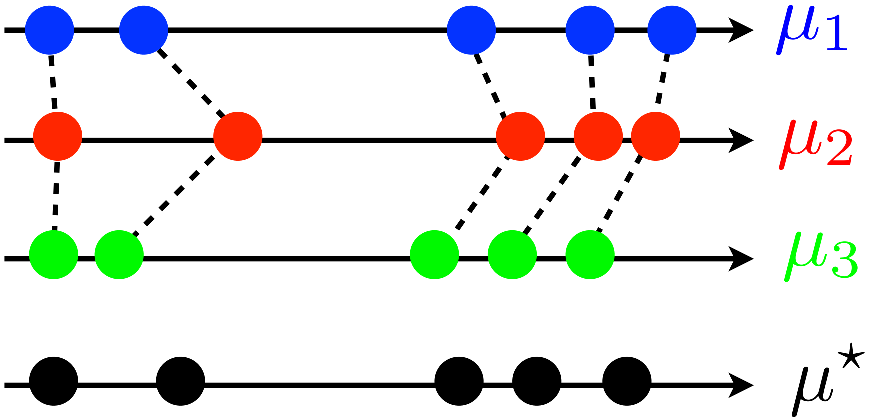
Discrete 1-D point clouds: $\mu_i = \sum_{k=1}^{N} \delta_{X_i(k)}$,

Ordering the points: $X_i(k) \leqslant X_i(k+1)$

$\rightarrow O(N\log(N))$ operations.

Barycenter: $\mu^\star = \sum_k \delta_{X^\star(k)}$

where $X^\star(k) = \sum_i \rho_i X_i(k)$

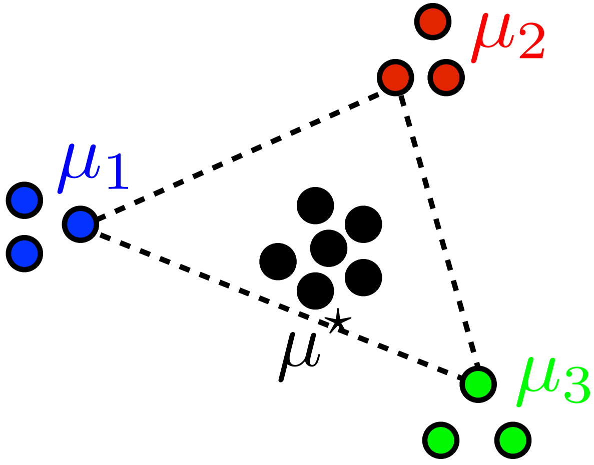


Continuous distributions: $C_{\mu^\star}^{-1} = \sum_i \rho_i C_{\mu_i}^{-1}$

$\rightarrow$ averaging the inverse cumulatives.

Linear Programming Resolution

Discrete setting: $\forall\, i = 1, \ldots, L, \quad \mu_i = \sum_k \delta_{X_i(k)}$

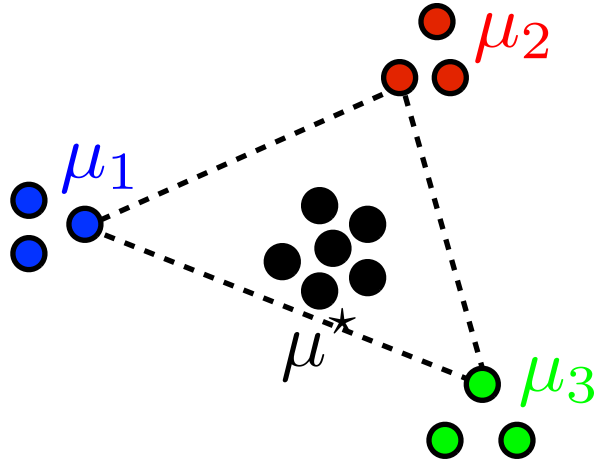


Linear Programming Resolution

Discrete setting: $\forall\, i = 1, \ldots, L, \quad \mu_i = \sum_k \delta_{X_i(k)}$

L-way interaction cost:

$$C(k_1, \ldots, k_L) = \sum_{i,j} \rho_i \rho_j \|X_i(k_i) - X_j(k_j)\|^2$$



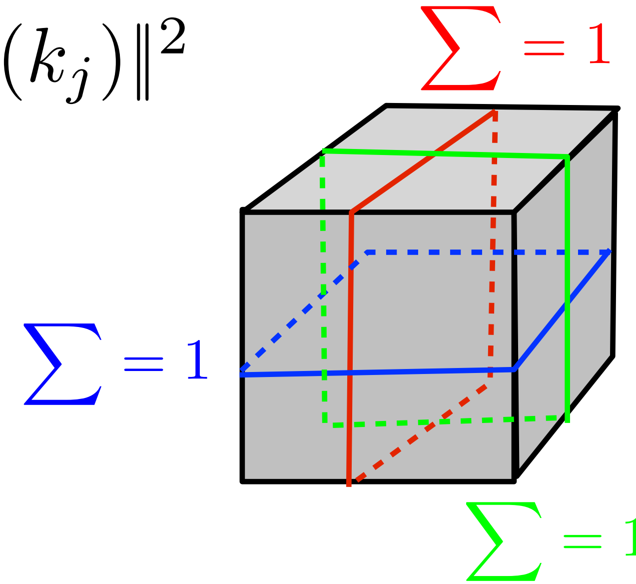
Linear Programming Resolution

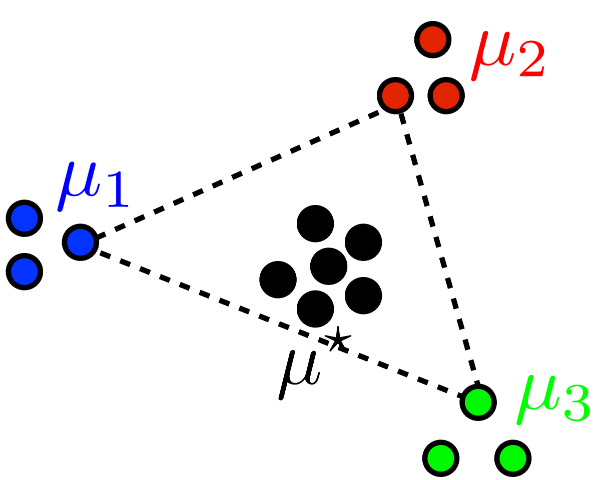
Discrete setting: $\forall\, i = 1, \ldots, L, \quad \mu_i = \sum_k \delta_{X_i(k)}$

L-way interaction cost:

$$C(k_1, \ldots, k_L) = \sum_{i,j} \rho_i \rho_j \|X_i(k_i) - X_j(k_j)\|^2$$

L-way coupling matrices: $P \in \mathcal{P}_L$





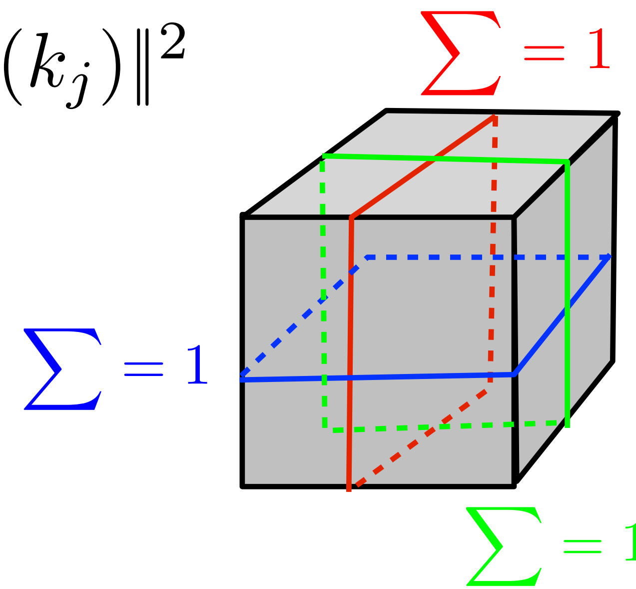
Linear Programming Resolution

Discrete setting: $\forall\, i = 1, \ldots, L, \quad \mu_i = \sum_k \delta_{X_i(k)}$

L-way interaction cost:

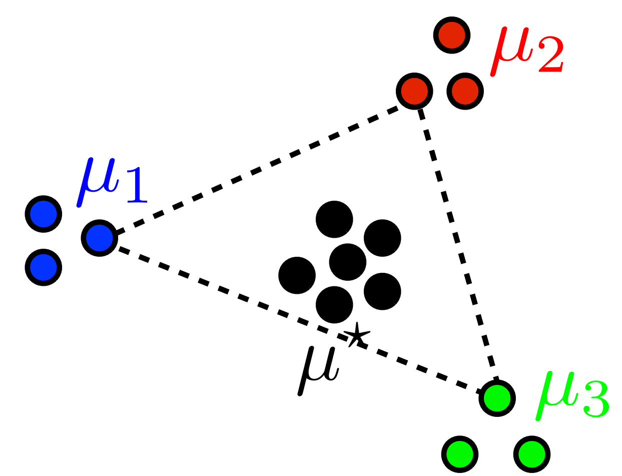
$$C(k_1, \ldots, k_L) = \sum_{i,j} \rho_i \rho_j \|X_i(k_i) - X_j(k_j)\|^2$$

L-way coupling matrices: $P \in \mathcal{P}_L$



Linear program:

$$\min_{P \in \mathcal{P}_L} \sum_{k_1, \ldots, k_L} P_{k_1, \ldots, k_L} C(k_1, \ldots, k_L)$$



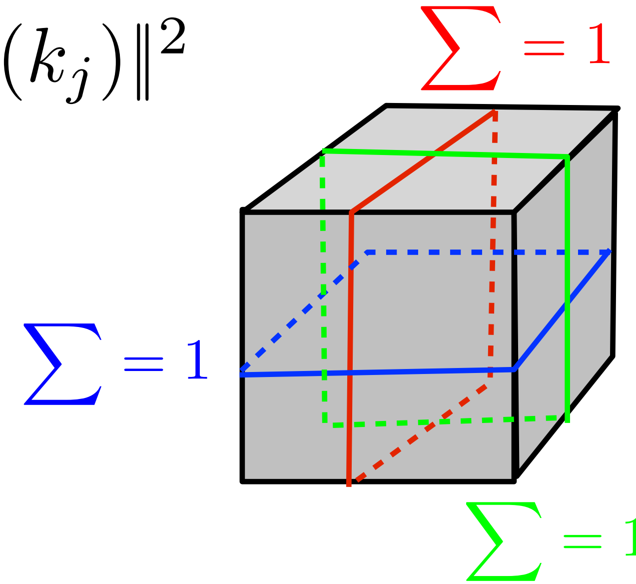
Linear Programming Resolution

Discrete setting: $\forall\, i = 1, \ldots, L, \quad \mu_i = \sum_k \delta_{X_i(k)}$

