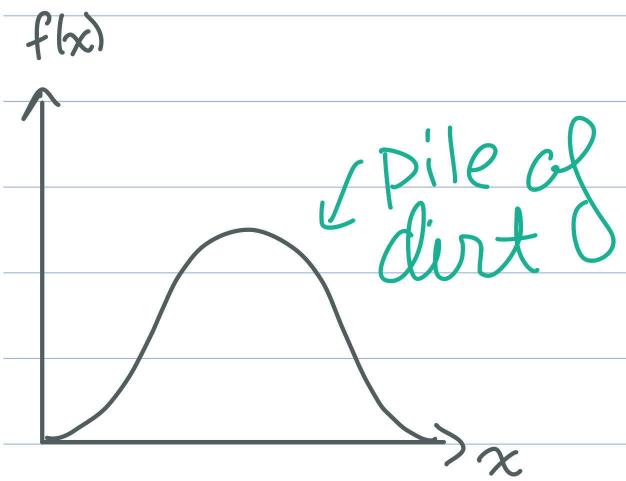
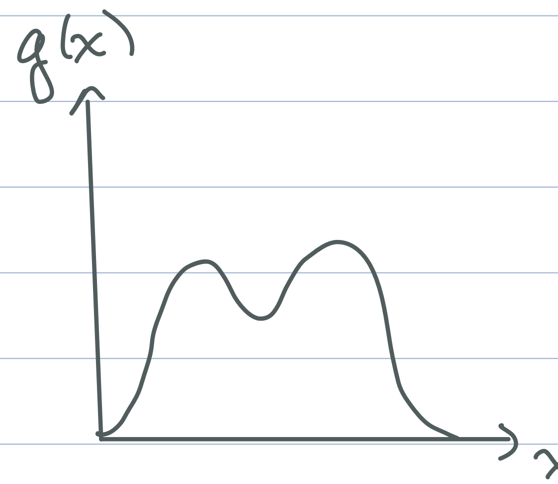
Math 260R: Optimal Transport
Prof. Katy Craig
(No office hours on Friday)

Gaspard Monge, 1781
"On cuttings and embankments"



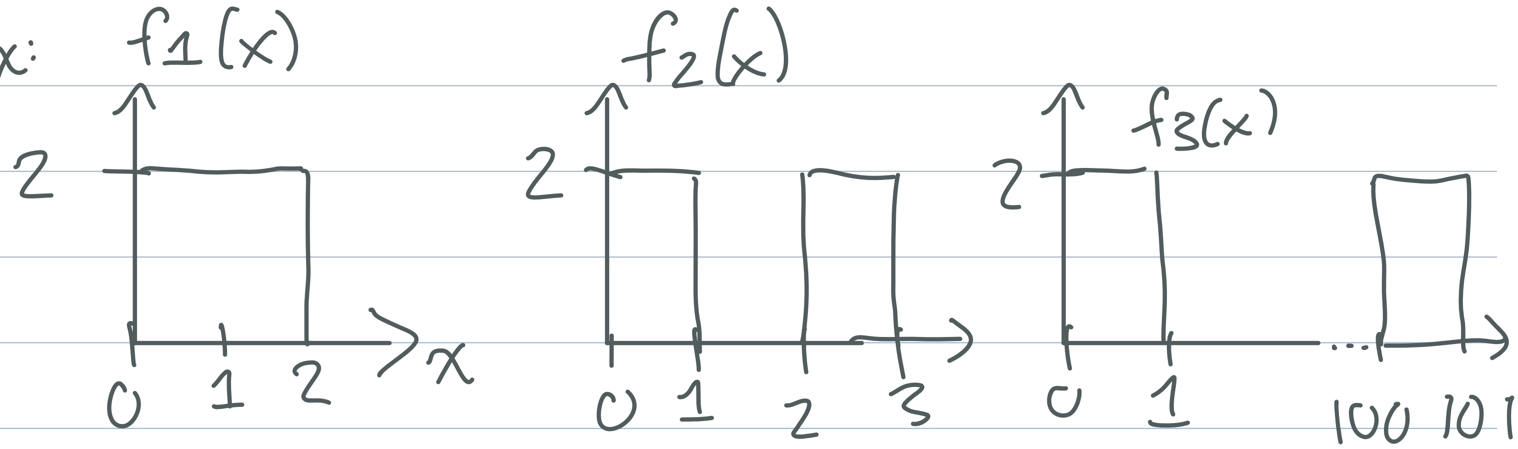


Q: How can we rearrange the dirt in f to look like g in the most efficient way?

