Some Connections between Sobolev–type Inequalities and Isoperimetry *

Serguei G. Bobkov
Department of Mathematics
Syktyvkar University
167001 Syktyvkar
Russia

Christian Houdré
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
USA

July 26, 1995

*Key Words: Sobolev inequalities, Isoperimetry, Geometric measure theory, Orlicz space, uniform distribution on the Sphere, Gaussian measure.
The first named author was supported in part by NSF DMS 92-09130, by the Grant 93–011–1454 of the Russian Foundation for Scientific Research, by the International Scientific Fond, Grant No NXZ000, by the Grant Center for Fundamental Sciences at Sankt–Petersburg University, Grant No 94-1.4-57, and by the Grant Center for Research in Mathematics at Novosibirsk University.
The second named author was supported in part by an NSF Mathematical Sciences Post–Doctoral Fellowship and by a NSF–NATO Fellowship while at CEREMADE, Université Paris Dauphine, Place de Lattre de Tassigny, 75775 Paris Cedex France and at CERMA, ENPC, La Courtine, 93167 Noisy Le Grand, Cedex, France.
Both authors are very grateful to the Center for Stochastic processes, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514-3260, the Department of Statistics, Stanford University, Stanford, CA 94305, and Le CERMA, Ecole Nationale des Ponts et Chaussées, La Courtine, 93167 Noisy Le Grand Cedex, France for their hospitality and support while part of this research was carried out.

1 Introduction

Motivation, Examples, Statements of Results

It is well known that the Sobolev inequality

$$(1.1) \qquad \int_{\mathbf{R}^n} |\nabla f(x)| d\mu(x) \ \geq \ n\omega_n^{1/n} \left(\int_{\mathbf{R}^n} |f(x)|^{\frac{n}{n-1}} d\mu(x) \right)^{\frac{n-1}{n}}, \ n \geq 2,$$

where f is a compactly supported smooth function on the Euclidean space $\mathbf{R}^n$, where μ is the Lebesgue measure in $\mathbf{R}^n$, and where ω_n is the volume of the unit ball in $\mathbf{R}^n$, is equivalent to the *isoperimetric property* of the balls in $\mathbf{R}^n$. This isoperimetric property states that, among all the compact sets of a fixed volume, the balls have the least surface area. This can also be written as

$$(1.2) \qquad \mu^+(A) \geq n\omega_n^{1/n} (\mu(A))^{\frac{n-1}{n}},$$

where $A \subset \mathbf{R}^n$ is compact, and where μ^+ is the surface measure.

With the best constant $c = n\omega_n^{1/n}$, (1.1) is due to Federer and Fleming [Fed–Fle, Remark 6.6] and to Maz'ya [Ma1, Theorem 6] who apparently were the first to point out the equivalence between (1.1) and (1.2) (see Osserman [Oss, p.1192], Federer [Fed, p.510], [Ma2, p.69]). Equality in (1.1) is only possible asymptotically for a sequence of smooth functions converging to the indicator function of an arbitary ball.

Rothaus [Rot, Theorem 1] found a natural extension of the equivalence between (1.1) and (1.2) for (X, d, μ) where X is a Riemannian manifold equipped with its Riemannian metric d and its Riemannian measure μ. His approach which, as the one in [Fed–Fle], is based on Federer's co–area formula will be generalized to an abstract setting. Before presenting such a generalization, let us introduce our framework and recall some definitions.

Throughout the paper, (X, d, μ) is a metric space equipped with a separable non–atomic Borel probability measure μ, i.e., $\mu(X_1) = 1$ for some separable Borel set $X_1 \subseteq X$ and $\mu(\{x\}) = 0$, for all $x \in X$.

Let $A \subseteq X$ be a Borel set, and let $\mu^+(A)$ denotes *the surface measure* of A. More precisely, and as suggested in [Oss, p.1189], $\mu^+(A)$ is the *lower outer 1-dimensional Minkowski μ–content* of the boundary of A, (see also Burago and Zalgaller [Bur–Zal, p.69]) which is defined by

$$(1.3) \qquad \mu^+(A) \ = \ \liminf_{h \to 0^+} \frac{\mu(A^h) - \mu(A)}{h},$$

where $A^h = \{x \in X : \exists a \in A, d(x,a) < h\}$ is the open h–neighbourhood of A. Analogously, the appropriate definition of the *modulus of gradient* is

$$(1.4) \qquad |\nabla f(x)| \ = \ \limsup_{d(x,y)\to 0^+} \frac{|f(x)-f(y)|}{d(x,y)}.$$

The function $|\nabla f|$ is always Borel measurable whenever f is continuous on X. Of course, if the function f, defined on $X = \mathbf{R}^n$ or on a submanifold $X \subset \mathbf{R}^n$, is differentiable at a point $x \in X$, then (1.4) defines $|\nabla f(x)|$ in the usual sense. Conversely, by Stepanoff's theorem [Fed, p.218], if for almost all $x \in X$ (with respect to the Lebesgue measure), the right side in (1.4) is finite, then f is differentiable almost everywhere. Thus, (1.4) may be used without any confusion for locally Lipschitz functions, i.e., for functions f such that for any $x \in X$, there exists a ball around x where the restriction of f to this ball is Lipschitz. In our abstract framework, we apply the definition (1.4) to functions which are Lipschitz on every ball, i.e., functions whose restriction to any ball is Lipschitz; and when X is locally compact, locally Lipschitz functions are Lipschitz on every ball.