L-way interaction cost:

$$C(k_1, \ldots, k_L) = \sum_{i,j} \rho_i \rho_j \|X_i(k_i) - X_j(k_j)\|^2$$

L-way coupling matrices: $P \in \mathcal{P}_L$



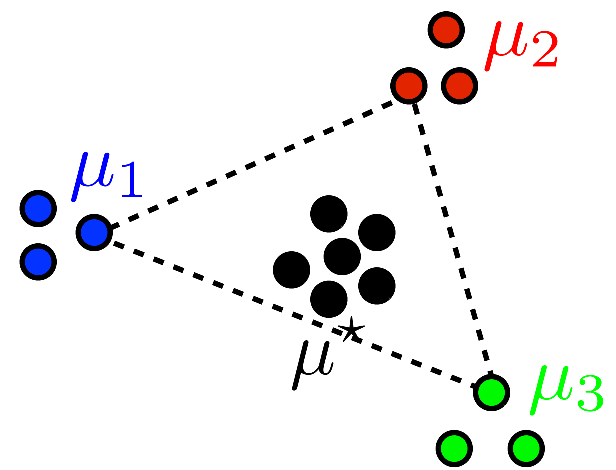
Linear program:

$$\min_{P \in \mathcal{P}_L} \sum_{k_1, \ldots, k_L} P_{k_1, \ldots, k_L} C(k_1, \ldots, k_L)$$

Barycenter:

$$\mu^\star = \sum_{k_1, \ldots, k_L} P_{k_1, \ldots, k_L} \delta_{X^\star(k_1, \ldots, k_L)}$$

$$X^\star(k_1, \ldots, k_L) = \sum_{i=1}^L \rho_i X_i(k_i)$$



L-ways Assignments

Numerical issues: $\rightarrow \mu^\star$ is a **weighted** cloud.

$\rightarrow$ up to N^L points.

Point clouds of fixed size: $$\mu = \sum_{k=1}^{N} \delta_{X(k)} \in \Theta_N$$

L-ways Assignments

Numerical issues: $\rightarrow \mu^\star$ is a **weighted** cloud.

$\rightarrow$ up to N^L points.

Point clouds of fixed size: $\mu = \sum_{k=1}^{N} \delta_{X(k)} \in \Theta_N$

Restricted barycenter: $\min_{\mu \in \Theta_N} \sum_i \rho_i W_2(\mu_i, \mu)^2$

$$\Longleftrightarrow \min_{P \in \tilde{\mathcal{P}}_L} \sum_{k_1,\ldots,k_L} P_{k_1,\ldots,k_L} C(k_1, \ldots, k_L)$$

$$\tilde{\mathcal{P}}_L = \left\{ P^\sigma \setminus \sigma = (\sigma_1, \ldots, \sigma_L) \in (\Sigma_N)^L \right\}$$

$$P^\sigma_{k_1,\ldots,k_L} = \begin{cases} 1 & \text{if} \quad \exists a, \forall\, i, k_i = \sigma_i(a), \\ 0 & \text{otherwise.} \end{cases}$$

L-ways Assignments

Numerical issues: $\rightarrow$ $\mu^\star$ is a **weighted** cloud.

$\rightarrow$ up to N^L points.

Point clouds of fixed size: $\mu = \sum_{k=1}^{N} \delta_{X(k)} \in \Theta_N$

Restricted barycenter: $\min_{\mu \in \Theta_N} \sum_i \rho_i W_2(\mu_i, \mu)^2$

$$\Longleftrightarrow \min_{P \in \tilde{\mathcal{P}}_L} \sum_{k_1,\ldots,k_L} P_{k_1,\ldots,k_L} C(k_1,\ldots,k_L)$$

$$\tilde{\mathcal{P}}_L = \left\{ P^\sigma \setminus \sigma = (\sigma_1,\ldots,\sigma_L) \in (\Sigma_N)^L \right\}$$

$$P^\sigma_{k_1,\ldots,k_L} = \begin{cases} 1 & \text{if} \quad \exists a, \forall\, i, k_i = \sigma_i(a), \\ 0 & \text{otherwise.} \end{cases}$$

$\rightarrow$ $\mathcal{P}_L$ is not the convex hull of $\tilde{\mathcal{P}}_L$.

$\rightarrow$ L-way assignment is NP-hard.

Sliced Wasserstein Barycenter

Sliced-barycenter: μ_X that solves

$$\min_X \sum_i \rho_i \, SW(\mu_X, \mu_{X_i})^2$$

Sliced Wasserstein Barycenter

Sliced-barycenter: μ_X that solves

$$\min_X \sum_i \rho_i \, SW(\mu_X, \mu_{X_i})^2$$

Gradient descent: $E_i(X) = SW(\mu_X, \mu_{X_i})^2$

$$X^{(\ell+1)} = X^{(\ell)} - \tau_\ell \sum_{i=1}^{L} \rho_i \nabla E_i(X^{(\ell)})$$

Sliced Wasserstein Barycenter

Sliced-barycenter: μ_X that solves

$$\min_X \sum_i \rho_i \, SW(\mu_X, \mu_{X_i})^2$$

Gradient descent: $E_i(X) = SW(\mu_X, \mu_{X_i})^2$

$$X^{(\ell+1)} = X^{(\ell)} - \tau_\ell \sum_{i=1}^{L} \rho_i \nabla E_i(X^{(\ell)})$$

Advantages:

μ_X is a sum of N Diracs.

Smooth optimization problem.

Disadvantage:

Non-convex problem

$\rightarrow$ local minima.

Sliced Wasserstein Barycenter

Sliced-barycenter: μ_X that solves

$$\min_X \sum_i \rho_i \, SW(\mu_X, \mu_{X_i})^2$$

Gradient descent: $E_i(X) = SW(\mu_X, \mu_{X_i})^2$

$$X^{(\ell+1)} = X^{(\ell)} - \tau_\ell \sum_{i=1}^{L} \rho_i \nabla E_i(X^{(\ell)})$$

Advantages:

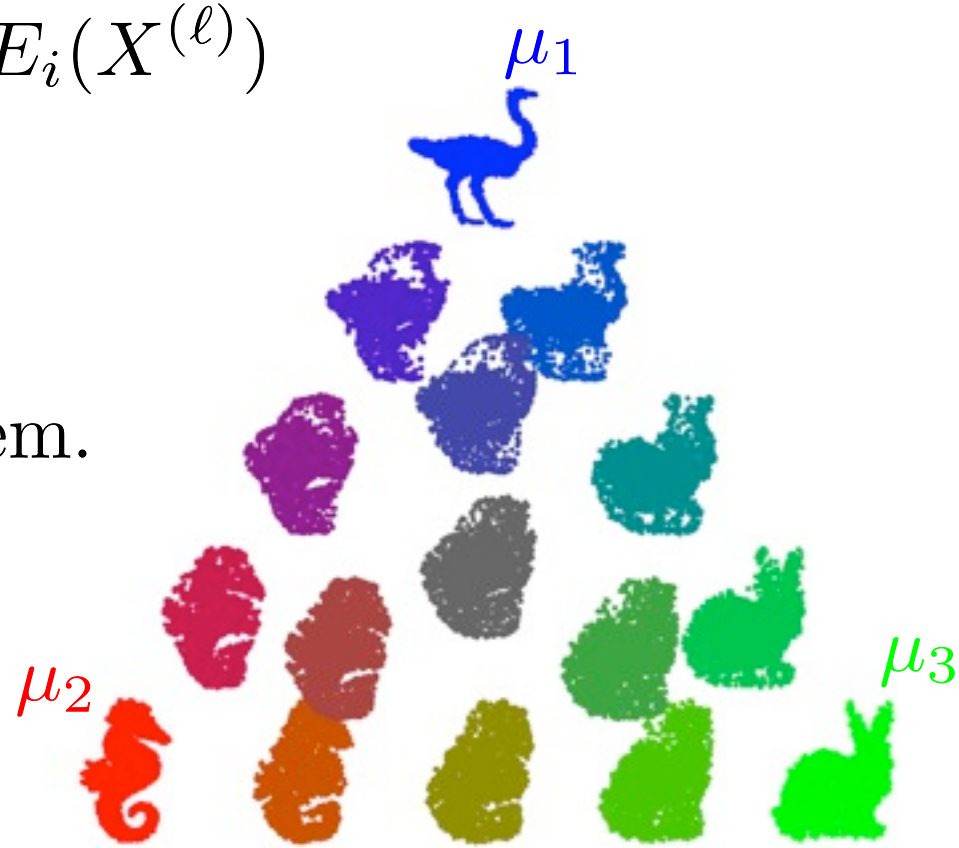
μ_X is a sum of N Diracs.

Smooth optimization problem.

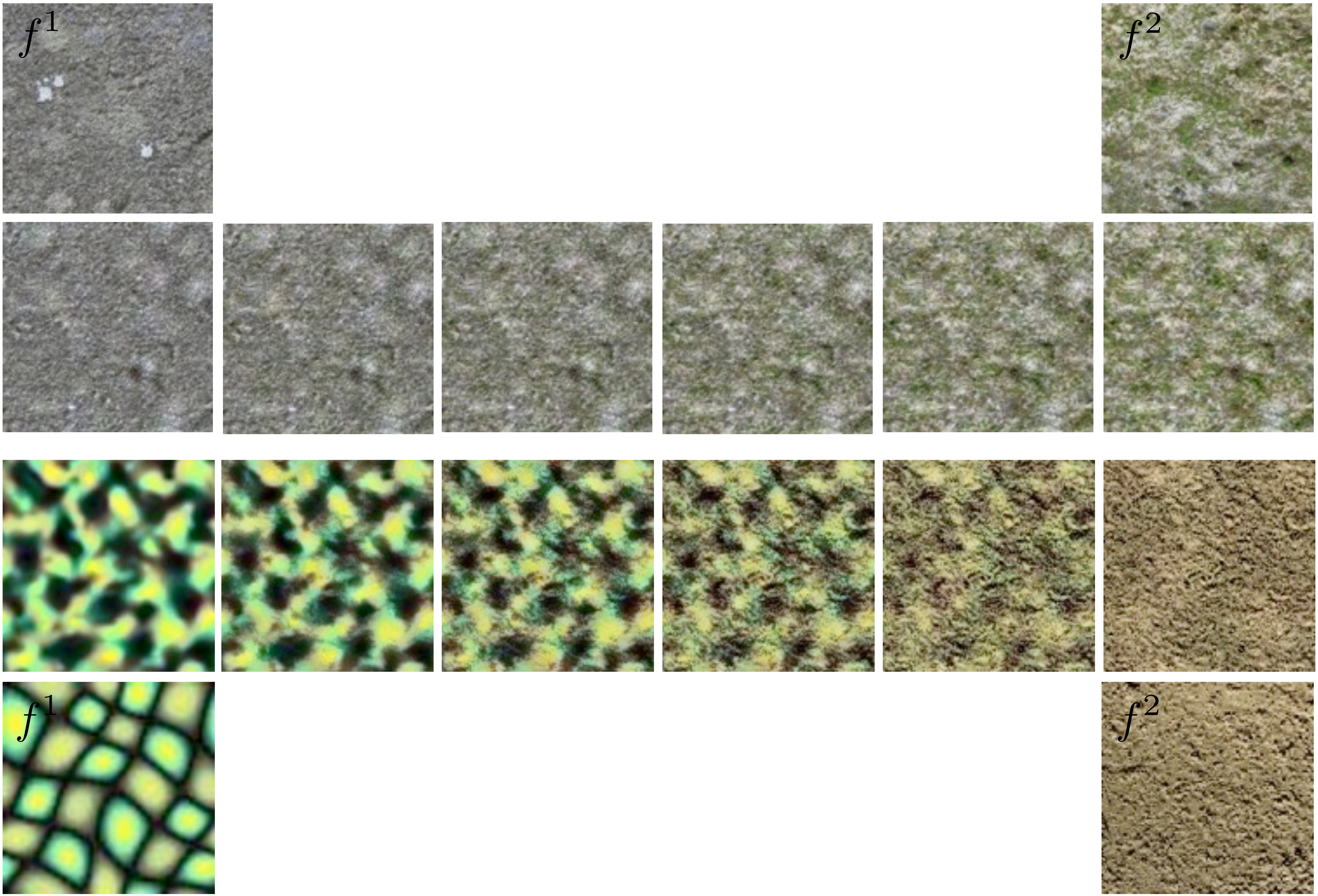
Disadvantage:

