CARLETON UNIVERSITY

SCHOOL OF
MATHEMATICS AND STATISTICS

HONOURS PROJECT

TITLE: An Introduction to Optimal Transport

AUTHOR: Jacob Turner-Hechtman

SUPERVISOR: Dr. Abbas Momeni

DATE: May 5th, 2020

An Introduction to Optimal Transport

Jacob Turner-Hechtman

May 2020

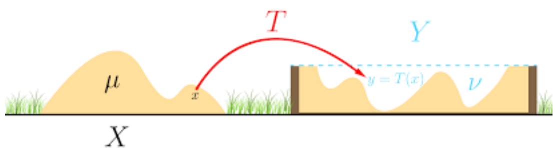
Supervisor: Dr. Abbas Momeni
Second Reader: Dr. David Amundsen

To my friends and family who supported me all the way.
With special thanks to Dr. Lily Hechtman and Dr. Peter Hechtman,
without which this would not be possible.

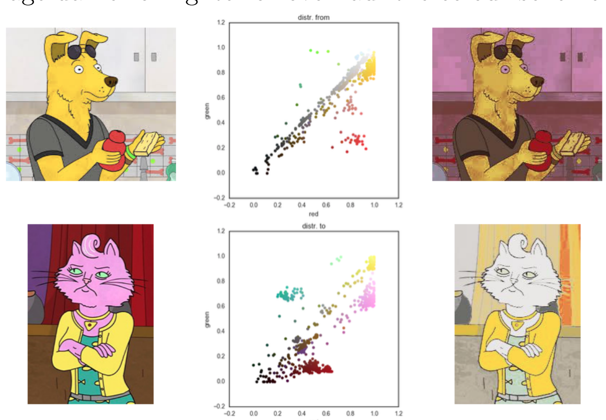
1 Introduction

There are many applications to optimal transport theory. Transforming one probability distribution into another at minimal cost can be applied to computer vision and machine learning, however the problem traces it's roots back to the 1700's.

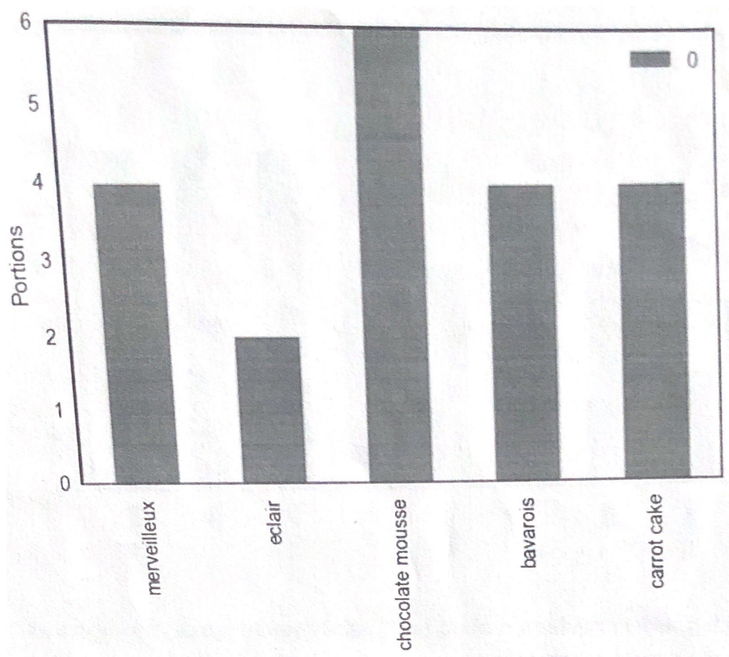
In 1781 French engineer Gaspard Monge (see [4]) was interested in the most effective way to redistribute mass. More specifically he wanted to reshape a pile of soil using minimal effort. His problem would remain unsolved until the late 20th century however and now has applications in other fields such as statistics, economics, geometry, and image processing.



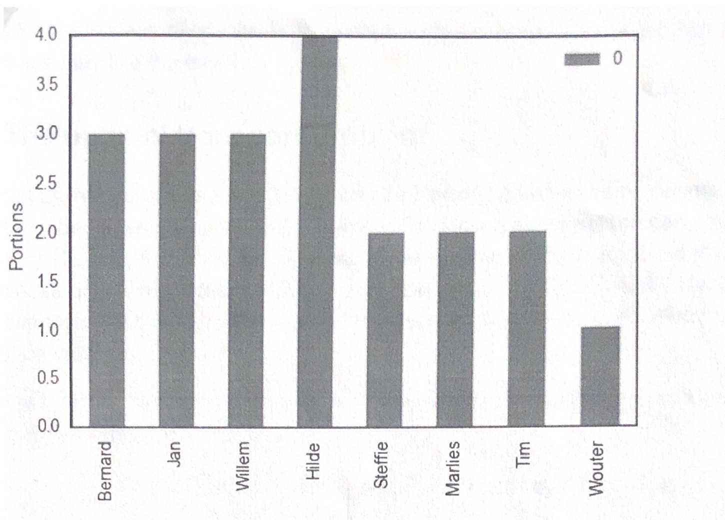
An every day example of optimal transport is its use in colour transfer, that is changing the colour scheme of an image. It is something used in animation or graphics and can be used to make an image darker or lighter or overhaul the colour scheme completely.



A more hands on example is the distribution of desserts at a dinner party. If we have a number of desserts say, a merveilleux, some eclairs, mousse, bavarois, and carrot cake. We can cut all of these into portions and have twenty shares.



If we have a number of professors at our party we can allow certain professors to help themselves to more dessert and distribute accordingly.



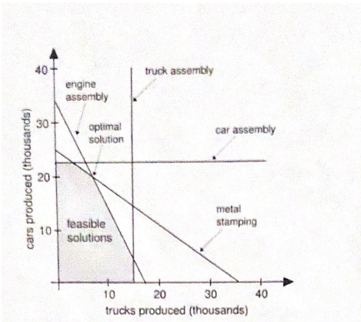
We can distribute these desserts the optimal way. How exactly can we divide desserts between professors to make everyone as happy as possible? We can ask everyone to rank the desserts on a scale from -2 to 2, with -2 being something they hated and 2 being their favourite. The preferences are given below in a table:

	merveilleux	eclair	chocolate mousse	bavarois	carrot cake
Bernard	2.0	2	1	0	0
Jan	0.0	-2	-2	-2	2
Willem	1.0	2	2	2	-1
Hilde	2.0	1	0	1	-1
Steffie	0.5	2	2	1	0
Marlies	0.0	1	1	1	-1
Tim	-2.0	2	2	1	1
Wouter	2.0	1	2	1	-1

Here we can see that most people like eclairs and mousse the best and merveilleux is a more polarizing dessert. This method also accounts for other factors like lactose intolerance or allergies. We can now divide these desserts in such a way that people get their portions of the kinds they like the most.

Data scientists can use optimal transport to transform one distribution into another. Lets say we are given a data set, that is, a set of points in some space that have the same weight. Suppose we have a second set of points with a different weight. Suppose also that we wish to interpolate between the two sets of data, we can simply take a weighted average between each point in set 1 and its analogs in set 2.

