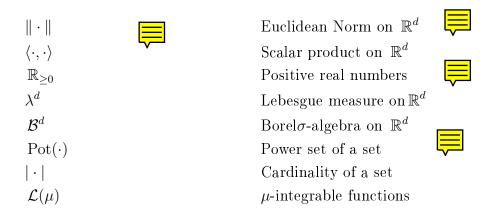
Nomenclature



TODO: Motivation for Optimal Transport & recall some definitions from measure theory

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

$$\tilde{\mu}(A) = 0 \Rightarrow \mu(A) = 0 \quad \forall A \in \mathcal{A}.$$

and denote this by $\mu \ll \tilde{\mu}$. Let $(X, \mathcal{A}, \overline{\mathcal{U}})$ 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

$$T_{\#}\mu(B) = \mu(T^{-1}(B)) \quad \forall B \in \mathcal{U}.$$

TODO:

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 $(S, \text{Pot}(S), \nu)$, we want to minimize

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

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

Recall that as $(S, \text{Pot}(S), \nu)$ is a finite probability space, we can write ν as a finite sum of Dirac measures

$$\nu = \sum_{s \in S} \nu_s \delta_s$$
 where $\nu_s \in \mathbb{R}_{\geq 0}$ and such that $\sum_{s \in S} \nu_s = 1$. (0.2)

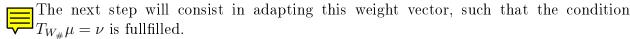
Thus, the problem we are considering translates to finding a measurable map T that minimizes the functional in (0.1) and such that we have $\mu(T^{-1}(s)) = \nu_s$ for all $s \in S$. For any $s \in S$, we will call $\mu(T^{-1}(s))$ the capacity of s.

TODO: ** Prove existence and uniqueness of a solution

As shown in [AHA], finding a measurable map that minimizes the above functional is



equivalent to finding the maximum of a concave function and thus, can be solved 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 the same capacities. This optimal solu= T_W is inspired geometrically and constructed using a predefined weight vector W.



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

$$\operatorname{reg}(s) := \{ x \in \mathbb{R}^d : ||x - s|| \le ||x - \tilde{s}|| \quad \forall \tilde{s} \in S \setminus \{s\} \}.$$

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

Remark. Note that the partition of \mathbb{R}^d generated by the Voronoi diag is given by 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

$$T(x) = s \Leftrightarrow x \in \operatorname{reg}(s).$$
 (0.3)

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 egion 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 0.2 (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

$$pow_W(x, s) := ||x - s||^2 - W(s).$$

We call W the weight function on S.

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

$$\operatorname{reg}_W(s) := \{ x \in \mathbb{R}^d : \operatorname{pow}_W(x, s) \le \operatorname{pow}_W(x, \tilde{s}) \quad \forall \tilde{s} \in S \setminus \{s\} \}$$

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 the definition of power diagrams becomes clear by looking at spheres around $s \in S$ with positive radius

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

whenever $W: S \to \mathbb{R}_{>0}$. The power function $pow(\cdot, s)$ for a fixed $s \in S$, 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 $s \in \mathbb{S}^{d-1}_{\sqrt{W}}(s)$ jus, 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 even meet empty. Nevertheless, the power diagram still partitions \mathbb{R}^d in a convex polyhedron?

By replacing reg with reg_W in (0.3) 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

$$(T_W(x) = s \Leftrightarrow x \in \operatorname{reg}_W(s)) \Leftrightarrow T_W(x) = \min_{s \in S} ||x - s||^2 - W(s) \tag{0.4}$$

for all points $x \in \mathbb{R}^d$ which don't share different regions. In fact, power functions even minimize the functional (0.1) for a fixed predefined weight vector W. We will prove this in Lemma 0.4. As a consequence, the natural question which remains to be clarified is how to fix a choice of the weight vector W, such that it fullfills the condition $T_{W_{\mu}}\mu = \nu$.

We recall the change of variables theorem from the measure theory.

Theorem (change of variables). Let (X, \mathcal{A}, μ) be a measure space, (Y, \mathcal{U}) a measurable space and $T: X \to Y$ a measurable function. For a measurable function $f: Y \to \mathbb{R}^d$ we have $f \in \mathcal{L}(T_{\#}\mu) \Leftrightarrow f \circ T \in \mathcal{L}(\mu)$ and when one of this statements is true then

$$\int_{T^{-1}(B)} f \, dT_{\#} \mu = \int_{B} f \circ T \, d\mu \quad \text{for all } B \in \mathcal{U}.$$
 (0.5)