Q': Why do we care?

A': The amount of effort it takes to rearrange one pile of dirt to look like another provides a notion of distance that is useful in PDE, geometry, statistics, machine learning, ...

Ex:



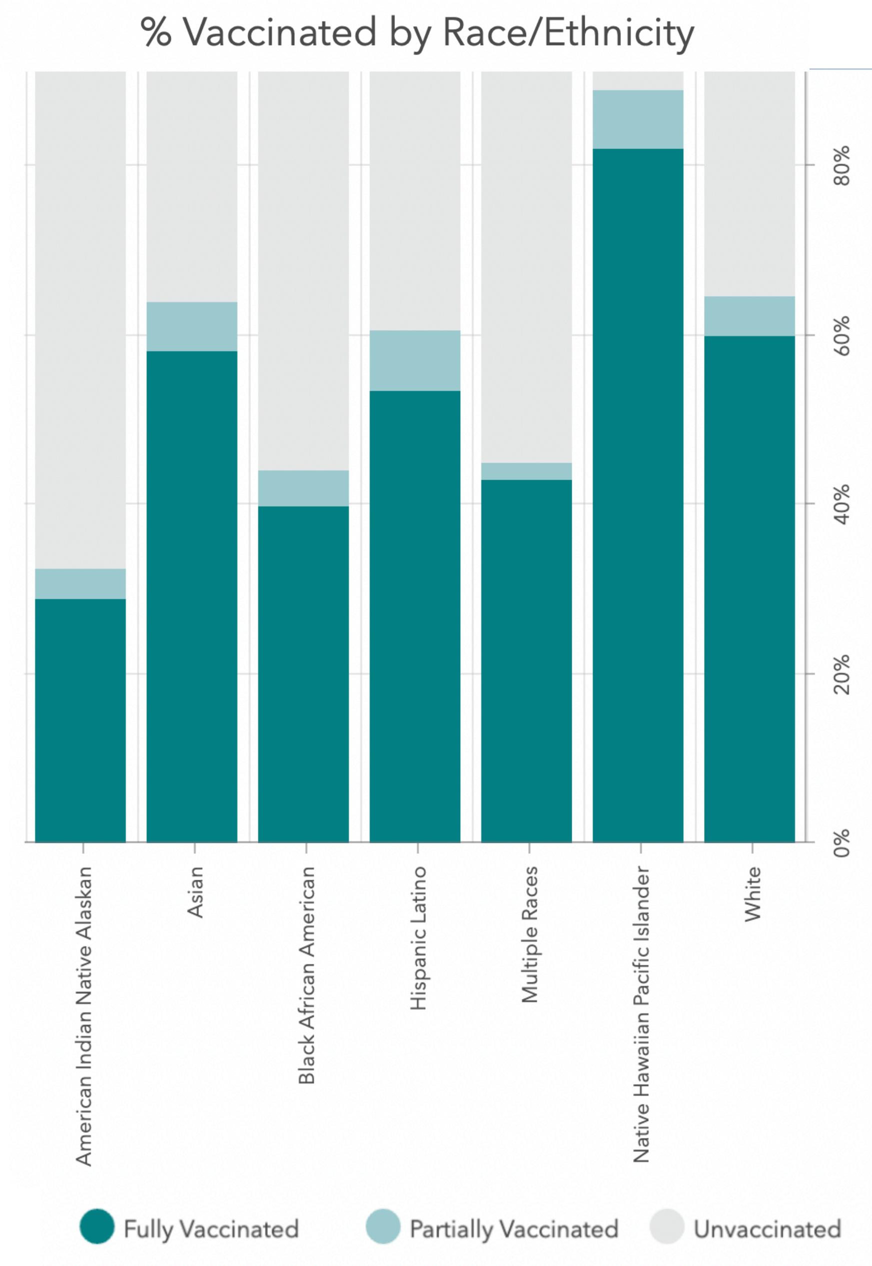
If we measure distance in the "usual way" (L^p norms, statistical divergences, ...)

$$\int |f_1(x) - f_2(x)|\,dx = 4 = \int |f_1(x) - f_3(x)|\,dx$$

... isn't f_1 "more similar" to f_2 than f_3?

