University of Heidelberg
Department of Mathematics and Computer Science
Image & Pattern Analysis Group

Master-Thesis

Semi-discrete Optimal Transport

Name:	Daniel Gonzalez
Enrolment number:	3112012
Supervisor:	Prof. Christoph Schnörr
Date:	23. Februar 2020

Ich versichere, dass ich diese Master-Arbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe und die Grundsätze und Empfehlungen "Verantwortung in der Wissenschaft" der Universität Heidelberg beachtet wurden.

Abgabedatum: 23. Februar 2020

Zusammenfassung

Die Zusammenfassung muss auf Deutsch **und** auf Englisch geschrieben werden. Die Zusammenfassung sollte zwischen einer halben und einer ganzen Seite lang sein. Sie soll den Kontext der Arbeit, die Problemstellung, die Zielsetzung und die entwickelten Methoden sowie Erkenntnisse bzw. Ergebnisse bersichtlich und verstndlich beschreiben.

Abstract

The abstract has to be given in German **and** English. It should be between half a page and one page in length. It should cover in a readable and comprehensive style the context of the thesis, the problem setting, the objectives, and the methods developed in this thesis as well as key insights and results.

Table of Contents

Probably not necessary - Introduce nomenclature and conventions in the text

Nomenclature and conventions

$\|\cdot\|$	Euclidean Norm on $\mathbb{R}^d$
$\langle\cdot,\cdot\rangle$	Euclideani Inner product on $\mathbb{R}^d$
$\mathbb{R}_{\geq 0}$	Non-negative real numbers
λ^d	Lebesgue measure on $\mathbb{R}^d$
$\mathcal{B}^d$	Borel σ-algebra on $\mathbb{R}^d$
$\mathrm{Pot}(\cdot)$	Power set of a set
$\vert\cdot\vert$	Cardinality of a set
$\mathcal{L}(\mu)$	μ-integrable functions with values in $\mathbb{R}$

TODO: Reminders from measure theory. Create an introduction and motivation.

Let $(X, \mathcal{A})$ be a measurable space and $\mu, \tilde{\mu}$ measures on X. We say that μ is *absolutely continuous* with respect to $\tilde{\mu}$, if

$$\tilde{\mu}(A) = 0 \Rightarrow \mu(A) = 0 \qquad \text{for all } A \in \mathcal{A}$$

and denote this by $\mu \ll \tilde{\mu}$.
Let $(X, \mathcal{A}, \mu)$ be a measure space and $(Y, \mathcal{U})$ a measurable space. For a measurable function $T : X \to Y$, we denote by $T_{\#}\mu$ the *pushforward measure* on Y induced by T, i.e. the measure on Y given by

$$T_{\#}\mu(B) = \mu(T^{-1}(B)) \qquad \text{for all } B \in \mathcal{U}\,.$$

TODO: Introduction for Optimal Transport

1 Introduction

1.1 Overview/Motivation

1.2 Related Work

- Je nach dem, wie relevant die andere Dokumente sein könnten, relevante Arbeiten citieren.
- Basic reference, e.g Vill, conference works

1.3 Contribution/Organization of work

- Description of the approach.
- Distribution of this work, briefly description of the chapters.

1.4 Background

- Change of variable formulas, etc.

2 Semi-dicrete Optimal Transport

TODO: Introduction Optimal Transport

2.1 Problem

TODO: Introduction

Problem von SDOT im Allgemein

2.2 Approach

TODO: Introduction

2.3 Characterization of the Optimal Transport Map

Adapt to the above sections

For a given finite set $S \subset \mathbb{R}^d$, we want to study the following problem:
Given two probability spaces $(\mathbb{R}^d, \mathcal{B}^d, \mu)$ and $(S, \mathrm{Pot}(S), \nu)$, we want to minimize

$$\int_{\mathbb{R}^d} \|x - T(x)\|^2 \, \mathrm{d}\mu \tag{2.1}$$

over all measurable maps $T : \mathbb{R}^d \to S$ satisfying $T_{\#}\mu = \nu$.
Recall that as we are working with a finite probability space $(S, \mathrm{Pot}(S), \nu)$, we can write ν as a finite sum of Dirac measures

$$\nu = \sum_{s \in S} \nu_s \delta_s \quad \text{where } \nu_s \in \mathbb{R}_{\geq 0}, \text{ such that } \quad \sum_{s \in S} \nu_s = 1. \tag{2.2}$$

