OPTIMAL MASS TRANSPORT AND THE ISOPERIMETRIC INEQUALITY

KENNETH DEMASON

ABSTRACT. We provide an introduction to optimal mass transport, which has proved in recent years to be a powerful tool in studying geometric inequalities. In particular, we show a clever application to the isoperimetric inequality.

CONTENTS

1. INTRODUCTION

Optimal mass transport has an interesting history. It started off as a problem on "excavations and embankments" – how to transport soil efficiently during the building of forts. The original formulation was presented in 1781 by Gaspard Monge while the field of analysis was blooming.

Due to its difficult nature, no significant progress on the problem was made. In 1947, L.V. Kantorovich revisited it and applied it to economics. In this setting, the transport problem becomes one of shipping goods – given production sites and destinations, how can we efficiently ship goods so each destination receives the amount of goods it needs?

In order to progress transportation theory, Kantorovich established a weaker version of Monge's problem. This enabled him to work with a larger class of objects and obtain results. Later on, mathematicians such as Yann Brenier and Robert McCann were able to use Kantorovich's theory in specific contexts to find nice solutions to Monge problems.

Optimal mass transport has found continued use since then. In particular, one can use it to prove the isoperimetric problem. This long standing problem was effectively solved in the 1800s. Optimal mass transport can be used to prove it in a simple, clean way.

This paper focuses on developing the theory of optimal mass transport and showcases two applications of it in solving the isoperimetric problem.

Section 2 starts by presenting Monge's problem and several examples. These examples highlight the many ways Monge's problem can fail to have solutions. We then study Kantorovich's weaker formulation of Monge's problem and how it addresses several of the previous issues. Finally, we analyze a specific case and show Brenier and McCann's nice solutions to Monge's problem.

Section 3 starts with a history of the isoperimetric problem, which provides some nice background to the technical challenges. Following this, we provide two different proofs of the isoperimetric inequality using optimal mass transport.

Background in measure theory is assumed. The author recommends Chapters 2 and 8 of [Mag12]. All information, unless otherwise specified, comes from [Amb00], [Vil03], and [FMP10].

2. Optimal Mass Transport

2.1. **The Monge Problem.** We begin with the following picture. Suppose we have a pile of dirt that we wish to transfer into some hole. We can impose a "cost" associated with moving a speck of dirt – say, the distance it travels. Is there a way to move the pile into the hole such that the cost is minimized? Is this mapping unique? This is the basic formulation of the Monge problem.

In modern language, we can state it as follows. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be two nonnegative $L^1(\mathbb{R}^n)$ functions such that

$$\int_{\mathbb{R}^n} f \; dx = \int_{\mathbb{R}^n} g \; dy = 1.$$

Imagine f as a distribution of dirt – choosing a point $x \in \mathbb{R}^n$, $f(x)$ tells you how much dirt lies on top of x. Imagine g as our target "hole" – choosing a point $y \in \mathbb{R}^n$, $g(y)$ tells you how much dirt can fit in the "hole" at y. It is convenient to normalize the total amount of dirt as 1 so that f, g give rise to probability measures $\mu = f dx, \nu = g dy$. These measures are precisely those which are absolutely continuous with respect to the Lebesgue measure. Sometimes, the Monge problem is stated for probability measures in general – we will move freely between the two formulations. The use of x and y is simply cosmetic – x is used for the source space while y is used for the target space.

We wish to find a map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that two conditions hold:

i) The transport condition: For any $E \subset \mathbb{R}^n$ Borel,

$$\int_{T^{-1}(E)} f \; dx = \int_E g \; dy.$$

This essentially tells us that no dirt lying above $T^{-1}(E)$ is lost in the transportation, and that all the dirt lying above $T^{-1}(E)$ exactly fills up the hole at E. Such a map fulfilling this condition is called a *transport map.*

ii) The cost condition: Define the cost of T to be

$$C(T) = \int_{\mathbb{R}^n} |T(x) - x| f(x)\ dx.$$

The $|T(x)-x|$ part gives the transport cost for moving mass, while $f(x)dx$ tells us how much mass is being moved. The cost of T is then interpreted as the cost of moving all the mass.

For each Monge problem there is an associated intrinsic cost given by

$$M(f,g) = \inf\{C(T) \mid T \text{ is a transport map}\}.$$

This is called the *Monge cost*, and will sometimes be written as $M(\mu,\nu)$. The Monge cost measures the minimum theoretical cost. The cost condition requires that T satisfies $C(T) = M(f,g)$. That is, T actually realizes the theoretical minimum cost. So, T is as efficient as possible.

Such a map T is called an *optimal transport map*. We remark that the transport condition may be restated as $T_{\#}\mu = \nu$, where $T_{\#}\mu(E) := \mu(T^{-1}(E))$. That is, ν is the push-forward of μ under T. To see this, one appeals to the well-known change of variables formula between μ and ν,

$$\int_{T(E)} \varphi\ d\nu = \int_E \varphi \circ T\ d\mu$$

for measurable φ. In light of this, we can rewrite the Monge cost as follows:

$$M(\mu,\nu) = \inf_{T_{\#}\mu=\nu} \int_{\mathbb{R}^n} |T(x) - x|\ d\mu.$$

Interestingly, there exists a "duality principle" used to solve optimal mass transport problems. Let μ, ν be probability measures on $\mathbb{R}^n$ such that $T_{\#}\mu = \nu$. Let $u : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz with $\mathrm{Lip}(u) \leq 1$. Consider the quantity

$$\int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu.$$

Applying the above change of variables formula with $u = \varphi$ yields

$$\int_{\mathbb{R}^n} u \circ T\ d\mu - \int_{\mathbb{R}^n} u\ d\mu.$$

Then, combining integrals and applying the Lipschitz condition reveals

$$\begin{aligned} \int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu &= \int_{\mathbb{R}^n} (u(T(x)) - u(x))\ d\mu(x) \\ &\leq \int_{\mathbb{R}^n} |T(x) - x|\ d\mu(x). \end{aligned}$$

So, we see that

$$\inf_{T_{\#}\mu=\nu} \left\{ \int_{\mathbb{R}^n} |T(x) - x|\ d\mu \right\} \geq \sup \left\{ \int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu \ \middle|\ u : \mathbb{R}^n \to \mathbb{R},\ \mathrm{Lip}(u) \leq 1 \right\}.$$

Thus, if we can find a pair (T, u) where $T_{\#}\mu = \nu$ and u is Lipschitz with $\mathrm{Lip}(u) \leq 1$ such that

$$\int_{\mathbb{R}^n} |T(x) - x|\ d\mu = \int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu,$$

then T must be optimal. If not, we would have some $\tilde{T}$ with a lower cost than T. In turn,

$$\begin{aligned}\sup_{\mathrm{Lip}(u)\leq 1}\left\{\int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu\right\} \quad &\geq \quad \int_{\mathbb{R}^n} u\ d\nu - \int_{\mathbb{R}^n} u\ d\mu = \int_{\mathbb{R}^n} |T(x)-x|\ d\mu \\ &> \quad \int_{\mathbb{R}^n} |\tilde{T}(x)-x|\ d\mu \geq \inf_{T_{\#}\mu=\nu}\left\{\int_{\mathbb{R}^n} |T(x)-x|\ d\mu\right\}\end{aligned}$$

contradicting the above $\inf, \sup$ inequality.

Let us look at some examples.

1) Here is a nice, first example. Let $N \in \mathbb{N}$. Consider $f = \chi_{[0,N]}$ and $g = \chi_{[1,N+1]}$. An obvious transport is to move $[0,N]$ laterally to $[1,N+1]$ via $T(x) = x+1$. This is depicted below

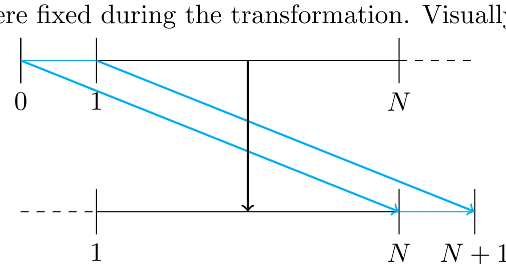
We now want to apply the duality principle to show this is optimal. First let us compute the cost of T:

$$\begin{aligned} C(T) \quad &= \quad \int_{\mathbb{R}} |T(x)-x|\chi_{[0,N]}\ dx \\ &= \quad \int_0^N |x+1-x|\ dx = N.\end{aligned}$$

Now consider the 1-Lipschitz function $u(x) = x$. We have that

$$\begin{aligned}\int_{\mathbb{R}} u(x)\chi_{[1,N+1]}\ dx - \int_{\mathbb{R}} u(x)\chi_{[0,N]}\ dx \quad &= \quad \int_1^{N+1} x\ dx - \int_0^N x\ dx \\ &= \quad \frac{x^2}{2}\bigg|_1^{N+1} - \frac{x^2}{2}\bigg|_0^N = N.\end{aligned}$$

Thus, by the duality principle, T is optimal. Is it unique? It turns out that T is *not* unique. To see this, imagine $[0,N+1]$ as a bookshelf, with books having length 1. Start with N books and one empty slot at $[N,N+1]$. Pick up the book at $[0,1]$ and move it to the space at $[N,N+1]$. After the transformation, we still have N books, but an empty slot at $[0,1]$. All other books were fixed during the transformation. Visually,



We can describe this map as

$$T(x) = \begin{cases} x+N & x \in [0,1] \\ x & x \in (1,N] \end{cases}$$

Let us check that this map is also optimal. First, we compute the cost

$$\begin{aligned} C(T) &= \int_{\mathbb{R}} |T(x)-x| \chi_{[0,N]} \, dx \\ &= \int_0^1 |x+N-x| \, dx + \int_1^N |x-x| \, dx = N. \end{aligned}$$

We have already shown that the minimum cost is N, hence T is optimal.