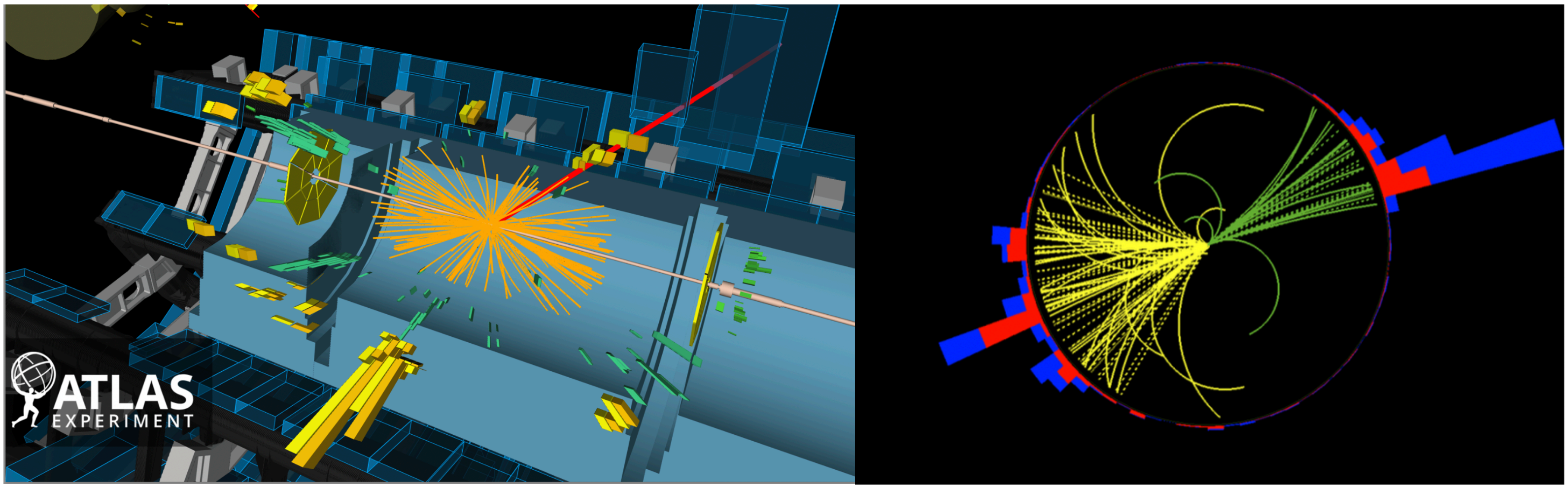
Moral: common notions of distance between functions do not ~~endow~~ endow independent variable with a spatial interpretation.

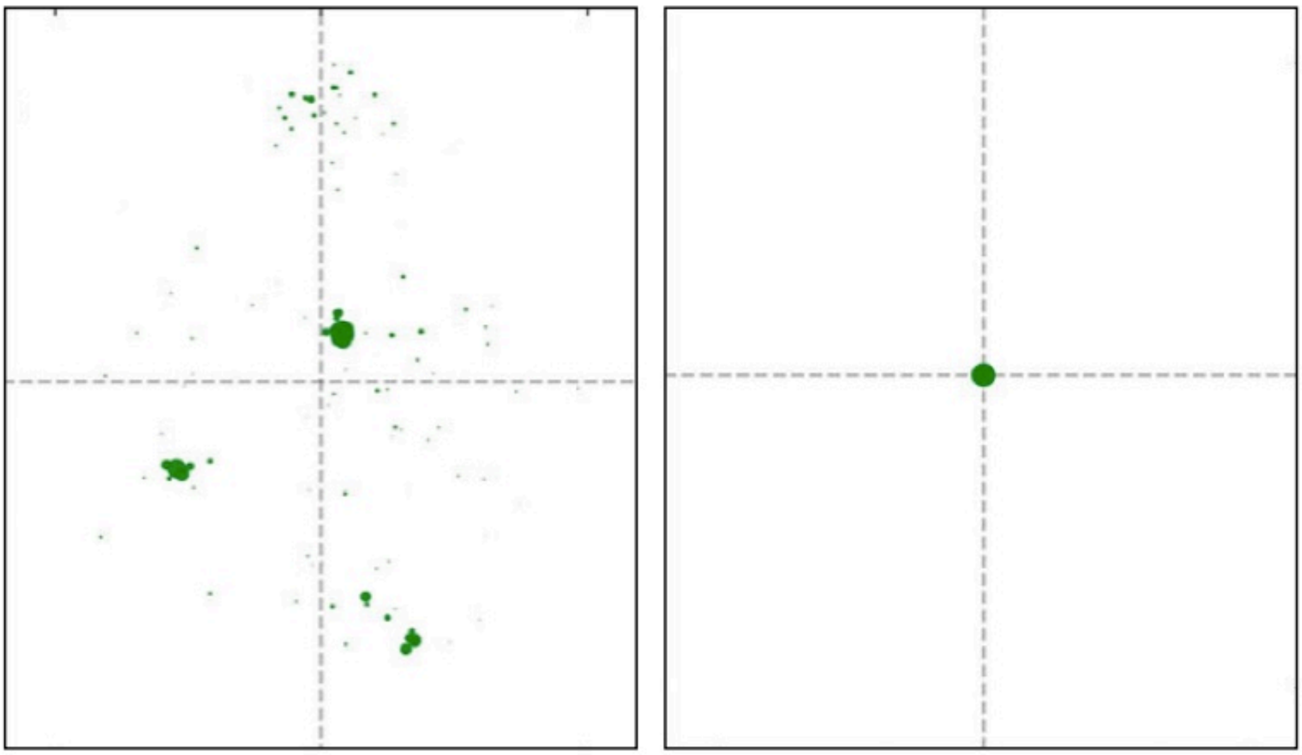
For certain data sets, this makes sense:



