A Multi-layer Approximation of Semi-discrete Optimal Transport

Arthur Leclaire
IMB, CNRS UMR 5251

Joint work with J. Rabin (Ensicaen)

SSVM
Hofgeismar
Wednesday, July 3rd, 2019

Semi-discrete Optimal Transport

Let us consider two probability measures on $X, Y \subset \mathbb{R}^D$

- $\mu(dx) = \rho(x)dx$ **absolutely continuous** measure on X
- $\nu = \sum_{j=1}^{J} \nu_j \delta_{y_j}$ **discrete** probability measure on $Y = \{y_j,\ 1 \leqslant j \leqslant J\}$

We consider the following **semi-discrete optimal transport** problem

$$\inf \int_X \|x - T(x)\|^2 d\mu(x) \qquad \text{(SD-OT)}$$

where inf is taken over all measurable maps $T : X \to Y$ such that $\nu = T_\sharp \mu$.

Recall the definition of the push-forward measure

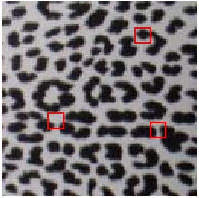
$$\forall A \in \mathcal{B}(\mathbb{R}^D), \quad T_\sharp \mu(A) = \mu(T^{-1}(A)).$$

Optimal Transport for Imaging

- Image matching [Rabin et al., 2009]
- Color transfer [Rabin et al., 2011], [Bonneel et al., 2015]
- Image denoising [Lellmann et al., 2014]
- Image segmentation [Papadakis et al., 2015], [Schmitzer, Schnörr, 2015]
- Image transport [Fitschen et al. 2015], [Maas et al., 2015]
- Shape interpolation/registration [Solomon et al., 2015], [Feydy et al., 2017]
- Reflectance function (BRDF) interpolation [Solomon et al., 2015]
- Texture synthesis and mixing [Xia et al., 2014], [Tartavel et al., 2016]

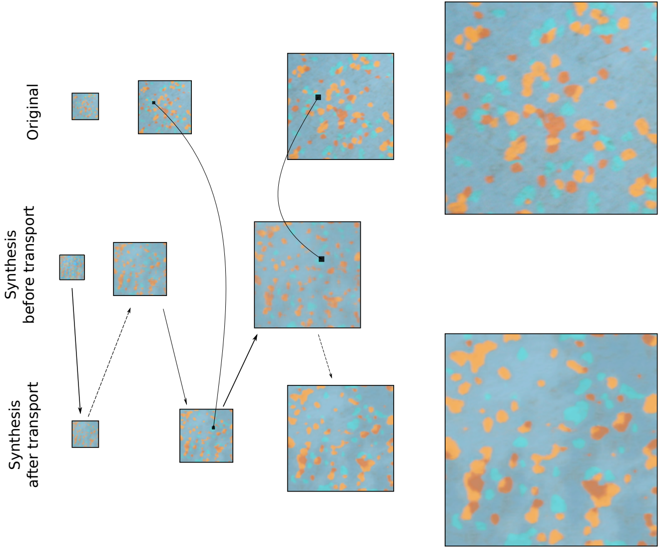
$\longrightarrow$ Here we will use optimal transport to **compare probability distributions of patches**

- Discrete [Gutierrez et al., 2017]
- Semi-discrete [Galerne et al., 2018]