We are also able to reformulate the Monge minimization problem to a maximisation problem. This is part of the section called the dual problem. An application of duality can be building cars and truck on an assembly line with capacity constraints. If we have to build cars and trucks but have to be aware of a max capacity we must figure out a way to maximize our profits.



2 Notation and Background Material

Through out this paper We will use the following notation:

Characteristic functions. Let $A \subset \mathbb{R}^d$. The *characteristic function* $\chi_A : \mathbb{R}^d \to \{0, 1\}$ is

$$\chi_A := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Preimage. Let $T : X \to Y, B \subseteq Y$. The *preimage* of B under T is the set

$$T^{-}1(B) := \{x \in X : T(x) \in B\}$$

Probability densities. Let $X \subseteq \mathbb{R}^d$. We say that $f : X \to [0, \infty)$ is a *probability density* on X if $\int_X f(x)dx = 1$

Push-forward. Let $X, Y \subseteq \mathbb{R}^d$ and $T : X \to Y$ Let f be a probability density on Y. We say that g is the *push-forward* of f under T and write $g = T\#f$ if

$$\int_B g(y)dy = \int_{T^{-1}(B)} f(x)dx$$

$\forall B \subseteq Y$. In other words, the mass of the set B with respect to the density g equals the mass of the set $T^{-1}(B)$ with respect to the density f.

We will compute a quick example of the push-forward using a function where g is the push-forward of f and conversely we will also see an example where this is not the case. We will also be using a lemma proved in the next section:

Let $T : \mathbb{R} \to \mathbb{R}$ be the translation $T(x) = x + 1$.Let $f = \chi_{[0,1]}$ and $g = \chi_{[1,2]}$ be probability densities on $\mathbb{R}$. Show that $T\#f = g$. Define $S : \mathbb{R} \to \mathbb{R}$ by $S(x) = 2x$ show that $S\#f \neq g$.

Proof. Using Lemma 3.4 we need to show

$$\int_Y \phi(y)g(y)dy = \int_X \phi(T(x))f(x)dx$$

Let $X = Y = \mathbb{R}$ let $\phi : Y \to \mathbb{R}$ be bounded

$$\int_X \phi(T(x))f(x)dx = \int_0^1 \phi(x+1)dx \text{ since } \chi_A := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

where $A = [0,1]$. Using change of variables, take $y = x + 1 = T(x)$, $dy = dx$ then we have

$$\int_0^1 \phi(x+1)dx = \int_1^2 \phi(y)dy = \int_Y \phi(y)g(y)dy$$

By definition. So $T\#f = g$.
Let $S(x) = 2x$ and let $X = Y = \mathbb{R}$ let $\phi : Y \to \mathbb{R}$ be bounded then we have

$$\int_X \phi(S(x))f(x)dx = \int_0^1 \phi(2x)dx$$

Using change of variables take $y = 2x = S(x)$, $dy = 2dx$ then we have $\int_0^2 \phi(y)\frac{dy}{2}$ equals the previous integral. So,

$$\int_X \phi(S(x))f(x)dx \neq \int_Y \phi(y)g(y)dy$$

and $S\#f \neq g$. □

We will now see a few definitions, theorems, and examples regarding the convexity and concavity of certain functions, this will help us later on when computing costs and transport maps.

Definition 2.2 (Convex and concave functions). Let $I \subseteq \mathbb{R}$ be an interval (possibly unbounded). We say that $h : I \to \mathbb{R}$ is *convex* if for all $\lambda \in (0,1), x, y \in I, x \neq y$,

$$h((1-\lambda)x + \lambda y) \leq (1-\lambda)h(x) + \lambda h(y)$$

We say that h is *strictly convex* if the inequality is strict. We say that h is *concave* if $-h$ is convex and *strictly concave* if $-h$ is strictly convex. Convexity of h means that the graph of h on the interval (x, y) lies below the line joining $h(x)$ to $h(y)$ for all $x, y \in I$. Strict convexity means the graph lies strictly below the line. For concavity and strict concavity, the same is true with *below* replaced by *above.*

Examples of convex and concave functions of $\mathbb{R}$. The function $h_1(x) = x^2$ is strictly convex, $h_2(x) = |x|$ is convex but not strictly convex, $h_3(x) = ax + b, a, b \in \mathbb{R}$ is both convex and concave (but strictly not strictly convex or concave), $h_4(x) = -x^2$ is strictly concave, $h_5(x) = x^3$ is neither convex nor concave.

We will use the following theorem and for proof see [3]

Theorem 2.4 (Second-derivative test). Let $h : I \to \mathbb{R}$ be twice differentiable. Then h is convex if and only if $h"(x) \geq 0 \ \forall x \in I$ if $h"(x) > 0$ with $\forall x \in I$ then h is strictly convex.

We can find an example of a strictly convex function $h : \mathbb{R} \to \mathbb{R}$ such that $h''(x) = 0$ for some $x \in \mathbb{R}$.

Consider the function $h(x) = x^4$ let $\lambda \in (0,1)$ and $(x,y) \in I, \ x \neq y$

$$
\begin{aligned}
h((1-\lambda)x + \lambda y) &= ((1-\lambda)x + \lambda y)^4 \\
&= (((1-\lambda)x + \lambda y)^2)^2 \\
&< ((1-\lambda)(x^2) + \lambda(y^2))^2 && \text{follows from } x^2 \text{is strict convex} \\
&< ((1-\lambda)(x^2)62) + \lambda(y^2)^2) \\
&= (1-\lambda)x^4 + \lambda y^4
\end{aligned}
$$

so we have $h((1-\lambda)x + \lambda y) < (1-\lambda)h(x) + h(y)$ which means $h(x) = x^4$ is strictly convex. We have that $h''(x) = 12x^3$ and $h''(0) = 0$ so we have shown that strictly convex does not imply $h'' > 0$

A few quick examples using the second derivative test: We can show that $h_6(x) = xlogx$ with $x \in (0, \infty)$ is strictly convex and we can Show that $h_7(x) = x^{\frac{1}{2}}$ with $x \in (0, \infty)$ is strictly concave.

Proof. For $h_6(x) = xlogx$ we use the second derivative test. $h_6'(x) = logx + 1$ and $h_6''(x) = \frac{1}{x}$ which is strictly greater than zero in our domain, so by the second derivative test $h_6(x)$ is strictly convex.

To show that $h_7(x) = x^{\frac{1}{2}}$ for $x \in (0, \infty)$ is strictly concave we will show that $-h_7(x) = -x^{\frac{1}{2}}$ is strictly convex. We will use the second derivative test.

$-h_7'(x) = \frac{-1}{2}x^{\frac{-3}{2}}$ and $-h_7''(x) = \frac{3}{2}x^{\frac{-5}{2}}$ which is strictly greater than zero in our domain so $-h_7(x) = x^{\frac{1}{2}}$ is strictly convex which means $h_7(x) = x^{\frac{1}{2}}$ is strictly concave.

□

We will use the following inequality in the next section which will indeed prove quite helpful, see [3].