This example highlights the intimate relationship between cost and mass. In the first map, we moved a lot of mass a short distance. In the second map, we moved a small amount of mass a large distance. Yet, both maps were optimal.

We can interpolate between the two results. Imagine moving a book of length $[0,d]$ to $[n+1-d, n+1]$, and then shifting the rest over – that is, take $(d,n]$ to $[1, n+1-d)$. Each T_d is optimal too, hence there are infinitely many optimal transport maps. I think some care is necessary here. In particular, I don't like mapping a half-open interval of the form (] to one of the form [). Let me know.

2) In some cases, we get lucky and can explicitly compute an optimal transport map. For this example, let $\mu = \chi_{[-1,1]}dx$ and $\nu = \delta_{-1} + \delta_1$. Here, δ_x is the dirac measure at x. Note that these are not normalized to 1 to avoid complicating the calculations, but one can easily normalize them. Imagine this setup as having dirt at $[-1,1]$ and holes at $1,-1$. A possible transport is to divide the dirt into $[-1,0)$ and $[0,1]$, then move these into the holes at $-1, 1$ respectively. Explicitly,

$$T(x) = \begin{cases} -1 & x \in [-1,0) \\ 1 & x \in [0,1] \end{cases}$$

We show that T is optimal, and is the unique optimal transport map (up to a.e. equivalence). First, the cost of T is

$$C(T) = \int_{-1}^1 |T(x)-x| \, dx = \int_{-1}^0 |x+1| \, dx + \int_0^1 |x-1| \, dx = 1.$$

Now consider the 1-Lipschitz function $u(x) = |x|$. Then

$$\int_{\mathbb{R}} |x| \, d(\delta_{-1}+\delta_1) - \int_{-1}^1 |x| \, dx = 1+1-1 = 1.$$

Thus, T is optimal. To show T is the unique optimal transport map, note that any transport map must be such that $T([-1,1]) = \{-1,1\}$. Therefore we can write any transport map as

$$T_F(x) = \begin{cases} -1 & x \in F \\ 1 & x \in F^c \end{cases}$$

where $F \subset [-1,1]$, and the complement is viewed as the relative complement in $[-1,1]$. Note that, in order to be a transport map, it must be that $1 = \nu(\{1\}) = \mu(T_F^{-1}\{1\}) = \mu(F)$. So, $|F| = 1$. Since F and F^c are disjoint, we also have $|F^c| = 1$.

Now, computing the cost gives

$$C(T_F) = \int_{F^c} |x+1| \; dx + \int_F |x-1| \; dx.$$

We want to minimize each integral simultaneously. Looking at $\int_F |x-1| \; dx$, we see that $|x-1|$ is monotone decreasing on $[-1,1]$. Thus, we should take $F = [0,1]$ (recalling that F must be a measure 1 set). Now, minimizing $\int_{F^c} |x+1| \; dx$, we see that $|x+1|$ is monotone increasing on $[-1,1]$. So, ideally we would have $F^c = [-1,0)$. Choosing $F = [0,1]$, this is indeed the case. Hence $C(T_F)$ is minimized for $F = [0,1]$. Is there a more precise way to say this..?

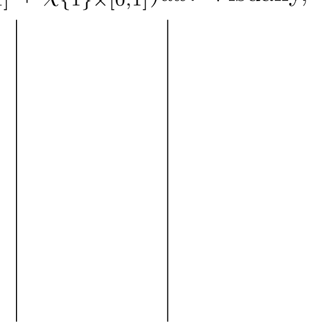
3) It can even be that uniqueness fails so spectacularly that all transport maps are optimal. For example, let μ a probability measure on $\mathbb{R}^2$ supported on $x_1 = 0$ and $\nu = 1/2(\delta_{(-1,0)} + \delta_{(1,0)})$. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a transport map – hence, it takes values in $\{(1,0), (-1,0)\}$. But, for any $(0,a)$ we have that $|(0,a) - T(0,a)|$ is either $|(0,a) - (1,0)|$ or $|(0,a) - (-1,0)|$. In both cases, the distances are equal and equal to $\sqrt{1+a^2}$. Importantly, this does not depend on T. Hence,

$$C(T) = \int_{\{x_1=0\}} |T(x) - x| \; d\mu = \int_{\{x_1=0\}} |x - (1,0)| \; d\mu$$

and thus all transport maps have the same cost. In particular they are all optimal.

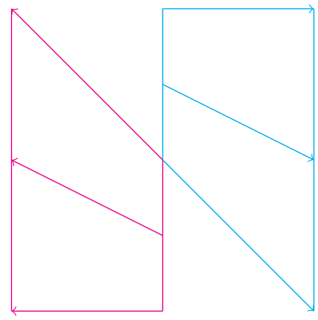
4) We may even fail so miserably that we can't find a single transport map. Let $\mu = \delta_0$ and $\nu = 1/2(\delta_{-1} + \delta_1)$. Consider $E = \{1\}$. Then $\nu(E) = 1/2$, but $\mu(T^{-1}(E))$ is either 0 or 1 for any E. Hence, $\nu(E) \neq \mu(T^{-1}(E))$, and there exists no T for which $\nu = T_{\#}\mu$.

5) Finally, sometimes failure is so subtle that we can find a minimizing sequence of transport maps, but no optimal transport map. Let $\mu = \chi_{\{0\}\times[0,1]}dx$ and $\nu = 1/2(\chi_{\{-1\}\times[0,1]} + \chi_{\{1\}\times[0,1]})dx$. Visually, we have the following



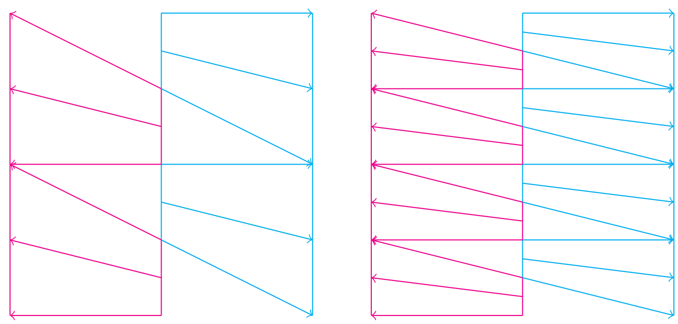
All the dirt is concentrated evenly on $\{0\} \times [0,1]$, and we want to transfer it evenly to the two holes on either side, at $\{1\} \times [0,1]$ and $\{-1\} \times [0,1]$. The "transfer evenly" condition is assumed in order to satisfy the transport condition – that is, we don't want to send most of our dirt to one point in the hole, and spread the rest to the rest of the hole.

Let us define a transport map T_1 visually as follows:



So, T_1 spreads the top part of our pile evenly into the right hole, and the lower part of our pile evenly into the left pile.

Now define T_2 and T_3 visually as well



At this point we can start to compute the cost. For each n, we subdivide $\{0\} \times [0,1]$ into 2^n evenly spaced regions $\{0\} \times [i/2^n, (i+1)/2^n]$ for $i = 0, 1, ..., 2^n - 1$. Then for even i, T_n stretches out $\{0\} \times [i/2^n, (i+1)/2^n]$ by a factor of 2 and carries it onto $\{-1\} \times [i/2^n, (i+2)/2^n]$. T_n acts similarly for odd i onto the right hole. Due to the symmetry of T_n, we need only compute the cost to move $\{0\} \times [0, 1/2^n]$ and multiply this cost by 2^n.

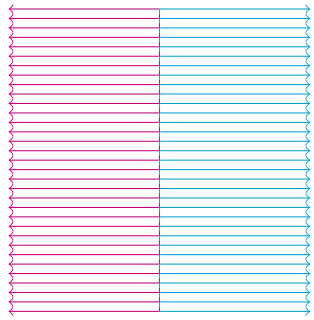
We have that T_n sends $(0, x)$ to $(-1, 2x)$ for $0 \leq x \leq 1/2^n$. Thus the cost for transporting $\{0\} \times [0, 1/2^n]$ is

$$\begin{aligned} C &= \int_0^{1/2^n} \sqrt{1+x^2}\, dx = \frac{1}{2}\left(x\sqrt{1+x^2} + \operatorname{arcsinh}(x)\right)\Big|_0^{1/2^n} \\ &= \frac{1}{2}\left(\frac{1}{2^n}\sqrt{1+\frac{1}{2^{2n}}} + \operatorname{arcsinh}\left(\frac{1}{2^n}\right)\right). \end{aligned}$$

