Lecture 2

Recall: Gaspard Monge, 1781

Q: How can we rearrange the dirt in $\mu$ to look like $\nu$ in the most efficient way?

$\mu$ \hspace{1cm} $\nu$

Dirt: Probability measures

$(X,d)$ metric space\footnote{where dirt lives} e.g. $(\mathbb{R}^d, d)$

$\mathcal{B}(X)$ Borel $\sigma$-algebra

smallest $\sigma$-algebra containing all open sets

$\mathcal{M}(X)$ (Borel) measures on $X$

Given $\mu \in \mathcal{M}(X), \mu(B) =$ how much dirt is in $B$, $A, B \in \mathcal{B}(X)$

To have any hope of rearranging $\mu$ to look like $\nu$, we must have $\mu(X) = \nu(X)$ \textquotedblleft some amount of dirt\textquotedblright
WLOG, always suppose $\mu(X) = \nu(X) = 1$, i.e. they are probability measures.

$\mathcal{P}(X) = \{ \mu \in \mathcal{M}(X) : \mu(X) = 1 \}$

**WARNING: HORRIBLE NOTATION ABUSE**

For $(X, d) = (\mathbb{R}^d, 1, 1)$, let $L^d$ denote Lebesgue measure $\mu$ absolutely continuous with $L^d$, i.e. $L^d(B) = 0 \Rightarrow \mu(B) = 0$.

If $\mu \in \mathcal{P}(X)$ satisfies $\mu \ll L^d$, by the Radon-Nikodym theorem, $\exists$ $f \in L^1(\mathbb{R}^d)$ s.t.

$$d\mu = f \, dL^d \iff d\mu(x) = f(x) \, dx$$

Radon-Nikodym derivative, $\frac{d\mu}{dx}$

Instead, we denote R-N derivative by $\mu(x)$.

$$d\mu(x) = \mu(x) \, dx$$

**Rearranging piles of dirt**

**Def (transport map):** Given $\mu, \nu \in \mathcal{P}(X)$ a measurable function $t : X \rightarrow X$ transports $\mu$ onto $\nu$ if...
\[ \mathcal{N}(B) = \mu(t'(B)) \quad \forall \ B \in \mathcal{B}(\mathbb{R}) \]

We call \( \mathcal{N} \) the push-forward of \( \mu \) under \( t \), writing \( \mathcal{N} = t#\mu \), and we call \( t \) the transport map.

A dict that starts at location \( x \) in \( \mu \) is sent to location \( t(x) \) in \( \mathcal{N} \).

Sanity check: if \( \mu \in \mathcal{P}(\mathbb{R}) \), \( t: \mathbb{R} \to \mathbb{R} \) measurable, is \( \mathcal{N} = t#\mu \) a probability measure?

\[ \mathcal{N}(x) = \mu(t^{-1}(x)) = \mu(x) = 1 \quad \checkmark \]

Ex: (Translation/Dilation): Suppose \((X,d) = (\mathbb{R}^d, 1,1)\). Fix \( a > 0 \), be \( \mathbb{R}^d \) and define \( t(x) = ax + b \)

\[ \mathcal{N} \quad \text{dilation} \quad \text{translation} \]

\[ 1 \xrightarrow{\mathcal{N}} 3 \]

\[ t(x) = 2x + 1 \]

\[ 3 \xrightarrow{\mathcal{N}} 7 \]
\[ t'(y) = \frac{y-b}{a} \]

By def \( \nu(B) = \mu(t'(B)) = \mu(\frac{B-b}{a}) \quad \forall B \in \mathcal{B}(X) \)

above \( \nu(B) = \mu(\frac{B-b}{a}) \)

**Lemma**: (equi-characterization of transport map)
Given \( \mu, \nu \in \mathcal{P}(X) \) and \( t:X \to X \) measurable, then \( t^\# \mu = \nu \) if and only if

\[ \int f(t(x)) \, d\mu(x) = \int f(y) \, d\nu(y) \quad \forall f \text{ bdd}, \]

\[ \text{by def of } f \text{ measurable.} \]

Bit of notation: for any \( B \in \mathcal{B}(X) \), define the indicator function

\[ 1_B(x) := \begin{cases} 0 & \text{if } x \notin B \\ 1 & \text{if } x \in B \end{cases} \]

Note: \( \forall B \in \mathcal{B}(X), 1_B \) is a bdd, meas fn.
Pg: Assume \((*)\). In particular, \(f = 1_B\),
\[
\mu(t^{-1}(B)) = \int 1_B(t(x)) d\mu(x) = \int 1_B(y) d\nu(y) = \nu(B)
\]
\[
1_B(t(x)) = \begin{cases} 
0 & \text{if } x \notin t^{-1}(B) \\
1 & \text{if } x \in t^{-1}(B)
\end{cases}
\]
Thus, by defn, \(\nu = t^# \mu\).

Assume \(\nu = t^# \mu\). By \((***)\) we know \((*)\) holds for \(f = 1_B\) \(\forall B \in B(E)\). By linearity, \((***)\) holds for \(f = \sum_{j=1}^\infty c_j 1_{B_j}\), \(c_j \in \mathbb{R}\), \(B_j \in B(E)\).  

For any \(\mu\), \(f\) measurable, \(\mu\)-integrable function \(f\), \(E\) a sequence of simple functions so that
\[
\lim_{n \to \infty} f_n(x) = f(x) \text{ pointwise}
\]
Thus, by the dominated convergence thm
\[
\lim_{n \to \infty} \int_E f_n(t(x)) d\mu(x) = \int_E \lim_{n \to \infty} f_n(y) d\nu(y) = \int_E f(y) d\nu(y)
\]
\[
\int_E f(t(x)) d\mu(x)
\]
Finally, to conclude for all bdd, meas \( f \), we can consider \( f^+ = \max \{ f, 0 \} \) and \( f^- = \max \{ -f, 0 \} \) separately.

\[
\]

Remark: If \( \mu, \nu \in \mathcal{P}(\mathbb{X}) \) satisfy
\[
\int f \, d\mu = \int f \, d\nu \quad \text{if} \quad f \text{ bdd, meas}
\]
then \( \mu = \nu \).

Now you know everything you need to know about rearranging \( \mu \) to look like \( \nu \) via transport maps. How to do this in the most efficient way?

**Monge's Optimal Transport Problem:**
Given \( \mu, \nu \in \mathcal{P}(\mathbb{X}) \), solve
\[
\min_{t : \mathbb{X} \to \mathbb{X}} \left\{ \int |t(x) - x| \, d\mu(x) : t \# \mu = \nu \right\}
\]
how far dirt is moved
how much dirt
how much dirt
ensure s t
rearranges \( \mu \) to
look like \( \nu \)
If it solves the Monge problem, we call it the optimal transport map.

We'll see many optimization problems of this form:

\[
\min \, F(\mathbf{t}) \\
\text{subject to } \mathbf{t} \in \mathcal{C},
\]

where \( \mathcal{C} \) is the constraint set, and \( F(\mathbf{t}) \) is the objective function.

Unfortunately, this is a horrible problem!