Theorem 2.7 (Jensen's inequality). Let $h : I \to \mathbb{R}$ be convex, let f be a probability density on $[a,b]$ and let $u : [a,b] \to I$ be bounded. then

$$
h(\int_a^b u(x)f(x)dx) \leq \int_a^b h(u(x))f(x)dx.
$$

3 The Monge Problem

We are now in a position to state Monge's optimal transport problem in modern mathematical language:

Definition 3.1 (The Monge problem). Let $X, Y \subseteq \mathbb{R}^d$. Let f be a probability density on Y. Let $c : X \times Y \to [0, \infty)$ be continuous. The *Monge problem* is to find a transport map $T : X \to Y$ satisfying $T\#f = g$ such that T minimises the cost function

$$M(T) := \int_X c(x, T(x)) f(x) dx$$

The *optimal transport cost* $\tau_c(f, g)$ of transporting f to g with cost function c is defines by

$$\tau_c(f, g) := \inf_{T\#f=g} M(T).$$

We write *inf* in the definition of $\tau_c(f, g)$ rather than *min* since the minimum may not exist. We will consider the following fundamental questions: Does there exist an optimal transport map T? If so, is it unique? Can we find an explicit expression for T? The answers to these questions will depend on the cost c and the probability densities f and g.

We also begin with a few quick remarks on the Monge Problem:

Remark 3.2(Physical interpretation). Let's interpret Definition 3.1 in terms of Gaspard Monge's original problem of redistributing (transporting and reshaping) a pile of sand or soil to form an embankment with minimal effort: $X = Y = \mathbb{R}^3$; $c(x, y)$ is the cost of moving sand from point x to y (a natural choice is $c(x, y) = |x - y|$); f is the density of the original pile of sand, i.e., $\int_A f(x)dx$ is the mass of sand occupying the set A in the original pile; g is the density of the target distribution (the embankment), i.e., $\int_B g(y)dy$ is the mass of sand occupying the set B in the embankment; $\int_X f(x)dx = \int_Y g(y)dy = 1$ is the total mass of sand (normalized without loss of generality to be 1); T is the transport map-sand at point x in the original pile in transported to point $T(x)$ in the embankment; and the total cost of moving the sand is $M(T)$. The constraint $T\#f = g$ represents conservation of mass- no sand is created or lost on the transportation process:

$$\int_{T^{-1}(B)} f(x)dx = \int_B g(y)dy$$

for all B in Y. Which means that the mass of sand transported from the original pile to B equals the mass of the sand in the embankment at B.

Remark 3.3(The Monge problem for more general densities). The assumptions $\int_X f(x)dx = 1, \int_Y g(y)dy = 1$ are not strictly necessary. The Monge problem can also be defined if f and g simply have the same total mass, not necessarily equal to 1: $\int_X f(x) = \int_Y g(y)dy$. If $\int_X f(x)dx \neq \int_Y g(y)dy$ then there does not exist any admissible map T satisfying $T\#f = g$ and so $\tau_c(f, g) = \infty$

The following is a useful lemma which makes checking the push forward much easier, indeed it had proved useful in the previous section and we will work through another example here:

Lemma 3.4(Equivalent formulation of the push-forward constraint). Let $X, Y \subseteq \mathbb{R}^d$ and $T : X \to Y$. Let f be a probability density on X and g be a probability density on Y. Then $T\#f = g$ if and only if

$$\int_Y \phi(y)g(y) = \int_X \phi(T(x))f(x)dx$$

(3.1) for all bounded functions $\phi : Y \to \mathbb{R}$

Proof. Suppose that equation (3.1) holds. Let $B \subseteq \mathbb{R}^d$ and choose $\phi = \chi_B$. Then (3.1) reduces to (2.1) and so $T\#f = g$. We just sketch the proof of the other direction. Suppose that $T\#f = g$. then the equation (3.1) holds for all characteristic functions $\phi = \chi_B$ for B in Y. By linearity of the integral, it also holds for all simple characteristic functions of the for $\phi = \sum_{i=1}^N a_i \chi_B$ for $a_i \in \mathbb{R}$ and $B_i \subseteq Y$. For general bounded functions $\phi : Y \to \mathbb{R}$ equation (3.1 can be proved by approximating ϕ by a sequence of simple functions.

□

A quick application of lemma 3.4:

Define $T : \mathbb{R} \to \mathbb{R}$ by $T(x) = 2 - x$. let $f = \chi_{[0,1]}$ and $g = \chi_{[1,2]}$. We can then show that $T\#f = g$.

Proof. Let $X = Y = \mathbb{R}$ and let $\phi : Y \to \mathbb{R}$ be bounded. We need to show that

$$\int_Y \phi(y)g(y)dy = \int_X \phi(T(x))f(x)dx$$

we have

$$\int_X \phi(T(x))f(x)dx = \int_0^1 \phi(2-x)dx$$

We use change of variables to get $y = 2 - x = T(x), dy = -dx$.

$$\int_0^1 \phi(2-x)dx = \int_2^1 -\phi(y)dy = \int_1^2 \phi(y)dy = \int_Y \phi(y)g(y)dy$$

□

We can also compare transport costs and find the ones that are optimal:

Let $X = [0,1], Y = [1,2]$ let $f(x) = 1, x \in X$ and $g(y) = 1, y \in Y$. We compare the transport cost of three different transport maps. Let T_1 be the translation $T_1(x) = x+1$, which transports all the mass the same distance, 1. Let $T_2(x) = 2 - x$ which flips or reflects the mass about the point $x = 1$. The point $x = 1$ is transported distance 0 while the point $x = 0$ is transported distance 2. We could also combine translation and flipping, e.g.,

$$T_3(x) = \begin{cases} x + \frac{3}{2} & \text{if } x \in [0, \frac{1}{2}] \\ 2 - x & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$

Which of these maps, if any, are optimal? The answer depends of course on the cost c. Let $c(x,y) = h(y-x)$. We will compare costs $h(s) = s^2$ (which is strictly convex), $h(s) = |s|$, (which is convex but not strictly convex), and $h(s) = |s|^{\frac{1}{2}}$, (which is concave for $s \geq 0$). If $h(s) = s^2$, then

$$M(T_1) = \int_0^1 (T_1(x) - x)^2 f(x)dx = \int_0^1 1^2 dx = 1$$

$$M(T_2) = \int_0^1 (T_2(x) - x)^2 f(x)dx = \int_0^1 (2-2x)^2 dx = \frac{4}{3}$$

$$M(T_3) = \int_0^1 (T_3(x) - x)^2 f(x)dx = \int_0^{\frac{1}{2}} (\frac{3}{2})^2 dx + \int_{\frac{1}{2}}^1 (2 - 2x)^2 dx = \frac{31}{24}$$

We will check the values in the table in the next example:

$h(s)$	$M(T_1)$	$M(T_2)$	$M(T_3)$
s^2	1	$\frac{4}{3}$	$\frac{31}{24}$
$\lvert s\rvert$	1	1	1
$\lvert s\rvert^{\frac{1}{2}}$	1	$\frac{2^{\frac{3}{2}}}{3}$	$\frac{3^{\frac{1}{2}}}{2^{\frac{3}{2}}} + \frac{1}{3}$

