Statistical OT Lecture 3: One-Dimensional Transport and Convex Analysis Background

1 One-Dimensional OT

The high-level takeaway is that in 1D the structure of the OT problem simplifies considerably, and many things (maps, Wasserstein distances) can be computed in closed form.

We will begin with a proof of the following theorem in the case when $p = 1$, for 1D optimal transport. The full proof of the following theorem (when $p \neq 1$) will be a HW question.

Theorem 1. Suppose that μ, ν are probability measures on $\mathbb{R}$, then

1. The OT coupling between μ, ν for any strictly convex cost $c(x, y) = h(x - y)$ for h strictly convex, is given by the joint distribution of the pair of random variables $(F_\mu^\dagger(U), F_\nu^\dagger(U))$.

   If μ is absolutely continuous then this coupling can also be written as the distribution of $(X, F_\nu^\dagger(F_\mu(X)))$, and this coupling is realized by the OT map $T_0(X) = F_\nu^\dagger(F_\mu(X))$.

2. The Wasserstein distance for $p \geq 1$ is given by:

$$W_p^p(\mu, \nu) = \int_0^1 |F_\mu^\dagger(u) - F_\nu^\dagger(u)|^p du.$$

3. Furthermore, we can write:

$$W_1(\mu, \nu) = \int_0^\infty |F_\mu(t) - F_\nu(t)| dt.$$

Proof. We focus on the case when $p = 1$. The main fact that requires proof is the first part of the first claim. The remaining claims follow from this. We begin by lower bounding the cost of any coupling, and then showing that the coupling given in the theorem statement attains the lower bound. Given any coupling γ, we can write:

$$\begin{aligned}
\int |x - y| d\gamma &= \int \int_{-\infty}^{\infty} \left[\mathbb{I}_{x \leq t < y} + \mathbb{I}_{y \leq t < x}\right] dt d\gamma \\
&= \int_{-\infty}^{\infty} \gamma(X \leq t < Y) + \gamma(Y \leq t < X) dt \\
&= \int_{-\infty}^{\infty} [\gamma(X \leq t) - \gamma(X \leq t, Y \leq t) + \gamma(Y \leq t) - \gamma(X \leq t, Y \leq t)] dt \\
&= \int_{-\infty}^{\infty} [F_\mu(t) + F_\nu(t) - 2\gamma(X \leq t, Y \leq t)] dt.
\end{aligned}$$

Now, we observe that $\gamma(X \leq t, Y \leq t) \geq \min(F_\mu(t), F_\nu(t))$ so we have that,

$$\begin{aligned}
\int |x - y| d\gamma &\geq \int_{-\infty}^{\infty} [F_\mu(t) + F_\nu(t) - 2\min\{F_\mu(t), F_\nu(t)\}] dt \\
&= \int_{-\infty}^{\infty} |F_\mu(t) - F_\nu(t)| dt.
\end{aligned}$$

Let us consider the coupling we have constructed and denote it by γ_0. We have that:

$$\begin{aligned}
\gamma_0(X \leq t, Y \leq t) &= P(F_\mu^\dagger(U) \leq t, F_\nu^\dagger(U) \leq t) \\
&= P(U \leq F_\mu(t), U \leq F_\nu(t)) = \min\{F_\mu(t), F_\nu(t)\}.
\end{aligned}$$

This shows that the coupling we have constructed attains the lower bound exactly, and thus is a minimum cost coupling. Furthermore, our proof also shows that:

$$W_1(\mu, \nu) = \int_0^\infty |F_\mu(t) - F_\nu(t)| dt.$$

□

2 Background on Convex Analysis

There are many deep connections between OT and convex analysis and throughout the course we will encounter concepts in convex analysis. Today, we will cover some basic background (you can get a slightly more elaborate treatment of this material either in my Convex Optimization lecture notes on my website, or a comprehensive treatment from Rockafellar's book on Convex Analysis).

A function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is convex if for every x, y and $t \in [0, 1]$ we have that:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

A set C is convex if for every $x, y \in C$, $t \in [0, 1]$, $(1-t)x + ty \in C$. For a convex function f, we will define its domain to be the set $\mathrm{dom}(f) = \{f < \infty\}$. A convex function f is *closed* if its epigraph is a closed set, and a convex function is *proper* if it is not identically ∞.

There are other more restrictive characterizations of convexity that are often useful:

1. If f is continuously differentiable, then f is convex if for every x, y:

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

2. If f is twice-continuously differentiable then f is convex if for every x:

$$\nabla^2 f(x) \succeq 0.$$

Convex functions are differentiable (Lebesgue) almost everywhere in the – a fact that is known as Rademacher's theorem. Intuitively, convex functions are always continuous, and though they can be non-differentiable the points at which they are non-differentiable has measure 0.

2.1 Subgradients

Even when a (convex) function is not differentiable an analogue of the first-order condition can still hold. A vector g is a subgradient of f at x if for every $y \in \mathbb{R}^d$:

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

The collection of such vectors g is called the subdifferential of f at x which we denote by $\partial f(x)$. The subdifferential can be empty (for instance, this is the case if f is strictly concave), but a nice property of convex functions is that in the interior of the domain of f, $\partial f(x)$ is non-empty. Furthermore, if f is differentiable at x then $\partial f(x)$ is a singleton and equal to $\nabla f(x)$.

2.2 Fenchel Conjugates

For any function $f : \mathbb{R}^d \to \mathbb{R} \cup \infty$, we can define its Fenchel conjugate by:

$$f^*(y) = \sup_x \left[x^T y - f(x) \right].$$

As a supremum of affine functions, f^* is a closed, convex function (even if f is not). For some intuition, the Fenchel conjugate provides a "dual" description of the function f – it effectively describes the collection of half-spaces which lie below the function f. If f is in fact closed and convex, $f^{**} = f$.

Directly from the definition of f^* we can see that the Fenchel-Young inequality holds:

$$f(x) + f^*(y) \geq x^T y.$$

Suppose that f is convex, then the following two facts are true:

1. If $y \in \partial f(x)$, then $f(x) + f^*(y) = x^T y$.

   Proof: We know that for any $z \in \mathbb{R}^d$:

   $$f(z) \geq f(x) + y^T(z - x),$$

   so we obtain that

   $$x^T y - f(x) \geq z^T y - f(z),$$

   for every z, so we conclude that,

   $$x^T y - f(x) \geq \sup_z [z^T y - f(z)] = f^*(y),$$

   and together with the Fenchel-Young inequality we see that $f(x) + f^*(y) = x^T y$.

2. Conversely, if for some y we know that $f^*(y) + f(x) = x^T y$ then we can conclude that $y \in \partial f(x)$.

   Proof: Let us consider for any $z \in \mathbb{R}^d$

   $$f(z) - f(x) - y^T(z - x) = f(z) - y^T z + f^*(y) \geq 0,$$

   using the Fenchel-Young inequality. So we conclude that $y \in \partial f(x)$.

These facts also highlight that ∂f and ∂f^* behave as inverse maps to each other. This is easiest to understand when f and f^* are both differentiable, proper, closed convex functions. Then if for some pair, (x, y) we have that,

$$f(x) + f^*(y) = x^T y,$$

we can conclude that, $y = \nabla f(x)$, and $x = \nabla f^*(y)$, i.e. that $(\nabla f)^{-1} = \nabla f^*$.