Then the cost of T_n is

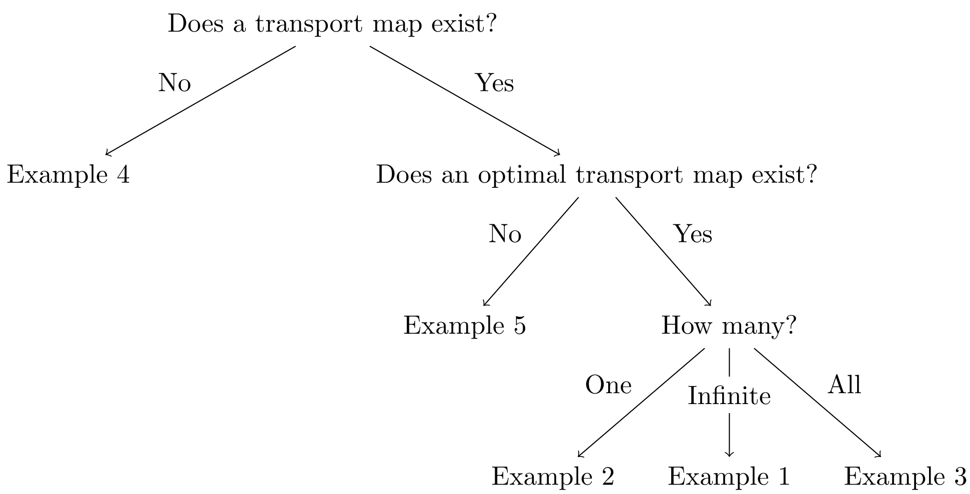
$$C(T_n) = 2^n C = \frac{1}{2}\left(\sqrt{1+\frac{1}{2^{2n}}} + 2^n \operatorname{arcsinh}\left(\frac{1}{2^n}\right)\right).$$

Observe that $C(T_n)$ is monotone decreasing and approaches 1. But if we try to take a limit of T_n we get something like



where each point $(0, x) \in \{0\} \times [0, 1]$ is split in half and sent to $(1, x)$ and $(-1, x)$. Each split portion travels a distance of 1, so in total it is as if $(0, x)$ travels a distance of 1. This rationale shows that $C(T)$ would be 1, as expected. However, $\lim T_n$ is not a map! We cannot send a point simultaneously to two different points. Although $\lim T_n$ is not a map, it actually turns out to be the correct picture in mind for solving this optimal mass transport problem.

As we have seen, even with such a simplistic formulation, the Monge problem is incredibly difficult to solve in general. The following flowchart summarizes the possible outcomes.



More issues arise by investigating the transport condition. Suppose T is a Lipschitz, injective map (which is a strong assumption to make!). By the area formula and injectivity of T, for Borel $\varphi : \mathbb{R}^n \to [-\infty, \infty]$ nonnegative or $L^1(\mathbb{R}^n)$, we get the following change of variables formula

$$\int_E \varphi(y)\, dy = \int_{T^{-1}(E)} \varphi(T(x)) |\det(\nabla T(x))|\, dx.$$

The following is definitely true

$$\int_{T(E)} \varphi(y)\, dy = \int_E \varphi(T(x)) |\det(\nabla T(x))|\, dx.$$

However in Francesco's OMT notes, he writes the above. What justifies moving from the blue to what is above? My thought was to just do $E \mapsto T^{-1}(E)$ in the blue. If T is surjective then $T(T^{-1}(E)) = E$, but we only assume injectivity... On the other hand, the cost condition guarantees that

$$\int_E g\ dy = \int_{T^{-1}(E)} f\ dx.$$

Now applying $\varphi = g$ in the change of variables gives

$$\int_{T^{-1}(E)} f\ dx = \int_E g\ dy = \int_{T^{-1}(E)} g(T(x))|\det(\nabla T(x))|\ dx.$$

which holds for all Borel E. Hence, $f(x) = g(T(x))|\det(\nabla T(x))|$ a.e. on $\{f > 0\}$. But this only used the transport condition, so it holds for all transport maps! Thus we may take a minimizing sequence of transport maps $\{T_j\}_{j=1}^{\infty}$; that is, such that

$$C(T_j) \to M(\mu, \nu).$$

If $T_j \to T$ and T is an optimal transport map, then we would need

$$g(T_j(x))|\det(\nabla T_j(x))| \to g(T(x))|\det(\nabla T(x))|.$$

and therefore $\nabla T_j \to \nabla T$ too. Unlike solving variational problems like minimizing Dirichlet energy, we do not a priori have any good control over ∇T_j.

Naturally, one might ask: If the Monge problem is so difficult to solve, why care about it? First, it turns out that relaxing some conditions of the Monge problem produces reasonable existence conditions. As in Example 5, if we could "split mass", we would have a solution – this is known as the Monge-Kantorovich formulation. It turns out that using different cost functions, like $c(x, y) = |x - y|^2$ instead of $c(x, y) = |x - y|$, also helps. Second, even in the cases where we cannot find an explicit solution, it is still useful to find transport maps. We will see an example of this in Section 3 with the Knothe map.

2.2. **The Monge-Kantorovich Formulation.** The notion of "splitting mass" will now be formalized. We thus deviate from transport maps, and look instead the so-called transport plans.

Definition 2.1. Let μ, ν be probability measures on X, Y respectively. We call a probability measure γ on $X \times Y$ a *transport plan* if

$$\pi_{0\#}\gamma = \mu \quad \text{and} \quad \pi_{1\#}\gamma = \nu.$$

In this context, we call μ and ν the *marginals* of γ.

How do we interpret this? The product space $X \times Y$ is seen as follows: there is mass at points $x \in X$ and holes at $y \in Y$. The pair (x, y) tells us that mass from x can be sent to y. The measure $d\gamma(x, y)$ tells us how much mass was sent to y from x. In general then for $A \subset X$, the marginal condition gives

$$\int_{A\times Y} d\gamma(x, y) = \int_A d\mu(x).$$

That is, there is an amount of mass at A, given by the RHS. This mass needs to be conserved no matter where it is sent, so that none is lost. The integral on the LHS looks at all the mass that came from A, so that equality implies none is lost.

Next let us look at $X \times \{y\}$ with a fixed $y \in Y$. For each $x \in X$, some amount of mass (possibly none) is sent to y. Integrating over X tells us how much mass in total from X is sent to the point y in the hole Y. We need each part of the hole to be filled, and $\nu(y)$ tells us how much mass can fit there. In general, for $B \subset Y$ the second marginal condition guarantees

$$\int_{X \times B} d\gamma(x, y) = \int_B d\nu(y).$$

That is, all the mass sent to B from X is equal to the amount of space available at B in Y.

The important part of γ is its support, which dictates where mass is sent to. That is to say if $d\gamma(x, y) = 0$ then no mass is transferred from x to y. It is visualized as follows.

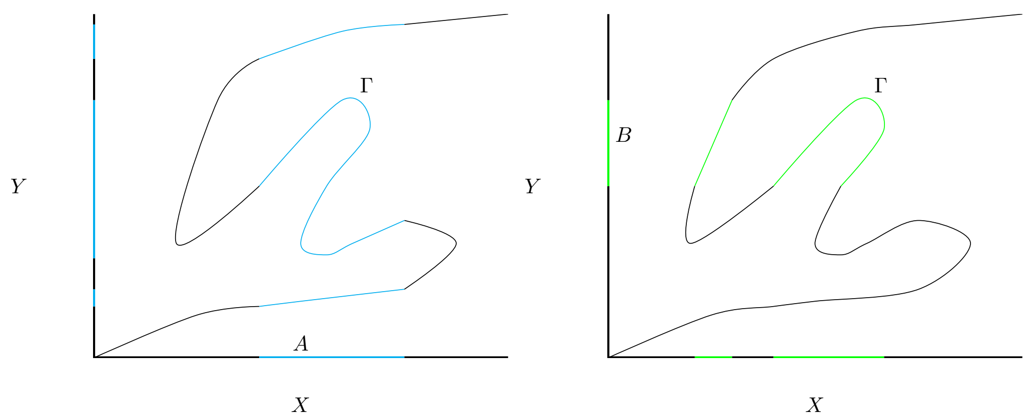


Figure 2.2. *Visualization of the support* Γ *of a probability measure* γ *on* $X \times Y$. *The blue and green regions are* $(A \times Y) \cap \Gamma$ *and* $(X \times B) \cap \Gamma$ *respectively.*

Let γ be some probability measure on $X \times Y$. The above figure shows two copies of its support Γ. Let us first look at the left panel. We have some subset $A \subset X$ shaded in blue with mass $\mu(A)$. Where does this mass go? Recall that each point $(x, y) \in \Gamma$ tells us that some mass went from x to y. The set $A \times Y$ is interpreted as all the possible locations in Y where mass from A is sent. Intersecting this with Γ tells us where mass is actually sent. Now, by projecting this onto Y, we see where the mass from A goes. Importantly, this does *not* say that $\mu(A) = \nu(\pi_1((A \times Y) \cap \Gamma))$, meaning we do not need to fill up the hole at $\pi_1((A \times Y) \cap \Gamma)$ solely with mass from A. Rather, it says that $\mu(A) = \gamma((A \times Y) \cap \Gamma))$, meaning all the mass in Y from A is equal to the total mass from A – none has been lost.

Let us now look at the right panel. We have some subset $B \subset Y$ of the hole, which has been filled up with mass $\nu(B)$. Where did this mass come from? In other words, we want to find all the $(x, y) \in \Gamma$ such that $y \in B$. This is precisely $(X \times B) \cap \Gamma$, shaded in green. Projecting onto X tells us where the mass came from. Importantly, this does *not* say that $\nu(B) = \mu(\pi_0((X \times B) \cap \Gamma)$, meaning that all the mass at $\pi_0((X \times B) \cap \Gamma)$ is not necessarily sent to B. Instead we have that B is filled up entirely from mass at $\pi_0((X \times B) \cap \Gamma)$, i.e. $\nu(B) = \gamma((X \times B) \cap \Gamma)$.