For the cost $h(s) = s^2$ we have $M(T_1) < M(T_3) < M(T_2)$. Therefore the translation T_1 is the best map amongst these three maps. In fact we can prove it is the best map amongst all admissible map as follows: Let T be any admissible map, $T\#f = g$. Since h is convex, then by Jensen's inequality, Theorem 2.7,

$$\begin{aligned}
M(T) &= \int_0^1 h(T(x) - x)f(x)dx \\
&\geq h(\int_0^1 (T(x) - x)f(x)dx) \\
&= h(\int_0^1 T(x)f(x)dx - \int_0^1 xf(x)dx) \\
&= h(\int_1^2 yg(y)dy - \int_)^1 xf(x)dx) \\
&= h(\frac{3}{2} - \frac{1}{2}) \\
&= h(1) \\
&= \int_0^1 h(1)f(x)dx \\
&= \int_0^1 h(T_1(x) - x)f(x)dx \\
&= M(T_1)
\end{aligned}$$

Therefore $M(T) \geq M(T_1)$ for all admissible transport maps T and so T_1 is an optimal transport map and the optimal transport cost is $\tau_c(f, g) = M(T_1) = 1$. In fact it can be shown that T_1 is the *unique* optimal transport map. In the argument above we only used the convexity of h, but not the explicit form of h. Therefore the translation T_1 is an optimal transport map for any convex cost. For the cost $h(s) = |s|$ we have $M(T_1) = M(T_2) = M(T_3) = 1$. Can we do better than this? The answer is no since the map $h(s) = |s|$ is convex and so the argument above shows that $\tau_c(f, g) = M(T_1) = 1$. Therefore all three transport maps T_1, T_2, T_3 are optimal. It turns out that any admissible transport map T is optimal: $M(T) = 1$ for all T. This is shown later. The lack of uniqueness is due to the lack of convexity of h.

Finally, consider the cost $h(s) = |s|^{\frac{1}{2}}$ which is concave for $s \geq 0$. In this case $M(T_2) < M(T_3) < M(T_1)$ and flipping the mass is better than translating it. We will see below that T_2 is an optimal transport map whereas T_1 is the worst transport map.

We will now check the values in the table. we can also use Jensen's inequality to prove that T_1 is the worst transport map for the concave cost $h(s) = |s|^{1/2}$

Proof. Solutions for $h(s) = s^2$

$$M(T_1) = \int_0^1 (T_1(x)) - x)^2 f(x)dx = \int_0^1 1dx = 1$$

$$M(T_2) = \int_0^1 (T_2(x)) - x)^2 f(x)dx = \int_0^1 (2-2x)^2 dx = \int_0^1 (4 - 8x + 4x^2)dx = \frac{4}{3}$$

$$M(T_3) = \int_0^1 (T_3(x)) - x)^2 f(x)dx = \int_0^{\frac{1}{2}} (\frac{3}{2})^2 dx + \int_{\frac{1}{2}}^1 (2-2x)^2 dx = \frac{9}{8} + \frac{1}{6} = \frac{31}{24}$$

Solutions for $h(s) = |s|$

$$M(T_1) = \int_0^1 |T_1(x) - x| f(x)dx = \int_0^1 1dx = 1$$

$$M(T_2) = \int_0^1 |T_2(x) - x| f(x)dx = \int_0^1 |2-2x|dx = \int_0^1 (2-2x)dx = 1$$

$$M(T_3) = \int_0^1 |T_3(x) - xS| f(x)dx = \int_0^{\frac{1}{2}} (\frac{3}{2})dx + \int_{\frac{1}{2}}^1 |2-2x|dx = \frac{3}{4} + \frac{1}{4} = 1$$

Solutions to $h(s) = |s|^{\frac{1}{2}}$

$$M(T_1) = \int_0^1 |T_1(x) - x|^{\frac{1}{2}} f(x)dx = \int_0^1 |1|^{\frac{1}{2}} dx = 1$$

$$M(T_2) = \int_0^1 |T_2(x) - x| f(x)dx = \int_0^1 |2-2x|^{\frac{1}{2}} dx = \int_0^1 (2-2x)^{\frac{1}{2}} dx$$

Letting $u = 2 - 2x, du = -2dx$ we now have:

$$\frac{-1}{2} \int_0^1 (u)^{\frac{1}{2}} du = \frac{2}{3} 2^{\frac{1}{2}}$$

$$M(T_3) = \int_0^1 |T_3(x) - x|^{\frac{1}{2}} f(x)dx = \int_0^{\frac{1}{2}} (\frac{3}{2})^{\frac{1}{2}} dx + \int_{\frac{1}{2}}^1 |2-2x|^{\frac{1}{2}} dx = \frac{1}{2}^{\frac{1}{2}} + \frac{1}{3}$$

We will now show that T_1 is the worst transport map for the concave cost $h(s) = |s|^{\frac{1}{2}}$ Recall Jensen's Inequality:

$$h(\int_a^b u(x)f(x)dx) \leq \int_a^b h(u(x)f(x)dx$$

since h is concave we have that -h is convex. Consider and admissible Transport map T with $T\#f = g$. We then have:

$$-M(T) = -\int_0^1 |T(x)-x|^{\frac{1}{2}} f(x) dx with u(x) = T(x) - x$$
$$\geq -(\int_0^1 |T(x)-x| f(x)dx)^{\frac{1}{2}}$$

by Jensen's inequality

$$= -(\int_0^1 T(x)f(x)dx - \int_0^1 xf(x)dx)^{\frac{1}{2}}$$
$$= -(\int_1^2 yg(y)dy - \int_0^1 xf(x)dx)^{\frac{1}{2}}$$

using 3.1 and 3.4 with

$$\phi(y) = y$$
$$= -(\frac{3}{2} - \frac{1}{2})^{\frac{1}{2}}$$
$$= -(1)^{\frac{1}{2}} = -h(1)$$

so we have $-M(T) \geq -M(T_1)$ which implies $M(T) \leq M(T_1)$ and so T_1 is the worst transport map for the concave cost $h(s) = |s|^{\frac{1}{2}}$

□

Given enough information we can also compute an optimal transport map, consider $X = [0,1], Y = [1,2], f = \chi_{[0,1]}, g = \chi_{[1,2]}, c(x,y) = h(|y-x|)$ with $h(s) = (s+1)log(s+1), s \geq 0$. We can then find $T(x)$

Proof. Let $T_1(x) = x+1$. We have that $h(s) = (s+1)log(s+1)$ is strictly convex by the second derivative test, $h'(s) = log(s+1)+1$ and $h''(s) = \frac{1}{s+1} > 0$ for $s \geq 0$ so we can apply Jensen's inequality and we can see from the previous example that $M(T) \leq M(T_1)$. So $T_1(X) = X+1$ is an optimal transport map.