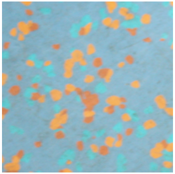


Patches 11×11

Patch Optimal Transport for Texture Synthesis

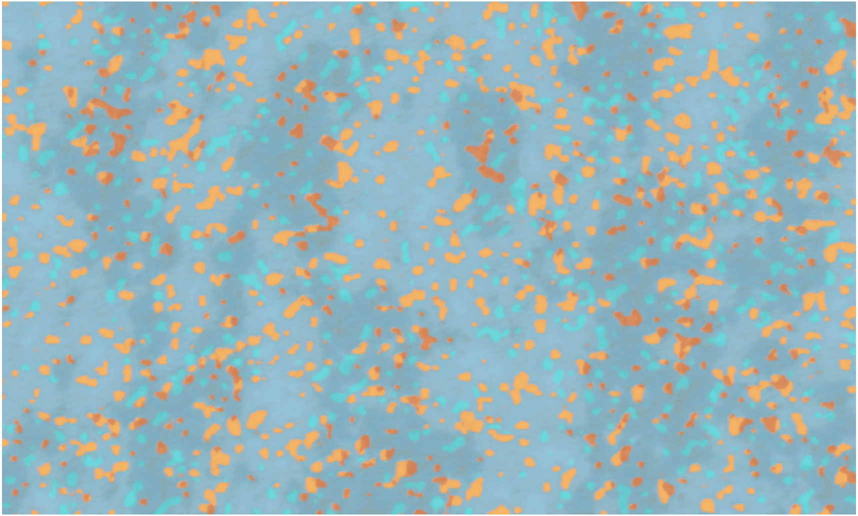


Synthesis with 3 $\times$ 3 patches



Original 256 $\times$ 256

Synthesis with 3 $\times$ 3 patches

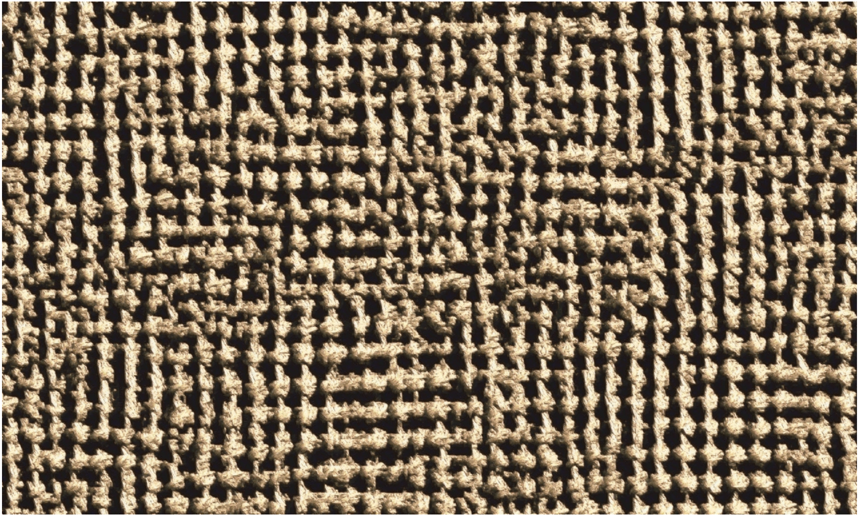


Synthesis 1280 $\times$ 768 (4 scales, $\approx$ 1 s)

Synthesis with 3 × 3 patches

Original 192 × 192

Synthesis with 3×3 patches



Synthesis 1280×768 (4 scales, $\approx$ 1 s)

Outline

Semi-discrete Optimal Transport

Multi-Layer Approximation

Results

Power cells

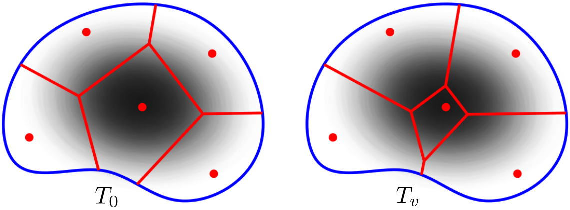
To solve this problem, for $v \in \mathbb{R}^J$ we consider the mapping

$$T_{Y,v}(x) = \underset{y_j}{\text{Argmin}} \, \|x - y_j\|^2 - v(j)$$

NB: When $v = 0 \rightarrow$ true nearest-neighbor (NN).

This mapping corresponds to a “power diagram” (or “Laguerre diagram”)

$$\text{Pow}_v(y_j) = \{ \, x \in \mathbb{R}^D \mid \forall k \neq j, \; \|x - y_j\|^2 - v(j) < \|x - y_k\|^2 - v(k) \, \}.$$



[Credits Kitagawa et al. 2017]

Dual Problem

The following theorem is due to [Aurenhammer, Hoffmann, Aronov, 1998]. See also [Kitagawa, Mérigot, Thibert, 2017].

Theorem

A solution to (SD-OT) is given by $T_{Y,v}$ where v maximizes the $\mathcal{C}^1$ concave function

$$H(v) = \int_{\mathbb{R}^D} \big(\min_j \|x - y_j\|^2 - v(j) \big) \, d\mu(x) + \sum_j \nu_j v(j),$$

whose gradient is given by $\dfrac{\partial H}{\partial v(j)} = -\mu(\mathrm{Pow}_v(y_j)) + \nu_j$.

NB: H is not strictly concave.

Corollary

The following statements are equivalent

- *v is a global maximizer of H*
- *$T_{Y,v}$ is an optimal transport map between μ and ν*
- $(T_{Y,v})_\sharp \mu = \nu$

Related works

NUMERICAL SCHEMES FOR SEMI-DISCRETE OPTIMAL TRANSPORT

- Least-squares assignment [Aurenhammer, Hoffmann, Aronov, 1998]
- Numerical solution based on L-BFGS [Mérigot, 2011], [Lévy, 2015]
- Iterative scheme to get an ε-approximate solution [Kitagawa, 2014]
- Stochastic gradient descent [Genevay, Cuturi, Peyré, Bach, 2016]
  Convergence analysis in [Bercu, Bigot, 2019]
- Damped Newton algorithm
  [Kitagawa, Mérigot, Thibert, 2017] [Mérigot, Meyron, Thibert, 2017]

MULTI-SCALE NUMERICAL SCHEMES

- Discretized [Oberman, Ruan, 2015], [Schmitzer, 2016]
- Semi-discrete [Mérigot 2011]

Stochastic Optimization

Writing $H(v) = \mathbb{E}_{X\sim\mu}[h(X, v)]$ where

$$h(x, v) = \left(\min_j \|x - y_j\|^2 - v(j)\right) + \sum_j \nu_j v(j) \ ,$$

$$\frac{\partial h}{\partial v(j)}(x, v) = -\mathbf{1}_{\mathrm{Pow}_v(y_j)}(x) + \nu_j \ .$$

We maximize with average stochastic gradient ascent [Genevay et al., 2016]:

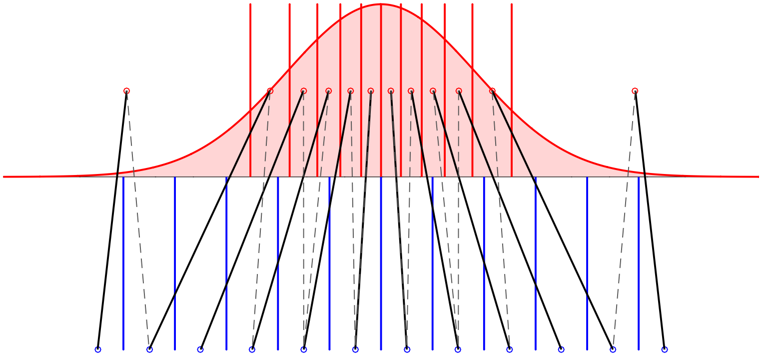
$$\begin{cases} \tilde{v}^k &= \tilde{v}^{k-1} + \gamma_k \nabla_v h(x^k, \tilde{v}^{k-1}) \quad \text{where } x^k \sim \mu \\ v^k &= \frac{1}{k}(\tilde{v}^1 + \ldots + \tilde{v}^k). \end{cases}$$

Theorem (Convergence Guarantee)

For $\gamma_k = \frac{C}{\sqrt{k}}$, *we have* $\quad \max H - \mathbb{E}[H(v^k)] = \mathcal{O}\left(\frac{\log k}{\sqrt{k}}\right)$.

