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:	4. Februar 2020

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

Abgabedatum: 4. 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.

Inhaltsverzeichnis

Optimal Transport

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

$$\int_{\mathbb{R}^d} \|x - T(x)\|^2 \, \mathrm{d}\mu \tag{0.1}$$

over all measurable maps $T : \mathbb{R}^d \to S$, which satisfy $T_{\#}\mu = \nu$.

This problem can be solved following the approach from [AHA] in two steps.
First, we find a minimizer of the functional (0.1) over a special class of measurable functions, which depend on predifined *weights*. This weights are defined for every point in S and have a geometrical interpretation, which we will describe below. Having this, we focus in adapting the weights in such a way that the functions fullfill the condition $T_{\#}\mu = \nu$. This approach is inspired geometrically by the study of a special kind of diagrams, the so called *power diagrams*, which are a generalization of Voroni diagrams. In order to get this geometric intuition, we first recall the necesary concepts for this approach.

Definition 0.1 (Voronoi Diagrams)**.** Let $S \subset \mathbb{R}^d$ be a finite set. We define for every point $s \in S$, the region of s as

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

This defines a partition of $\mathbb{R}^d$ in convex regions which is called the Voronoi diagram of S.

The Voronoi Diagram of S induce naturally a map $T : \mathbb{R}^d \to S$, assigning each point in $\mathbb{R}^d$ the corresponding point in S where is located, i.e

$$T(x) = s \quad \Leftrightarrow \quad x \in \mathrm{reg}(s). \tag{0.2}$$

Note that by defintion, some points in $\mathbb{R}^d$ may belong to more than one region. By convention, T assigns those points to an arbitrary point in S of a region where it is located. We call it, the induce assigment.

A generalization of this concepts arise by changing the euclidean distance in the defintion of the regions. This allow us to vary the induced assignment. One practial way to do this is by using the power function, which uses weights on the points of S.

Definition 0.2 (Power functions)**.** 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

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

W is called weight function or weight vector on S.

Similarly as in the Voronoi Diagram, we can define regions on $\mathbb{R}^d$ by using the power function with weights W. We call

$$\mathrm{reg}_W(s) \coloneqq \{x \in \mathbb{R}^d : \mathrm{pow}_W(x, s) \leq \mathrm{pow}_W(x, \tilde{s}) \quad \forall \tilde{s} \in S \setminus \{s\}\}.$$

the *power region* of s and the induced partition of R^d is called the *power diagram* of S with weights W.
The geometric intution of the power diagrams appears by looking at the spheres on $s \in S$

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

when $W : S \to \mathbb{R}_{>0}$. For a fixed $s \in S$ delivers the power function $\mathrm{pow}(\cdot, s)$ a negative (resp. positive) value whenever $x \in \mathbb{R}^d$ is inside (resp. outside) the sphere $\mathbb{S}^{d-1}_{\sqrt{W}}(s)$. 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, power cells of points $s \in S$ may not contain the point s or may be even empty. Nevertheless, the power diagram still partioned $\mathbb{R}^d$ in convex polyhedron.

By replacing reg with reg_W in (0.2) we obtain an assignment $T_W : \mathbb{R}^d \to S$. Similarly as by Voronoi diagrams, we assign those points who share diffent $\mathrm{reg}_(s)$ an arbitrarly of this s. We call this map, the *power assignment* of S with weights W.
This maps have an optimization property, which we recall in the following Lemma.

We recall the change of variables theorem of the measure theory. 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 $f : Y \to \mathbb{R}^d$ holds $f \in \mathcal{L}(t_\#) \Leftrightarrow f \circ t \in \mathcal{L}(\mu)$ and then

$$\int_{t^{-1}(B)} f \, \mathrm{d}t_\#\mu = \int_B f \circ t \, \mathrm{d}\mu \quad \text{for all } B \in \mathcal{U}\,. \tag{0.3}$$

Lemma 0.3. *Let* $(\mathbb{R}^d, \mathcal{A}, \mu)$ *be a Probabilty space, s.t* $\mu \ll \lambda^d$. *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. The power assignment is defined such that $\mathrm{pow}_W(x, T_W(x)) \leq \mathrm{pow}_W(x, s)$ over all $s \in S$. Consequently, T_W minimizes