□

In fact, we can go even further and say that *every* admissible transport map is optimal (Note that these values are taken from 5 and section 2):

Let $X, Y \subset \mathbb{R}$ be bounded and $c(x,y) = h(y-x)$ $h : X \to Y$ is a linear function i.e., if $T : X \to Y$, $T\#f = g$ then $M(T) = \mathcal{T}_c(f,g)$.

Proof. We now compute $M(T)$

$$
\begin{aligned}
M(T) &= \int_X c(x, T(x)) f(x) dx && \text{Letting y=T(x)} \\
&= \int_X h(T(x) - x) f(x) dx \\
&= \int_X (h(T(x)) f(x) - h(x) f(x)) dx \\
&= \int_X h(T(x)) f(x) dx - \int_X h(x) f(x) dx \\
&= \int_Y h(y) g(y) dy - \int_X h(x) f(x) dx
\end{aligned}
$$

So we have $M(T) = \mathcal{T}_c(f, g)$

□

Here we have a quick example in non-uniqueness for non-strictly convex costs taken from [5] and section 2: Let $X = [0, 2], Y = [1, 3],\ f = \frac{1}{2}\chi_{[0,2]}, g = \frac{1}{2}\chi_{[1,3]},\ c(x, y) = h(y - x), h(s) = |s|$. Let $T_1(x) = x + 1$ and $T_2(X) = \begin{cases} x + 2 & \text{if } x \in [0, 1] \\ x & \text{if } x \in (1, 2] \end{cases}$ We can the show that T_1 and T_2 are both optimal transport maps $M(T_1)$=$M(T_2)$=$\tau_c(f, g)$

Proof.

$$
\begin{aligned}
M(T_1) &= \int_0^2 c(x, T_1(x)) f(x) dx \\
&= \int_0^2 |(T_1(x) - x)| \frac{1}{2} dx \\
&= \int_0^2 |x + 1 - x| \frac{1}{2} dx \\
&= \int_0^2 \frac{1}{2} dx \\
&= 1
\end{aligned}
$$

$$
\begin{aligned}
M(T_2) &= \int_0^2 c(x, T_2(x)) f(x) dx \\
&= \int_0^1 |x + 1 - x| \frac{1}{2} dx + \int_1^2 |x + 1 - x| \frac{1}{2} dx \\
&= \int_0^1 (2) \frac{1}{2} dx \\
&= 1
\end{aligned}
$$

So we have $M(T_1) = M(T_2)$

Indeed we have that these maps are both optimal since for any admissible map T with $T\#f = g$ we have:

$$\begin{aligned}\int_0^2 |T(x) - x|f(x)dx &\geq |\int_0^2 (T(x) - x)f(x)dx| \\ &= |\int_0^2 T(x)f(x)dx - \int_0^2 xf(x)dx| \\ &= |\int_1^3 yg(y)dy - \int_0^2 xf(x)dx| \\ &= |\frac{1}{2}\int_1^3 ydy - \frac{1}{2}\int_0^2 xdx| \\ &= 1\end{aligned}$$

□

here we have an example in non-existence for a strictly concave cost with overlapping masses:

Let $X = [0,1], Y = [0,2], f = \chi_{[0,1]}, g = \frac{1}{2}\chi_{[0,2]}, c(x,y) = |x-y|^{\frac{1}{2}}$. In this case it can be shown that there does not exist any optimal transport map: the infimim in the definition of $\tau_c(f,g)$ is not attained. It turns out that

$$\tau_c(f,g) = \tau_c(\frac{1}{2}\chi_{[0,1]}, \frac{1}{2}\chi_{[1,2]})$$

In other words, the cost of transporting f to g is the same as the cost of transporting $\frac{1}{2}\chi_{[0,1]}$ to $\frac{1}{2}\chi_{[1,2]}$. The problem here is that c is strictly concave and the masses f and g overlap; f and g are both positive on the interval $[0,1]$. Whenever c is a strictly concave metric, as it is here, then it turns out that it is best to leave 'common mass' where it is. But since any function must take a single value at every point, it is not possible to find a function T that both leaves the common mass $\frac{1}{2}\chi_{[0,1]}$ fixed ($T(x) = xon[0,1]$ and transports the mass $\frac{1}{2}\chi_{[0,1]}$ to $\frac{1}{2}\chi_{[1,2]}$ ($T([0,1]) = [1,2]$).

We can also observe quadratic transport under translations (this is taken from 6 and remark 2.19): Let $X = Y = \mathbb{R}$ and c be the quadratic cost $c(x,y) = (x-y)^2$. For $a \in \mathbb{R}$, define the translation $\tau_a : \mathbb{R} \to \mathbb{R}$ by $\tau_a(x) = x - a$. Let $f \circ \tau_a$ denote the composition $(f \circ \tau_a)(x) = f(\tau_a(x)) = f(x-a)$. In this exercise we show that

$$\tau_c(f \circ \tau_a, g \circ \tau_b) = \tau_c(f,g) + (b-a)^2 + 2(b-a)(m_g - m_f)$$

where $a, b \in \mathbb{R}$ and

$$m_f = \int_\infty^\infty xf(x)dx$$

and

$$m_g = \int_\infty^\infty yg(y)dy$$

are the centres of mass of f and g.

(i) Let $T\#f = g$. Define $S : \mathbb{R} \to \mathbb{R}$ by $S(x) = T(x-a) + b$. Show that $S\#(f \circ \tau_a)$.

(ii) Show that

$$\tau_c(f \circ \tau_a, g \circ \tau_b) \leq \tau_c(f, g) + (b-a)^2 + 2(b-a)(m_g - m_f)$$

(iii) Use a similar argument to show

$$T_c(f \circ \tau_a, g \circ \tau_b) \geq \tau_c(f, g) + (b-a)^2 + 2(b-a)(m_g - m_f)$$

(iv) Use (3.2) to give an alternative proof that $\tau_c(\chi_{[0,1]}, \chi_{[1,2]}) = 1$

Proof. (i) $S\#(f \circ \tau_a)(x)dx = g \circ \tau_b$ if and only if $\int_Y \phi(y)(g \circ \tau_b)(y)dy = \int_X \phi(S(x))(f \circ \tau_a)(x)dx$.

$$\int_X \phi(S(x)(f \circ \tau_a)(x)dx = \int_{-\infty}^{\infty} \phi(T(x-a)+b)f(x-a)dx$$

Let $x_1 = x - a$ then $x_1 + a = x$ and $dx_1 = dx$

$$= \int_{\infty}^{\infty} \phi(y+b)g(y)dy$$

Let $y_1 = y + b$ then $y_1 = y + b$ and $dy_1 = dy$

$$= \int_{\infty}^{\infty} \phi(y_1)g(y_1 - b)dy_1$$
$$= \int_{\infty}^{\infty} \phi(y_1)(g \circ \tau_b)(y_1)dy_1$$

as required.

