Isoperimetric inequalities with practical applications

Zola Mahlaza
Supervisor: Dr J. Ratzkin

Abstract

This paper serves as an introduction to isoperimetric inequalities. It provides an intuitive approach as opposed to rigor. The classical isoperimetric inequality states that amongst all regions in a plane enclosed by a Jordan curve with a fixed perimeter, the circular region has the maximal area. Isoperimetric inequalities specify the relation between two or more geometric quantities. Surprisingly, these inequalities have outstanding applications in a variety of fields. This paper focuses on isoperimetric inequalities such as the Gaussian isoperimetric inequality, which is used in Information Theory for decoding error probabilities for the Gaussian channel[1].

Contents

0.1 Introduction

"Everyone knows what a curve is, until he has studied enough mathematics to become confused through the countless number of possible exceptions"
— F. Klein, *Methods of modern mathematical physics*

Higher mathematics is abstract. The abstract concepts it presents may seem inapplicable to real-world problems to the layman. This paper aims to demystify isoperimetric inequalities. These inequalities have become powerful tools in modern mathematics. A popular isoperimetric inequality is known as the classical isoperimetrical inequality. It was proposed by Zenodorus, a Greek mathematician. This document exposes the applications of isoperimertic inequalities in modern fields. An obvious application of isoperimetric inequalities is in area optimization. This case was exploited by Queen Dido of Carthage. We will not describe Queen Dido's application of the inequality, interested readers might find [8] informative. This singular application, area optimization, seems

insignificant when one considers the complexity of proving the classical isoperimetric inequality. This document will provide applications which are specific to computers and computer networks as they have become ubiquitous tools in the 21st century.
In computer networks, data is transferred from sender to receiver as a sequence of on and off signals via a communication channel. The issue with data transfer is that it is sometimes done through unreliable or noisy channels hence loss of data is inevitable. This poses the challenge of how to determine if data contains errors at the receiver's side. Information theory, a subject mostly attributed to Claude E. Shannon, presents concepts such as error correction and detection which allow the detection of error and restoration of data.

Theorem 1 *Among all sets in $\mathbb{R}^n$ with prescribed Gaussian measure, the half-spaces have minimal Gaussian perimeter*

At first glance, this theorem seems divorced from reality. Readers who are confused by this theorem should not worry, an intuitive will be given in section 0.3, and an application of it will be presented in section 0.4

0.2 Classical isoperimetric inequality

We now present the classical isoperimetric inequality, however, we begin a preliminary definitions.

Definition 1 (Jordan arc) *Let $f : [0..1] \to \mathbb{R}^2$ be a path from $a = (x_1, y_1) tob = (x_2, y_2)$ where $a, b \in \mathbb{R}^2$. We say that f is a Jordan arc iff f in one-to-one, with the exception that $a = b$ is allowed.*

Definition 2 (Jordan curve) *Let f be a Jordan arc from a to b where $a, b \in \mathbb{R}^2$. f is a Jordan curve iff $a = b$*

Intuitively, a Jordan curve is a closed curve which does not intersect itself.

Theorem 2 *Among all regions of area A in the plane, enclosed by a Jordan curve C with fixed perimeter L, we have,*

$$4\pi A \leq L^2$$

Equality holds if and if the curve is a circle.

An attempt to demystify the inequality, before proof, is as follows. The area of a circle is given by equation $A_{circle} = \pi r^2$ Using the equation of the circumference, $L = 2\pi r$, we have

$$A_{circle} = \pi(\frac{L}{2\pi})^2$$
$$A_{circle} = (\frac{L^2}{4\pi})$$

Hence, for any region of area A enclosed by a curve of length L we have that $A \leq (\frac{L^2}{4\pi}) \implies 4\pi A \leq L^2$.

Starting at this point, where there will be no confusion, we will refer to the classical isoperimetric inequality as simply the isoperimetric inequality, this is done for brevity. Jakob Steiner, a Swiss mathematician, is perhaps the most mentioned mathematician in texts whose subject matter are isoperimetric inequalities. This is largely because he is one very first modern day mathematician to produce significant contributions to the proof of the isoperimetric inequality. We will begin by exploring his contributions. We now establish a few facts before looking into Steiner's first approach towards proving the isoperimetric inequality.

- The circle encloses maximal area out of all closed curves in a plane of equal length.
- The circle has the smallest perimeter out of all closed curves in a plane with equal areas.