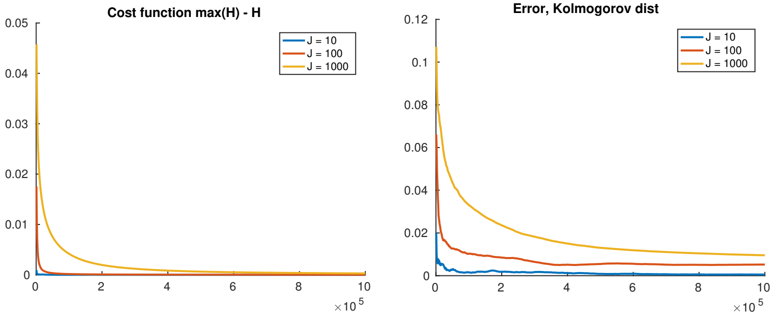
Remark: See also [Bercu, Bigot, 2019] for a.s. convergence and asymptotic normality.

An example in 1D



Power diagram : from Gaussian to discrete uniform.
Blue: True Voronoi cells (NN assignment T_0)
Red: Power cells (optimal assignment $T_{Y,v}$)

ASGD for Semi-discrete OT
Convergence in dimension 1



$\max(H) - H(v^k)$ $\quad$ $d_{\mathrm{Kol}}(T_{Y,v^k}\sharp\mu, \nu)$

Transport in 1D from Gaussian to discrete uniform on J points

Outline

Semi-discrete Optimal Transport

Multi-Layer Approximation

Results

Multi-layer decomposition of the target

As in [Mérigot, 2011], we recursively compute for $\ell = 0, 1, \ldots, L-1$

$$\nu^\ell = \sum_{j \in J^\ell} \nu_j^\ell \delta_{y_j^\ell}.$$

Initialization $\nu^0 = \nu$
For $\ell = 0, \ldots, L-2$,

1. Compute clusters (C_j^ℓ) (k-means) :

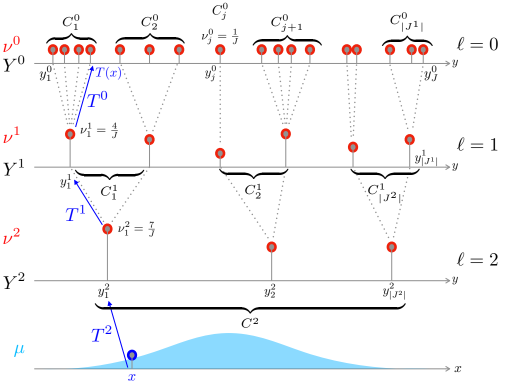
$$Y^\ell = \bigsqcup_{j \in J^{\ell+1}} C_j^\ell$$

   and gather centroids

$$Y^{\ell+1} = \{y_j^{\ell+1}, j \in J^{\ell+1}\}$$

2. Gather the weights

$$\forall j \in J^{\ell+1},\ \nu_j^{\ell+1} = \nu^\ell(C_j^\ell)$$



Multi-layer map

We will simultaneously estimate transport maps

$$T^\ell(x) = T_{C_j^\ell, v_j^\ell}(x) \quad \text{defined on } \mathcal{L}_j^{\ell+1} \quad \text{with parameters } v_j^\ell$$

for all scales $\ell = 0, \ldots, L-1$ and all clusters C_j^ℓ.

Computing $T(x)$ amounts to **tracing back a hierarchy of power cells**:
for $\ell = L-1, \ldots, 0$, if $T^{\ell+1}(x) = y^{\ell+1}_{j^{\ell+1}}$ then

$$j^\ell(x) = \underset{j | y_j^\ell \in C^\ell_{j^{\ell+1}}}{\text{Argmin}} \|x - y_j^\ell\|^2 - v^\ell_{j^{\ell+1}(x)}(j)$$

$$T^\ell(x) = y^\ell_{j^\ell(x)}.$$

Eventually, we set $T(x) = T^0(x)$.

Illustration - 1st layer

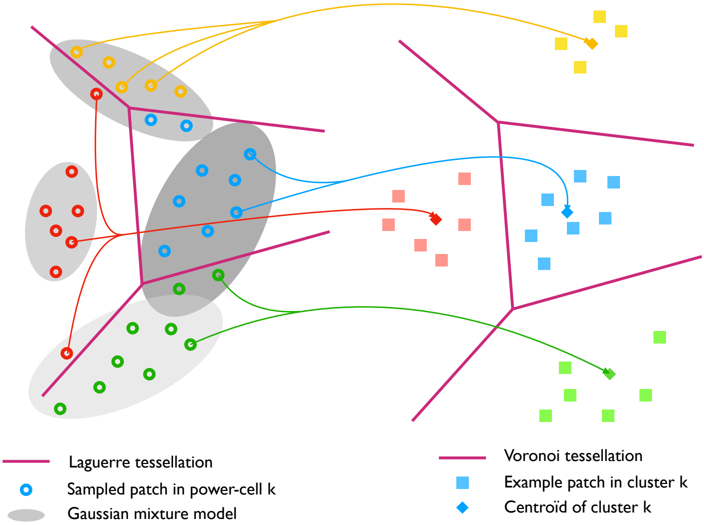
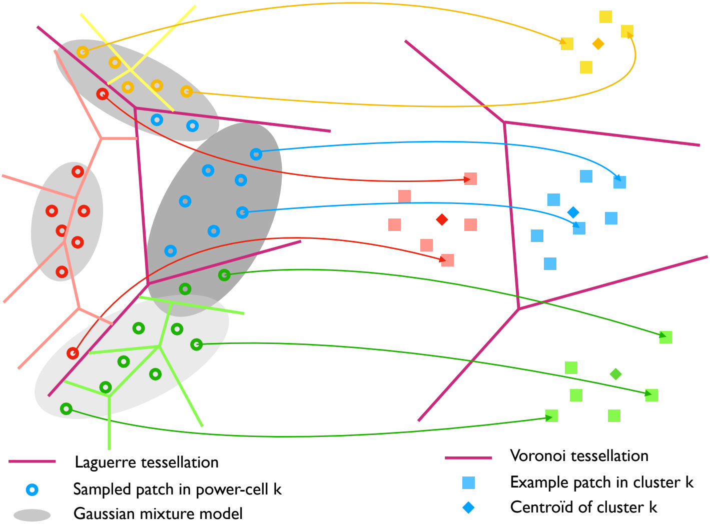