Non-convex problem

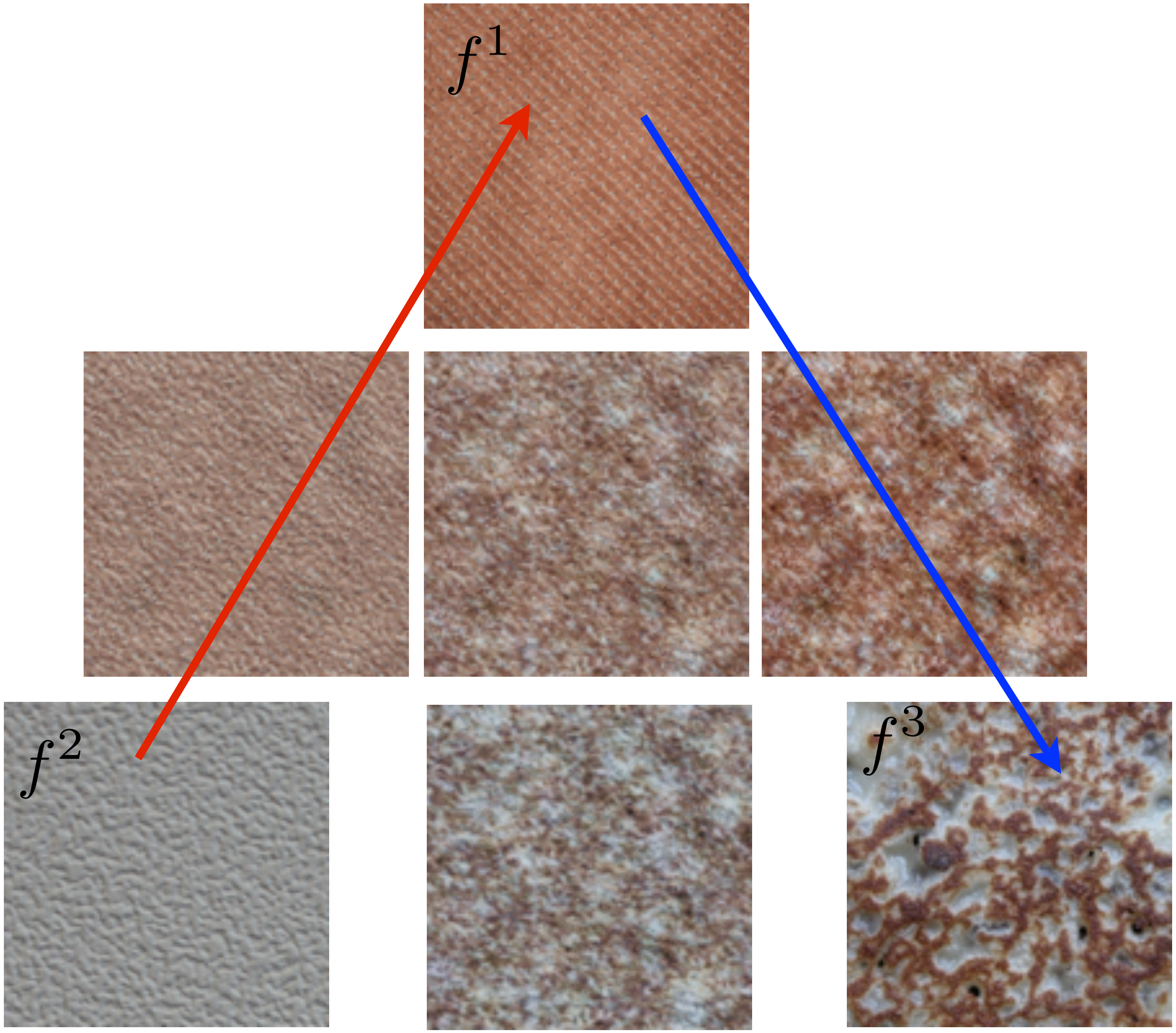
$\rightarrow$ local minima.