(ii)Let T be an optimal transport map that transports f to g so we have $\tau_c(f,g) = \int_{-\infty}^{\infty} |T(x) - x|^2 f(x)dx$ and let $S(x) = T(x-a)+b)$ then $S\#(f \circ \tau_a) = (g \circ \tau_b)$

$$\begin{aligned}
\tau_c(f \circ \tau_a, g \circ \tau_b) &\leq M(S) \\
&= \int_{-\infty}^{\infty} (S(x) - x)^2 (f \circ \tau_a)(x)dx \\
&= \int_{-\infty}^{\infty} (T(x-a) + b - x)^2 f(x-a)dx
\end{aligned}$$

Let $x_1 = x - a$,

$$\begin{aligned}
&= \int_{-\infty}^{\infty} (T(x_1) + b - x_1 - a)^2 f(x_1)dx \\
&= \int_{-\infty}^{\infty} ((T(x_1) - x_1)^2 + (b-a)^2 + 2(b-a)(T(x_1) - x_1))f(x_1)dx_1 \\
&= \int_{-\infty}^{\infty} (T(x_1) - x_1)^2 f(x_1)dx_1 + \int_{-\infty}^{\infty} (b-a)^2 f(x_1)dx_1 + \int_{-\infty}^{\infty} 2(b-a)(T(x_1) - x_1))f(x_1)dx_1 \\
&= \tau_c(f,g) + (b-a)^2 + 2(b-a)(\int_{-\infty}^{\infty} (T(x_1)f(x_1)dx_1 - \int_{-\infty}^{\infty} (x_1)f(x_1)dx_1)
\end{aligned}$$

Since $T\#f = g$,

$$= \tau_c(f,g) + (b-a)^2 + 2(b-a)(\int_{-\infty}^{\infty} yg(y)dy - \int_{-\infty}^{\infty} (x_1)f(x_1)dx_1)$$
$$= \tau_c(f,g) + (b-a)^2 + 2(b-a)(m_g - m_f)$$

(iii) We will use a similar argument as in (ii). Let T be an optimal transport map that transports $f \circ \tau_a$ to $g \circ \tau_b$ so $\tau_c(f \circ \tau_a, g \circ \tau_b) = \int_{-\infty}^{\infty} |T(x) - x|^2 f \circ \tau_a(x)dx$

and let $S(x) = T(x+a) - b$ by similar argument in part (i) we have $S\#f = g$. So we have that

$$\tau_c(f,g) \leq M(S)$$
$$= \int_{-\infty}^{\infty} (S(x) - x)^2 f(x)dx$$
$$= \int_{-\infty}^{\infty} (T(x+a) - b - x)^2 f(x)dx$$

Let $x_1 = x + a$,

$$= \int_{-\infty}^{\infty} (T(x_1) - b - x_1 + a)^2 f(x_1 - a)dx_1$$
$$= \int_{-\infty}^{\infty} (T(x_1) - x_1)^2 f(x_1 - a)dx_1 - \int_{-\infty}^{\infty} (b-a)f(x_1 - a)dx_1 - \int_{-\infty}^{\infty} 2(b-a)(T(x_1) - x_1)f(x_1 - a)dx_1$$
$$= \tau_c(f \circ \tau_a, g \circ \tau_b) - (b-a) - 2(b-s)\left(\int_{-\infty}^{\infty} (T(x_1)f(x_1 - a)dx_1 - \int_{-\infty}^{\infty} (x_1)f(x_1 - a)dx_1\right)$$

Since $T\#f = g$,

$$= \tau_c(f \circ \tau_a, g \circ \tau_b) - (b-a) - 2(b-s)\left(\int_{-\infty}^{\infty} yg(y)dy - \int_{-\infty}^{\infty} (x_1)f(x_1 - a)dx_1\right)$$
$$= \tau_c(f \circ \tau_a, g \circ \tau_b) - (b-a) - 2(b-a)(m_g - m_f)$$

as required.

(iv) we use the previous parts with $f = g = \chi_{[0,1]} a = 0 b = 1$ with $f \circ \tau_a = \chi_{[0,1]} g \circ \tau_b = \chi_{[1,2]}$ and $m_g = m_f = \frac{1}{2}$

$\tau_c(\chi_{[0,1]}, \chi_{[1,2]}) = 0 + (1-0)^2 + 2(1-0)(\frac{1}{2} - \frac{1}{2})$ □

4 The Dual Problem

In this section we will see that the Monge minimisation problem can be reformulated as a maximisation problem. This is not just a mathematical curiosity, it was an important step along the road to solving Monge's problem. Throughout this section we assume that $X, Y \subset \mathbb{R}^d$ are compact. Let $C(X)$ denote the set of continuous , real-valued functions on X

We begin with an important theorem for duality:

Theorem 4.1 (Kantorovich Duality Theorem). Let f be a probability density on X, g be a probability density on Y and $c : X \times Y \to \mathbb{R}$ be continuous. Define $D : C(X) \times C(Y) \to \mathbb{R}$ by

$$D(\phi, \psi) = \int_X \phi(x) f(x) dx + \int_Y \psi(y) g(y) dy$$

If $\phi \in C(X), \psi \in C(Y)$ we say that $\phi \bigoplus \psi \leq c$ if and only if $\phi(x) + \psi(y) \leq c(x, y)$ for all $x \in X, y \in Y$. Then

$$\tau_c(f, g) = \inf_{T\#f=g} M(T) = \sup_{\phi \bigoplus \psi \leq c} D(\phi, \psi)$$

Moreover, the supremum is a maximum, i.e., there exists an admissible pair (ϕ, ψ) such that $\tau_c(f, g) = D(\phi, \psi)$ and we say that (ϕ, ψ) is an optimal Kantorovich potential pair.

Proof. We will limit ourselves to the proof of the duality equality:

$$\inf_{T\#f=g} M(T) \geq \sup_{\phi \bigoplus \psi \leq c} D(\phi, \psi)$$

Let T satisfy $T\#f = g$ and (ϕ, ψ) satisfy $\phi \bigoplus \psi \leq c$ then

$$\begin{aligned}
D(\phi, \psi) &= \int_X \phi(x) f(x) dx + \int_Y \psi(y) g(y) dy \\
&= \int_X \phi(x) f(x) dx + \int_X \psi(T(x)) f(x) dy \\
&= \int_X (\phi(x) + \psi(T(x))) f(x) dx \\
&\leq \int_X c(x, T(x)) f(x) dx \\
&= M(T)
\end{aligned}$$

We record for future use that

$$D(\phi, \psi) = \int_X (\phi(x) + \psi(T(x))) f(x) dx \leq M(T)$$

(4.1)

In particular

$$D(\phi, \psi) \leq M(T)$$

for all T satisfying $T\#f = g$ and all (ϕ, ψ) satisfying $\phi \bigoplus \psi \leq c$. Taking the supremum over all admissible (ϕ, ψ) gives

$$\sup_{\phi \bigoplus \psi \leq c} D(\phi, \psi) \leq M(T)$$

Then taking the infimum gives

$$\sup_{\phi \bigoplus \psi \leq c} D(\phi, \psi) \leq \inf_{T\#f=g} M(T)$$

as required. □

Remark 4.2 (the constraint $\phi \bigoplus \psi \leq c$). By examining the proof of Theorem 4.1 more closely we see that the constraint $\phi(x) + \psi(y) \leq c(x,y)$ does not need to hold for all $x \in X, y \in Y$, but only for x, y such that $f > 0$ in a neighbourhood of y.