Thus, the problem we are considering translates to find a measurable map T that minimizes the functional in (2.1) and such that $\mu(T^{-1}(s)) \coloneqq \mu(T^{-1}(\{s\})) = \nu_s$ for all $s \in S$. For any $s \in S$, we call $\mu(T^{-1}(s))$ the *capacity* of s.

TODO: cite/mention existence of solution

Following the exposition in [Aurenhammer et al., 1998], we will now show that finding a minimizer of the functional (2.1) is equivalent to finding the maximum of a concave function and thus, can be computed with standard optimization methods. The formulation of this optimization problem is done in two steps. First, we find a minimizer over all measurable functions with prescribed equal capacities. This optimal solution T_W will be inspired geometrically and constructed using a predefined *weight vector* W. The next step will consist in adapting this weight vector such that the condition $T_{W\#}\mu = \nu$ is fullfilled.
The motivation behind this approach is inspired geometrically by studying a generalization of *Voroni diagrams*. For clarity, we recall the definition of these diagrams and review the necessary concepts needed for this approach.

Definition 2.1 (Voronoi Diagrams)**.** Let $S \subset \mathbb{R}^d$ be a finite set. We define for every point $s \in S$

$$\mathrm{reg}(s) \coloneqq \{x \in \mathbb{R}^d : \|x - s\| \leq \|x - \tilde{s}\| \quad \text{for all } \tilde{s} \in S \setminus \{s\}\}.$$

We call this the *region* or *cell* of the point s. The partition of $\mathbb{R}^d$ created by the union of the regions of all points is called the Voronoi diagram of S.

Remark. Note that the partition of $\mathbb{R}^d$ generated by the Voronoi diagram of $S \subset \mathbb{R}^d$, consists of convex regions. Such a partition induces naturally a map $T : \mathbb{R}^d \to S$, which assign each point in $\mathbb{R}^d$ the corresponding point in S of the cell where it is located, i.e

$$T(x) = s \quad \Leftrightarrow \quad x \in \operatorname{reg}(s). \tag{2.3}$$

By definition, some points in $\mathbb{R}^d$ may belong to more than one region. By convention, T assigns those points an arbitrary one in S of a region where it is located. We call T, the by the Voronoi diagram *induced assigment*.

A generalization of the presented concepts arise when using another distance function for the definition of the regions. One application of this amounts to using the *power function* with weights W, which we now define.

Definition 2.2 (Power function)**.** Let $S \subset \mathbb{R}^d$ be a finite set and $W : S \to \mathbb{R}$ a function on S. The power function with weights W is defined as

$$\operatorname{pow}_W(x, s) \coloneqq \|x - s\|^2 - W(s).$$

We call W *weight function* on S or simply *weights* of S.

Remark. For simplicity of notation we will write sometimes the weight function
$W : S \to \mathbb{R}^d$ as a vector in $\mathbb{R}^{|S|}$. We call then W *weight vector* of S.

As in the case of Voronoi diagrams, we can define regions on $\mathbb{R}^d$ by using the power function with weights W. For a point $s \in S$ we call

$$\operatorname{reg}_W(s) \coloneqq \{x \in \mathbb{R}^d : \operatorname{pow}_W(x, s) \leq \operatorname{pow}_W(x, \tilde{s}) \quad \text{for all } \tilde{s} \in S \setminus \{s\}\}$$

the *power region* (or *power cell*) of s with weights W. Power regions also create a partition of $\mathbb{R}^d$ which is called the *power diagram* of S with weights W.
The geometric intution behind this definition becomes clear by looking at spheres around $s \in S$ with positive radius $\sqrt{W(s)}$

$$\mathbb{S}^{d-1}_{\sqrt{W}}(s) \coloneqq \{x \in \mathbb{R}^d : \|x - s\| = \sqrt{W(s)}\}.$$