source: publichealthsbc.org

For other data sets, disregarding the spatial interpretation of independent variable throws away important information:





t^1 jet W^1 jet

https://www.google.com/url?sa=i&url=https%3A%2F%2Fdatasets.activeloop.ai%2Fdocs%2Fml%2Fdatasets%2Ffashion-mnist-dataset%2F&psig=AOvVaw1KXqxVNf9FmZL16cfZlSJh&ust=

Optimal transport provides a notion of distance between

[functions, data distributions, measures]

that preserves the spatial interpretation of the independent variable.

Over the past 20 years, this has had an enormous impact in:

1. PDE: two Fields medals
   Villani (2010), Figalli (2018)

2. Geometry: novel characterization of Ricci curvature in terms of convexity of entropy

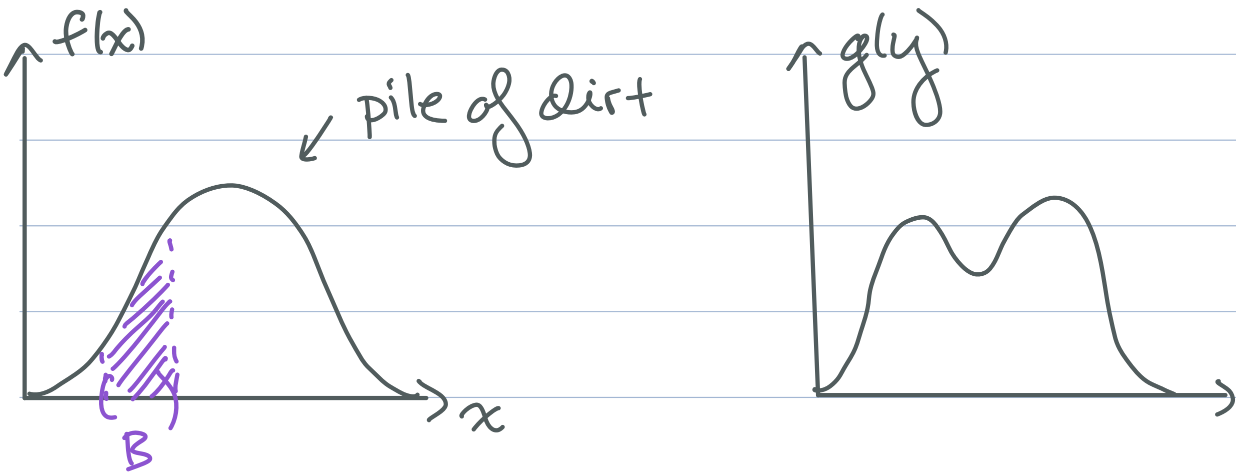
3. Statistics: sampling

4. Machine learning: neural networks, normalizing flows, ...

Back to original optimal transport problem...

source measure μ target measure ν

$d\mu(x) = f(x)\, d\lambda(x)$ $d\nu(y) = g(y)\, d\lambda(y)$



amount of dirt in region $B = \int_B f(x)\, d\lambda(x) = \mu(B)$

Q: How can we rearrange the dirt in f to look like g in the most efficient way?

Rmk: This problem only makes sense if $\int f(x)\, d\lambda(x) = \int g(y)\, d\lambda(y)$

$\mu(X) = \nu(Y)$

WLOG, assume $\mu(X) = \nu(Y) = \underline{1}$

It turns out that the "right" mathematical setting to solve this problem is more general...