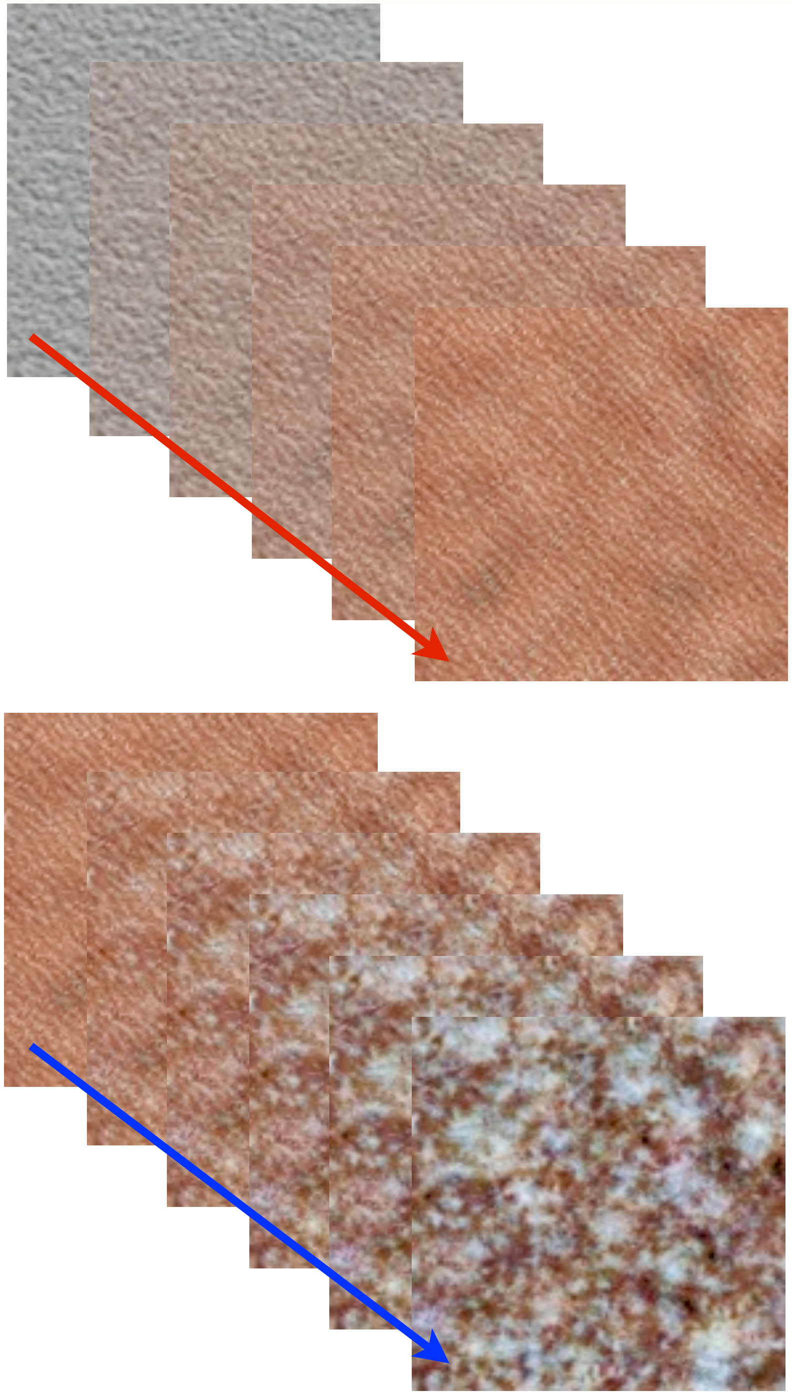


Barycenters of 2 Textures

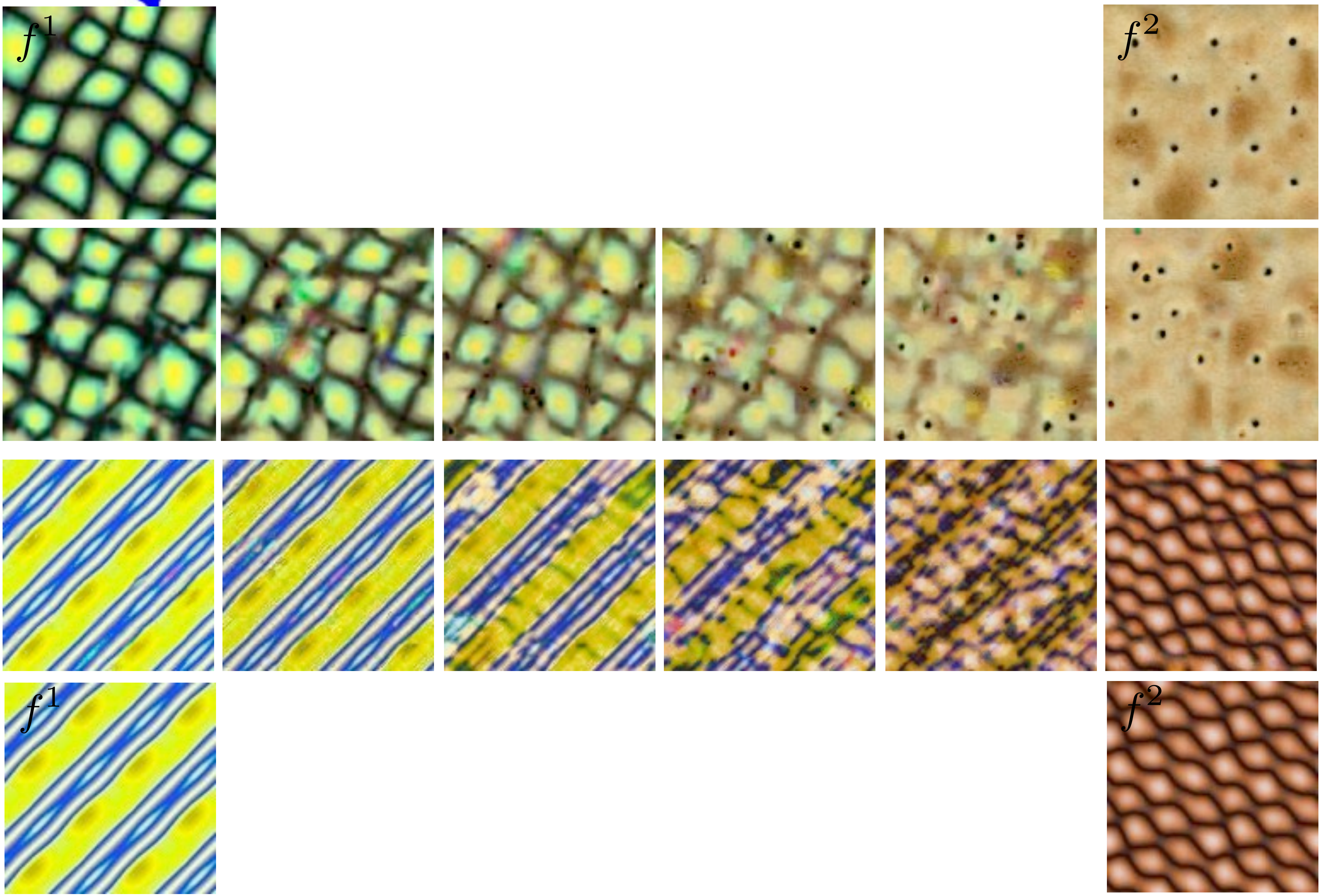


Barycenters of 3 Textures





Higher-order Synthesis



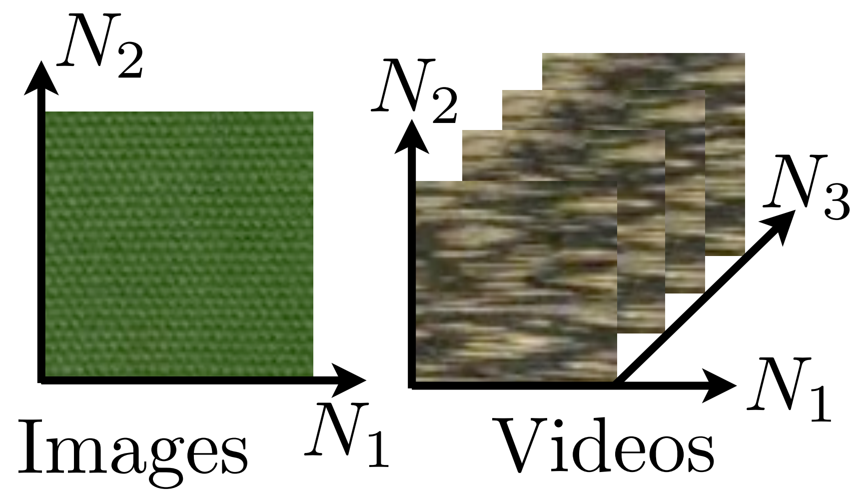
Overview

- Wasserstein Distance
- Sliced Wasserstein Distance
- Color Transfer
- Regularized Color Transfer
- Texture Synthesis
- Wasserstein Barycenter for Texture Mixing
- **Gaussian Texture Models**

Gaussian Texture Model

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$

$d = 1$ (grayscale), $d = 3$ (color)



Gaussian Texture Model

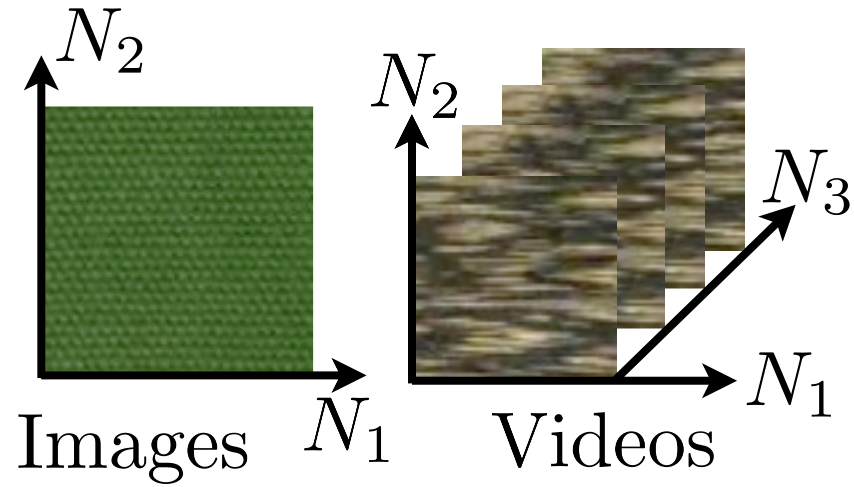
Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$

$d = 1$ (grayscale), $d = 3$ (color)

Gaussian model:

$$X \sim \mu = \mathcal{N}(m, \Sigma)$$

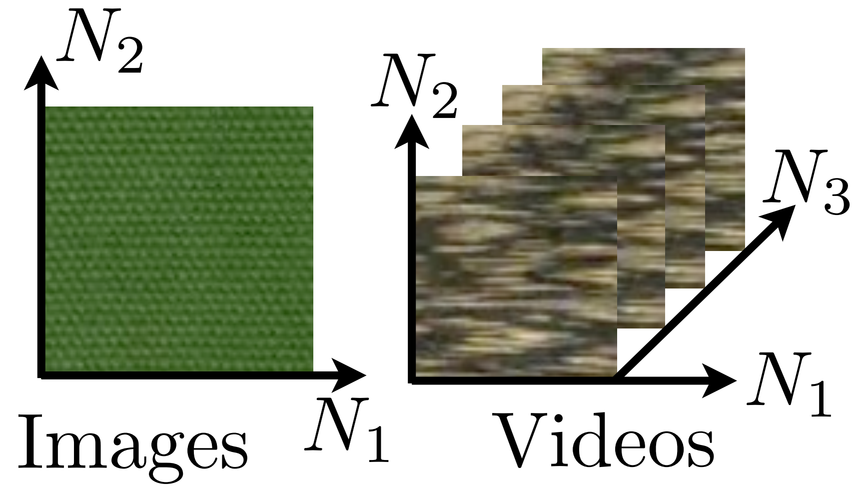
$$m \in \mathbb{R}^{N \times d}, \Sigma \in \mathbb{R}^{Nd \times Nd}$$



Gaussian Texture Model

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$

$d = 1$ (grayscale), $d = 3$ (color)



Gaussian model:

$$X \sim \mu = \mathcal{N}(m, \Sigma)$$

$$m \in \mathbb{R}^{N \times d}, \Sigma \in \mathbb{R}^{Nd \times Nd}$$

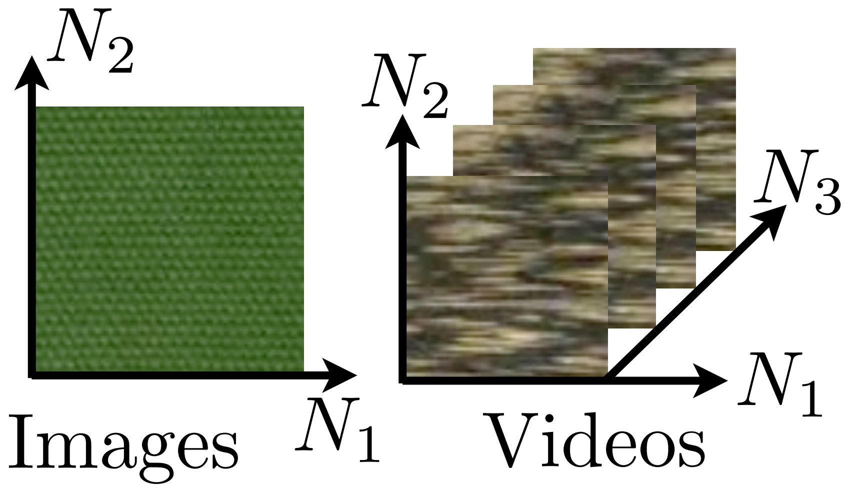
Texture analysis: from $f_0 \in \mathbb{R}^{N \times d}$, learn (m, Σ).

$\rightarrow$ highly under-determined problem.

Gaussian Texture Model

Input exemplar: $f_0 \in \mathbb{R}^{N \times d}$

$d = 1$ (grayscale), $d = 3$ (color)



Gaussian model:

$$X \sim \mu = \mathcal{N}(m, \Sigma)$$

$$m \in \mathbb{R}^{N \times d}, \Sigma \in \mathbb{R}^{Nd \times Nd}$$

Texture analysis: from $f_0 \in \mathbb{R}^{N \times d}$, learn (m, Σ).

$\rightarrow$ highly under-determined problem.

Texture synthesis:

given (m, Σ), draw a realization $f = X(\omega)$.

$\rightarrow$ Factorize $\Sigma = AA^*$ (e.g. Cholesky).

$\rightarrow$ Compute $f = m + Aw$ where w drawn from $\mathcal{N}(0, \mathrm{Id})$.

Spot Noise Model [Galerne et al.]

Stationarity hypothesis: $\quad$ (periodic BC) $\quad X(\cdot + \tau) \sim X$

Spot Noise Model [Galerne et al.]

Stationarity hypothesis: (periodic BC) $X(\cdot + \tau) \sim X$

Block-diagonal Fourier covariance:

$$y = \Sigma f \quad \text{computed as} \quad \hat{y}(\omega) = \hat{\Sigma}(\omega) \hat{f}(\omega)$$

$$\text{where} \quad \hat{f}(\omega) = \sum_x f(x) e^{\frac{2ix_1\omega_1\pi}{N_1} + \frac{2ix_2\omega_2\pi}{N_2}}$$

Spot Noise Model [Galerne et al.]

Stationarity hypothesis: (periodic BC) $X(\cdot + \tau) \sim X$

Block-diagonal Fourier covariance:

$$y = \Sigma f \quad \text{computed as} \quad \hat{y}(\omega) = \hat{\Sigma}(\omega)\hat{f}(\omega)$$

$$\text{where} \quad \hat{f}(\omega) = \sum_x f(x) e^{\frac{2ix_1\omega_1\pi}{N_1} + \frac{2ix_2\omega_2\pi}{N_2}}$$

Maximum likelihood estimate (MLE) of m from f_0:

$$\forall\, i, \quad m_i = \frac{1}{N} \sum_x f_0(x) \in \mathbb{R}^d$$

Spot Noise Model [Galerne et al.]