For a fixed $s \in S$, the power function $\operatorname{pow}(\cdot, s)$ returns a negative (resp. positive) value, whenever $x \in \mathbb{R}^d$ is inside (resp. outside) the sphere $\mathbb{S}^{d-1}_{\sqrt{W}}(s)$ and zero when s lies on the sphere. Thus, increasing (resp. decreasing) the values of the weights $W(s)$ on each point s would expand (resp. schrink) the power cells.

Remark. Unlike Voronoi diagrams, the power cells of a point $s \in S$ may not contain the point s or may even be empty. Nevertheless, the power diagram still partitions $\mathbb{R}^d$ in convex polyhedron.

A small plot of this case

By replacing reg with reg_W in (2.3) we obtain a map $T_W : \mathbb{R}^d \to S$ depending on the weight vector W. Similarly as with Voronoi diagrams, we assign those points sharing different cells, an arbitrarly point $s \in S$ of those shared regions. We call this map, the *power assignment* of S with weights W.
Power assignments have a natural optimization property, since by definition it holds

$$(T_W(x) = s \;\Leftrightarrow\; x \in \text{reg}_W(s)) \quad \Leftrightarrow \quad T_W(x) = \arg\min_{s\in S} \|x - s\|^2 - W(s) \tag{2.4}$$

for all points $x \in \mathbb{R}^d$ which don't share different regions. In fact, power assignments even minimize the functional (2.1) for prescribed capacities. We will prove this in Lemma 2.4. As a consequence, the natural question which remains to be clarified is how to choose the weight vector W, such that the condition $T_{W\#}\mu = \nu$ is fullfilled.

For the proof of Lemma 2.4 we recall the change of variables theorem from measure theory.

Necessary? Cite it?

Theorem (change of variables)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space, $(Y, \mathcal{U})$ a measurable space and $T : X \to Y$ a measurable function.*
For a measurable function $f : Y \to \mathbb{R}^d$ the following are equivalent

(i) $f \in \mathcal{L}(T_\#\mu)$

(ii) $f \circ T \in \mathcal{L}(\mu)$

In case any of these statements is true we have also

$$\int_{T^{-1}(B)} f \; \mathrm{d}T_\#\mu = \int_B f \circ T \; \mathrm{d}\mu \quad \textit{for all } B \in \mathcal{U}\,. \tag{2.5}$$

Proof. Measure theory, e.g p.191 [J.E] □

Remark 2.3. Note that as the power region with weights W of a point $s \in S \subset \mathbb{R}^d$ is either an empty set or a convex polyhedra, it is measurable with respect to the Lebesgue measure λ^d on $\mathbb{R}^d$. Denoting by $\text{int}(B)$ the interior of a set $B \in \mathcal{B}$ with respect to the standard topology, we get

$$\lambda^d(\text{reg}_W(s)) = \lambda^d(\text{int}(\text{reg}_W(s))).$$

For a probability space $(\mathbb{R}^d, \mathcal{B}^d, \mu)$ such that $\mu \ll \lambda^d$, we have then

$$\mu(\mathrm{reg}_W(s)) = \mu(\mathrm{int}(\mathrm{reg}_W(s))) \qquad \text{and} \qquad \sum_{s\in S} \mu(\mathrm{reg}_W(s)) = 1.$$

Lemma 2.4. *Let $(\mathbb{R}^d, \mathcal{B}, \mu)$ be a probability space, such that $\mu \ll \lambda^d$. Let S be a finite subset of $\mathbb{R}^d$ with weights W and $\zeta : S \to \mathbb{R}_{\geq 0}$ be a function on S. Then, the power assignment T_W minimizes*

$$\int_{\mathbb{R}^d} \|x - T(x)\|^2 \, \mathrm{d}\mu$$

over all measurable maps $T : \mathbb{R}^d \to S$ with capacities $\mu(T^{-1}(s)) = \zeta(s)$ for all $s \in S$.