Illustration - 2nd layer



Estimation of the Multi-layer Map

For all $\ell \in \{0, \dots, L-1\}$ and $j \in J^{\ell+1}$, $T_{C_j^\ell, v_j^\ell}$ should realize the OT between

- the restriction of μ to the Laguerre cell $\mathcal{L}_j^{\ell+1}$
- the restriction of ν^ℓ to C_j^ℓ.

At the current scale, we seek to maximize

$$H^\ell(v^\ell) = \sum_{j^{\ell+1} \in J^{\ell+1}} \int_{\mathcal{L}_{j^{\ell+1}}^{\ell+1}} \min_{\substack{j \\ y_j^\ell \in C_{j^{\ell+1}}^\ell}} \left(\|x - y_j^\ell\|^2 - v_{j^{\ell+1}}^\ell(j) \right) d\mu(x) + \sum_{\substack{j \\ y_j^\ell \in C_{j^{\ell+1}}^\ell}} v_{j^{\ell+1}}^\ell(j) \nu_j^\ell .$$

We simultaneously optimize all the H^ℓ's with stochastic gradient descent.

NB: The global problem is not convex anymore since $\mathcal{L}_{j^{\ell+1}}^{\ell+1}$ depends on the weights at the previous scale.

Estimating Multi-layer Transport

Algorithm : Computing multi-layer transport map

- Multi-layer decomposition $\{\nu^\ell, \ell = 0, \ldots, L-1\}$ of ν
- Set $v_j^\ell \leftarrow 0, \forall \ell, j$ (weights initialization)
- Set $n_j^\ell \leftarrow 0, \forall \ell, j$ (number of visits in cluster C_j^ℓ)
- For $t = 1, \ldots, T$,

  $\rightarrow$ Draw a sample $x \sim \mu$

  $\rightarrow$ For $\ell = L-1, \ldots, 0$

  Compute the corresponding cluster index: $j \leftarrow j^\ell(x)$

  $n_j^\ell \leftarrow n_j^\ell + 1$

  $g \leftarrow \nabla_v h_{C_j^\ell, \nu_j^\ell}(x, \tilde{v}_j^\ell)$

  $\tilde{v}_j^\ell \leftarrow \tilde{v}_j^\ell + \frac{C}{\sqrt{n_j^\ell}} g$

  $v_j^\ell \leftarrow v_j^\ell + \frac{1}{n_j^\ell} \left(\tilde{v}_j^\ell - v_j^\ell\right)$

Outline

Semi-discrete Optimal Transport

Multi-Layer Approximation

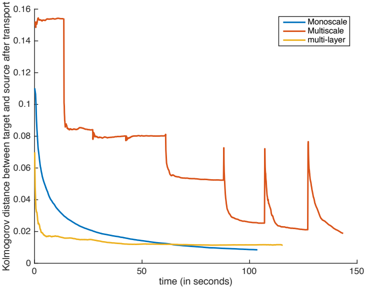
Results

1D case : Gaussian to uniform

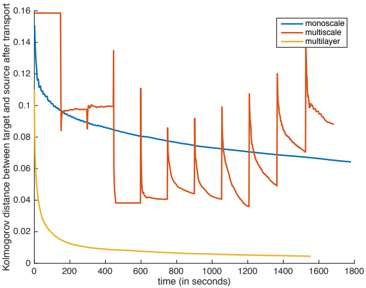
We approximate the optimal transport between

- μ : normalized Gauss distribution $\mathcal{N}(0, 1)$
- ν : discrete uniform on J equally spaced points in $(-1, 1)$.

We compare with naive multiscale algorithm (in red)



$J = 1000$ points



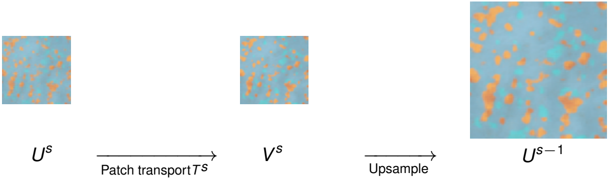
$J = 10\,000$ points

Application to Texture Synthesis

- Model based on optimal transport in the space of $w \times w$ patches
- Start from a Gaussian model and apply OT maps to reimpose statistics of the input texture at several resolutions
- μ : current (continuous) patch distribution at resolution s
- ν : exemplar empirical patch distribution at resolution s
  ... actually, a subsampled version

$$\nu = \frac{1}{J} \sum_{j=1}^{J} \delta_{p_j}$$

where $p_1, \ldots, p_J$ are randomly chosen patches in the exemplar image.