Stationarity hypothesis: (periodic BC) $X(\cdot + \tau) \sim X$

Block-diagonal Fourier covariance:

$$y = \Sigma f \quad \text{computed as} \quad \hat{y}(\omega) = \hat{\Sigma}(\omega)\hat{f}(\omega)$$

$$\text{where} \quad \hat{f}(\omega) = \sum_x f(x) e^{\frac{2ix_1\omega_1\pi}{N_1} + \frac{2ix_2\omega_2\pi}{N_2}}$$

Maximum likelihood estimate (MLE) of m from f_0:

$$\forall\, i, \quad m_i = \frac{1}{N}\sum_x f_0(x) \in \mathbb{R}^d$$

MLE of Σ: $\quad \Sigma_{i,j} = \frac{1}{N}\sum_x f_0(i+x)^* f_0(j+x) \in \mathbb{R}^{d\times d}$

Spot Noise Model [Galerne et al.]

Stationarity hypothesis: (periodic BC) $X(\cdot + \tau) \sim X$

Block-diagonal Fourier covariance:

$$y = \Sigma f \quad \text{computed as} \quad \hat{y}(\omega) = \hat{\Sigma}(\omega)\hat{f}(\omega)$$

$$\text{where} \quad \hat{f}(\omega) = \sum_x f(x) e^{\frac{2ix_1\omega_1\pi}{N_1} + \frac{2ix_2\omega_2\pi}{N_2}}$$

Maximum likelihood estimate (MLE) of m from f_0:

$$\forall\, i, \quad m_i = \frac{1}{N}\sum_x f_0(x) \in \mathbb{R}^d$$

MLE of Σ: $\displaystyle \Sigma_{i,j} = \frac{1}{N}\sum_x f_0(i+x)^* f_0(j+x) \in \mathbb{R}^{d\times d}$

$$\iff \quad \forall\, \omega \neq 0, \quad \hat{\Sigma}(\omega) = \hat{f}_0(\omega)\hat{f}_0(\omega)^* \in \mathbb{C}^{d\times d}$$

Σ is a spot noise $\iff \forall\, \omega \neq 0$, $\hat{\Sigma}(\omega)$ is rank-1.

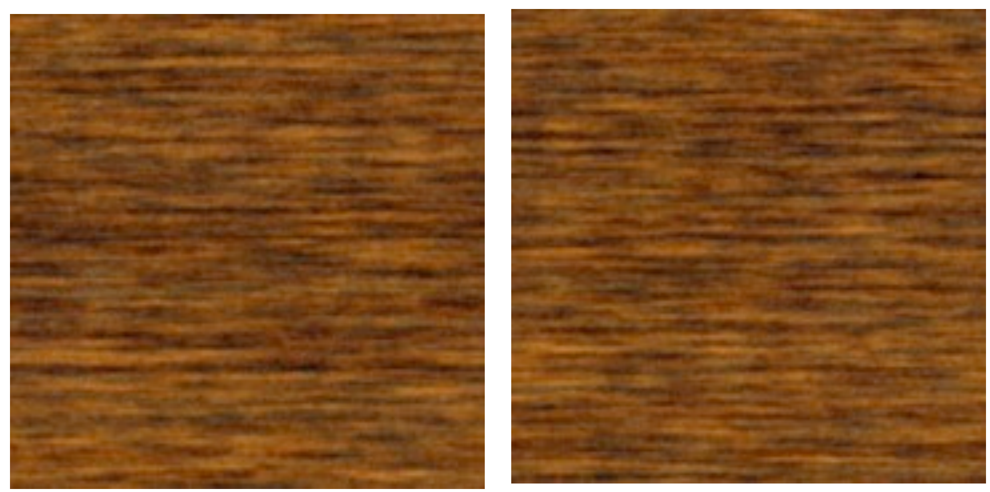
Example of Synthesis

Synthesizing $f = X(\omega), X \sim \mathcal{N}(m, \Sigma)$:

$$\forall\, \omega \neq 0, \quad \hat f(\omega) = \underbrace{\hat f_0(\omega)}_{\in\, \mathbb{C}^d}\underbrace{\hat w(\omega)}_{\in\, \mathbb{C}} \qquad w \sim \mathcal{N}(N^{-1}, N^{-1/2}\mathrm{Id}_N)$$

$\rightarrow$ Convolve each channel with the **same** white noise.

Input $f_0 \in \mathbb{R}^{N \times 3}$



Realizations f

Gaussian Wasserstein Distance

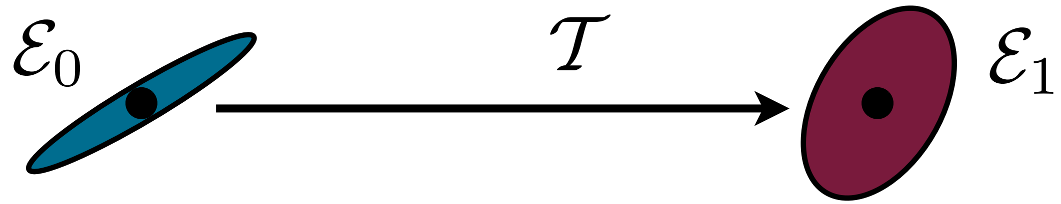
Input distributions (μ_0, μ_1) with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

Ellipses: $\mathcal{E}_i = \{u \setminus (m_i - x)^* \Sigma_i^{-1} (m_i - x) \leqslant c\}$

Gaussian Wasserstein Distance

Input distributions (μ_0, μ_1) with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

Ellipses: $\mathcal{E}_i = \{u \setminus (m_i - x)^* \Sigma_i^{-1} (m_i - x) \leqslant c\}$



Unique transport: $\ker(\Sigma_0) \cap \mathrm{Im}(\Sigma_1) = \{0\}$

L^2 *optimal transport:* affine map

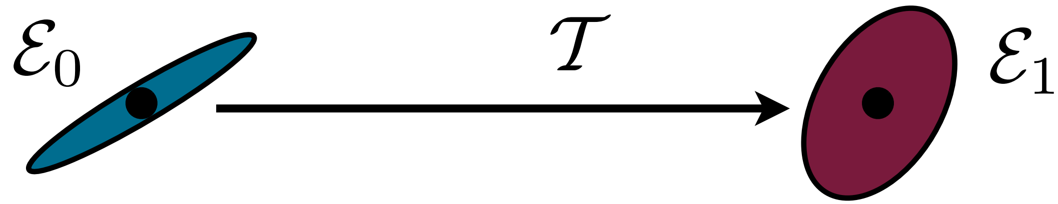
$$\mathcal{T} : u \mapsto Tu + m_1 - m_0$$

$$T = \Sigma_1^{1/2} \Sigma_{0,1}^{+} \Sigma_1^{1/2} \qquad \Sigma_{0,1} = (\Sigma_1^{1/2} \Sigma_0 \Sigma_1^{1/2})^{1/2}$$

Gaussian Wasserstein Distance

Input distributions (μ_0, μ_1) with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

Ellipses: $\mathcal{E}_i = \{u \setminus (m_i - x)^* \Sigma_i^{-1} (m_i - x) \leqslant c\}$



Unique transport: $\ker(\Sigma_0) \cap \mathrm{Im}(\Sigma_1) = \{0\}$

L^2 *optimal transport:* affine map

$$\mathcal{T} : u \mapsto Tu + m_1 - m_0$$

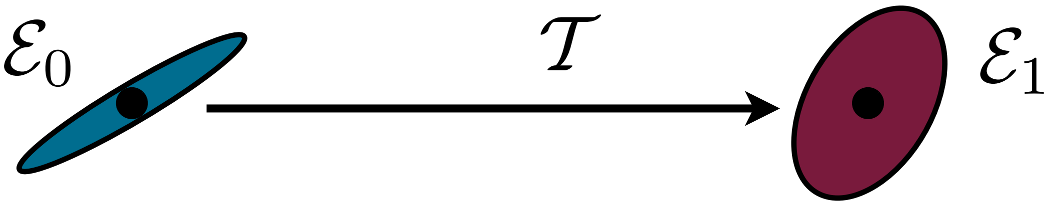
$$T = \Sigma_1^{1/2} \Sigma_{0,1}^{+} \Sigma_1^{1/2} \qquad \Sigma_{0,1} = (\Sigma_1^{1/2} \Sigma_0 \Sigma_1^{1/2})^{1/2}$$

One has $\mathcal{T} \sharp \mu_0 = \mu_1$ and $\mathcal{T} = \nabla \varphi$ where φ is convex.

Gaussian Wasserstein Distance

Input distributions (μ_0, μ_1) with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

Ellipses: $\mathcal{E}_i = \{ u \setminus (m_i - x)^* \Sigma_i^{-1} (m_i - x) \leqslant c \}$



Unique transport: $\ker(\Sigma_0) \cap \mathrm{Im}(\Sigma_1) = \{0\}$

L^2 *optimal transport:* affine map

$$\mathcal{T} : u \mapsto Tu + m_1 - m_0$$

$$T = \Sigma_1^{1/2} \Sigma_{0,1}^{+} \Sigma_1^{1/2} \qquad \Sigma_{0,1} = (\Sigma_1^{1/2} \Sigma_0 \Sigma_1^{1/2})^{1/2}$$

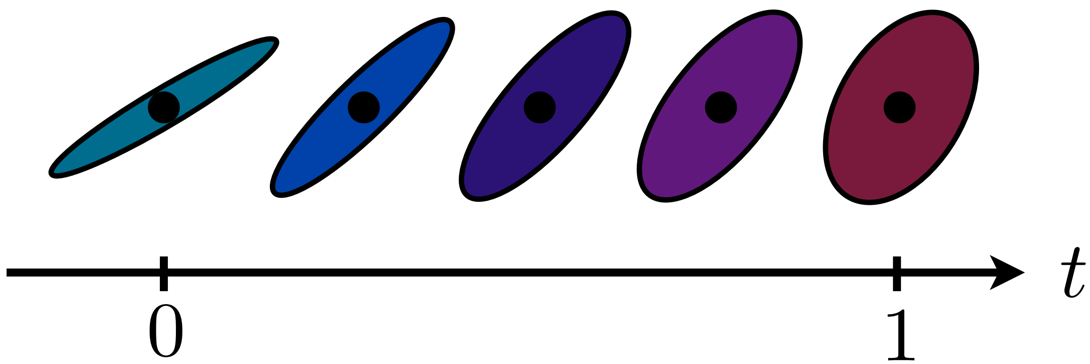
One has $\mathcal{T} \sharp \mu_0 = \mu_1$ and $\mathcal{T} = \nabla \varphi$ where φ is convex.