Proof. Using the minimality condition (2.4) of power assignments, we see that for a fixed $x \in \mathbb{R}^d$ holds

$$\mathrm{pow}_W(x, T_W(x)) \leq \mathrm{pow}_W(x, s) \quad \text{for all } \; s \in S.$$

Consequently, T_W minimizes

$$\int_{\mathbb{R}^d} \mathrm{pow}_W(x, T(x)) \, \mathrm{d}\mu = \int_{\mathbb{R}^d} \|x - T(x)\|^2 \, \mathrm{d}\mu - \int_{\mathbb{R}^d} W(T(x)) \, \mathrm{d}\mu$$

over all measurable maps $T : \mathbb{R}^d \to S$. Using the fact that $\mathbb{R}^d = \bigcup_{s\in S} \mathrm{reg}_W(s)$ together with remark 2.3, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} W(T(x)) \, \mathrm{d}\mu &= \sum_{s\in S} \int_{\mathrm{reg}_W(s)} W(T(x)) \, \mathrm{d}\mu \\
&\overset{(2.5)}{=} \sum_{s\in S} \int_s W \, \mathrm{d}T_{\#}\mu \\
&= \sum_{s\in S} \mu(T^{-1}(s)) W(s) \\
&= \sum_{s\in S} \zeta(s) W(s)
\end{aligned}$$

which is constant for a fixed ζ and W. □

The natural question to handle next, is how to choose W such that the condition $T_{W\#}\mu = \nu$ holds. As we will show below, this question can be equivalently formulated as finding the maximum of a concave function. In order to achive this, we recall the original setting of our original problem and introduce some definitions.

Let $(\mathbb{R}^d, \mathcal{B}, \mu)$ and $(S, \mathrm{Pot}(S), \nu)$ be two probability spaces such that $\mu \ll \lambda^d$. As in equation (2.2), we write the measure ν as a finite sum of Dirac measures $\nu = \sum_{s\in S} \nu_s \delta_s$.

For $\mathcal{F} \coloneqq \{f : \mathbb{R}^d \to S : f \text{ is measurable}\}$, define

$$L : \mathcal{F} \times \mathbb{R}^{|S|} \to \mathbb{R}, \quad (T, W) \mapsto \int_{\mathbb{R}^d} \mathrm{pow}_W(x, T(x)) \, \mathrm{d}\mu.$$

This map has important properties, which we will show below and use for formulation of the concave optimization problem. For a map $T \in \mathcal{F}$, let

$$\zeta_T : S \to \mathbb{R}, \quad s \mapsto \mu(T^{-1}(s)) \tag{2.6}$$

be the vector of capacities induced by T and

$$Q : \mathcal{F} \to \mathbb{R}, \quad T \mapsto \int_{\mathbb{R}^d} \|x - T(x)\|^2 \, \mathrm{d}\mu$$

be the functional that we want to minimize. As shown in the proof of Lemma 2.4, it holds

$$L(T, W) = Q(T) - \langle \zeta_T, W \rangle. \tag{2.7}$$

And hence, $L_T \coloneqq L(T, \cdot)$ defines a linear function on $\mathbb{R}^{|S|}$ for any fixed $T \in \mathcal{F}$.

Recall that for a given $W \in \mathbb{R}^{|S|}$ and $x \in \mathbb{R}^d$, the minimality condition for power assignments (2.4) states

$$\mathrm{pow}_W(x, T_W(x)) \leq \mathrm{pow}_W(x, s) \quad \text{for all } s \in S.$$

Consequently, for a fixed $W \in \mathbb{R}^d$ we must have

$$T_W = \arg\min_{T \in \mathcal{F}} L(T, W). \tag{2.8}$$

This two properties lead us to the definition of a smooth concave function.