We will represent the piles of dirt as measures instead of functions.

(X, d_X), (Y, d_Y) metric spaces

$\mathcal{B}(X)$ Borel σ-algebra

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

$\mathcal{P}(X)$ (Borel) probability measures on X

What does it mean to "rearrange" one probability measure to look like another?

Def: (transport map) Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a measurable function $t : X \to Y$, we say t <u>transports μ to ν</u> if

$$\nu(B) = \mu(t^{-1}(B)), \ \forall \ B \in \mathcal{B}(Y).$$

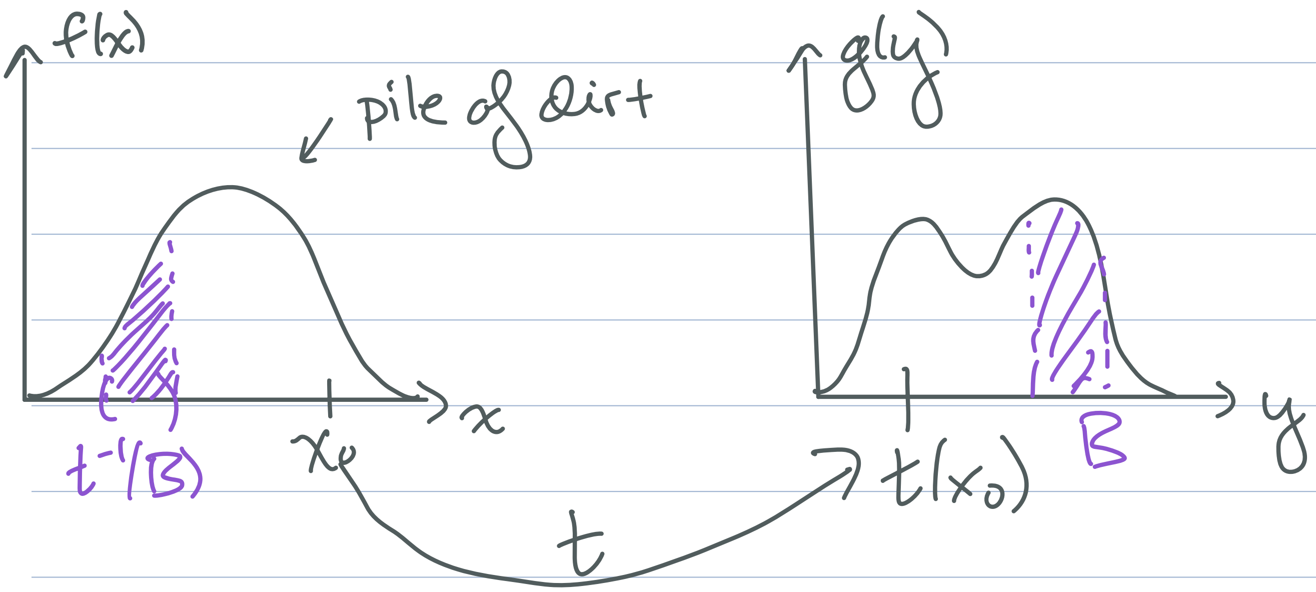
We call ν the pushforward of μ under t, written $\nu = t_\# \mu$, and we call t a transport map from μ to ν.

source measure μ

$$d\mu(x) = f(x)\, d\lambda(x)$$

target measure ν

$$d\nu(y) = g(y)\, d\lambda(y)$$



Informally, "mass starting at location x_0 in μ is sent to $t(x_0)$ in ν." See Exercise 1.

Ex (translation / dilation)