Wasserstein L^2 distance:

$$W_2(\mu_0, \mu_1)^2 = \mathrm{tr}\,(\Sigma_0 + \Sigma_1 - 2\Sigma_{0,1}) + \|m_0 - m_1\|^2,$$

Gaussian Wasserstein Geodesics

$$\forall\, t \in [0,1], \quad \mathcal{T}_t = (1-t)\mathrm{Id} + t\mathcal{T}$$



OT geodesic: $\mu_t = \mathcal{T}_t \sharp \mu_0 = \mathcal{N}(m_t, \Sigma_t)$

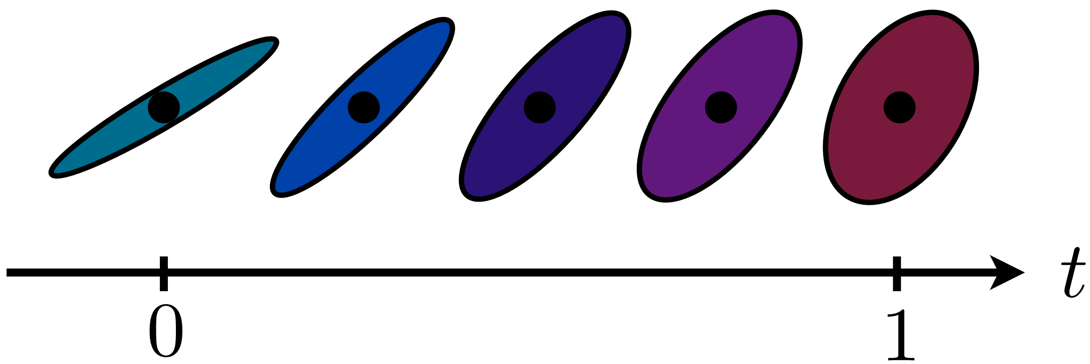
$$m_t = (1-t)m_0 + tm_1$$

$$\Sigma_t = [(1-t)\mathrm{Id} + tT]\Sigma_0[(1-t)\mathrm{Id} + tT]$$

$\longrightarrow$ the set of Gaussians is geodesically convex.

Gaussian Wasserstein Geodesics

$$\forall\, t \in [0,1], \quad \mathcal{T}_t = (1-t)\mathrm{Id} + t\mathcal{T}$$



OT geodesic: $\mu_t = \mathcal{T}_t \sharp \mu_0 = \mathcal{N}(m_t, \Sigma_t)$

$$m_t = (1-t)m_0 + t m_1$$

$$\Sigma_t = [(1-t)\mathrm{Id} + tT]\Sigma_0[(1-t)\mathrm{Id} + tT]$$

$\longrightarrow$ the set of Gaussians is geodesically convex.

Variational caracterization: (W_2 is a geodesic distance)

$$\mu_t = \underset{\mu}{\operatorname{argmin}}\ (1-t)W_2(\mu_0,\mu)^2 + tW_2(\mu_1,\mu)^2$$

Geodesic of Spot Noises

Theorem: Let for $i = 0, 1$, $\mu_i = \mu(f^{[i]})$ be spot noises, i.e. $\hat{\Sigma}_i(\omega) = \hat{f}^{[i]}(\omega)\hat{f}^{[i]}(\omega)^*$. Then $\forall\, t \in [0, 1]$, $\mu_t = \mu(f^{[t]})$

$$f^{[t]} = (1-t)f^{[0]} + tg^{[1]}$$

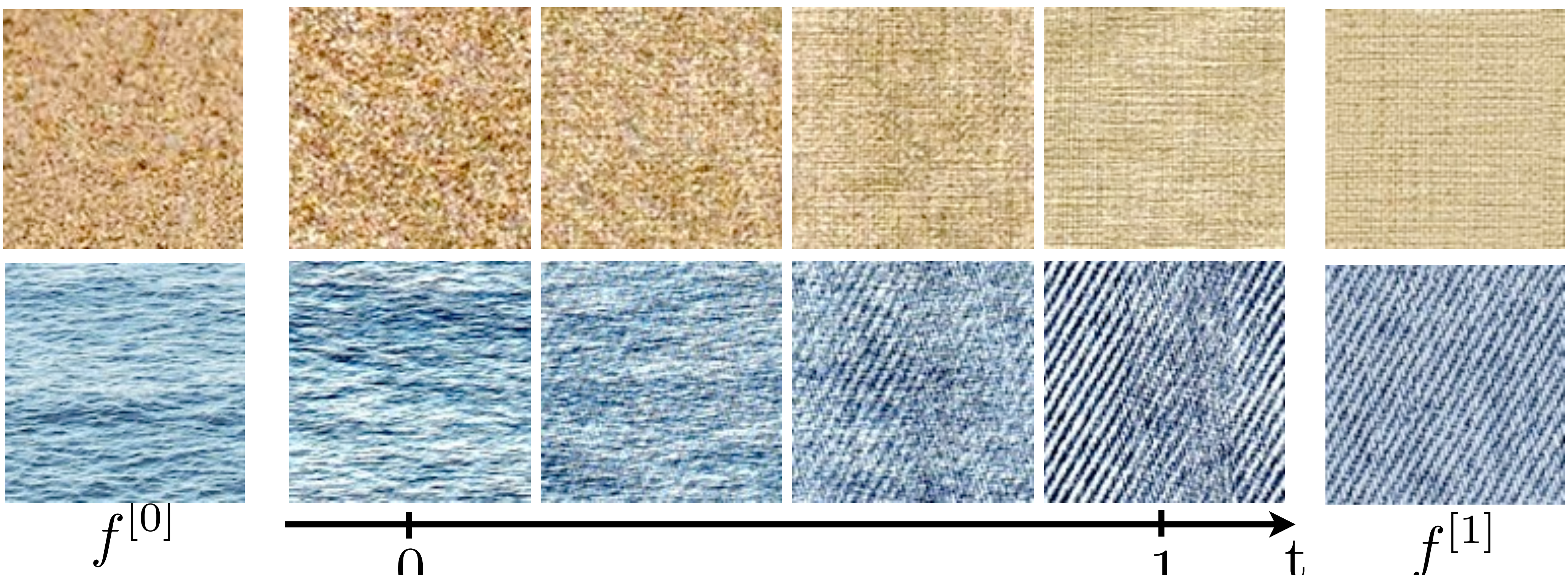
$$\hat{g}^{[1]}(\omega) = \hat{f}^{[1]}(\omega)\frac{\hat{f}^{[1]}(\omega)^*\hat{f}^{[0]}(\omega)}{|\hat{f}^{[1]}(\omega)^*\hat{f}^{[0]}(\omega)|}$$

Geodesic of Spot Noises

Theorem: Let for $i = 0, 1$, $\mu_i = \mu(f^{[i]})$ be spot noises, i.e. $\hat{\Sigma}_i(\omega) = \hat{f}^{[i]}(\omega)\hat{f}^{[i]}(\omega)^*$. Then $\forall\, t \in [0, 1]$, $\mu_t = \mu(f^{[t]})$

$$f^{[t]} = (1-t)f^{[0]} + tg^{[1]}$$

$$\hat{g}^{[1]}(\omega) = \hat{f}^{[1]}(\omega)\frac{\hat{f}^{[1]}(\omega)^*\hat{f}^{[0]}(\omega)}{|\hat{f}^{[1]}(\omega)^*\hat{f}^{[0]}(\omega)|}$$



Information Geometry Barycenters

Input distributions: $\mu_i = \mathcal{N}(0, \Sigma_i)$

Rao	Wasserstein
$d(\mu_0, \mu_1)^2 =$ $\\|\log(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})\\|^2$	$\mathrm{tr}\left(\Sigma_0 + \Sigma_1 - 2\Sigma_{0,1}\right)$ $\Sigma_{0,1} = (\Sigma_1^{1/2}\Sigma_0\Sigma_1^{1/2})^{1/2}$
Diagonal $\Sigma_i = \mathrm{diag}(\lambda_k^i)_k$ $d(\mu_0, \mu_1)^2 = \sum_k \log(\lambda_k^0/\lambda_k^1)^2$	$\sum_k \left(\sqrt{\lambda_k^0} - \sqrt{\lambda_k^1}\right)^2$

$\rightarrow$ Rao's distance requires full rank covariances.

$\rightarrow$ Case $m_i \neq 0$?

Information Geometry Barycenters

Input distributions: $\mu_i = \mathcal{N}(0, \Sigma_i)$

Rao	Wasserstein
$d(\mu_0, \mu_1)^2 =$ $\lVert \log(\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2}) \rVert^2$	$\mathrm{tr}\,(\Sigma_0 + \Sigma_1 - 2\Sigma_{0,1})$ $\Sigma_{0,1} = (\Sigma_1^{1/2} \Sigma_0 \Sigma_1^{1/2})^{1/2}$
Barycenter $\mu_t = \mathcal{N}(0, \Sigma_t)$ $\Sigma_t = \Sigma_0^{1/2} (\Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2})^t \Sigma_0^{1/2}$	$\Sigma_t = \mathcal{T}_t \Sigma_0 \mathcal{T}_t$ $\mathcal{T}_t = (1-t)\mathrm{Id} + t \Sigma_1^{1/2} \Sigma_{0,1}^{+} \Sigma_1^{1/2}$
Diagonal $\Sigma_i = \mathrm{diag}(\lambda_k^i)_k$ $d(\mu_0, \mu_1)^2 = \sum_k \log(\lambda_k^0 / \lambda_k^1)^2$ $\lambda_k^t = (\lambda_k^0)^{1-t} (\lambda_k^1)^t$	$\sum_k \left(\sqrt{\lambda_k^0} - \sqrt{\lambda_k^1} \right)^2$ $\sqrt{\lambda_k^t} = (1-t)\sqrt{\lambda_k^0} + t\sqrt{\lambda_k^1}$

→ Rao's distance requires full rank covariances.

→ Case $m_i \neq 0$?

