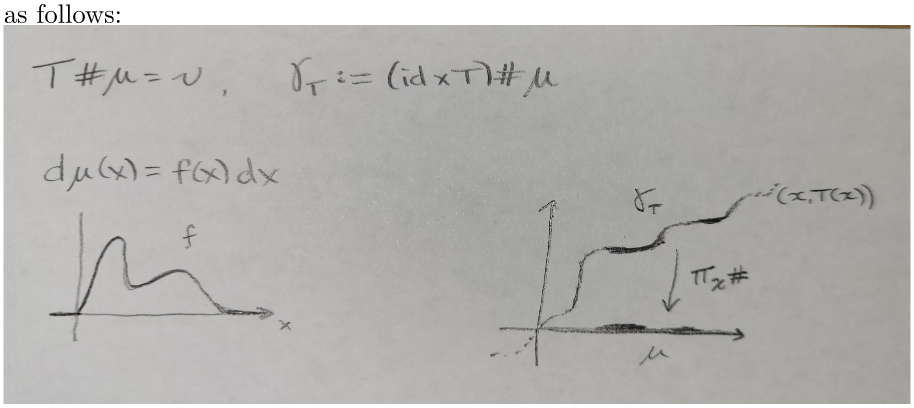
Optimal Transport Seminar 7/26/23

Benjamin Faktor

Connection between Monge and Kantorovich

Every transport map T induces a transport plan γ_T, via $\gamma_T := (\mathrm{id} \times T)\#\mu$, whose Kantorovich cost is the same as the Monge cost of T. It can be visualized as follows:



Thus there are more transport plans than transport maps, so the solution to the Kantorovich problem is no larger than that of the Monge problem:

$$\begin{aligned}(K) \;=\; \min_{\gamma \in \Pi(\mu,\nu)} \int c(x,y)d\gamma(x,y) &\leq \min_{T\#\mu=\nu} \int c(x,y)d\gamma_T(x,y) \\ &= \min_{T\#\mu=\nu} \int c(x,T(x))d\mu(x) \;=\; (M).\end{aligned}$$

In fact, under certain conditions the solutions are the same.

Theorem 1. *Suppose μ is of the form $d\mu(x) = f(x)dx$ and $\Omega \subseteq \mathbb{R}^d$ is compact. Then the Kontorovich problem is a relaxation of the Monge problem. That is,*

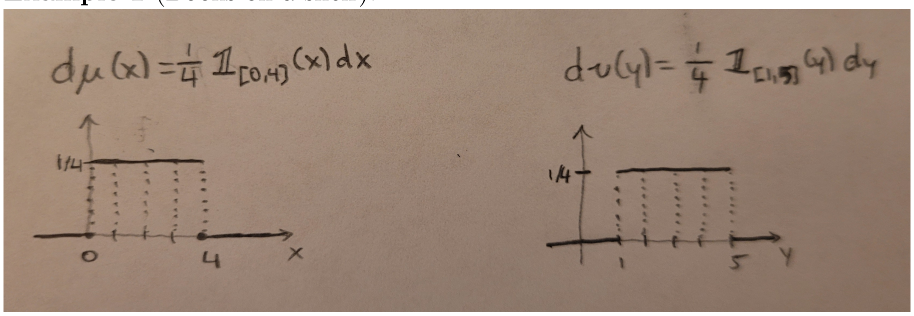
$$\min_{\gamma \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} c(x,y)d\gamma(x,y) = \min_{T\#\mu=\nu} \int_{\Omega} c(x,T(x))d\mu(x)$$

Remark 1. *μ being of the form $d\mu(x) = f(x)dx$ means μ does not concentrate mass on $d-1$ dimensional subsets.*

Brenier's Theorem

The cost $c(x,y) = |x-y|$ gives rise to a nice dual formulation. One problem with this cost is the non-uniqueness of the optimal transport map.

Example 1 (Books on a shelf)**.**



Definition 1. *A function $u : \mathbb{R}^d \to \mathbb{R}$ is called "convex" if for every $x, y \in \mathbb{R}^d$ and $\alpha \in [0,1]$,*

$$u((1-\alpha)x + \alpha y) \leq (1-\alpha)u(x) + \alpha u(y).$$

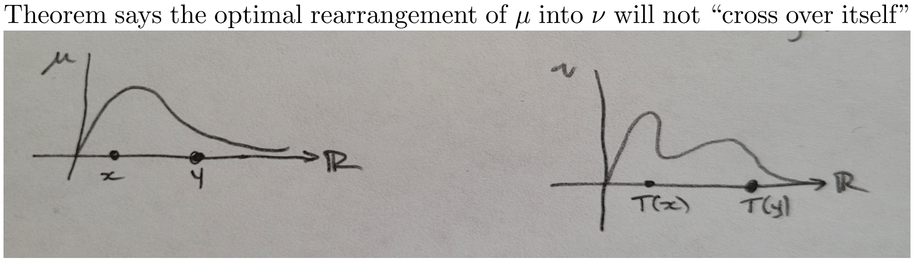
Remark 2. *Equivalently, a function $u : \mathbb{R}^d \to \mathbb{R}$ is convex if and only if its Hessian is positive semi-definite.*