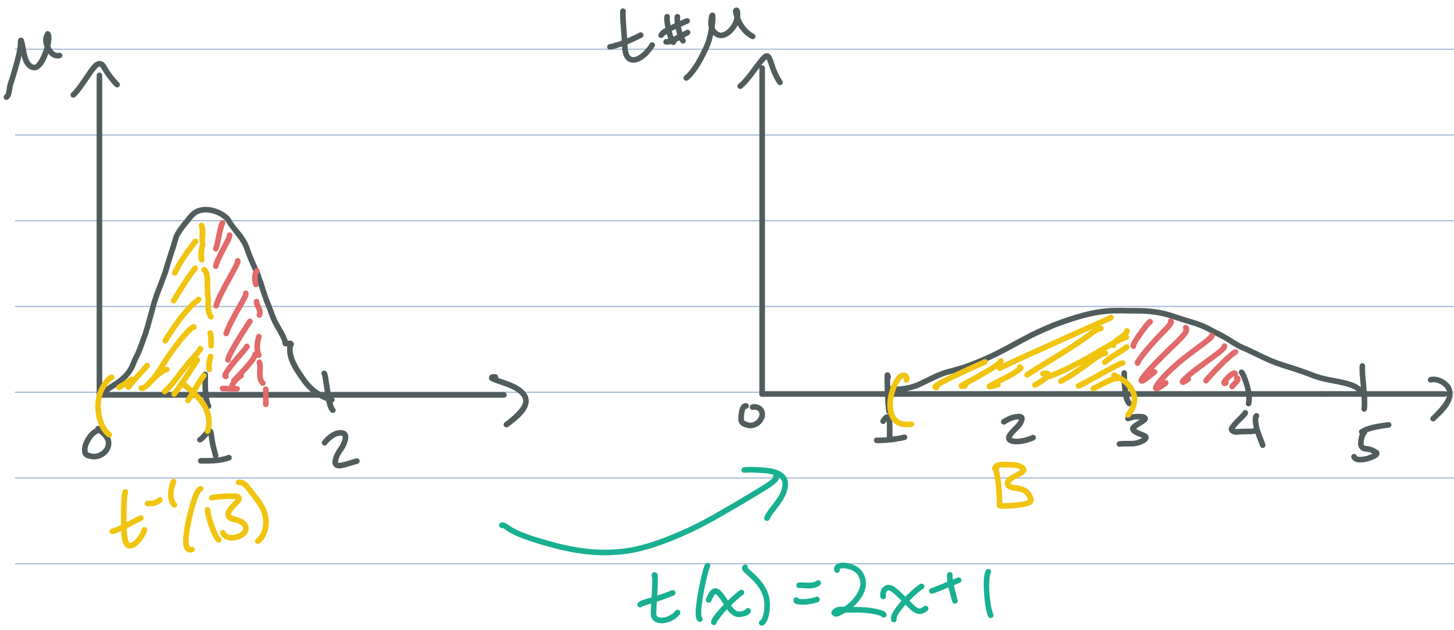
Suppose $(X, d_X) = (Y, d_X) = (\mathbb{R}^d, |\cdot|)$

Fix $a > 0$, $b \in \mathbb{R}^d$, and let $t(x) = ax + b$.

Thus, for any $\mu \in P(\mathbb{R}^d)$, $t\#\mu$ satisfies

$$(t\#\mu)(B) = \mu(t^{-1}(B)) = \mu\left(\frac{B-b}{a}\right)$$

$$= \mu\left(\left\{\frac{y-b}{a} : y \in B\right\}\right)$$



Lemma (equiv characterization of transp map): Given $\mu \in \mathcal{P}(X)$ and $t: X \to Y$ measurable,

$$t\#\mu = \nu \iff \int \varphi(t(x))\,d\mu(x) = \int \varphi(y)\,d\nu(y) \quad \forall \varphi \in L^1(\nu)$$

$$\parallel$$

$$\int \varphi(y)\,d(t\#\mu)(y)$$

Pf: Exercise 2

Lemma (change of variables formula)

Suppose

- $f \in L^1(\lambda^d)$, $f \geq 0$, $\int f\,d\lambda^d = 1$
- μ is given by $d\mu(x) = f(x)\,d\lambda^d(x)$
- $t \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ is injective and satisfies $|\det(Dt)|(x) \neq 0 \ \forall x \in \mathbb{R}^d$.

Then $\nu := t\#\mu$ satisfies $d\nu(y) = g(y)\,d\lambda^d(y)$

where

$$g(y) = \begin{cases} \left(\frac{f}{|\det(Dt)|}\right)\circ t^{-1}(y) & \text{if } y \in t(\mathbb{R}^d) \\ 0 & \text{if } y \notin t(\mathbb{R}^d) \end{cases}$$

Pf: Exercise 3

Application: Normalizing Flows

Reference: Kobyzev, Prince, Brubaker '21

e.g. μ uniform prob meas or μ is a Gaussian

Problem: Given a reference measure $\mu \in \mathcal{P}(X)$, about which we know everything, and given a target measure $\nu \in \mathcal{P}(Y)$, from which we have samples $\{y_i\}_{i=1}^n$, find $t: X \to Y$ "nice" so that $t\#\mu \approx \nu$.