It is important to realize that the above statements are equivalent. Given closed curves which in a plane have the same area, the optimal curve is the one which encloses the same area with the smallest perimeter. Given that a curve has the smallest perimeter among all curves of equal area, one can increase the perimeter to make it roughly equal the other curves. This increase in perimeter will increase the area hence the curve encloses maximal area.

Theorem 3 *An inscribed angle in a semicircle is a right angle*

Proof 1 *Let* $\vec{OA} = u$, $\vec{OC} = v$ *and* $\vec{OB} = w$. *Then* $\vec{AB} = w - u$ *and* $\vec{BC} = v - w$ $\vec{AB} \cdot \vec{BC} = (w - u) \cdot (v - w)$. *We know that* $u = -v$ *since* u *lies in the same semicircle opposite* v. $\vec{AB} \cdot \vec{BC} = (w + v) \cdot (v - w) = w \cdot v - w \cdot w + v \cdot v - v \cdot w$ *This means that* $\vec{AB} \cdot \vec{BC} = |v|^2 - |w|^2 = 0$. *We then conclude that* AB *and* BC *are perpendicular.*

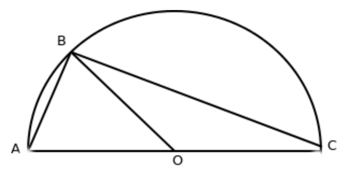


Figure 0.1: Illustration of mentioned sides in following proof

Theorem 4 *Out of all possible triangles with two sides of given length, the triangle of maximum area is the right triangle with the given sides as the perpendicular sides.*

Proof 2 *Given $\triangle ABC$ then $A(ABC) = \frac{1}{2}ab\sin(A\hat{B}C)$ where $a = AB$, $b = BC$. We know that a and b are fixed hence maximum area is achieved when $\sin A\hat{B}C$ is maximum. Hence the maximum area is achieved when $A\hat{B}C = 90$.*

Steiner's approach was an attempt to prove Dido's problem, so it considered maximizing an area which was bounded on one side by a straight line. We now aim to prove that a semi-circle has maximal area amongst all given curves.

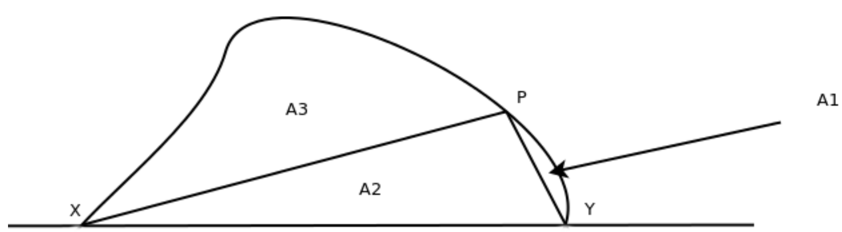


Figure 0.2: Illustration of Steiner's approach

Consider a Jordan curve which is bounded on a side by a straight line. Pick an arbitrary point, P, on the curve. X and Y are the points where the curve meets the straight line. Let $A2$ be the area of $\triangle XPY$ and $A1$, $A2$ be areas of the other sections within the curve which not enclosed by $\triangle XPY$. The idea is to move X and/or Y along the same line they lie in with XP and PY fixed, such that $X\hat{P}Y = 90$. The resulting triangle, $\triangle XPY$, will have a larger area by theorem 4. This process is repeated, and as the iterations increase the curve becomes a semicircle. One can gather intuition using theorem 3 to be convinced of this fact. It is obvious that the area can not decrease, however, it could increase. This approach is normally introduced in texts as the four-hinge problem. This approach is nice and intuitive, however, it only works in 2 dimensions. In an attempt to produce more general and complete ways of solving the isoperimetric inequality, we introduce the following concepts.

- Steiner symetrization.
- Calculus of variations.

Steiner symetrization

Steiner symetrization is a symetrization technique which is also due to Jakob Steiner. It's power lies in in two facts:

- It preserves area/volume[2]
- Perimeter does not increase under Steiner symmetrization

Definition 3 (Steiner symetrization) *Let $L \subset \mathbb{R}^n$ be a compact convex set, and let $u \in \mathbb{R}^n$ be a unit vector. Think of L as a family of line segments parallel to u. Translate each of these line segments along the direction u until they are all balanced symmetrically around the plane $u^\perp$. The result is a new convex set $s_u L$, called the Steiner symmetrization of L with respect to the direction u.[3]*

Let C be a Jordan curve. Let H be a vertical hyperplane in Euclidean space that cuts curve C in two opposite points. P is the region enclosed by curve C. We then slice P using lines which are perpendicular to H. Let $L = \{x + \delta p : p \perp H, \delta \in \mathbb{R}\ \forall x \in H\}$ be the set of all such lines.