OT Barycenters

Input distributions $(\mu_i)_{i \in I}$ with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

$$\mu^\star = \underset{\mu}{\operatorname{argmin}} \sum_{i \in I} \rho_i W_2(\mu_i, \mu)^2$$

OT Barycenters

Input distributions $(\mu_i)_{i \in I}$ with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

$$\mu^\star = \underset{\mu}{\operatorname{argmin}} \sum_{i \in I} \rho_i W_2(\mu_i, \mu)^2$$

Theorem: [Carlier, Agueh] If Σ_0 has full rank,

$\mu^\star$ is unique and $\mu^\star = \mathcal{N}(m^\star, \Sigma^\star)$ where

$$m^\star = \textstyle\sum_{i \in I} \rho_i m_i$$

$$\Sigma^\star = \Phi(\Sigma^\star) \quad \text{where} \quad \Phi(\Sigma) = \sum_{i \in I} \rho_i \left(\Sigma^{1/2} \Sigma_i \Sigma^{1/2} \right)^{1/2}$$

OT Barycenters

Input distributions $(\mu_i)_{i\in I}$ with $\mu_i = \mathcal{N}(m_i, \Sigma_i)$.

$$\mu^\star = \underset{\mu}{\operatorname{argmin}} \sum_{i\in I} \rho_i W_2(\mu_i, \mu)^2$$

Theorem: [Carlier, Agueh] If Σ_0 has full rank,

$$\mu^\star \text{ is unique and } \mu^\star = \mathcal{N}(m^\star, \Sigma^\star) \text{ where}$$

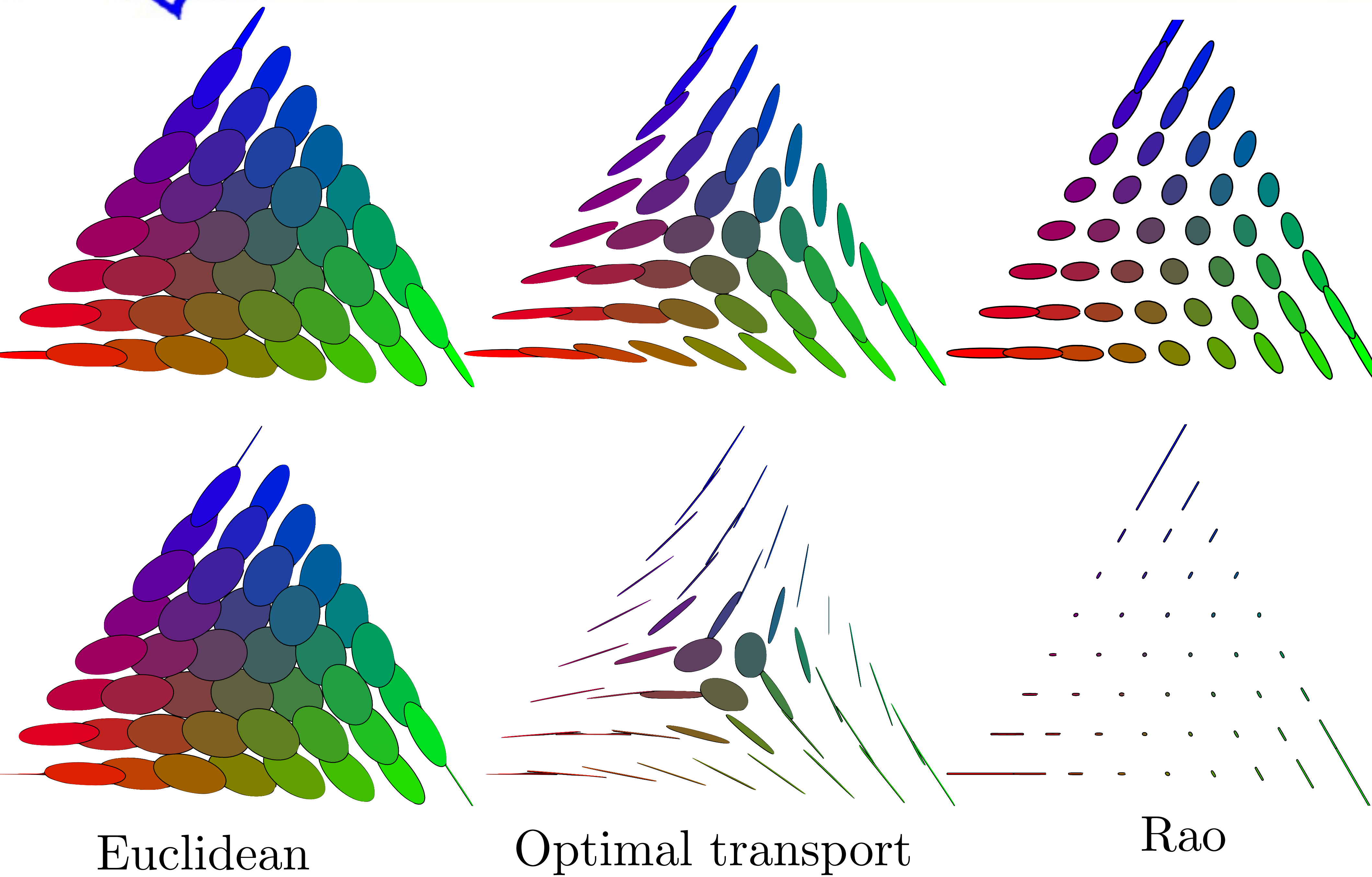
$$m^\star = \textstyle\sum_{i\in I} \rho_i m_i$$

$$\Sigma^\star = \Phi(\Sigma^\star) \quad \text{where} \quad \Phi(\Sigma) = \sum_{i\in I} \rho_i \left(\Sigma^{1/2} \Sigma_i \Sigma^{1/2} \right)^{1/2}$$

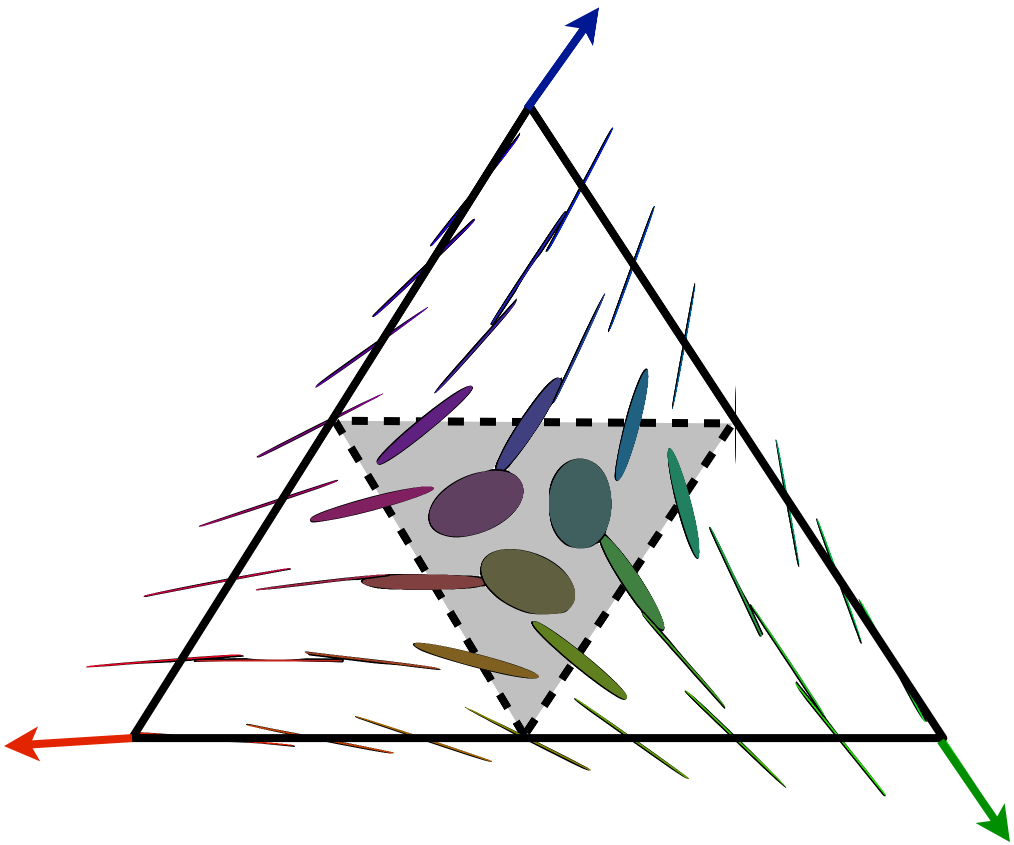
Numerical scheme: $\Sigma^{(\ell+1)} = \Phi(\Sigma^{(\ell)})$

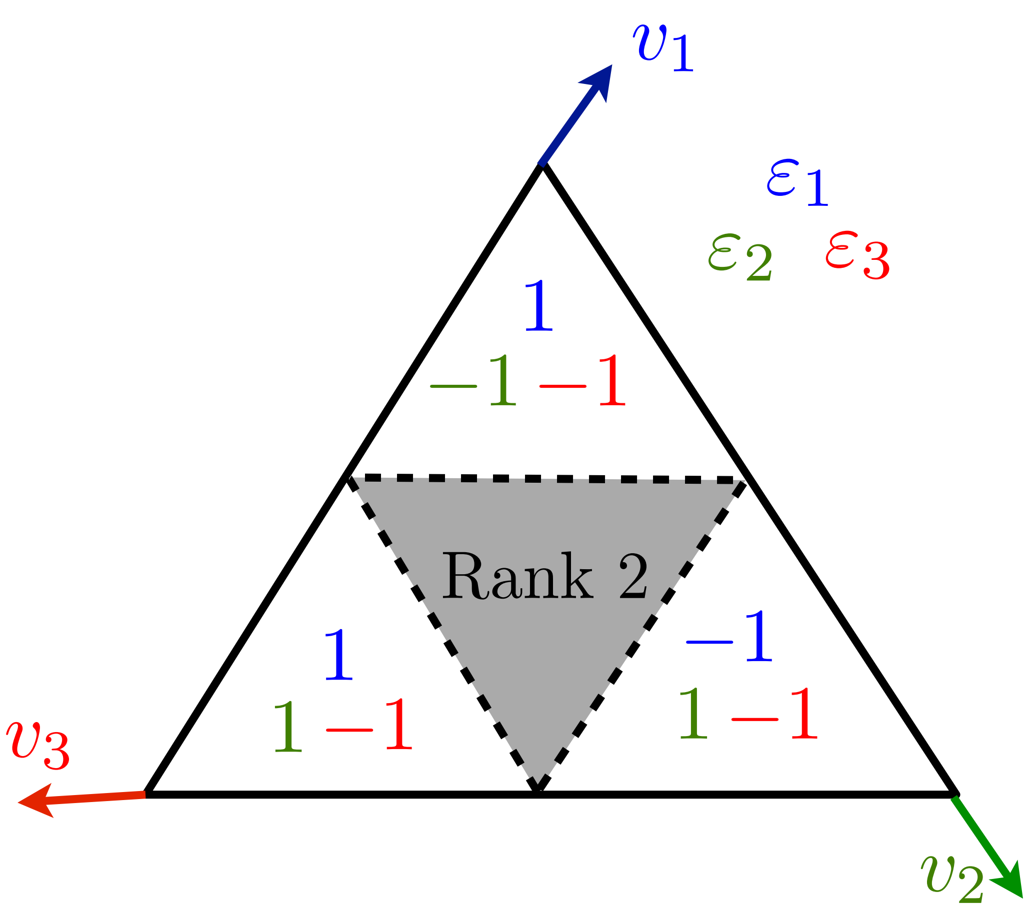
Conjecture: $\Sigma^{(\ell)} \to \Sigma^\star$.