Theorem 2.5. *The function*

$$F : \mathbb{R}^{|S|} \to \mathbb{R}, \quad W \mapsto L(T_W, W) = L_{T_W}(W)$$

is smooth and concave

Proof. **TODO: Argue why F is differentiable (Proof was wrong) see: [Bruno Levy]** We show now the convexity of F. Let $\alpha \in [0, 1]$ and $W_1, W_2 \in \mathbb{R}^{|S|}$, then

Probably better argumentation with envelopes

$$
\begin{aligned}
F(\alpha W_1 + (1-\alpha)W_2) &= L_{T_{\alpha W_1+(1-\alpha)W_2}}(\alpha W_1 + (1-\alpha)W_2) \\
&\overset{\text{lin}}{=} L_{T_{\alpha W_1+(1-\alpha)W_2}}(\alpha W_1) + L_{T_{\alpha W_1+(1-\alpha)W_2}}((1-\alpha)W_2) \\
&\overset{(2.8)}{\geq} L_{T_{\alpha W_1}}(\alpha W_1) + L_{T_{(1-\alpha)W_2}}((1-\alpha)W_2) \\
&\overset{\text{lin}}{=} \alpha L_{T_{\alpha W_1}}(W_1) + (1-\alpha)L_{T_{(1-\alpha)W_2}}(W_2) \\
&\overset{(2.8)}{\geq} \alpha L_{T_{W_1}}(W_1) + (1-\alpha)L_{T_{W_2}}(W_2) = \alpha F(W_1) + (1-\alpha)F(W_2)
\end{aligned}
$$

□

Consider now the function

$$
\begin{aligned}
H^\nu : \mathbb{R}^{|S|} \to \mathbb{R}, \quad W \mapsto F(W) + \langle \nu, W\rangle &= Q(T_W) - \langle \zeta_{T_W}, W\rangle + \langle \nu, W\rangle \\
&= \langle \nu - \zeta_{T_W}, W\rangle + Q(T_W).
\end{aligned}
$$

This function is concave and differentiable as a sum of concave differentiable functions. Furthermore we have

$$
\nabla H^\nu(W) = \nabla F(W) + \nu \overset{\text{Thm 2.5}}{=} -\zeta_{T_W} + \nu \tag{2.9}
$$

and hence also

$$
T_{W\#}\mu(s) = \mu(T_W^{-1}(s)) = \zeta_{T_W}(s) = \nu_s \quad \text{for all } s \in S \quad \Leftrightarrow \quad \nabla H^\nu(W) = 0.
$$

Thus, because of Lemma 2.4, we see that finding a solution of our original problem is indeed equivalent to finding a maximum of the concave function H^ν. Regarding (2.9), we have

$$
\frac{\partial H^\nu}{W(s)} = -\mu(\mathrm{reg}_W(s)) + \nu_s.
$$

TODO: motivate this part, connect with what follows

This optimization problem has a probabilistic interpretation. If we realize X as a random variable of distribution μ, i.e $X \sim \mu$. Then, defining

$$
h_W^\nu(x) := \min_{s\in S} \mathrm{pow}_W(x,s) + \langle W, \nu\rangle = \mathrm{pow}_W(x, T_W(x)) + \langle W, \nu\rangle,
$$

we have

$$\begin{aligned}\mathbb{E}[h_W^\nu(X)] &= \int_{\mathbb{R}^d} \mathrm{pow}_W(x, T_W(x))\ \mathrm{d}\mu + \int_{\mathbb{R}^d} \langle W, \nu\rangle\ \mathrm{d}\mu \\ &= L(T_W, W) + \langle W, \nu\rangle = H^\nu(W).\end{aligned}$$

In other words, our problem can be stated as maximizing the expected value of $h_W^\nu(X)$.