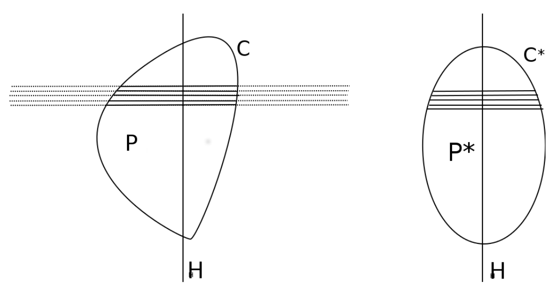


Figure 0.3: Steiner symmetrization of region P

Steiner symmetrization takes all line segments $\mathcal{L} \in L \cap P$ and maps them such that they are symmetrical about H to obtain P^*. The measure of each line, $\mathcal{L}$, is preserved.

Properties

Theorem 5 *Let $K \subset \mathbb{R}^n$ be compact and convex[3]. Let H be a hyperplane that intersects K and splits it into two sections. Again, let $v \in \mathbb{R}^n$ and $c \perp H$.The volume of K after Steiner symmetrization about L in the direction of v is preserved, that is, $V(K) = V(s_v K)$.*

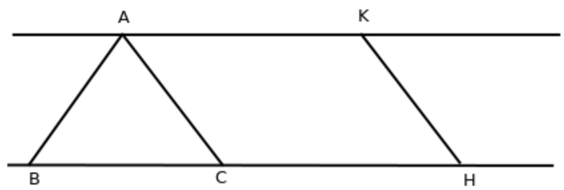
A discussion prior to the proof might be helpful. It should be noted that $\dim(H) = n - 1$ by the definition of a hyperplane. Our previous discussion revealed that the length of the slices should be unchanged thus it is should be clear that Steiner symmetrization preserves volume/area.

Proof 3 *$\forall h \in \mathbb{R}^n$ where $(h_1, h_2, ..., h_{n-1}) \in H$ consider the line $S_n = \{h + v\beta : \beta \in \mathbb{R}\}$. We know that the measure of each slice is preserved by Steiner symmetrization, denote the measure of each slice of the region K by M_n. Formally, this means $\int_{h_n \in K \cap S_n} dh = M_n = \int_{h_n \in S_v(K) \cap S_n} dh$*

$$\begin{aligned}
V(K) &= \int_{(h_1,h_2,...,h_{n-1}) \in H} \int_{h_n \in K \cap S_n} dh dh_1 dh_2 ... dh_{n-1} \\
&= \int_{(h_1,h_2,...,h_{n-1}) \in H} (M_n) dh_1 dh_2 ... dh_{n-1} \\
&= \int_{(h_1,h_2,...,h_{n-1}) \in H} \int_{h_n \in S_v(K) \cap S_n} dh dh_1 dh_2 ... dh_{n-1} \\
&= V(S_v(K))
\end{aligned}$$

Theorem 6 *If BH and AK are segments, each of given length, lying on fixed parallel lines, the sum $BA + KH$ is at minimum when there is symmetry, and $BA = KH$.*

Theorem 7 *Steiner symetrization does not increase perimeter, that is, $P(C^*) \leq P(C)$*



Proof 4 *Consider the lines* $\pi_1, \pi_2, ..., \pi_n$ *which are perpendicular to* H. *Let* π_k *cut* C *at* P_k *and* Q_k, *and also cut* C^* *at* $P\prime_k$ *and* $Q\prime_k$. *Using the previous theorem,*

$$P\prime_1 P\prime_2 + Q\prime_1 Q\prime_2 \leq P_1 P_2 + Q_1 Q_2 \tag{1}$$

We can create similar pairs for all lines π_k,π_{k+1} *where* $1 \leq k \leq n-1$. *Then if* K *is a polygon* $P_1 P_2 P_3 ... P_{n-1} P_n Q_n Q_{n-1} ... Q_2 Q_1 P_1$ *constructed in this way and inscribed in C, and if* $\hat{K}$ *is the corresponding polygon* $P\prime_1 P\prime_2 ... P\prime_{n-1} P\prime_n Q\prime_n Q\prime_{n-1} ... Q\prime_2 Q\prime_1 P\prime_1$ *inscribed in* C^*. *Naturally, it follows that*