We conclude from this analysis that the first marginal condition allows for mass to be split. It allows for cases where Γ is not a graph. Indeed, if Γ is not a graph, then in general it does not pass the vertical line test. If this occurs at x then this precisely means there are multiple points in y where mass from x is sent to. We can also see that the second marginal condition is just the transport condition; all the mass sent from X into a part of the hole $B \subset Y$ must completely fill it up.

To put this into practice, let us revisit Example 5 from the previous subsection. Recall that mass is moved evenly from $X = \{0\}\times[0,1]$ to $Y = \{-1\}\times[0,1]\cup\{1\}\times[0,1]$. Ideally mass at each point $(0,x)$ for $x \in [0,1]$ is sent to both $(1,x)$ and $(-1,x)$. The corresponding transport plan γ will have mass concentrated evenly on two lines in the product space $X \times Y$. These lines correspond to $f_1(x) = (x,1,x)$ and $f_{-1}(x) = (x,-1,x)$. To visualize the product space, we consider X simply as the interval $[0,1]$, otherwise $X \times Y \subset \mathbb{R}^4$.

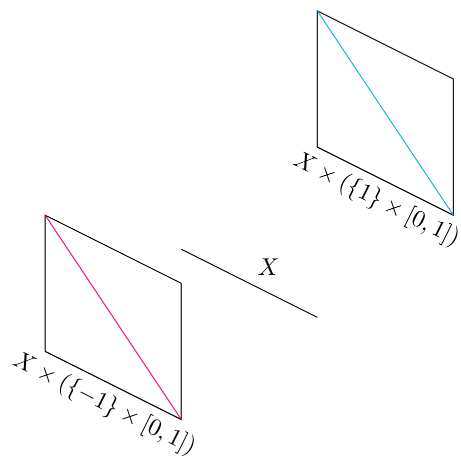


Figure 2.3. *Visualization of $X \times Y$ with a copy of X. The support Γ of the given transport plan γ is the union of the red and blue lines.*

The above figure shows $X \times Y$. There are two connected components of Γ, corresponding to the red and blue regions. Recall that Γ tells us where mass is sent. Let us restrict our attention to transporting mass from a subset $A \subset X$, for example $A = \{0\} \times [1/4, 1/2]$.

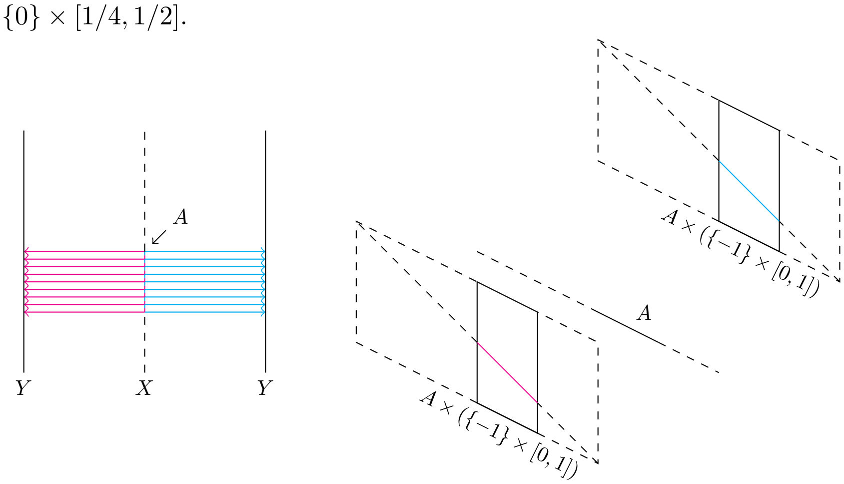


Figure 2.4. *Visualization of where mass is sent from A (left). Visualization of $(A \times Y) \cap \Gamma$ (right). The dashed lines in each show the rest of $X, X \times Y$, and Γ.*

On the left, we can see that mass in A is split and sent left and right to Y. This is seen in the right panel. Thus the advantage of considering the product space $X \times Y$ is revealed, since we can easily see where mass is sent. Is everything clear here? I feel like I need to add more, but I'm not sure how to write this part.

Let us now consider the subset $B \subset Y$ given by $B = \{-1\} \times [3/4, 1] \cup \{1\} \times [1/4, 1/2]$.

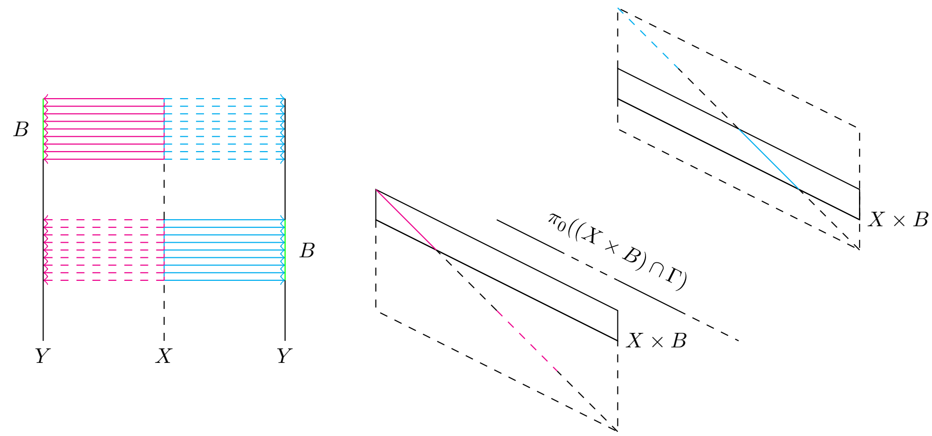


Figure 2.5. *Visualization of where all mass in B comes from (left). The dashed arrows tell us where leftover mass is sent. Visualization of $(X \times B) \cap \Gamma$ (right).*

On the left, we can see all the parts of X which send mass to B. Note that these regions of X send mass to other parts of Y not included in B. To the right, we can see this fact. The dashed blue and red regions show where the rest of the mass from $\pi_0((X \times B) \cap \Gamma)$ is sent. These are not included in $(X \times B) \cap \Gamma$ in the product space, highlighting the fact that $\nu(B) \neq \mu((\pi_0(X \times B) \cap \Gamma))$ in general.

Restrict attention now to $X = Y = \mathbb{R}^n$. As with the Monge problem, there is a corresponding Monge-Kantorovich optimization problem. Namely, we minimize the cost associated to a transport plan. Define $\Pi(\mu, \nu, c)$ to be the set of transport plans γ. Then the Kantorovich cost is

$$K(\mu, \nu, c) = \inf_{\gamma \in \Pi(\mu,\nu,c)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y)\ d\gamma(x, y).$$

where $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ is the cost of sending unit mass at x to y.

Two natural questions come to mind after giving this formulation:

(1) Does this actually generalize the Monge problem? For example, given an optimal transport map, can we realize it as an optimal transport plan?
(2) Do optimal transport plans exist?

The answer to the first question is a firm yes, and we have seen why already. In discussing the support Γ of a transport plan γ, we saw that the first condition, which allows us to "split mass", is really only necessary when Γ cannot be written as a graph. We then have the following result:

Theorem 2.6. *Every transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ induces a transport plan γ_T given by*

$$\gamma_T := (\mathrm{Id}_{\mathbb{R}^n} \times T)_{\#}\mu.$$

Conversely, if a transport plan γ is concentrated on a γ-measurable graph Γ, then it is induced by a transport map.

Proof. To verify the first part, we first prove the following. If $f, g : \mathbb{R}^n \to \mathbb{R}^n$ are measurable functions then

$$(f \circ g)_{\#}\mu = f_{\#}g_{\#}\mu.$$

To see this, let $E \subset \mathbb{R}^n$ be Borel. Then,

$$f_{\#}(g_{\#}\mu(E)) = (g_{\#}\mu)(f^{-1}(E)) = \mu(g^{-1}(f^{-1}(E))) = \mu((f \circ g)^{-1}(E)).$$

Next, observe that $\pi_0 \circ (\mathrm{Id}_{\mathbb{R}^n} \times T) = \mathrm{Id}_{\mathbb{R}^n}$. It follows that

$$\pi_{0\#}\gamma_T = \pi_{0\#}(\mathrm{Id}_{\mathbb{R}^n} \times T)_{\#}\mu = (\mathrm{Id}_{\mathbb{R}^n})_{\#}\mu = \mu.$$

Similarly, $\pi_1 \circ (\mathrm{Id}_{\mathbb{R}^n} \times T) = T$ so that

$$\pi_{1\#}\gamma_T = \pi_{1\#}(\mathrm{Id}_{\mathbb{R}^n} \times T)_{\#}\mu = T_{\#}\mu = \nu$$

since T is a transport plan.