2.4 Basic Algorithm

TODO: Something

Image = 256 x 256 x 3

Over iterations nscales= $4 - 1 : -1 : 0$

1. scale image (2^{iter}) i.e in first iter=3 : $256/2^3 = 32$

2. then generate on with 2 dim more, (i.e scale im with $256/2^2 = 64$)

- In first iteration, i.e iter = 3: estime adsn model on scaled image (32x32x3)

In view auf Multi-scale scaling Approach

TODO: Something

3 Multi-Layer Approach

3.1 To Define, Muli-scale?

- Decomposition of the target distribution (K-means algorithm - Clustering method)
- Sub/Up-sampling procedure
- Multi-layer model

Multi-layer approach

Target measure decomposition

IDEA

Decompose the target measure $\nu = \sum_{s \in S} \nu_s \delta_s$ at different scales using the K-means algorithm (Lloyd's algorithm) $\longrightarrow$ Generate a finite sequence of discrete probabity measures $\{\nu_l\}_{l=0,\dots,L}$ with decreasing support and such that ν_{l+1} should be a *similar* to ν_l.

MORE PRECISELY

Using a clustering algorithm, generate a finite sequence of finite sets $\{S = S^0, \dots, S^L\}$ (to use as support for the discrete probabilty distributions) such that $|S^l| < |S^{l+1}|$ for all $l \in \{0, \dots, L-1\}$ (and $|S^L| = 1$). Then, define $\nu^0 := \nu$ and use successively transport maps to define the distributions on the supports S^l, i.e. define measurable maps

$$\pi_l : S^l \to S^{l+1} \quad \text{and set} \quad \nu^{l+1} := \pi_{l\#}\nu^l.$$

Thus, we get get successibly probabilty measures $\nu^{l+1} = \sum_{s \in S^{l+1}} \nu_s^{l+1} \delta_s$ supported on S^{l+1} satisfying

$$\nu_s^{l+1} = \nu^{l+1}(s) = \nu^l(\pi_l^{-1}(s)) = \sum_{p \in \pi_l^{-1}(s)} \nu_p^l.$$

Idea: Using Lloyd's algorithm we get: $\pi_l : x \mapsto \arg\min_{s \in S^{l+1}} \|x - s\|^2$.

Multi-layer Transport Map - With 2 Layers (as in Paper)

Reminder from last chapter (is written different):

$$h^{\nu}(x, W) : \mathbb{R}^d \times \mathbb{R}^{|S|} \to \mathbb{R},$$
$$(x, W) \mapsto \min_{s \in S} \mathrm{pow}_W(x, s) + \langle W, \nu \rangle = (\min_{s \in S} \|x - s\|^2 - W(s)) + \langle W, \nu \rangle$$

Then we have

$$\nabla_W h^{\nu} = \nu - \mathbb{1}^S_{T_W(x)=s} .$$

Where

$$\mathbb{1}^S_{T_W(x)=s} : S \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = T_W(x) \\ 0 & \text{else} \end{cases} .$$

Sketch of the algorithm:

Given: μ (Target distribution), ν (Source distribution), $L = 2$ (number of layers).

- Decompose target measure: $\{\nu^l\}_{l=0,1}$, $\{S^l\}_{l=0,1}$ as above.
- Set $W^l = 0$, for $l = 0, 1$. (Weights to be computed)
- Set $n^l : S^l \to 0$, for $l = 0, 1$. (Number of visits of points in S^l)

Apply ASGD. At each iteration: sample $x \sim \mu$ and then:

1. (L=1: first layer) Compute

   $$\tilde{s} = \arg\min_{s \in S^1} \|x - s\|^2 - W^1(s).$$

   I.o.w. compute $T_{W^1}(x)$. If $W^1 = 0$ (as in the first iterations), this is equivalent to computing a Least-squares.