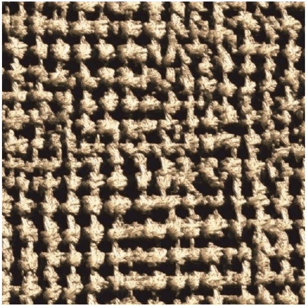


Improvement with Multi-layer Transport

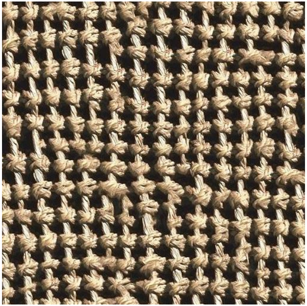
- Problem: the patch space is high-dimensional
  $\longrightarrow$ for color 3×3 patches : $D = 27$
  $\longrightarrow$ for color 7×7 patches : $D = 147$
- Need for a sufficient subsampling of the target empirical measure.
- Work at four resolutions with adapted subsampling strategy:
  $J = 1000, 2000, 4000, 16000$
- 2-layer transport allows to deal with such target measures by using
  $J^1 = 10, 10, 20, 40$ clusters
- Technical point: enhance details at the last scale with one additional transport on 3×3 patches

Visual Results

Original

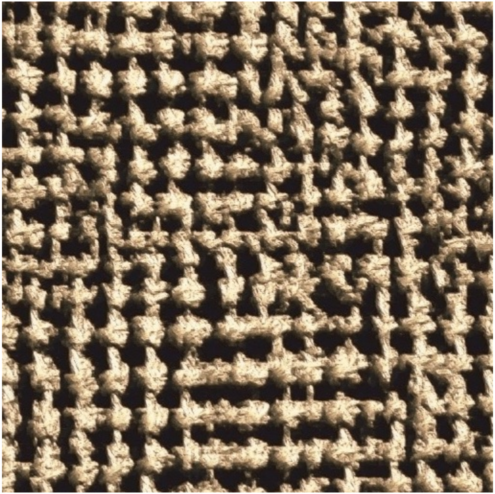


1-layer OT
3×3 patches



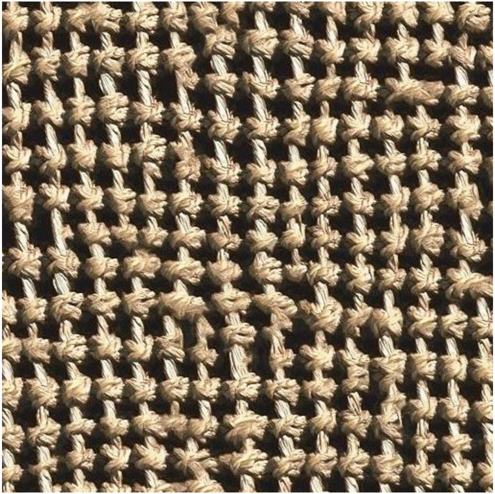
2-layer OT
7×7 patches

Visual Results



1-layer OT
3×3 patches

Visual Results

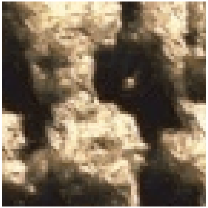


2-layer OT
7×7 patches

Visual Results

Exemplar

2-layer (7×7)



1-layer (3×3)

1-layer (7×7)

Visual Results

Original

1-layer OT
3×3 patches

2-layer OT
7×7 patches

Visual Results

Original

1-layer OT
3×3 patches

2-layer OT
7×7 patches

Visual Results

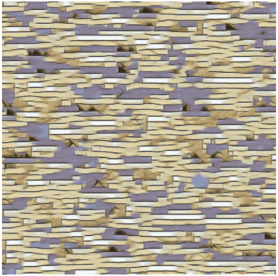
Original

1-layer OT
3×3 patches

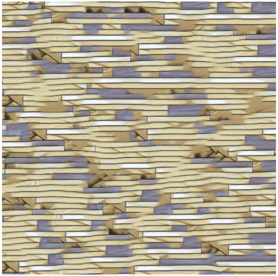
2-layer OT
7×7 patches

Visual Results

Original



1-layer OT
3×3 patches



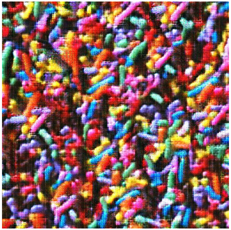
2-layer OT
7×7 patches

Comparison

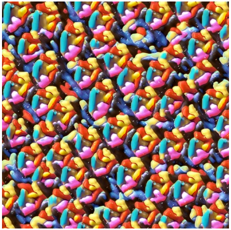

Original


2-layer OT (7×7)

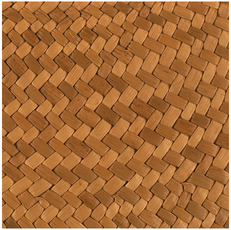

1-layer OT (3×3)


[Jetchev et al.]
Spatial GAN

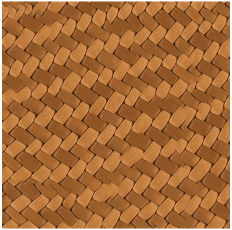

[Gatys et al.]


[Ulyanov et al., 2017]
Texture Networks

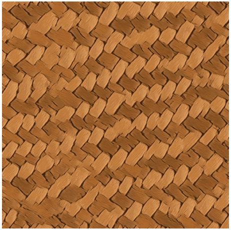
Comparison



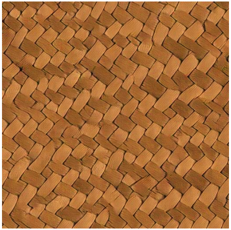
Original



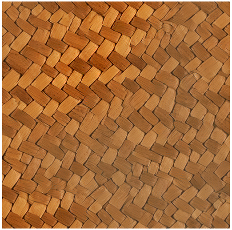
2-layer OT (7×7)



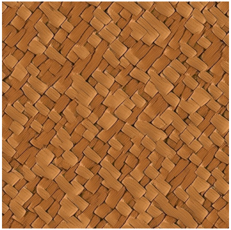
1-layer OT (3×3)



[Jetchev et al.]
Spatial GAN



[Gatys et al.]