$$P(\hat{K}) \leq P(K) \tag{2}$$

Let $\hat{K}$ *be a polygon of that satisfies (2) inscribed in* C^*, *and such that*

$$P(\hat{K}) \geq P(C^*) - \epsilon, \epsilon \in \mathbb{R} \tag{3}$$

It the follows that

$$\begin{aligned} P(C^*) &\leq P(\hat{K}) + \epsilon \\ &\leq P(K) + \epsilon \\ &\leq P(C) + \epsilon \\ P(C^*) &\leq P(C) \end{aligned}$$

The last inequality follows from the fact that ϵ *is arbitrary.*

The properties of Steiner symmetrization reveal that the volume is unchanged after symmetrization, however, the perimeter is decreased except when the region is symetrical about the hyperplane in the direction of the symmetrization vector. Furthermore, if the set is already symmetrical about the plane in every direction then its the ball/circle. In conclusion, Steiner symmetrization will reduce the perimeter of any set with the exception of a circle but preserve the volume/area. One can futher conclude using the bullet points in section 0.2, page 4 that the circle encloses maximal area among all curves of equal perimeter. Unfortunately, Steiner's ingenious contributions are incomplete. The problem is that they fail to prove existence of a solution, that is, they fail to show that there exists a curve which encloses maximal area among all curves with the same perimeter. The reason might be unclear, consider the following paradox.

Paradox 1 (Perron) *Let N be the largest integer. Suppose $N > 1$ then $N^2 > N$. This contracts the definition of N, hence $N = 1$*

We now look at other methods for a proof of existence of a curve that encloses maximal area among curves of equal perimeter.

Calculus of variations

Calculus of variations is a branch of mathematics which is very closely related to the isoperimetric inequalities. Its focus is the optimization of physical inequalities. A hefty amount of the problems considered in calculus of variations have origins in physics where one has to minimize the energy associated to the problem under consideration. Its creation began with the brachistone curve problem which is was formulated by Johann Bernoulli in 1696 [5]. The word brachistone comes from greek, it means shortest time. This paper concerns itself with isoperimetric inequalities hence an attempt at stating and proving the brachistone problem will not be made. The problem can be restated such that it fits a problem that can be solved using the calculus of variations.

Problem

Among all curves $y = f(x)$ where $y(a) = y_a$, $y(b) = y_b$ and

$$\int_a^b \sqrt{1 + y\prime(x)^2} dx = L$$

find one for which $\int_a^b y(x) dx$ is maximum.

Lax's Proof

We now present a proof of the isoperimetric inequality for completeness and also to intrigue the interested reader. This proof is due to Peter Lax. It is arguably the most concise of proofs of the isoperimetric inequality In an attempt to make this proof more understandable, we be begin by presenting a simple proposition. In attempt at simplifying the proof, we will prove that the semi-circle encloses maximal area out of all curves with the same perimeter, bounded on one side by a straight line.

Proposition 1 *Let $a, b \in \mathbb{R}$ then $ab \leq \frac{a^2+b^2}{2}$*

Proof 5 *Take $a, b \in \mathbb{R}$ then $(a-b)^2 \geq 0$ and equality is achieved iff $a = b$.*

$$\begin{aligned} a^2 - 2ab + b^2 &\geq 0 \\ a^2 + b^2 &\geq 2ab \\ \frac{a^2+b^2}{2} &\geq ab \end{aligned}$$

Proof 6 (Lax) *Let $x(s), y(s)$ be the parametric representation of the curve. Let s be the arc length where $0 \leq s \leq \pi$. Suppose that the curve is positioned such that $(x(0), y(0))$ and $(x(\pi), y(\pi))$ lie on the x-axis, that is, $y(0) = y(\pi) = 0$. The area enclosed the curve is given by equation 1.*

$$A = \int_0^\pi y \frac{dx}{ds} ds \tag{4}$$

Using proposition 2, we get

$$A = \int_0^\pi y \frac{dx}{ds} ds \leq \frac{1}{2} \int_0^\pi (y^2 + (\frac{dx}{ds})^2) ds \tag{5}$$

We know that s is the arc length so,

$$\begin{aligned} ds^2 &= dx^2 + dy^2 \\ 1 &= (\frac{dx}{ds})^2 + (\frac{dy}{ds})^2 \end{aligned}$$

