Statistical OT Lecture 4: Brenier's Theorem

1 Cyclical Monotonicity

In the last class we talked about subgradients. One property that subgradients have is that they define a *monotone* operator. Suppose we have $T \subseteq \mathbb{R}^d \times \mathbb{R}^d$ which we think of as a set-valued operator (for each x it outputs a set $T(x) \subseteq \mathbb{R}^d$).

This is a monotone operator – for any x, y, and $u \in T(x), v \in T(y)$ we have that,

$$(u - v)^T (x - y) \geq 0.$$

The subdifferential of a convex function defines a monotone operator. Notice that for any x, y and $g_x \in \partial f(x)$ and $g_y \in \partial f(y)$ we have that:

$$f(y) \geq f(x) + g_x^T (y - x)$$
$$f(x) \geq f(y) + g_y^T (x - y).$$

Summing these inequalities shows that the subdifferential is a monotone operator.

One can ask if this fact has a converse, i.e. if I give you a monotone operator T can you construct a convex function whose subdifferential is equal to T? When you think about this you will hit two quick barriers:

1. If I take the subdifferential of a convex function, view it as a set-valued map, and simply drop some of the output values, this will still be a monotone operator. (We are simply taking a subset of T and checking the monotonicity condition.) So maybe the best converse we could hope for is that:

   $$T \subseteq \partial\varphi,$$

   for come convex φ.

2. Even this turns out to be false. Consider for instance the map $T(x) = Ax$ where $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. This is a monotone operator, but is not the gradient of a convex function.

It is a beautiful result of Rockafellar's that a strengthening of being a monotone operator does in fact permit a nice converse. This strengthening is called cyclical monotonicity. Concretely, given an operator $T \subseteq \mathbb{R}^d \times \mathbb{R}^d$, suppose for any $k \geq 2$ we consider a set of points $(x_1, y_1), \ldots, (x_k, y_k) \subseteq T$. Throughout the rest of the lecture we'll use (cyclical) notation where $x_{k+1} = x_1$.

Then T is cyclically monotone if:

$$\sum_{i=1}^{k} y_i^T (x_i - x_{i+1}) \geq 0.$$

This condition generalizes the definition of a monotone operator (which is just this definition with $k = 2$). It is simple to verify that for a convex function φ, its subdifferential $\partial\varphi$ is always cyclically monotone.

The deep result of Rockafellar is that given any T which is cyclically monotone, there exists a proper, closed convex function φ such that:

$$T \subseteq \partial\varphi.$$

Proof: The proof of this result is slick and explicit. Lets fix an $(x_0, y_0) \in T$. For any $x \in \mathbb{R}^d$ define the (convex) function:

$$\varphi(x) = \sup_{k \geq 0} \sup_{(x_i, y_i) \in T, i \in \{1, \ldots, k\}} \left\{ (x_1 - x_0)^T y_0 + (x_2 - x_1)^T y_1 + \ldots + (x - x_k)^T y_k \right\}.$$

Notice that this is the supremum of affine functions, and is closed and convex. By cyclical monotonicity we see that $\varphi(x_0) \leq 0$. It is also easy to see that $\varphi(x_0) \geq 0$, by taking $k = 1$ and choosing $(x_1, y_1) = (x_0, y_0)$. So we see that this is a proper convex function.

Finally, take any $(x, y) \in T$, any $z \in \mathbb{R}^d$ and we see that:

$$\begin{aligned} \varphi(z) &\geq \sup_{k \geq 0} \sup_{(x_i, y_i) \in T, i \in \{1, \ldots, k\}} \left\{ (x_1 - x_0)^T y_0 + (x_2 - x_1)^T y_1 + \ldots + (x - x_k)^T y_k + (z - x)^T y \right\} \\ &= \varphi(x) + (z - x)^T y, \end{aligned}$$

which shows that $y \in \partial\varphi(x)$.

2 Brenier's Theorem

We now turn our attention to the proof of the following result:

Theorem 1. Let μ, ν be two distributions with finite second moments, and μ has a density. If γ_0 is an optimal coupling for the squared Euclidean cost, i.e.:

$$\int \|x - y\|^2 d\gamma_0 = W_2^2(\mu, \nu),$$

then there is a convex function $\varphi_0 : \mathbb{R}^d \to \mathbb{R}$ such that γ_0 is the joint distribution of $(X, \nabla\varphi_0(X))$ where $X \sim \mu$.

It turns out the theorem is also true without any moment assumptions (but one needs a bit of care). One (direct) consequence of the theorem is that Monge's problem (of finding an OT map) has a solution in this setting, and this map is given by the gradient of a convex function.

Before we prove this result, let us notice that we've already done some of the work. Concretely, given an OT coupling γ_0, if we can show that $\text{supp}(\gamma_0)$ is a cyclically monotone set, then we can apply Rockafellar's theorem to conclude that there is a convex function φ_0 such that $\text{supp}(\gamma_0) \subseteq \partial\varphi_0$, i.e. $\gamma_0(Y \in \partial\varphi_0) = 1$. Then we can note that convex functions are (classically) differentiable Lebesgue almost everywhere in the interior of their domain (this is known as Rademacher's theorem), and that the boundary of their domain has 0 Lebesgue measure. Then, since μ has a density, we can conclude that φ_0 is differentiable μ almost everywhere and:

$$\gamma_0(Y = \nabla\varphi_0(X)) = 1.$$

This is Brenier's theorem. To actually complete the proof we need to only show one fact – that the support of γ_0 is cyclically monotone.