[Ulyanov et al., 2017]
Texture Networks

Conclusion

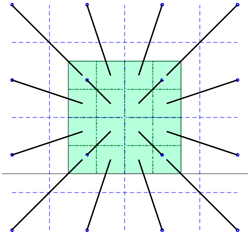
- We proposed a multi-layer approximation of semi-discrete OT
- It improves the *practical* convergence speed of the stochastic algorithm
- ... but suffers from a bias which does not vanish
- No proof of convergence for now
- Helps to improve a texture synthesis model
  (better visual results, and faster!)
- For texture synthesis, why don't we optimize all the model?...
  $\rightarrow$ Wasserstein GAN?

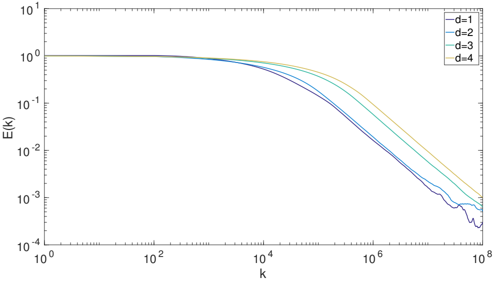
SOURCE CODES are available at

`www.math.u-bordeaux.fr/~aleclair/texto/multilayer.php`

THANKS FOR YOUR ATTENTION

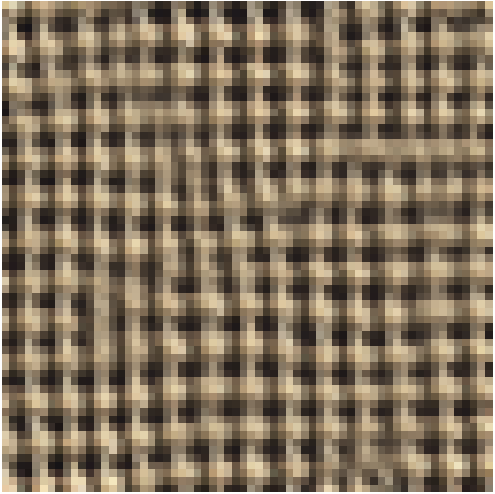
ASGD for Semi-discrete OT
Convergence in dimension > 1





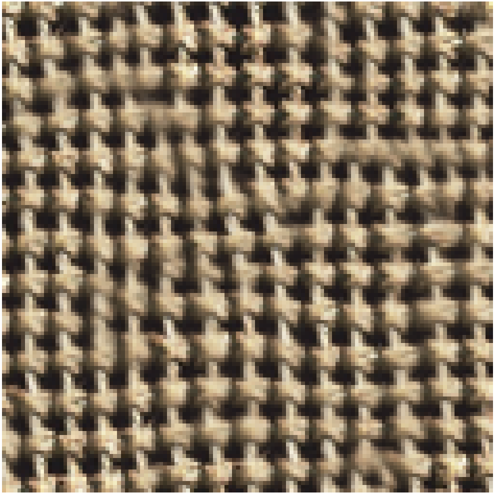
$$E(k) = \frac{\|v^k - v^\star\|}{\|v^\star\|}$$

From Coarse to Fine Resolution



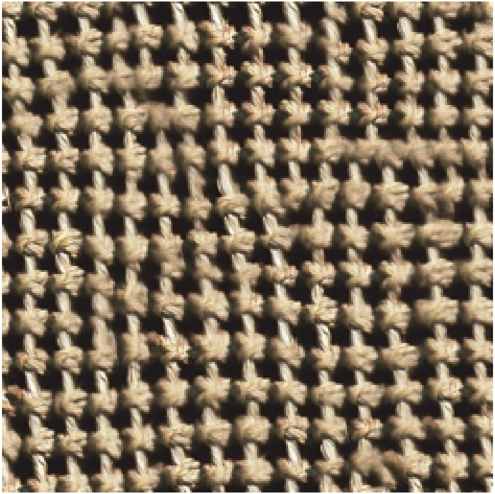
Scale 3 (7×7)

From Coarse to Fine Resolution



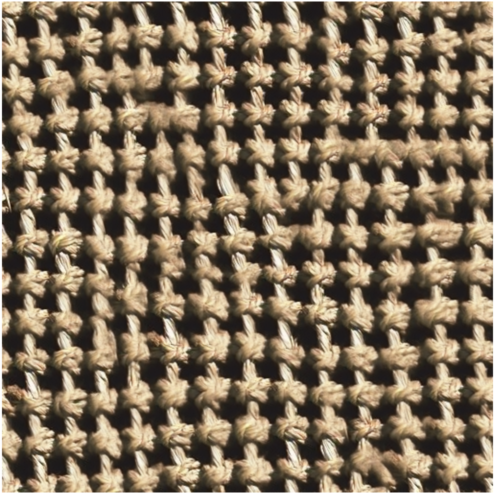
Scale 2 (7×7)

From Coarse to Fine Resolution



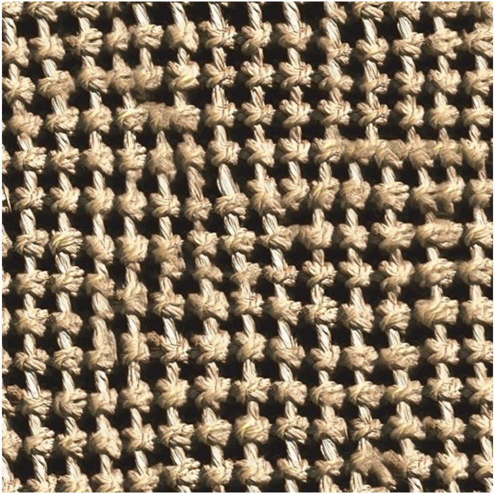
Scale 1 (7×7)