To simplify notation, let $\frac{dx}{ds} = \tilde{x}$ and $\frac{dy}{ds} = \tilde{y}$. Hence equation 2 becomes

$$A \leq \frac{1}{2} \int_0^\pi (y^2 - \tilde{y}^2 + 1) ds \qquad (\tilde{x}^2 = 1 - \tilde{y}^2)$$

We ensured that $y = 0$ when $s = 0$ and $s = \pi$, so we can factor y such that $y(s) = v(s)\sin s$ where $v(s)$ is bounded and differentiable. The derivative of y with respect to arc length then becomes

$$\tilde{y} = \tilde{v}\sin s + u\cos s$$

Hence

$$A \leq \frac{1}{2}\int_0^{\pi} (v^2(\sin^2 s - \cos^2 s) - 2v\tilde{v}\sin s\cos s - \tilde{v}^2\sin^2 s + 1)ds \qquad (6)$$

Observing that $2v\tilde{v} = \frac{d}{ds}v^2$, we integrate by parts equation 3 to obtain

$$A \leq \frac{1}{2}\int_0^{\pi} (1 - \tilde{v}^2\sin^2 s)ds \qquad (7)$$

Inequality (7) is $\leq \frac{\pi}{2}$ and equality only holds when $\tilde{v} = 0$. Recall that we said that $ab = \frac{a^2+b^2}{2}$ iff $a = b$ in proposition 1, in this case this means that $A = \frac{1}{2}\int_0^{\pi}(1 - \tilde{v}^2\sin^2 s)ds$ iff $y = \tilde{x} = \sqrt{1 - \tilde{y}^2}$ since we applied proposition 1 to obtain inequality (7). This means that $y(s) = \pm\sin s$ and $x = \int \tilde{x}ds = \int yds = \mp\cos s + l$, $l \in \mathbb{R}$. Hence the curve is a semi-circle. This proof can also be used to show that the circle encloses maximal area, more or less in the same fashion. However, the arc length will range from $0 \leq s \leq \pi$ and $\pi \leq s \leq 2\pi$.

0.3 Gaussian isoperimetric inequality

The Gaussian isoperimetric inequality was developed extensively in the study of the functional analytic aspects of probability theory[6]. Here, we only present it to illustrate it's power in Information theory. We begin with a definition of a measure. In 1,2 and 3 dimensional Euclidean space, there exists simple and intuitive concepts of measure. We speak of the length of a line segment, area under a curve and the volume of a sphere for instance. The concepts of length, area and volume are familiar concepts. It is only natural that there exists a generalization of the principal concept of measure in higher dimensions.

Theorem 8 *Among all sets in $\mathbb{R}^n$ with prescribed Gassian measure, half spaces have minimal Gaussian perimeter.*

Definition 4 *A set of sets S is said to be pairwise disjoint iff:* $\forall X, Y \in S : X \neq Y \implies X \cap Y = \emptyset$

It has been mentioned that measure is a generalisation of volume. Without an explicit definition of a measurable space, we can further explain what a measure is. A measure is a function which takes an element X(of a measurable space Γ) and returns a non-negative number which we will refer to as the "measure of X". We denote the measure by $\mu : \Gamma \to \mathbb{R}_{\geq 0} \cup \{-\infty, \infty\}$. Further, in addition to non-negativity, a measure satisfies the following properties.

1. $\mu(\emptyset) = 0$
2. **Countable Additivity** Let $\{S_n\} \subset \Gamma$ be a sequence of pairwise disjoint sets. Then, $\mu(\cup_{n=1}^{\infty} S_n) = \sum_{n=1}^{\infty} \mu(S_n)$

A Gaussian measure is probability measure. This means that is a special case of a measure. It satisfies non-negativity and countable additivity properties. However, it has an additional property that t is normalized, that is, $\forall A \in \Gamma : 0 \leq \mu(A) \leq 1$.

One can split n dimensional space with a $n-1$ dimensional hyperplane. A half-space is simply the space which remains after the section of the space which falls on one side of the hyperplane is removed.