The proof of the converse is beyond the scope of these notes, but can be found in [Amb00]. The key point is that Γ is a graph, so that we can construct a transport map that does not split mass. $\square$

To answer the second question, observe first that there always exist transport plans. Clearly, $\gamma = \mu \times \nu$ is a suitable transport plan. In contrast, a general Monge problem may not have a transport map. Moreover, Theorem 2.6 tells us that for each transport map T we have a corresponding transport plan γ_T. Note that

$$C(\gamma_T) = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\ d\gamma_T(x,y) = \int_{\mathbb{R}^n} c(x, T(x))\ d\mu = C(T)$$

so that γ_T and T have the same cost. This implies that $M(\mu,\nu,c) \geq K(\mu,\nu,c)$, since $K(\mu,\nu,c)$ is an infimum over a possibly larger set.

These facts suggest that the Monge-Kantorovich formulation is weaker than the Monge formulation. A tenant of analysis is, in the search for solutions, to pass weaker class of objects, solve there, and then upgrade the solution. The following theorem asserts that solutions to the Monge-Kantorovich problem exist.

Theorem 2.7. *Let μ, ν be probability measures on X, Y respectively. Let $c : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ be lower semi-continuous. Then there exists an γ optimal plan in $\Pi(\mu,\nu,c)$.*

Proof. A sketch of the proof is provided. The key idea is to use a form of convergence known as narrow convergence. This is slightly stronger than weak-$*$ convergence. Note that if μ_k is a sequence of Radon measures with $\sup_k \mu_k(B_r(0))$ for all $r > 0$, then there exists a convergent weak-$*$ subsequence.

Given a minimizing sequence $\{\gamma_j\} \subset \Pi(\mu,\nu,c)$, we can extract a weak-$*$ subsequence. This can be upgraded to a narrow convergent subsequence using the finiteness of μ,ν. Then, show that the convergent measure γ is in $\Pi(\mu,\nu,c)$ using several equivalent criterion for narrow convergence. Finally, it remains to show that γ is optimal – this follows from the fact that we chose a minimizing sequence. $\square$

So, we can solve in the weaker class of objects – transport plans. Can we upgrade these? If an optimal γ takes the special form in Theorem 2.6, that is $\gamma = (\mathrm{Id}_{\mathbb{R}^n} \times$

$T)_{\#}\mu$ for a transport map T, then

$$\begin{aligned} M(\mu,\nu,c) \geq K(\mu,\nu,c) &= \int_{\mathbb{R}^n\times\mathbb{R}^n} c(x,y)\ d((\mathrm{Id}_{\mathbb{R}^n}\times T)_{\#}\mu) \\ &= \int_{\mathbb{R}^n} c(x,T(x))\ d\mu(x) \geq M(\mu,\nu,c). \end{aligned}$$

Hence, T is actually optimal. This forms the central idea of Brenier's theory, which we will review next.

Before continuing, recall the useful duality principle in the Monge problem for finding optimal transport maps. It turns out there is a corresponding duality principle for the Monge-Kantorovich formulation. We state it here:

Theorem 2.8 (Duality principle)**.** *Let $\alpha,\beta:\mathbb{R}^n\to\mathbb{R}$ be Borel maps such that*

$$\alpha(x)+\beta(y)\leq c(x,y)$$

for all $x,y\in\mathbb{R}^n$. Define $\mathcal{A}$ the collection of pairs (α,β) of these maps. Then under the conditions for the Monge-Kantorovich formulation,

$$K(\mu,\nu,c)\leq \sup_{(\alpha,\beta)\in\mathcal{A}}\left\{\int_{\mathbb{R}^n}\alpha(x)\ d\mu(x)+\int_{\mathbb{R}^n}\beta(y)\ d\nu(y)\right\}.$$

Notice that this reduces to the original duality principle by taking $\alpha=u,\beta=-u$ and $c(x,y)=|x-y|$. Then the condition $\alpha(x)+\beta(y)\leq c(x,y)$ guarantees that u is a 1-Lipschitz function.

2.3. **Brenier Theory.** Here we specialize to the case of the quadratic cost $c(x,y)=|x-y|^2$ and μ,ν absolutely continuous probability measures with finite second moments. To give some intuition, we first define the notion of c-convexity:

Definition 2.9. A function $f:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ is said to be *c-convex* if for some function $\alpha:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ we have

$$f(x)=\sup_{y\in\mathbb{R}^n}\{\alpha(y)-c(x,y)\}$$

for all $x\in\mathbb{R}^n$ and f is not uniformly ∞.

Formally, we must allow f to take infinite values. Next, observe the following

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}|x|^2\ d\gamma(x,y)=\int_{\mathbb{R}^n}|x|^2\ d\mu(x)$$

if $\gamma\in\Pi(\mu,\nu,c)$. So, if μ has finite second moment, this is a constant independent of γ. Then, we have that

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}|x-y|^2\ d\gamma(x,y)=\int_{\mathbb{R}^n}|x|^2\ d\mu(x)+\int_{\mathbb{R}^n}|y|^2\ d\nu(y)-2\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\ d\gamma(x,y).$$

Since the second moments are finite and independent of γ, it follows that γ is optimal in $\Pi(\mu,\nu,|x-y|^2)$ iff it is optimal in $\Pi(\mu,\nu,-\langle x,y\rangle)$. The negative sign is used here by convention.

Now, c-convexity for $c=-\langle x,y\rangle$ is precisely convex and lower-semicontinuous. This follows since the sup of affine functions (of the form $a+\langle x,b\rangle$ for constants a,b) is convex and lower-semicontinuous. To see this, one can define the epigraph of a function f as follows.

Definition 2.10. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Then the *epigraph* of f is

$$\mathrm{Epi}(f) = \{(x,t) \in \mathbb{R}^{n+1} \mid t \geq f(x)\}.$$

A function is convex iff its epigraph is a convex subset of $\mathbb{R}^{n+1}$. Note that the epigraph of an affine function is a half-space and taking the sup amounts to intersecting these half-spaces. A detailed account of convex functions can be found here [Roc70]. Let me know if I go on too much here. I don't know how clear it is, and I may not need to introduce the epigraph at all.

So, c-convexity is a natural generalization of convexity. The role it plays takes significant time to establish, but the main takeaway is that when the cost is quadratic, convex functions are important.

Now, given a Kantorovich probblem $K(\mu,\nu,c)$ recall that if $\gamma = (\mathrm{Id}_{\mathbb{R}^n} \times T)_{\#}\mu$ is optimal for a transport map T, then T is optimal. Thus, we should study the structure of optimal plans. We have the following theorem to help us with this endeavor

Theorem 2.11 (Brenier)**.** *Let μ,ν be absolutely continuous probability measures on $\mathbb{R}^n$ with respect to Lebesgue measure. Further assume that μ,ν have finite second moments. Then for the cost $c(x,y) = |x-y|^2$ there exists a unique optimal transport plan of the form*

$$\gamma = (\mathrm{Id}_{\mathbb{R}^n} \times \nabla f)_{\#}\mu$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a convex function.

In this case, we call ∇f a Brenier map. Some slight care needs to be taken in order for this to make sense. Indeed, f may not be differentiable everywhere. Let F be its set of differentiability points, which is a Borel set. Then

$$(\nabla f)_{\#}(\mu)(\mathbb{R}^n) = \mu(\nabla f^{-1}(\mathbb{R}^n)) = \mu(F)$$

since ∇f is only defined on F. Then, it is possible that the pushforward is not a probability measure!

But if f is convex, it is locally Lipschitz, and hence by Rademacher $m(\Omega \backslash F) = 0$. Here, $\Omega = \mathrm{Int}(\mathrm{Dom}(f))$, where $\mathrm{Dom}(f) = \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. One can show that f is finite a.e. so that μ is concentrated on Ω. Since μ is absolutely continuous to Lebesgue measure, we see that $\mu(\mathbb{R}^n \setminus F) = 0$. So, μ is concentrated on F, and the pushforward is a probability measure.

One can ask whether we can weaken the hypotheses to omit finiteness of the second moments. Robert McCann proved that this was the case.

Theorem 2.12 (McCann)**.** *Let μ,ν be probability measures on $\mathbb{R}^n$ absolutely continuous to Lebesgue measure. Further suppose that μ does not give mass to small sets. Then, the conclusion of Brenier's theorem holds.*

The proof uses a nice implicit function theorem for convex functions (instead of differentiable functions!). The nontechnical requirement that "μ does not give mass to small sets" can be made technical by introducing $(n-1)$-rectifiable sets. This, however, is beyond the scope of the paper. We direct the interested reader to [Vil03] for proofs of these – see Theorem 2.32 in the book.

As a concluding remark, all of the above theory has been developed for probability measures absolutely continuous to Lebesgue measure. This was motivated by imagining the Radon-Nikodym derivatives $d\mu/dx$ and $d\nu/dx$ as a pile of dirt and a hole, respectively. This gives some physical interpretation to the problem, but it is not necessary.

3. The Isoperimetric Inequality

3.1. **History of the Isoperimetric Problem.** We begin our discussion of the isoperimetric inequality with a legend about its origin. We travel back several centuries, to 825 BC and the city of Tyre. Dido's husband Sychaeus was just murdered by her brother, and king of Tyre, Pygmalion. She flees Tyre with some followers and, heading westward, lands at what would become Carthage on the coast of north Africa. She bargains with a local ruler to obtain land, who sells her some oxhide. He explains that she can have all the land she can enclose within the oxhide.

Dido cut the oxhide into thin strips and sewed them together. She was now faced with the following problem: Dido wants to enclose the most area possible with a certain boundary length. She uses the strips to draw a semicircle bordering the coast, which maximizes the enclosed area. Effectively, Dido has solved the first isoperimetric problem, and with it founds the prosperous Carthage.