From Coarse to Fine Resolution



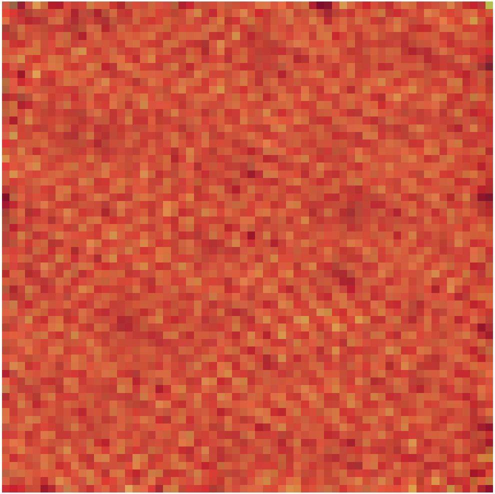
Scale 0 (7×7)

From Coarse to Fine Resolution



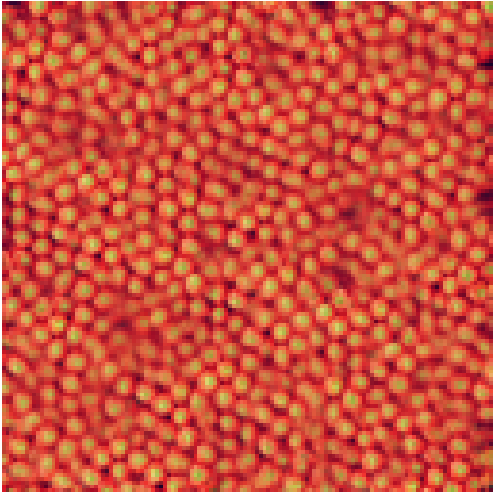
Scale 0 (3×3)

From Coarse to Fine Resolution



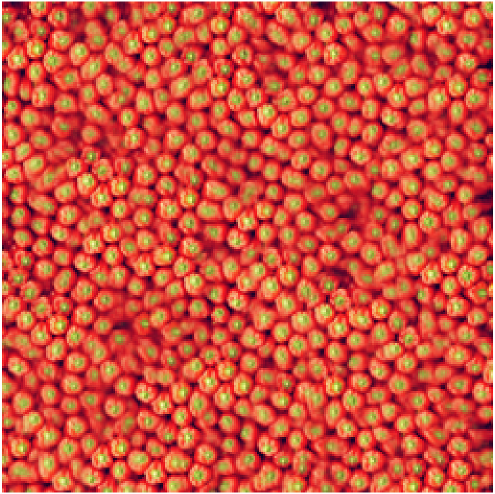
Scale 3 (7×7)

From Coarse to Fine Resolution



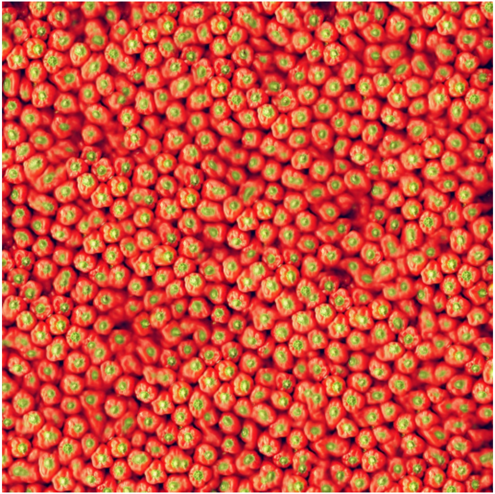
Scale 2 (7×7)

From Coarse to Fine Resolution



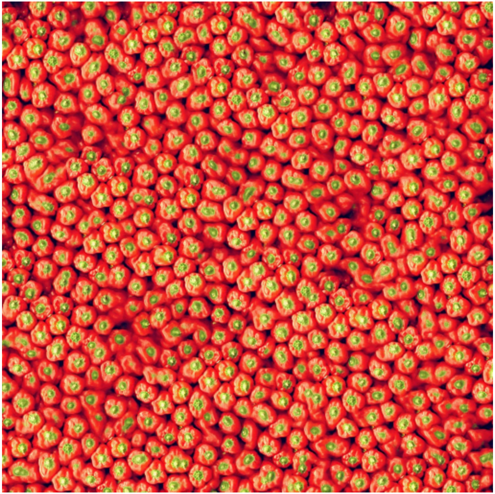
Scale 1 (7×7)

From Coarse to Fine Resolution



Scale 0 (7×7)

From Coarse to Fine Resolution



Scale 0 (3×3)