Indeed we can show that if $\phi \bigoplus \psi \leq c$ and $a \in \mathbb{R}$ then $(\phi+a) \bigoplus (\psi-a) \leq c$ and $D(\phi+a, \psi-a) = D(\phi, \psi)$. Therefore we can conclude that if (ϕ, ψ) is an optimal Kantorovich potential pair then so is $(\phi + a, \psi - a)$ for any $a \in \mathbb{R}$

Proof. $\phi(x) + a + \psi(x) - a = \phi(x) + \psi(x) \leq c(x,y)$ so indeed we have $(\phi + a) \bigoplus (\psi - a) \leq c$.
We now need to show $D(\phi + a, \psi - a) = D(\phi, \psi)$

$$\begin{aligned} D(\phi + a, \psi - a) &= \int_X (\phi(x) + a) f(x) dx + \int_Y (\psi(y) - a) g(y) dy \\ &= \int_X (\phi(x) f(x) dx + \int_Y (\psi(y) g(y) + a(\int_X f(x) dx - \int_Y g(y) dy \\ &= D(\phi, \psi) + a(1 - 1) \\ &= D(\phi, \psi) \end{aligned}$$

□

The following is a useful corollary of the Kantorovich Duality Theorem and will help us with a few examples later on:

Corollary 4.4 Let T be a continuous optimal transport map and (ϕ, ψ) be a differentiable optimal Kantorovich potential pair, in particular $M(T) = D(\phi, \psi) = \tau_c(f, g)$. Assume that c is differentiable in its first argument. Then

$$\phi(x_0) + \psi(T(x_0)) = c(x_0, T(x_0))$$

where $\triangledown \phi(x_0) = \triangledown_x c(x_0, T(x_0))$
for all $x_0 \in X$ such that $f > 0$ in a neighbourhood of x_0. In particular, we have equality in the inequality constraint $\phi(x) + \psi(y) \leq c(x,y)$ if mass is transported from x to $y = T(x)$

Proof. Suppose for contradiction that we can find $x_0 \in X, r > 0$ such that $f > 0$ on the ball $B_r(x_0)$ and $\phi(x_0) + \psi(T(x_0)) < c(x_0, T(x_0))$. Then by continuity of all of our functions we have $\phi(x) + \psi(T(x)) < c(x, T(x))$. in some neighbourhood of x_0 where $f > 0$. Combining this with equation (4.1) gives

$$D(\phi, \psi) = \int_X (\phi(x) + \psi(T(x))) f(x) dx < \int_X c(x, T(x)) f(x) dx = M(T)$$

But this contradicts the optimality of T and (ϕ, ψ) Take and $x_0 \in X$ such that $f > 0$ in a neighbourhood of x_0. Consider the map $F(x) = c(x, T(x_0)) - \phi(x), x \in X$. We have shown that

$$F(x) \geq \psi(T(x_0)) \forall x \in X$$

$F(x_0) = \psi(T(x_0))$ therefore x_0 is a minimum point of F and so

$$\bigtriangledown F(x_0) = 0 \iff \bigtriangledown_x c(x_0, T(x_0)) - \bigtriangledown \phi(x_0)$$

as required. $\square$

Remark 4.5(Why the dual problem is useful.) Given a transport map T, how do we know if it is optimal? The Kantorovich Duality Theorem gives us a way of checking. Given any admissible transport map T and admissible Kantorovich potential pair (ϕ, ψ) we have

$$D(\phi, \psi) \leq \tau_c(f, g) \leq M(T)$$

by theorem 4.1. Therefore if we can construct T and (ϕ, ψ) such that $M(T) = D(\phi, \psi)$, then

$$D(\phi, \psi) \leq \tau_c(f, g) \leq M(T) = D(\phi, \psi) \implies D(\phi, \psi) = \tau_c(f, g) = M(T)$$

and so T and (ϕ, ψ) must be optimal. Even if we cant construct such a T and (ϕ, ψ), then the duality gap $M(T) - D(\phi, \psi) \geq 0$ gives us an idea of how far T and (ϕ, ψ) are from being optimal.

Here we have a table with a few potential pairs and transport maps, we can compare and we will be verifying the values in the next example:

Let $X = [0, 1], Y = [1, 2], f = \chi_{[0,1]}, g = \chi_{[1,2]}$. The following table gives optimal Kantorovich potentials and transport maps for the costs from the previous section:

$h(s)$	$T(x)$	$\phi(x)$	$\psi(x)$	$D(\phi, \psi) = M(T)$
s^2	$T_1(x) = x + 1$	$-2x$	$2y - 1$	1
$\lvert s\rvert$	$T_1(x) = 2 - x$	$-x$	y	1
$\lvert s\rvert^{\frac{1}{2}}$	$T_2(2 - x)$	$\frac{1}{2}(2 - 2x)^{\frac{1}{2}}$	$\frac{1}{2}(2y - 2)^{\frac{1}{2}}$	$\frac{2^{\frac{3}{2}}}{3}$

For example, for the cost $h(s) = s^2$, then

$$D(-2x, 2y - 1) = \int_0^1 (-2x)dx + \int_1^2 (2y - 1)dy = -1 + \frac{1}{4}(9 - 1) = 1$$

Therefore $D(-2x, 2y - 1) = M(x + 1) = \tau_c(f, g) = 1$ which again verifies that the translation $T_1(x) = x + 1$ is an optimal transport map for the quadratic cost. We can derive the potentials as follows: By Corollary 4.4, if T_1 and (ϕ, ψ) are optimal then

$$\phi'(x) = c_x(x, T_1(x)) = 2(x - T_1(x)) = -2, x \in [0, 1]$$

Integrating gives $\phi(x) = -2x + a$. We can choose $a = 0$ By exercise 4.3. Using Corollary 4.4 again (and again assuming that T_1 is optimal) gives

$$\psi(T_1(x)) = c(x, T_1(x)) - \phi(x) = (T_1(x) - x)^2 + 2x = 1 + 2x$$

By setting $y = T_1(x) = x + 1$ we find that

$$\psi(y) = 1 + 2(y - 1) = 2y - 1$$

Remark 4.7 (Kantorovich potentials for metric costs). Consider the special case where $Y = X$ and c is a metric on X, e.g., $X = \mathbb{R}^d$ and $c(x, y) = |y - x|$ or $c(x, y) = h(|y - x|)$ for a concave function $h : [0, \infty) \to [0, \infty)$ with $h(0) = 0$. In this case it can be shown that there exists an optimal Kantorovich potential pair (ϕ, ψ) with $\psi = -\phi$ and where ϕ is 1- Lipschitz with respect to c which means that $|\phi(x) - \phi(y)| \leq c(x, y)$ for all $x, y \in X$. For instance, if $c(x, y) = |y - x|$ then we can choose optimal Kantorovich potentials (ϕ, ψ) such that $\psi(y) = -\phi(y)$ and $|\phi(x) - \phi(y)| \leq |x - y|$