Definition 5 *Let* $\mu : (R^n, \Gamma) \to [0, 1]$ *be a probability measure on the given measurable space. We say that* μ *is a Gaussian measure if it is defined by*

$$\mu(X) = \frac{1}{\sqrt{2\pi}^n} \int_X e^{-\frac{1}{2}\|x\|^2} d\mu(x) \tag{8}$$

where $\|\cdot\|$ *denotes the length of vectors in* $\mathbb{R}^n$*.*

In order to gather intuition on the Gaussian isoperimetric inequality, we begin by restating the classical isoperimetric inequality. Among all compact sets A in $\mathbb{R}^n$ with a smooth boundary ∂A and with a fixed volume, Euclidean balls are the ones with the minimal surface measure[6]. The surface measure we were concerned with was the perimeter. The notion behind the Gaussian isoperimetric inequality is similar. The measure, however, is different as we now focus on the Gaussian measure. Also there exists a Steiner symmetrization equivalent in the context of the Gaussian isoperimetric inequality. It is known as Gaussian symmetrization. A formal presentation of the Gaussian isoperimetric inequality is presented by both Ros[10] and Ledoux[6].

0.4 Applications

Information Theory

Information theory is a field which most texts credit Claude Shannon for its inception. This is largely due his 1948 two-part paper published in the Bell System Technical Journal. Pierce[9] presents a comprehensive history of information theory. The main focus of the field in the simplest view is data compression, error decoding and data recovery. The importance of these issues is quite obvious. Given the a stream of data, the symbols of the stream are mapped to another sequence. The resultant sequence of symbols is in bits as they represent analog signals which can be transferred over a physical communication channel.

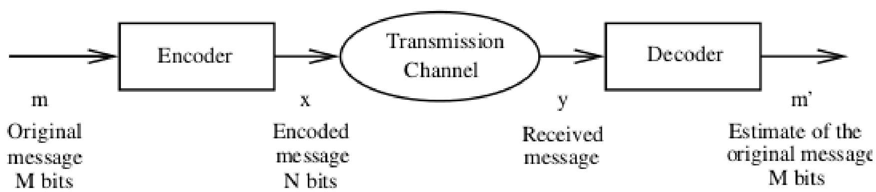


Figure 0.4: Data transmission process. Source http://goo.gl/lIAkI6

Given a set of points in Euclidean space. We can divide the space into regions such that around each point the borders of the regions are equidistant from the two nearest points. The regions are referred to as Voronoi regions.

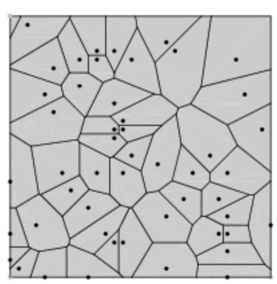


Figure 0.5: Voronoi diagram.

Tillich and Zémor [1] apply the Gaussian isoperimetric inequality to Voronoi regions of codes in Euclidean space and obtain a precise description of how the maximum-likelihood decoding error probability varies as a function of the minimum Euclidean distance.

Bibliography

[1] The Gaussian Isoperimetric Inequality and Decoding Error Probabilities for the Gaussian Channel, Jean-Pierre Tillich and Gilles Zmor, IEEE, Faculty of Computer Science and System Engineering, Okayama Prefectural University, Okayama, 719-1197, Japan, pg328

[2] Steiner symetrization and applications, Andrejs Treibergs, University of Utah, http://goo.gl/UUeQ7D

[3] The perimeter inequality under Steiner symetrization: case of equality, M Chelbík, A. Cianchi and N. Fusco, Annals of Mathematics, Seconds series, Vol. 162, No. 1, pg 525.

[4] Steiner symetrization and convergence, Dan Klein et all, Fifth international workshop on convex geometry: analytical aspects, http://goo.gl/fzrtlv, pg 2-3, June 2011

[5] Calculus of variations : lecture notes, Erich Miersemann, Department of mathematics, Leipzig University, Oct 2012

[6] Isoperimetric and Gaussian analysis, Michel Ledoux, Department of mathematics, Laboratoire de statistique et probabilitеś, UniversieṔaul-Sabatier

[7] The isoperimetric problem revisted: Extracting a short proof of sufficiency from Euler's 1744 Approach to Necessity, Richard Tapia, Department of Computational and Applied Mathematics, Rice University, April 2013.

[8] Popular lectures and addresses, W.T. Kelvin, Volume 1, Nature series, pg 571

[9] The early days of information theory, J. R. Pierce, IEEE transactions on information theory, Vol 19, No 1, Jan 1973.

[10] The isoperimetric problem, Antonio Ros, University of Granada, pg 28, http://www.ugr.es/ aros/isoper.pdf