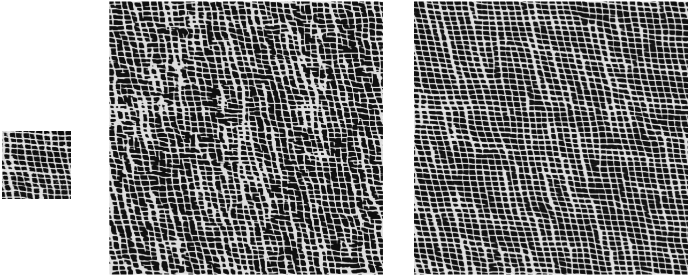
Visual Results

Original

1-layer OT

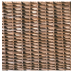
2-layer OT

Visual Results

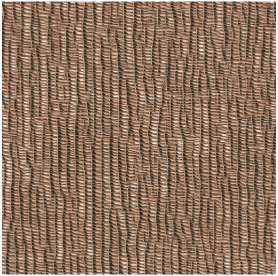


Original 1-layer OT 2-layer OT

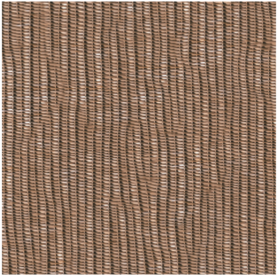
Visual Results



Original



1-layer OT



2-layer OT

Visual Results

Original

1-layer OT

2-layer OT

Visual Results

Original

1-layer OT

2-layer OT

Visual Results

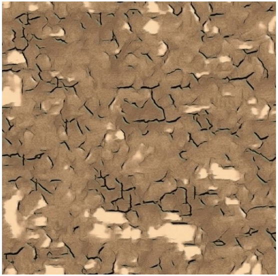
Original

1-layer OT

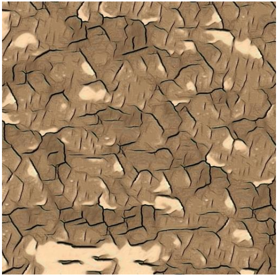
2-layer OT

Visual Results

Original

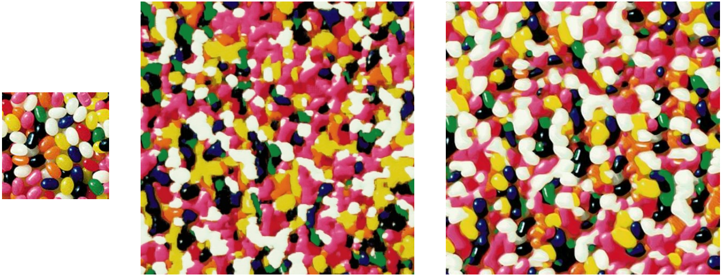


1-layer OT



2-layer OT

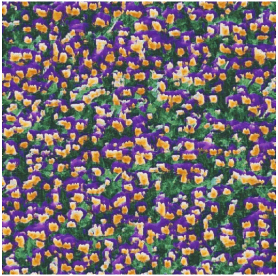
Visual Results



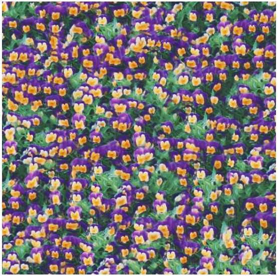
Original 1-layer OT 2-layer OT

Visual Results

Original



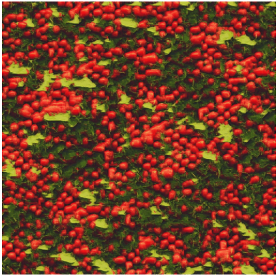
1-layer OT



2-layer OT

Visual Results

Original

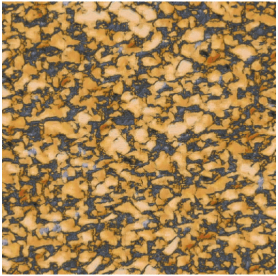


1-layer OT

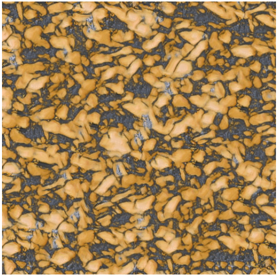
2-layer OT

Visual Results

Original



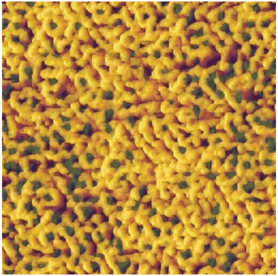
1-layer OT



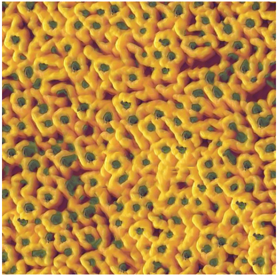
2-layer OT

Visual Results

Original



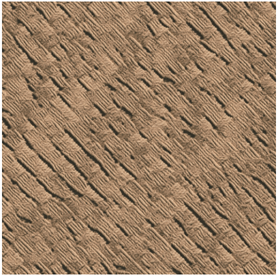
1-layer OT



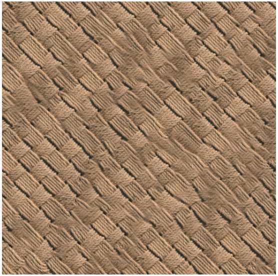
2-layer OT

Visual Results

Original

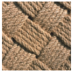


1-layer OT

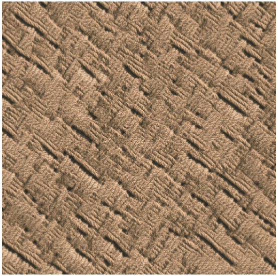


2-layer OT

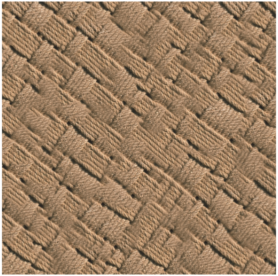
Visual Results



Original



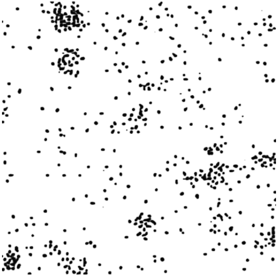
1-layer OT



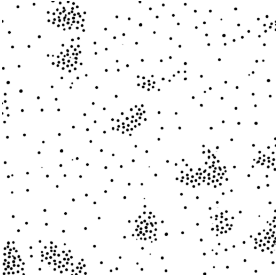
2-layer OT

Visual Results

Original



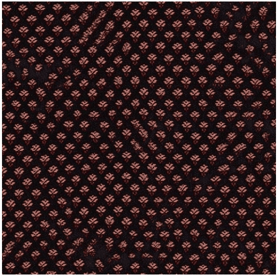
1-layer OT



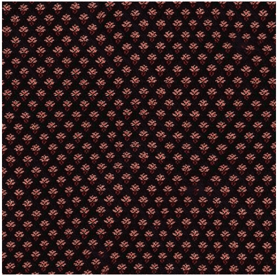
2-layer OT

Visual Results

Original



1-layer OT



2-layer OT