We will now fill in the missing details for our table

Proof. $D(-x, y) = \int_0^1 (-x)dx + \int_1^2 ydy = \frac{-1}{2} + \frac{4}{2} - \dfrac{1}{2} = 1$

Therefore $D(-x, y) = M(x + 1) = T_c(f, g) = 1$

$\phi'(x) = c_x(x, T_1(x) = (x - (x + 1) = -1$ for $x \in [0, 1]$ integrateing yields $\phi = -x + a$ we can set $a = 0$ by 4.3 and corollary 4.4 with being optimal we have, $\phi(T_1(x)) = c(x, T_1(x)) - \phi(x) = |T_1(x) - x| - \phi(x) = X + 1 - x + x = 1 + 1$

Let $y = T_1(x) = x + 1$ then $\phi(y) = 1 + x = 1 + y - 1 = y, y \in [1, 2]$

Consider the substitution $u = 2 - 2x$ and $v = 2y - s$ then we have $\frac{du}{-2} = dx$ and $\frac{dv}{2} = dy$

$$\begin{aligned} D(\frac{1}{2}u^{\frac{1}{2}}, \frac{1}{2}v^{\frac{1}{2}}) &= \int \frac{1}{2}u^{\frac{1}{2}} \frac{du}{-2} + \int \frac{1}{2}v^{\frac{1}{2}} \frac{dv}{2} \\ &= \frac{-1}{4}(\frac{u^{\frac{3}{2}}}{3}) + \frac{1}{4}(\frac{v^{\frac{3}{2}}}{3}) \\ &= \frac{-(2 - 2x)^{\frac{3}{2}}}{6} + \frac{(2y - 2x)^{\frac{3}{2}}}{6} \end{aligned}$$

and once we plug in our bounds we have $\frac{2^{\frac{1}{2}}}{3} + \frac{2^{\frac{1}{2}}}{3} = \frac{2(2^{\frac{1}{2}})}{3}$

Now we derive the potential pair $\phi'(x) = c_x(x, T_2(x) = 2(x - (2 - x)) = 2(2x - 2)$ so $\phi(x) = (2x - 2)^2 + a$ we can set $a = 0$

$$\begin{aligned} \psi(T_2(x)) &= c(x, T_2(x) - \phi(x) \\ &= | - (2 - x - x)^{\frac{1}{2}}| - \frac{1}{2}(2 - 2x)^{\frac{1}{2}} \\ &= (2 - 2x)^{\frac{1}{2}} - \frac{1}{2}(2 - 2x)^{\frac{1}{2}} \\ &= \frac{1}{2}(2 - 2x)^{\frac{1}{2}} \end{aligned}$$

let $y = 2 - x$ then $x = 2 - y$ $\psi(y) = \frac{1}{2}(2 - 2(2 - y))^{\frac{1}{2}} = \frac{1}{2}(2y - 2)^{\frac{1}{2}}$

□

We can now go back and derive an optimal Kantorovich potenial pair for the book shifting problem in Section 3.

Proof. we have $c(x, y) = h(y - x)h(s) = |s|$ and $T_1(x) = x + 1$ and (ϕ, ψ) are optimal. $\phi'(x) = c_x(x, T_1(x)) = sgn(x - x - 1) = sgn(-1) = -1$ integrating gives us $\phi(x) = -x + a$ we can set $a = 0$ by a previous example and use corollary 4.4 to get $\phi(x) = -x$.

$\psi(T_1(x)) = x + 1, \psi(y) = 1 + (y - 1) = y$ so we have $\phi(x) = -x$ and $\psi(y) = y$ □

We can also use Kantorovich potential pairs to to prove that T_2 is the worst transport map for the convex cost $h(s) = s^2$ from a previous example. This is equivalent to proving that T_2 is the best transport map for the concave cost $\check{h}(s) = -s^2$. We can verify this by constructing an optimal Kantorovich potential pair (ϕ, ψ) such that $D(\phi, \psi) = M(T_2)$ for the cost $\check{h}(s) = -s^2$

Proof. We will now show that $T_2(x) = 2 - x$ is the worst transport map for $\check{h}(s) = -s^2$. We begin by constructing an optimal Kantorovich pair ϕ, ψ

$\phi'(x) = c_x(x, T_2(x)) = 2(x - T_2(x)) = 2(x - (2 - x) = 2(2x - 2)$ so we have that $\phi(x) = 2(x - 1)^2 + a$ but we can set $a = 0$ as before to get $\phi(x) = 2(x - 1)^2$.

$$\begin{aligned} \psi(T_2(x)) &= c(x, T_2(x)) - \phi(x) \\ &= | - x - (2 - x)|^2 - 2(x - 1)^2 \\ &= (2x - 2)^2 - 2(x - 1)^2 \\ &= 2(x - 1)^2 \end{aligned}$$

Let $y = T_2(x) = 2 - x$ so we have $\phi(y) = 2(x - 1)^2 = 2(1 - y)^2 = 2(y - 1)^2$

we compute $D(\phi, \psi)$ to get: $\int_0^1 2(x - 1)^2 dx + \int_1^2 2(y - 1)^2 dy = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$ so we have $D(\phi, \psi) = M(T_2)$.

□

We will now return to our dinner party in the Introduction and distribute the desserts among the professors. We note that Jan is lactose intolerent and therefore can only eat the carrot cake, so we let all 3 of her portions be carrot cake. Next we move onto Bernard who also wants 3 portions and who has merveilleux and ecalair as his top 2 favourites. We will give him 1 eclair and 2 merveilleux and we can then give the second eclair to Willem. Now we must give 2 other portions to Willem and we note that his 2 other favourites are merveilleux and mousse, so we give him one of each. Hilde would like 4 portions and her favourite is merveilleux and her second favourite is a tie between eclair and bavarois. Since we are now out of eclair we can give her 2 merveilleux and 2 bavarois. Steffie only wants 2 portions and her favourites are a tie between ecalir and mousse. Luckily we still have one of her favourites so we will give her 2 mousse. Marlies would also only like 2 portions but she doesnt have a particular favourite and only gave a mid range score to eclair, mousse, and bavarois, so we can give her 1 bavarois and 1 mousse. Next is Tim who would also like 2 portions and who loves ecalir and mousse and who is the only other person who likes carrot cake, so we give him a carrot and a mousse. Lastly is Wouter who only wants 1 portion and luckly we have one mousse left which is one of his favourites so we give that one to him.

Works Cited

1. Bourne, D.P. "A Brief Introduction to Optimal Transport Theory" July 27th, 2018.http://www.maths.gla.ac.uk/ gb lamy/LMS/BourneLectures.pdf

2. Stock, Michiel "Notes on Optimal Transport" https://michielstock.github.io/OptimalTransport/

3. Stewat, James "Calculus" 8th edition May 19th 2015

4.G. Monge. Mémoire sur la théorie des déblais et des remblais. In Histoire de l'Académie Royale des Sciences de Paris, pages 666–704, 1781.

5. F. Santambrogio. Optimal Transport for Applied Mathematicians. Birkhäuser, 2015.

6.G. Peyré and M. Cuturi. Computational optimal transport. arXiv:1803.00567, 2018