The first step towards proving the isoperimetric problem came from the Greeks. Though his work is lost, Zenodorus (non-rigorously) proved the following in the 100s BC.

Theorem 3.1 (Zenodorus)**.** *The following hold:*

i) Among regular polygons with the same perimeter, that which has more sides has greater area.
ii) A circle with the same perimeter as a regular polygon has greater area.
iii) The regular n-gon maximizes area among all n-gons with the same perimeter.

See [Kli72] for an account of this. As far as the Greeks were concerned, this solved the isoperimetric problem. But there were crucial flaws – for one, there are more extravagant shapes than just polygons and circles. Another more glaring issue is that Zenodorus assumes the existence of a maximizer in his proof. The isoperimetric problem then lay dormant for many centuries while mathematicians were unable to resolve these.

In 1842, Jakob Steiner miraculously gave five different proofs of the improved isoperimetric problem: among all closed plane curves with a prescribed length, the circle bounds the greatest area. The five proofs are similar and contain many of the same ideas, namely they all revolve around techniques to increase the area while keeping the perimeter fixed. Doing this requires many symmetrization arguments, and the end result is a circle.

However, like the Greeks before him, Steiner presupposed existence of a maximizer. Indeed, he crucially assumed that his symmetrization arguments never halt, and we can keep performing them until (in the limit) we get a circle. Though this was obvious to Steiner, he never formally proved it.

In 1879, Weierstrass proved existence using the calculus of variations, finally completing a rigorous solution of the isoperimetric problem. In the decades afterwards, several mathematicians returned to Steiner's original proofs and showed they were valid (namely, the limit process holds). A detailed account of this history can be found here [Bla05].

3.2. **Solution Using Optimal Mass Transport.** We now turn to proving the isoperimetric problem in higher dimensions using optimal mass transport. We first note that the original statements of the isoperimetric problem had fixed perimeter, and we were trying to find a solution which maximized volume. We can reformulate this to instead look at sets with fixed volume, and minimize perimeter. Here is how we can see this: Suppose that Σ is a solution only to the above "dual" isoperimetric problem. Then, there exists a set $\tilde{\Sigma}$ with the same perimeter, but greater area (since Σ does not solve the classical isoperimetric problem). Now rescale $\tilde{\Sigma}$ so that it has the same area as Σ. It follows that the rescaled $\tilde{\Sigma}$ must have smaller perimeter than Σ, a contradiction. We now present the solution.

Theorem 3.2 (Isoperimetric Inequality)**.** *Let $E \subset \mathbb{R}^n$ be bounded with smooth boundary. Then*

$$P(E) \geq P(B_r)$$

where r is such that $\mathrm{Vol}(B_r) = \mathrm{Vol}(E)$*. Furthermore, equality holds iff $E = B_r$ up to translation and modification on a set of measure zero.*

The isoperimetric inequality as stated is readable and has the easy interpretation that, among all sets with the same volume, the ball minimizes perimeter. However, it is not scale invariant. That is, if we want to scale up E by some factor, we would need to find a new radius to compare B_r and E. We wish to avoid this, and we can.

Theorem 3.3 (Scale Invariant Isoperimetric Inequality)**.** *Let $E \subset \mathbb{R}^n$ be bounded with smooth boundary. Then,*

$$P(E) \geq n \,\mathrm{Vol}(E)^{(n-1)/n} \,\mathrm{Vol}(B_1)^{1/n}.$$

To see that this is scale invariant, consider $E \mapsto rE$. Then,

$$P(rE) = r^{n-1}P(E) \geq r^{n-1} n \,\mathrm{Vol}(E)^{(n-1)/n} \,\mathrm{Vol}(B_1)^{1/n} = n \,\mathrm{Vol}(rE)^{(n-1)/n} \,\mathrm{Vol}(B_1)^{1/n}.$$

We now show how to derive this form.

Proof. First, there exists an r such that $\mathrm{Vol}(E) = \mathrm{Vol}(B_r)$. Now, observe that

$$\mathrm{div}(\mathrm{Id}(x)) = \sum_{k=1}^{n} \frac{\partial \,\mathrm{Id}}{\partial x_k} = n$$

since all the partial derivatives are 1. Hence, by the divergence theorem we have

$$n \,\mathrm{Vol}(B_1) = \int_{B_1} \mathrm{div}(\mathrm{Id}) = \int_{\partial B_1} \langle x, \nu_{B_1}(x) \rangle = P(B_1)$$

since $\nu_{B_1}(x) = x$ and $\langle x, x \rangle = |x|^2 = 1$. Substituting this gives

$$P(E) \geq P(B_r) = r^{n-1}P(B_1) = r^{n-1} n \,\mathrm{Vol}(B_1).$$

We can actually find what r is! Since $\mathrm{Vol}(E) = \mathrm{Vol}(B_r)$, it follows that $r^n = \mathrm{Vol}(E)/\mathrm{Vol}(B_1)$. Hence,

$$P(E) \geq n \,\mathrm{Vol}(E)^{(n-1)/n} \,\mathrm{Vol}(B_1)^{1/n}$$

as desired. □

We remark that the two inequalities are obviously equivalent. For if $\mathrm{Vol}(E) = \mathrm{Vol}(B_1)$, then the RHS becomes $n \,\mathrm{Vol}(B_1) = P(B_1)$.

We present two different proofs of the isoperimetric inequality.

Proof 1. This proof was given first by Gromov using what is known as the Knothe map. Suppose we have measures $\mu = f(x)dx$ and $\nu = f(y)dy$ with f, g nonnegative. Let $F = \{f > 0\}$ and $G = \{g > 0\}$. Modify g on a set of measure zero so that G is closed. We first construct the Knothe map $T : F \to G$ component by component, and define each component pointwise. I will use the case $n = 2$ to visualize this construction.

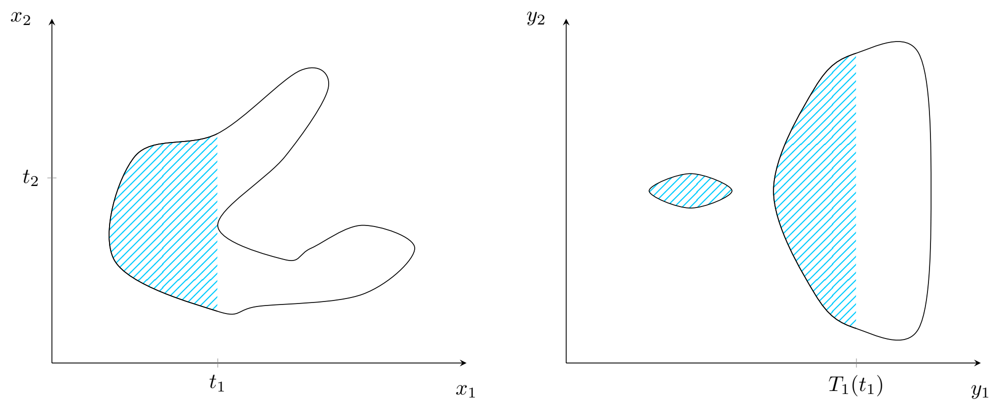


Figure 3.4. *The level sets $F = \{f > 0\}$ and $G = \{g > 0\}$, respectively. In blue we have $F \cap \{x_1 < t_1\}$ and $G \cap \{y_1 < T_1(t_1)\}$. Plotted are the first two coordinates of t and $T_1(t_1)$.*

Let T_1 be the first component of T. We define T_1 solely in terms of x_1, the first component of $x \in \mathbb{R}^n$. Fix an $t \in \mathbb{R}^n$ and read off its first component t_1. We can look at how much of f lies above the area in $F = \{f > 0\}$ to the left of t_1 relative to all of F. That is, we look at the quantity

$$\frac{\int_{F \cap \{x_1 < t_1\}} f}{\int_F f}$$

Then, there exists a $\tilde{t}_1$ such that

$$\frac{\int_{F \cap \{x_1 < t_1\}} f}{\int_F f} = \frac{\int_{G \cap \{y_1 < \tilde{t}_1\}} g}{\int_G g}$$

So, $\tilde{t}_1$ is such that the relative integrals are the same. Note that this $\tilde{t}_1$ doesn't have to be unique. See the right panel in Figure 3.4, for which any $\tilde{t}_1$ between the two connected components gives the same integral. But, there will always be a largest $\tilde{t}_1$ satisfying this – set $T_1(t_1)$ to be this number. By definition, T_1 is monotone increasing (by monotonicity of the integral). Furthermore, T_1 is continuous if the support of g is connected. The above can be restated as $\mu\{x_1 < t_1\} = \nu\{y_1 < T_1(t_1)\}$.