2-D Gaussian Barycenters



Rank-1 Wasserstein Barycenters





Rank-1 barycenter uu^*: $u = \sum_i \rho_i \varepsilon_i u_i$

$\longrightarrow$ Find suitable $(\varepsilon_i)_{i \in I} \in \{+1, -1\}^{|I|}$ sucht that

$$\forall\, i, \; \sum_j \rho_j \langle \varepsilon_i u_i, \, \varepsilon_j u_j \rangle \geqslant 0$$

Spot Noise Barycenters

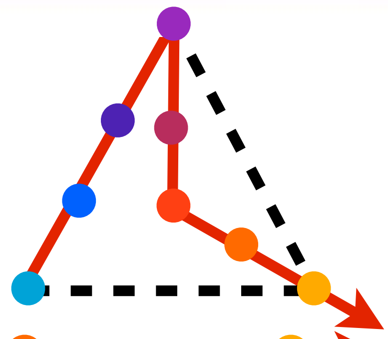
Barycenter $\Sigma^\star$: $\quad \hat{\Sigma}^\star(\omega) = \Phi(\hat{\Sigma}^\star(\omega))$

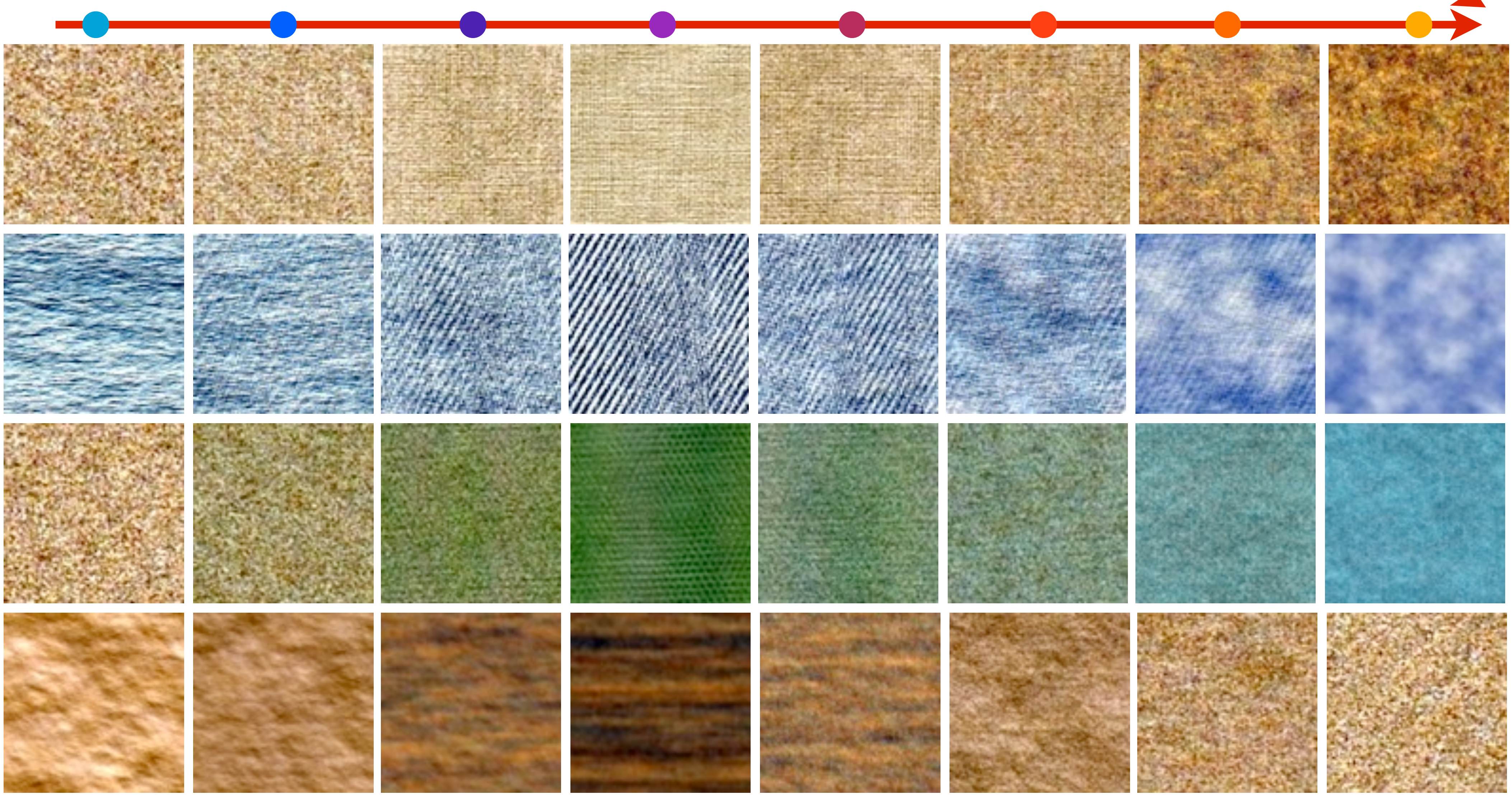
$$\Phi_\omega(\Sigma) = \sum_i \rho_i \left(\Sigma^{1/2} \hat{f}^{[i]}(\omega) \hat{f}^{[i]}(\omega)^* \Sigma^{1/2} \right)^{1/2}$$

Spot Noise Barycenters

$$\text{Barycenter } \Sigma^\star: \quad \hat{\Sigma}^\star(\omega) = \Phi(\hat{\Sigma}^\star(\omega))$$

$$\Phi_\omega(\Sigma) = \sum_i \rho_i \left(\Sigma^{1/2} \hat{f}^{[i]}(\omega) \hat{f}^{[i]}(\omega)^* \Sigma^{1/2} \right)^{1/2}$$

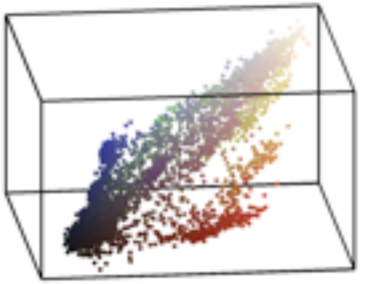




Conclusion

Image modeling with statistical constraints

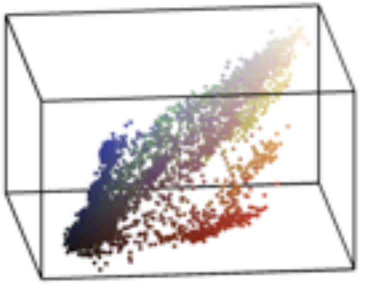
$\longrightarrow$ Colorization, synthesis, mixing, ...



Conclusion

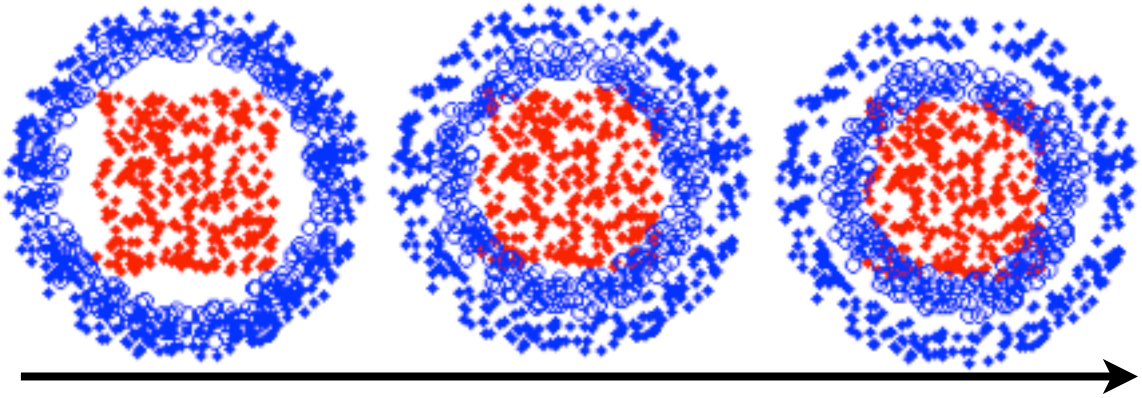
Image modeling with statistical constraints

$\longrightarrow$ Colorization, synthesis, mixing, ...



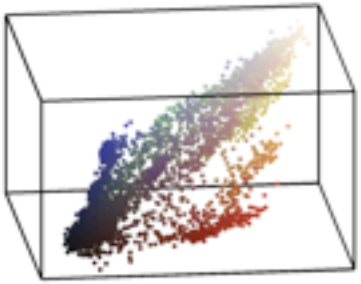
Wasserstein distance approach

$\longrightarrow$ Fast sliced approximation.

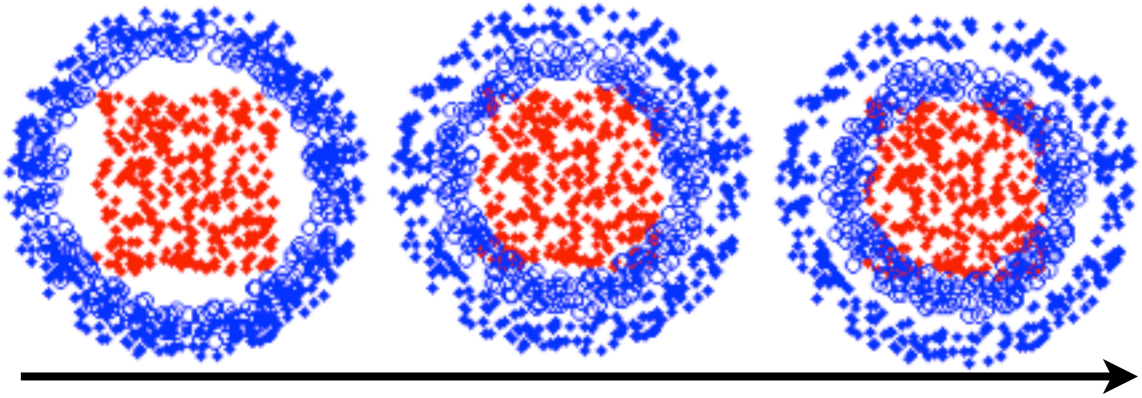


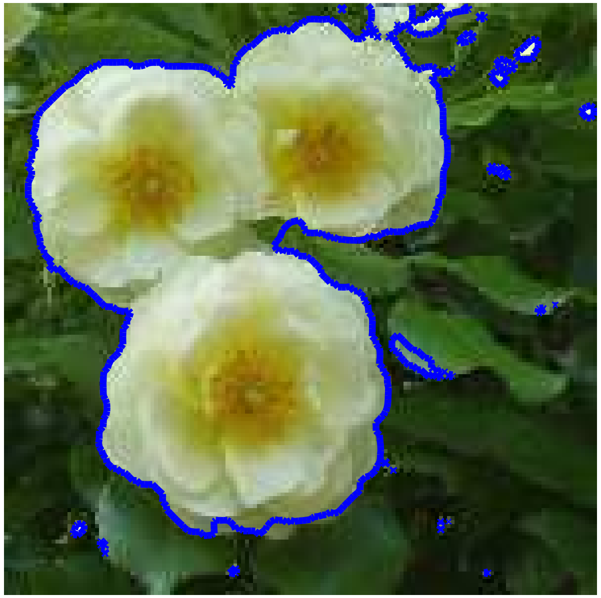
Conclusion

Image modeling with statistical constraints
$\longrightarrow$ Colorization, synthesis, mixing, ...



Wasserstein distance approach
$\longrightarrow$ Fast sliced approximation.





Extension to a wide range
of imaging problems.
$\longrightarrow$ Color transfert, segmentation, ...