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

Remark 0.3. Note that as the power region 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 \mathring{B} the interior of a set $B \in \mathcal{B}$ with respect to the standard topology, it holds

$$\lambda^d(\operatorname{reg}(s)) = \lambda^d(\operatorname{reg}(s)).$$

holds
$$\lambda^d(\operatorname{reg}(s)) = \lambda^d(\operatorname{reg}(s)).$$
 For a probability $\operatorname{ce}(\mathbb{R}^d, \mathcal{B}^d, \mu)$ such that $\mu \ll \lambda^d$, we have then
$$\mu(\operatorname{reg}(s)) = \mu(\operatorname{reg}(s)) \quad \text{and} \quad \sum_{s \in S} \mu(\operatorname{reg}(s)) = 1.$$

Lemma 0.4. 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 (0.4) of power assignments, it holds

$$pow_W(x, T_W(x)) \le pow_W(x, s)$$

■

for all $s \in S$. Consequently, T_W minimizes

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

over all measurable maps $T: \mathbb{R}^d \to S$. Using the fact that $\mathbb{R}^d = \bigcup_{s \in S} \operatorname{reg}(s)$ together with remark 0.3, we obtain

$$\begin{split} \int_{\mathbb{R}^d} \omega(T(x)) \ \mathrm{d}\mu &= \sum_{s \in S} \int_{\mathrm{reg}(s)} W(T(x)) \ \mathrm{d}\mu \\ &\stackrel{(0.5)}{=} \sum_{s \in S} \int_s W \ \mathrm{d}T_\# \mu \\ &= \sum_{s \in S} \mu(T^{-1}(s))W(s) \\ &= \sum_{s \in S} \zeta(s)W(s) \end{split}$$

which is constant for a fixed ζ and W.

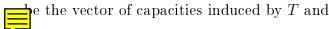
The natural question to handle next, is how to choose W, such that the condition $W_{\#}\mu = \nu$ holds. As we will show below, this question can be equivalently formulated as finding the maximum concave function. In order to achive this, we first recall the original setting of our originial problem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

Let $(\mathbb{R}^d, \mathcal{B}, \mu)$ and $(S, \text{Pot}(S), \nu)$ be two probability spaces such that $\mu \ll \lambda^d$. As in Equation (0.2), we write $\nu = \sum_{s \in S} \nu_s \delta_s$ as a finite sum of Dirac measures. For $\mathcal{F} := \mathbb{R}^d \to S : f$ is measurable}, define

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

This map has important properties, with we will show and then transfer them to the concave function of our reformulated problem. For a map $T \in \mathcal{F}$, let

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



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

be the functional that we want to study. As shown in Lemma 0.4, it holds

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

And hence, $L = L(T, \cdot)$ defines a linear function on $\mathbb{R}^{|S|}$ for a fixed $T \in \mathcal{F}$. Recall that for a given $W \in \mathbb{R}^{|S|}$ and a power assignment T_W , we have $\operatorname{pow}_W(x, T_W(x)) \leq W(x, s)$ for all $s \in S$. Consequently, for a fixed $W \in \mathbb{R}^d$ we must have

$$T_W(W) = \arg\min_{T \in \mathcal{F}} L(T, W). \tag{0.6}$$

We claim that

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

is concave.

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

$$f(\alpha W_{1} + (1 - \alpha)W_{2}) = L_{T_{\alpha W_{1} + (1 - \alpha)W_{2}}}(\alpha W_{1} + (1 - \alpha)W_{2})$$

$$= L_{T_{\alpha W_{1} + (1 - \alpha)W_{2}}}(\alpha W_{1}) + L_{T_{\alpha W_{1} + (1 - \alpha)W_{2}}}((1 - \alpha)W_{2})$$

$$\stackrel{(0.6)}{\geq} L_{T_{\alpha W_{1}}}(\alpha W_{1}) + L_{T_{(1 - \alpha)W_{2}}}((1 - \alpha)W_{2})$$

$$= \alpha L_{T_{\alpha W_{1}}}(W_{1}) + (1 - \alpha)L_{T_{(1 - \alpha)W_{2}}}(W_{2})$$

$$\stackrel{(0.6)}{\geq} \alpha L_{T_{W_{1}}}(W_{1}) + (1 - \alpha)L_{T_{W_{2}}}(W_{2}) = \alpha f(W_{1}) + (1 - \alpha)f(W_{2})$$

TODO: Prove formally that f is smooth

f is smooth and the gradient of f at W is given by $\nabla f(W) = -\zeta_{T_W}$. Recall that because of Lemma 0.4, to solve our problem (0.1) we need to find a weight vector W^* is fying $T_{W^*_{\#}}(\mu) = \nu$. In other words, it should hold $\mu(T_{W^*}^{-1}(s)) = \nu_s$ for all $s \in S$. Consider now the function

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

This function is concave and differentiable as a sum of concave differentiable functions. Furthermore we have $\nabla H(W) = \nu - \zeta_{T_W}$ and hence also

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

Thus we see that finding a solution of our original problem is in deed equivalent to

finding a maximum of the concave function H. We can compute

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

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



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

we have

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

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