It'll turn out to be important to study each component's partial derivatives. Since T_1 does not depend on $x_2, x_3, ..., x_n$, it follows that $\partial T_1/\partial x_i = 0$ for $i = 2, ..., n$. What about $\partial T_1/\partial x_1$? To answer this, we appeal to Fubini's theorem. Indeed,

$$\int_{\{x_1 < t_1\}} f = \int_{-\infty}^{t_1} \left(\int_{\mathbb{R}^{n-1}} f \ dx_2 ... dx_n \right) dx_1.$$

By taking the x_1-derivative of this and applying the fundamental theorem of calculus, we obtain

$$\frac{\partial}{\partial x_1}\int_{\{x_1<t_1\}} f = \int_{\mathbb{R}^{n-1}} f(t_1, x_2, ..., x_n)\, dx_2...dx_n = \int_{\{x_1=t_1\}} f(x_1, x_2, ..., x_n)\, dx_1...dx_n.$$

I'm concerned about having the integrals go from integrating against $dx_2...dx_n$ to $dx_1dx_2...dx_n$, but the above makes sense to me. On the other hand, if we do the same thing for the integral of g we get

$$\begin{aligned}\frac{\partial}{\partial x_1}\int_{\{y_1<T_1(t_1)\}} g &= \frac{\partial}{\partial x_1}\int_{-\infty}^{T_1(t_1)}\left(\int_{\mathbb{R}^{n-1}} g\ dy_2...dy_n\right) dy_1 \\ &= \frac{\partial T_1}{\partial x_1}(t_1)\int_{\mathbb{R}^{n-1}} g(T_1(t_1), y_2, ..., y_n)\ dy_2...dy_n \\ &= \frac{\partial T_1}{\partial x_1}(t_1)\int_{\{y_1=T_1(t_1)\}} g(y_1, y_2, ..., y_n)\ dy_1...dy_n\end{aligned}$$

by an application of the chain rule and FTC. Now, by definition of $T_1(t_1)$,

$$\frac{\partial}{\partial x_1}\left(\frac{\int_{F\cap\{x_1<t_1\}} f}{\int_F f}\right) = \frac{\partial}{\partial x_1}\left(\frac{\int_{G\cap\{y_1<T_1(t_1)\}} g}{\int_G g}\right).$$

where the denominators are just some constants. Hence, after rearranging,

$$\frac{\partial T_1}{\partial x_1}(t_1, x_2, ..., x_n) = \frac{\int_G g}{\int_F f}\,\frac{\int_{\{x_1=t_1\}} f}{\int_{\{y_1=T_1(t_1)\}} g}.$$

This partial derivative is positive due to monotonicity of T_1.

We now define T_2, which will depend on x_1 and x_2. Here's the idea: we used the ordering on $\mathbb{R}$ to define T_1. But, in $\mathbb{R}^n$, there is no such ordering. Having chosen t_1 and $T_1(t_1)$, we can look at $F_1 = F \cap \{x_1 = t_1\}$ and $G_1 = G \cap \{y_1 = T_1(t_1)\}$. These sets are $n-1$ dimension, so we can drop down a dimension and do almost the same thing with t_2 in place of t_1 and F_1, G_1 in place of F, G. See the figure below.

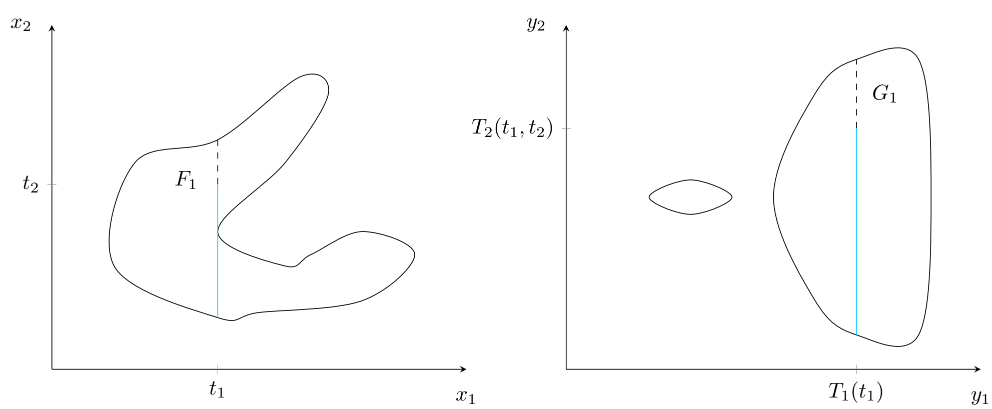


Figure 3.5. *The sets $F_1 = F \cap \{x_1 = t_1\}$ and $G_1 = G\cap\{y_1 = T_1(t_1)\}$ contained in F, G respectively. In blue we have the sets $F_1\cap\{x_2 < t_2\}$ and $G_1\cap\{y_2 < T_2(t_1, t_2)\}$. Plotted are the first two coordinates of t as well as $T_1(t_1)$ and $T_2(t_1, t_2)$.*

As before, it could be that

$$\int_{F_1} f \neq \int_{G_1} g$$

so we have to add a normalizing factor. How do I write this in terms of the measures μ and ν? Since they're absolutely continuous wrt Lebesgue measure and F_1, G_1 are $n-1$ dim sets, they're null sets. It's like I want to view μ as fdx^{n-1} instead of fdx^n, but I don't know how to write this in a nice way Thus, we define $T_2(t_1, t_2)$ as the greatest number such that

$$\frac{\int_{F_1 \cap \{x_2 < t_2\}} f}{\int_{F_1} f} = \frac{\int_{G_1 \cap \{y_2 < T_2(t_1, t_2)\}} g}{\int_{G_1} g}.$$

So, $F_1 \cap \{x_2 < t_2\}$ and $G_1 \cap \{y_2 < T_2(t_1, t_2)\}$ cover the same relative area of F_1, G_1 respectively. Figure 3.5 shows this property.

Now, similarly with T_1, we have that $x_2 \mapsto T_2(x_1, x_2)$ is monotone. However, $x_1 \mapsto T_2(x_1, x_2)$ could behave fairly poorly. Thus we will not investigate $\partial T_k / \partial x_i$ for $i < k$. We build all the components in a similar manner, so that T_k depends only on $x_1, ..., x_k$. Thus $\partial T_k / \partial x_i = 0$ for $i > k$. So the only partial derivatives we should try and study are the $\partial T_k / \partial x_k$. The computation of $\partial T_k / \partial x_k$ from earlier generalizes, so that

$$\frac{\partial T_k}{\partial x_k}(t_1, t_2, ..., t_k, x_{k+1}, ..., x_n) = \frac{\int_{G_{k-1}} g}{\int_{F_{k-1}} f} \frac{\int_{F_k} f}{\int_{G_k} g}.$$

where $F_{k-1} = F_{k-2} \cap \{x_{k-1} = t_{k-1}\}$ and $G_{k-1} = G_{k-2} \cap \{y_{k-1} = T_{k-1}(t_1, ..., t_{k-1})\}$, with $F_0 = F$ and $G_0 = G$. Once more, the first fraction comes the normalizing factors, which are constant with respect to x_k. What about $\partial T_n / \partial x_n$? Observe that F_{n-1} is a line, since it is the nontrivial intersection of $n-1$ orthogonal hyperplanes. A similar conclusion holds for G_{n-1}. Thus,

$$\int_{F_{n-1} \cap \{x_n < t_n\}} f = \int_{-\infty}^{t_n} f(t_1, ..., t_{n-1}, x_n)\ dx_n$$

is a one-dimensional integral. Differentiating this gives, by FTC

$$\frac{\partial}{\partial x_n} \int_{F_{n-1} \cap \{x_n < t_n\}} f = f(t_1, ..., t_{n-1}, t_n) = f(t).$$

Similarly,

$$\frac{\partial}{\partial x_n} \int_{G_{n-1} \cap \{y_n < T_n(t)\}} g = \frac{\partial T_n}{\partial t_n}(t) g(T_1(t_1), ..., T_n(t)).$$

Since $T_n(t)$ is defined to be such that

$$\frac{\int_{\{x_n < t_n\}} f}{\int_{F_{n-1}} f} = \frac{\int_{\{y_n < T_n(t)\}} g}{\int_{G_{n-1}} g},$$

we see that

$$\frac{\partial T_n}{\partial x_n}(t) = \frac{\int_{G_{n-1}} g}{\int_{F_{n-1}} f} \frac{f(t)}{g(T(t))}.$$

In total, we have that ∇T is an upper triangular matrix. Thus the determinant is

$$\begin{aligned}\det(\nabla T)(x) &= \prod_{k=1}^{n} \frac{\partial T_k}{\partial x_k} = \left(\frac{\int_G g}{\int_F f} \frac{\int_{F_1} f}{\int_{G_1} g} \right) \left(\frac{\int_{G_1} g}{\int_{F_1} f} \frac{\int_{F_2} f}{\int_{G_2} g} \right) \dots \left(\frac{\int_{G_{n-2}} g}{\int_{F_{n-2}} f} \frac{\int_{F_{n-1}} f}{\int_{G_{n-1}} g} \right) \left(\frac{\int_{G_{n-1}} g}{\int_{F_{n-1}} f} \frac{f(t)}{g(t)} \right) \\ &= \frac{\nu(G)}{\mu(F)} \frac{f(x)}{g(T(x))} > 0\end{aligned}$$

for a.e. $x \in F$. Since T maps F into G (should I say more here? Not sure how obvious it is), and $g > 0$ on G, we see that this is well defined. Clearly this construction depends on an initial choice of basis for $\mathbb{R}^n$ (because the half-spaces generated will be different). We note here that T is a transport map, albeit not an optimal transport map! Thus, it can still be fruitful to study transport maps in general.