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

over all measurable maps $T : \mathbb{R}^d \to S$. Using the fact that $\mathbb{R}^d = \bigcup_{s \in S} \mathrm{reg}(s)$ and that the boundaries of the power cells have zero λ^d-measure and thus also zero μ-measure, it holds

$$\begin{aligned}
\int_{\mathbb{R}^d} \omega(T(x)) \ \mathrm{d}\mu &= \sum_{s \in S} \int_{\mathrm{reg}(s)} W(T(x)) \ \mathrm{d}\mu \\
&\overset{(0.3)}{=} \sum_{s \in S} \int_s W \ \mathrm{d}T_{\#}\mu \\
&= \sum_{s \in S} \mu(T^{-1}(s)) W(s) \\
&= \sum_{s \in S} \zeta(s) W(s)
\end{aligned}$$

which is constant for a fixed ζ and W. $\square$

Let us have the same setting, i.e $(\mathbb{R}^d, \mathcal{A}, \mu)$ be a probility space s.t $\mu \ll \lambda^d$ and $(S, \mathrm{Pot}(S), \nu)$ another one. Let $\mathcal{F} := \{f : \mathbb{R}^d \to S : f \text{ is measurable}\}$, 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 two important properties which we will describe now. Let

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

be the vector of capacities induced by T and

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

As shown in Lemma 0.3, it holds

$$L(T, W) = Q(T) - \langle \zeta_T, W \rangle.$$

Thus, $L_T := L(T, \cdot)$ defines a linear function on $\mathbb{R}^{|S|}$ for a fixed $T \in \mathcal{F}$.
Recall that for a given $W \in \mathbb{R}^S$, the for the power assignment with weights W, holds $\text{pow}_W(x, T_W(x)) \leq \text{pow}_W(x, s)$, for all $s \in S$. Consequently, for a fixed $W \in \mathbb{R}^d$ holds $T_W = \arg\min_{T \in \mathcal{F}} L(T, W)$. We claim that

$$f : \mathbb{R}^{|S|} \to \mathbb{R}^d, \quad W \mapsto L_{T_W}(W)$$

is concave.

TODO: Rewrite properties for the proof

Proof. Let $\alpha \in [0, 1]$ and $W_1, W_2 \in \mathbb{R}^{|S|}$, then

$$\begin{aligned}
f(\alpha W_1 + (1 - \alpha)) &= L_{T_{\alpha W_1 + (1-\alpha) W_2}}(\alpha W_1 + (1 - \alpha) W_2) \\
&= L_{T_{\alpha W_1 + (1-\alpha) W_2}}(\alpha W_1) + L_{T_{\alpha W_1 + (1-\alpha) W_2}}((1 - \alpha) W_2) \\
&\geq L_{T_{\alpha W_1}}(\alpha W_1) + L_{T_{(1-\alpha) W_2}}((1 - \alpha) W_2) \\
&= \alpha L_{T_{\alpha W_1}}(W_1) + (1 - \alpha) L_{T_{(1-\alpha) W_2}}(W_2) \\
&\geq \alpha L_{T_{W_1}}(W_1) + (1 - \alpha) L_{T_{W_2}}(W_2) = \alpha f(W_1) + (1 - \alpha) f(W_2)
\end{aligned}$$

□

TODO: Prove f is smooth

f is smooth, and the gradient of f is given by $\nabla f = -\zeta_{T_W}$ TODO: CHECK THIS GENAUER Recall that to solve our problem (0.1), we need to find a weight vector W^* s.t $T_{W^*\#}(\mu) = \nu$. In other words $\mu(T_{W^*}^{-1}(s)) = \nu_s$. Consider the function

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

For this function holds $\nabla H(W) = \nu - \zeta_{T_W}$ and hence

$$\zeta_{T_W} = \nu \Leftrightarrow \nabla H(W) = 0.$$

Thus finding a solution of the problem (0.1) is equivalent to an optimization problem of a concave function. I.o.w we want find the maximum value of H.
This translates to:

$$\begin{aligned}\frac{\partial H}{W(s)} &= \frac{\partial f(W)}{W(s)} + \frac{\partial \langle \nu, W\rangle}{W(s)} \\ &= -\mu(T_W^{-1}(s)) + W(s) \\ &= -\mu(\mathrm{reg}(s)) + W(s).\end{aligned}$$

This has a probabilistic interpration. If we realize X as a random variable of distribution μ, i.e $X \sim \mu$. Then using

$$\min_{s \in S} \|x - s\|^2 - W(s) = \|x - T_W(x)\|^2 - W(s)$$

it holds with

$$h(x, \nu) = \min_{s \in S} \|x - s\|^2 - W(s) + \langle W, \nu \rangle$$

$$\begin{aligned}\mathbb{E}[h(X, W)] &= \int_{\mathbb{R}^d} \min_{s \in S} \|x - s\|^2 - W(s) \, \mathrm{d}\mu + \int_{\mathbb{R}^d} \langle W, \nu \rangle \, \mathrm{d}\mu \\ &= \int_{\mathbb{R}^d} \|x - T_W(x)\|^2 - W(s) \, \mathrm{d}\mu + \langle W, \nu \rangle = H(W)\end{aligned}$$

Literaturverzeichnis