2.1 Discrete Case

Before we actually prove Brenier's theorem, it is worth thinking about why cyclical monotonicity arises in OT (with quadratic cost).

In the discrete case where $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ we have already remarked that there is an OT map which is a matching of the points. Suppose we take an OT matching, we can argue that its support is always cyclically monotone. Conversely, given any cyclically monotone matching we can conclude that it must be an OT matching.

Optimality $\implies$ Cyclical Monotonicity: To simplify things let's relabel so that the OT coupling simply matches x_i to y_i. Suppose we are given k distinct points in the support of our OT coupling (with indices $j_1, \ldots, j_k$). We would like to show that,

$$\sum_{i=1}^{k} y_{j_i}^T (x_{j_i} - x_{j_{i+1}}) \geq 0.$$

Define the permutation τ to be the one which leaves the remaining indices unchanged, and shifts the indices in the k points by one, i.e. it matches $x_{j_{i+1}}$ with y_{j_i}. By optimality we have that,

$$\sum_{i=1}^{k} \|x_{j_i} - y_{j_i}\|^2 \leq \sum_{i=1}^{k} \|x_{j_{i+1}} - y_{j_i}\|^2,$$

which is exactly the fact we wanted to prove.

Cyclical Monotonicity $\implies$ Optimality: Now, suppose that the identity permutation is cyclically monotone (after relabeling). Given any other permutation τ we can decompose τ into cycles. Roughly, start from an index i, collect the indices $(i, \tau(i), \tau(\tau(i)), \ldots)$. This forms a cycle, then start from a new index and repeat this till you have exhausted all indices. This is a disjoint collection of cycles. For each of these cycles I, cyclical monotonicity of the identity permutation will show that,

$$\sum_{i \in I} \|x_i - y_i\|^2 \leq \sum_{i \in I} \|x_i - y_{\tau(i)}\|^2,$$

and putting these together we obtain that the identity permutation is optimal.

2.2 Proof of Brenier's Theorem

Recall that all we need to show is that the support of γ_0 is cyclically monotone. We note in passing that the support of γ_0 along the first marginal is the support of μ and along the second marginal is the support of ν. Our proof will mimic the discrete proof, but we'll need to be careful to ensure that our modified coupling is still valid. Suppose the support of γ_0 is not cyclically monotone. Then for some $k \geq 2$, we can find $\{(x_1, y_1), \ldots, (x_k, y_k)\} \in \text{supp}(\gamma_0)$ such that,

$$\sum_{i=1}^{k} \|x_i - y_i\|^2 > \sum_{i=1}^{k} \|x_{i+1} - y_i\|^2.$$

Now, there are neighborhoods U_i, V_i of x_i, y_i such that, $\gamma_0(U_i \times V_i) > 0$, and for which:

$$\sum_{i=1}^{k} \|\widetilde{x}_i - \widetilde{y}_i\|^2 > \sum_{i=1}^{k} \|\widetilde{x}_{i+1}^{'} - \widetilde{y}_i^{'}\|^2,$$

for any $x_i, x_i^{'} \in U_i, y_i, y_i^{'} \in V_i$, for $i \in \{1, \ldots, k\}$.

Now, we can intuitively imagine cutting out the pieces $U_i \times V_i$ and moving them to $U_{i+1} \times V_i$. More formally, let γ_i denote the conditional distribution of γ_0 restricted to $U_i \times V_i$. Let γ_i^1 denote its first marginal, and γ_i^2 denote the second marginal. Then, for some small enough (specified later on), $c > 0$, we set:

$$\gamma = \gamma_0 + \frac{c}{k} \sum_{i=1}^{k} \left[\gamma_{i+1}^1 \times \gamma_i^2 - \gamma_i \right].$$

We need to verify this is a valid coupling, and then argue that it is better than γ_0. For any set A,

$$\gamma(A \times \mathbb{R}^d) = \gamma_0(A \times \mathbb{R}^d) + \frac{c}{k} \sum_{i=1}^{k} \left[\gamma_{i+1}^1(B) - \gamma_i^1(B) \right] = \gamma_0(A \times \mathbb{R}^d) = \mu(A).$$

A similar argument works for the second marginal. Now, we also need to ensure that $\gamma(A) \geq 0$:

$$\gamma(A) \geq \gamma_0(A) - \frac{c}{k} \sum_{i=1}^{k} \gamma_i(A) = \gamma_0(A) - \frac{c}{k} \sum_{i=1}^{k} \frac{\gamma_0(A \cap U_i \times V_i)}{\gamma_0(U_i \times V_i)}.$$

So it suffices to choose $c < \min_i \gamma_0(U_i \times V_i)$ to ensure that $\gamma(A) \geq 0$.

Finally, let us evaluate the cost of γ:

$$\begin{aligned} \int \|x - y\|^2 d\gamma &= \int \|x - y\|^2 d\gamma_0 + \frac{c}{k} \sum_{i=1}^{k} \left[\int_{U_{i+1} \times V_i} \|x - y\|^2 d\gamma_{i+1}^1 \times \gamma_i^2 - \int_{U_i \times V_i} \|x - y\|^2 d\gamma_i \right] \\ &< \int \|x - y\|^2 d\gamma_0, \end{aligned}$$

contradicting the optimality of γ_0. So we conclude that the support of γ_0 must be cyclically monotone.