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

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

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

$$\nu = \sum_{s \in S} \nu_s \delta_s \quad \text{where } \nu_s \in \mathbb{R}_{\geq 0}, \text{ such that } \quad \sum_{s \in S} \nu_s = 1. \tag{0.2}$$

Thus, the problem we are considering translates to find a measurable map T that minimizes the functional in (0.1) and such that $\mu(T^{-1}(s)) := \mu(T^{-1}(\{s\})) = \nu_s$ 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 [?], we will now show that finding a minimizer of the functional (0.1) 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 0.1 (Voronoi Diagrams)**.** Let $S \subset \mathbb{R}^d$ be a finite set. We define for every point $s \in S$

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

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

$$T(x) = s \quad \Leftrightarrow \quad x \in \mathrm{reg}(s). \tag{0.3}$$

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 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 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

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

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

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

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 $\sqrt{W(s)}$

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

For a fixed $s \in S$, the power function $\mathrm{pow}(\cdot, s)$ 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 (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

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

for all points $x \in \mathbb{R}^d$ which don't share different regions. In fact, power assignments even minimize the functional (0.1) for prescribed capacities. We will prove this in Lemma 0.4. 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 0.4 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

$$\int_{T^{-1}(B)} f \ \mathrm{d}T_{\#}\mu = \int_B f \circ T \ \mathrm{d}\mu \quad \textit{for all } B \in \mathcal{U}\,. \tag{0.5}$$

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

Remark 0.3. 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 $\mathrm{int}(B)$ the interior of a set $B \in \mathcal{B}$ with respect to the standard topology, we get

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

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 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, 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 0.3, we obtain

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

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

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 (0.2), we write the measure ν as a finite sum of Dirac measures $\nu = \sum_{s\in S}\nu_s\delta_s$. For $\mathcal{F} := \{f : \mathbb{R}^d \to S : f \text{ is measurable}\}$, 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

$$\zeta_T : S \to \mathbb{R}, \quad s \mapsto \mu(T^{-1}(s)) \tag{0.6}$$

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 0.4, it holds

$$L(T, W) = Q(T) - \langle\zeta_T, W\rangle. \tag{0.7}$$

And hence, $L_T := L(T, \cdot)$ 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 (0.4) states

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

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

$$T_W = \arg\min_{T \in \mathcal{F}} L(T, W). \tag{0.8}$$

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

Theorem 0.5. *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

$$\begin{aligned}
F(\alpha W_1 + (1-\alpha)W_2) &= L_{T_{\alpha W_1 + (1-\alpha)W_2}}(\alpha W_1 + (1-\alpha)W_2) \\
&\overset{\text{lin}}{=} L_{T_{\alpha W_1 + (1-\alpha)W_2}}(\alpha W_1) + L_{T_{\alpha W_1 + (1-\alpha)W_2}}((1-\alpha)W_2) \\
&\overset{(0.8)}{\geq} L_{T_{\alpha W_1}}(\alpha W_1) + L_{T_{(1-\alpha)W_2}}((1-\alpha)W_2) \\
&\overset{\text{lin}}{=} \alpha L_{T_{\alpha W_1}}(W_1) + (1-\alpha) L_{T_{(1-\alpha)W_2}}(W_2) \\
&\overset{(0.8)}{\geq} \alpha L_{T_{W_1}}(W_1) + (1-\alpha) L_{T_{W_2}}(W_2) = \alpha F(W_1) + (1-\alpha) F(W_2)
\end{aligned}$$

□

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

$$\nabla H^\nu(W) = \nabla F(W) + \nu \overset{\text{Thm 0.5}}{=} -\zeta_{T_W} + \nu \tag{0.9}$$

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 0.4, we see that finding a solution of our original problem is indeed equivalent to finding a maximum of the concave function H^ν. Regarding (0.9), we have

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

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

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

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 $h_W^\nu(X)$.