Example 2. *$u : \mathbb{R} \to \mathbb{R}$, $u(x) = x^2$ is convex.*

Theorem 2 (Brenier)**.** *Let $c(x,y) = \|x-y\|^2$ and μ, ν have finite second moment, i.e. $\int_{\mathbb{R}^d} |x|^2 d\mu(x), \int_{\mathbb{R}^d} |y|^2 d\nu(y) < \infty$. Suppose further μ has the form $d\mu(x) = f(x)dx$. Then there exists a unique optimal transport map T of the form $T = \nabla u$ where u is a convex function. Conversely, if T is a transport map and $T = \nabla u$ then T is optimal.*

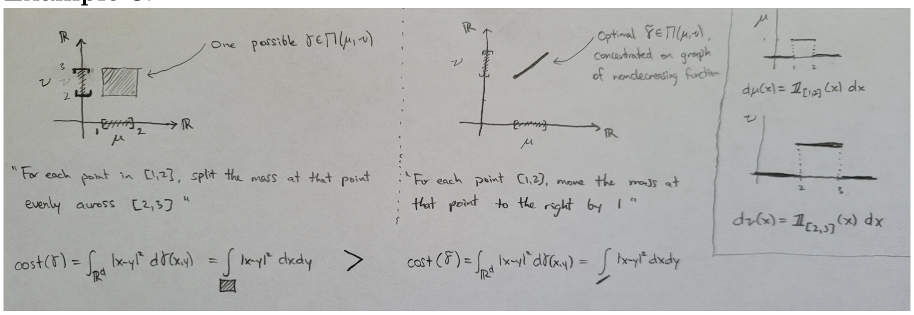
Remark 3. *The condition that $d\mu(x) = f(x)dx$ is necessary, for if μ concentrates mass on a $d-1$ dimensional subset uniqueness fails. Consider $d = 2$, $\mu = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}$, $\nu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,0)}$.*

Discussion of the special case $d = 1$.
Fact: The derivative of a convex function $u : \mathbb{R} \to \mathbb{R}$ is nondecreasing. Brenier's Theorem says the optimal rearrangement of μ into ν will not "cross over itself".



Extension to Kantorovich problem:
The optimal $\tilde{\gamma} \in \Pi(\mu,\nu) = \{\gamma : \pi_{\mathcal{X}}\#\gamma = \mu, \pi_{\mathcal{Y}}\#\gamma = \nu\}$ concentrates its mass on the graph $(x, T(x))$ of a function T which is the gradient of a convex function u. Namely this is via $\tilde{\gamma} = (\mathrm{id} \times T)\#\mu)$.

Example 3.



Translations and Dilations

In $\mathbb{R}^d$, a translation is the gradient of a convex function, namely

$$T_c(x) = x + c = \nabla \left[\frac{1}{2} \|x\|^2 + c \cdot x \right].$$

By Brenier's Theorem, whenever a translation is a transport map, it is necessarily an optimal map and $\gamma_T := (\mathrm{id} \times T)\#\mu$ is the optimal transport plan. In this case one computes that the 2-Wasserstein distance between the original and translated measure is

$$\begin{aligned} W_2(\mu, T_c\#\mu) &= \left(\int |x-y|^2 d[(\mathrm{id} \times T_c)\#\mu] \right)^{1/2} \\ &= \left(\int |x - T_c(x)|^2 d\mu \right)^{1/2} \\ &= \left(\int |c|^2 d\mu \right)^{1/2} \\ &= |c| \end{aligned}$$

as expected. Similarly scalings can be written as

$$T_a(x) = ax = \nabla \left[\frac{a}{2} \|x\|^2 \right],$$

so the 2-Wasserstein distance between a measure μ and its dilation $\nu = T_a\#\mu$ is

$$\begin{aligned} W_2(\mu, T_a\#\mu) &= \left(\int |x-y|^2 d[(\mathrm{id} \times T_a)\#\mu] \right)^{1/2} \\ &= \left(\int |x - T_a(x)|^2 d\mu \right)^{1/2} \\ &= |1-a| \left(\int |x|^2 d\mu \right)^{1/2}. \end{aligned}$$