As a remark, the above construction relied on having measures $\mu = f dx$ and $\nu = g dy$. This need not be the case – we can just as easily define $T : F \to G$ solely by knowing what F and G are. In this case, we define $T_{k+1}(t_1, \dots, t_{k+1})$ for $k = 0, \dots, n-1$ by

$$\frac{\mathcal{H}^{n-k}(F_k \cap \{x_{k+1} < t_{k+1}\})}{\mathcal{H}^{n-k}(F_k)} = \frac{\mathcal{H}^{n-k}(G_k \cap \{y_{k+1} < T_{k+1}(t_1, \dots, t_{k+1})\})}{\mathcal{H}^{n-k}(G_k)}.$$

The reader can verify that this constructs the same Knothe map as above with $f = \chi_F$ and $g = \chi_G$. Let me know if it's confusing about the above constructions, in particular with having everything defined in terms of F and G. I can rework everything so that they're not referenced at all

We now consider the case of $f = \chi_E$ and $g = \chi_{B_1}$, where $E \subset \mathbb{R}^n$ is bounded and such that $\mathrm{Vol}(E) = \mathrm{Vol}(B_1)$. Note that T transports E onto B_1. So, $T : E \to B_1$, and in particular $|T| \leq 1$. Moreover, by the above formula for the determinant,

$$(\det \nabla T)(x) = \frac{\mathrm{Vol}(B_1)}{\mathrm{Vol}(E)} \frac{\chi_E(x)}{\chi_{B_1}(T(x))} = \frac{\chi_E(x)}{\chi_{B_1}(T(x))} = 1$$

for a.e. $x \in E$. Trivially, we have that

$$\mathrm{Vol}(E) = \int_E 1 = \int_E \det \nabla T = \int_E (\det \nabla T)^{1/n}$$

since we're just modifying something equal to 1. Now, we can apply AM-GM, which states for $\lambda \geq 0$ that

$$\left(\prod_{k=1}^{n} \lambda_k \right)^{1/n} \leq \frac{1}{n} \sum_{k=1}^{n} \lambda_k.$$

The eigenvalues for ∇T are $\partial T_k / \partial x_k > 0$ since ∇T is upper triangular. Thus applying AM-GM to the eigenvalues λ_k gives

$$\mathrm{Vol}(E) \leq \frac{1}{n} \int_E \sum_{k=1}^{n} \frac{\partial T_k}{\partial x_k} = \frac{1}{n} \int_E \operatorname{div} T.$$

We now apply the divergence theorem to obtain

$$\mathrm{Vol}(E) \leq \frac{1}{n} \int_{\partial E} \langle T, \nu_E \rangle,$$

where ν_E is the outer unit normal of E. Since $|T| \leq 1$, and $|\nu_E| = 1$ by definition, we suggestively use Cauchy-Schwarz

$$\operatorname{Vol}(E) \leq \frac{1}{n}\int_{\partial E} |T||\nu_E| \leq \frac{1}{n}\int_{\partial E} 1 = \frac{1}{n}P(E).$$

We proved earlier that

$$n \operatorname{Vol}(B_1) = P(B_1).$$

Substituting this into the estimate for $\operatorname{Vol}(E)$ gives

$$P(B_1) = n \operatorname{Vol}(B_1) = n \operatorname{Vol}(E) \leq P(E)$$

as desired. The general case holds by scaling.

To solve the iff part of the theorem, note that if $P(B_1) = P(E)$ then the inequality from AM-GM is an equality. Hence, all the partial derivatives are equal, and in particular equal to 1. This is not enough to conclude that $E = B_1$ modulo the appropriate congruences. The details will be omitted, but the essence is to construct infinitely many Knothe maps in each direction $\nu \in S^{n-1}$ and force E to lie in between the two supporting hyperplanes of B_1 with unit normals $\nu, -\nu$. $\square$

Proof 2. One major complication in Gromov's proof is the need to use infinitely many transport maps to show that $P(E) = P(B_1)$ implies $E = B_1$ modulo the appropriate congruences. We present an alternative proof using Brenier maps. Note that the Knothe map had three key properties when $f = \chi_E$ and $g = \chi_{B_1}$,

i) $\det \nabla T = 1$,
ii) $|T| \leq 1$,
iii) $\operatorname{div} T \geq n(\det \nabla T)^{1/n}$.

Any map T with these properties can be used in the first proof to reach the same conclusion by following the exact same steps. Let $E \subset \mathbb{R}^n$ be bounded and such that $\operatorname{Vol}(E) = \operatorname{Vol}(B_1)$. Consider now the optimal transport problem sending $\mu = \chi_E / \operatorname{Vol}(E)dx$ to $\nu = \chi_{B_1}/\operatorname{Vol}(B_1)dy$ with cost $c(x,y) = |x-y|^2$. It follows that these are probability measures absolutely continuous to the Lebesgue measure. Note that μ, ν have finite second moments since E, B_1 are bounded. Thus, we can apply Brenier's theorem and conclude that there exists a convex $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that $(\nabla\varphi)_{\#}\mu = \nu$. Since φ is convex it follows that $\nabla^2\varphi$ is a positive semi-definite symmetric matrix. Thus, $|\det \nabla\varphi| = \det \nabla\varphi$. We saw previously that transport maps obey

$$f(x) = g(\nabla\varphi(x))|\det(\nabla^2\varphi)| = g(\nabla\varphi(x))\det(\nabla^2\varphi).$$

Rearranging this yields

$$\det(\nabla^2\varphi) = \frac{\chi_E(x)}{\chi_{B_1}(\nabla\varphi(x))}\frac{\operatorname{Vol}(B_1)}{\operatorname{Vol}(E)} = \frac{\chi_E(x)}{\chi_{B_1}(\nabla\varphi(x))}$$

so that condition i) is satisfied for a.e. $x \in E$, so long as $\nabla\varphi$ maps into B_1.

Now,

$$\operatorname{Vol}(E) = 1 = \nu(B_1) = \mu(\nabla\varphi^{-1}(B_1)) = \int_{\nabla\varphi^{-1}(B_1)\cap E} 1 \leq \operatorname{Vol}(E)$$

so that $\nabla\varphi$ maps E into B_1. Thus, $|\nabla\varphi| \leq 1$, and conditions i) and ii) are satisfied.

Finally, let λ_k be the eigenvalues of $\nabla^2\varphi$. Observe that the $\lambda_k > 0$ since φ is convex (in contrast to the previous example, where they were positive by monotonicity) Then,

$$\begin{aligned}\operatorname{div}(\nabla\varphi) &= \sum_{k=1}^{n} \partial^2\varphi/\partial x_k^2\varphi = \operatorname{tr}(\nabla^2\varphi) = \sum_{k=1}^{n}\lambda_k \\ &= n\left(\frac{1}{n}\sum_{k=1}^{n}\lambda_k\right) \geq n\left(\prod_{k=1}^{n}\lambda_k\right)^{1/n} = n(\det(\nabla^2\varphi))^{1/n}.\end{aligned}$$

So, condition iii) is satisfied. Thus, we can prove the isoperimetric inequality using the Brenier map $\nabla\varphi$.

We can now prove that equality holds iff $E = B_1$ modulo some congruences. Since $\nabla^2\varphi$ is symmetric it is diagonalizable. Assuming that $P(E) = P(B_1)$, we once more obtain that all the eigenvalues λ_k of $\nabla^2\varphi$ must be equal, and since $\det(\nabla^2)\varphi = 1$, in particular are all equal to 1. Thus $\nabla^2\varphi$ is similar to the identity matrix. Hence the transport map $\nabla\varphi$ is some translation. This proof easily shows the inequality is sharp. Crucially, symmetry was obtained from φ being a convex function. □

As a concluding remark, we can actually prove the scale invariant form of the isoperimetric inequality via a slight modification in the above proofs. Namely, we simply drop the assumption that $\operatorname{Vol}(E) = \operatorname{Vol}(B_1)$. Because of this property i) changes to $\det\nabla T = \operatorname{Vol}(B_1)/\operatorname{Vol}(E)$ while properties ii) and iii) remain the same. Then,

$$\begin{aligned}P(E) &= \int_{\partial E} 1 \geq \int_{\partial E}\langle T, \nu_E\rangle = \int_E \operatorname{div} T \geq n\int_E (\det\nabla T)^{1/n} \\ &= n\operatorname{Vol}(E)\frac{\operatorname{Vol}(B_1)^{1/n}}{\operatorname{Vol}(E)^{1/n}} \geq n\operatorname{Vol}(E)^{(n-1)/n}\operatorname{Vol}(B_1)^{1/n}\end{aligned}$$

still using Cauchy-Schwarz with property ii), the divergence theorem, and properties iii) and i) (modified), in that order.

4. Acknowledgements

To be added :)

References

[Amb00] Luigi Ambrosio. Lecture notes on optimal transport problems. `https://cvgmt.sns.it/media/doc/paper/1008/trasporto.pdf`, 2000.

[Bla05] Viktor Blasjö. The isoperimetric problem. *The American Mathematical Monthly*, 112(6):526–566, 2005.

[FMP10] A. Figalli, F. Maggi, and A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Inv. Math.*, 182(1):167–211, 2010.

[Kli72] Morris Kline. *Mathematical Thought from Modern to Ancient Times*. Oxford University Press, 1972.

[Mag12] Francesco Maggi. *Sets of Finite Perimeter and Geometric Variational Problems: An Introduction to Geometric Measure Theory*, volume 135 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2012.

[Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics. American Mathematical Society, 1970.

[Vil03] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.