2. Compute gradient

   $$g = \nabla_{W^1} h^{\nu^1} = \nu^1 - \mathbb{1}^{S^1}_{s=\tilde{s}}$$

   Where

   $$\mathbb{1}^{S^1}_{s=\tilde{s}} : S^1 \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = \tilde{s} \\ 0 & \text{else} \end{cases} .$$

3. Update W^1 as in Algorithm 1:

   $$W^1 \longleftarrow \text{Use gradient, gradient-step, iteration } (g, C, iter)$$

4. Update number of visits

   $$n^1(\tilde{s}) = n^1(\tilde{s}) + 1$$

5. (L=0: second layer) Compute

$$\tilde{\tilde{s}} = \arg\min_{s \in \pi_0^{-1}(\tilde{s})} \|x - s\|^2 - W^0(s)$$

I.o.w. compute $T_{W^0|_{\pi_0^{-1}(\tilde{s})}}|^{\pi_0^{-1}(\tilde{s})}(x) = T_{W^0}|^{\pi_0^{-1}(\tilde{s})}(x)$, where $T_{W^0}|^{\pi_0^{-1}(\tilde{s})}$ denotes the map T_{W^0} with restricted codomain $\pi_0^{-1}(\tilde{s})$.
Observations:

- This computations are faster than computing

$$T_{W^0}(x) = \arg\min_{s \in S^0} \|x - s\|^2 - W^0(s),$$

as $|\pi_0^{-1}(\tilde{s})| < |S^0|$.

- It may happend (yes? when?) that

$$T_{W^0}(x) \in S^0 \setminus \pi_0^{-1}(\tilde{s}).$$

Consequences?

6. Compute gradient

$$\tilde{g} = \nabla_{W^0|_{\pi^{-1}(\tilde{s})}} h^{\nu^0|_{\pi^{-1}(\tilde{s})}} = \nu^0|_{\pi^{-1}(\tilde{s})} - \mathbb{1}_{s=\tilde{\tilde{s}}}^{\pi^{-1}(\tilde{s})}$$

Where

$$\mathbb{1}_{s=\tilde{\tilde{s}}}^{\pi^{-1}(\tilde{s})} : \pi^{-1}(\tilde{s}) \to \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{if } x = \tilde{\tilde{s}} \\ 0 & \text{else} \end{cases}$$

7. Update W^0 as in Algorithm 2:

$$W^0 \longleftarrow \text{Use gradient, gradient-step, number of visits}(\tilde{g}, C, n^1)$$

Actually, only update entries on $\pi^{-1}(\tilde{s})$, i.e

$$W^0|_{\pi^{-1}(s)} \longleftarrow \text{Use gradient, gradient-step, number of visits}(\tilde{g}, C, n^1)$$

3.2 Algorithm (ASGD)

4 Application: Texture Synthesis

4.1 Patch distributions

4.1.1 Source Distribution

Gaussian Random Fields (Gaussian Synthesis)

Gaussian Mixture Models

Gaussian patches distribution and properties

4.1.2 Target Distribution

Emprirical patch distribution and simplifications

4.2 Optimal Transport in Patch Space

4.2.1 Local Transformations (Patch convolutions)

5 Conclusion

[]

Bibliography

[Aurenhammer et al., 1998] Aurenhammer, F., Hoffmann, F., and Aronov, B. (1998). Minkowski-type theorems and least-squares clustering. Algorithmica, 20(1):61–76.

[Galerne et al., 2010] Galerne, B., Gousseau, Y., and Morel, J.-M. (2010). Random phase textures: Theory and synthesis. IEEE Transactions on image processing, 20(1):257–267.

[Galerne et al., 2018] Galerne, B., Leclaire, A., and Rabin, J. (2018). A texture synthesis model based on semi-discrete optimal transport in patch space. SIAM Journal on Imaging Sciences, 11(4):2456–2493.

[Lévy, 2015] Lévy, B. (2015). A numerical algorithm for l2 semi-discrete optimal transport in 3d. ESAIM: Mathematical Modelling and Numerical Analysis, 49(6):1693–1715.

[McLachlan and Krishnan, 2007] McLachlan, G. J. and Krishnan, T. (2007). The EM algorithm and extensions, volume 382. John Wiley & Sons.

[Mérigot, 2011] Mérigot, Q. (2011). A multiscale approach to optimal transport. In Computer Graphics Forum, volume 30, pages 1583–1592. Wiley Online Library.

[Villani, 2003] Villani, C. (2003). Topics in optimal transportation. Number 58. American Mathematical Soc.