Next, following [Rot], let G be a non-empty set of pairs (g_1, g_2) of μ-integrable functions on X, and let $\mathcal{L}(\cdot)$ be a functional generated by G via

$$(1.5) \qquad \mathcal{L}(f) \ = \ \sup_{(g_1,g_2)\in G} \int_X (f^+ g_1 + f^- g_2) d\mu,$$

where we used the standard notation $f^+ = \max(f,0), f^- = \max(-f,0)$. The value $\mathcal{L}(f)$, finite or not, is defined for all f such that $f^+ g_1$ and $f^- g_2$ are μ-integrable whenever $(g_1, g_2) \in G$. If $g_1 = -g_2$, then (1.5) becomes

$$(1.6) \qquad \mathcal{L}(f) \ = \ \sup_{g\in G} \int_X fg d\mu,$$

where the sup is now taken over some set G of μ–integrable functions. Examples of functionals (1.6) include norms $\|\cdot\|_B$ in many Riesz (Banach ideal) spaces B such as Lebesgue spaces $B = L_\alpha(X,\mu)$ with $\mathcal{L}(f) = \|f\|_\alpha = (\int_X |f|^\alpha d\mu)^{\frac{1}{\alpha}}, 1 \le \alpha \le +\infty$. We remind the reader the following definition: a Banach space B of μ–measurable functions (with the usual identification of functions which coincide almost everywhere) is called a Riesz (or a Banach ideal) space, if for all μ–measurable f, g such that $|f| \le |g|$ and $g \in B$, we have $f \in B$ and $\|f\|_B \le \|g\|_B$. In a Riesz space, the existence of the representation (1.6) is equivalent to the property of order semicontinuity of the norm (see, e.g., Kantorovich and Akilov [Kan–Aki, Theorem 6, p.190]): for any pointwise non–decreasing sequence $\{f_n\} \subset B$, converging pointwise to a function $f \in B$, one has $\|f_n\|_B \to \|f\|_B$. In particular, any Orlicz norm can be represented in the form (1.6). Another class of functionals which admit this representation is the one given by

$$(1.7) \qquad \mathcal{L}(f) \ = \|f - m(f)\|_B,$$

where $m(f) = \int_X f d\mu$. This last class of functionals is of particular interest for probability measures because it is shift invariant, i.e., $\mathcal{L}(f + const) = \mathcal{L}(f)$ and such is also the modulus of gradient functional. It is also clear that (1.7) corresponds to (1.6) for the functions $g - m(g), g \in G$. Another interesting example of representation (1.6) is the functional $\mathcal{L}(f) = \inf_{a \in \mathbf{R}} ||f - a||_\alpha$. On the other hand, the functional

$$\mathcal{L}(f) = \int_X |f| \log |f| d\mu - \int_X |f| d\mu \log \int_X |f| d\mu = \sup_{\int_X e^g d\mu \leq 1} \int |f| g d\mu$$

admit the representation (1.5), but cannot be expressed in the form (1.6).

For (X, d, μ), and if χ_A is the indicator function of the set A, the extension of Rothaus' theorem can now be formulated as:

Theorem 1.1 *Let $c \geq 0$, then the following are equivalent:*

a) for all Borel measurable (or closed) sets $A \subset X$,

$$\mu^+(A) \ \geq \ c \max(\mathcal{L}(\chi_A), \mathcal{L}(-\chi_A)); \tag{1.8}$$

b) for all bounded Lipschitz functions f,

$$\int_X |\nabla f| d\mu \ \geq \ c\mathcal{L}(f). \tag{1.9}$$

Under (1.8), if f is Lipschitz (not necessarily bounded) and if $\mathcal{L}(f)$ is defined, then it is finite and moreover (1.9) holds.

Often, the inequality (1.9) extends to functions which are Lipschitz on every ball in X. So, roughly speaking, (1.9) is true for all functions if and only if it is true for indicator functions. It will also be clear that the above result, as well as many others, continue to hold for two measures, i.e., if μ in the right hand sides of (1.8) and (1.9) is replaced by another (non–atomic, separable) Borel probability measure ν. In fact, often, our results also continue to hold for infinite measures μ.

Part of our attention will be devoted to Theorem 1.1 for Orlicz spaces $L_N = L_N(X, \mu)$ with norm $\| \cdot \|_N$ given by:

$$||f||_N \ = \ \inf \left\{ \lambda > 0 : \int_X N\left(\frac{f(x)}{\lambda}\right) d\mu(x) \leq 1 \right\}, \tag{1.10}$$

where N is a non–negative, even, convex function on the real line such that $N(x) = 0$ only for $x = 0$ (such N will be called a *Young function*, although no polynomial behavior at infinity is assumed). Below, we will, in particular, discuss the following three questions:

- How to find the optimal constant c in (1.9) via (1.8), if $\mathcal{L}(f) = \|f - m(f)\|_N$?

• Does there exists a function N, as above, for which the (analytic) inequality (1.9) with $\mathcal{L}(f) = \|f - m(f)\|_N$ becomes equivalent to the (geometric) isoperimetric inequality in (X, d, μ). In other words, can (1.9) be equivalent to the isoperimetric problem in (X, d, μ) or, more precisely, to the extremal property of some sets in the isoperimetric problem?

• Does there exists a minimal Orlicz space which contains the Sobolev space $W = W(X, \mu)$ as an embedded space? So as not to consider problems of completeness, by W we mean the space (equipped with the norm $||f||_W = \int_X |\nabla f| d\mu$) of all μ–integrable functions f which are Lipschitz on every ball in X, and such that $m(f) = \int_X f d\mu = 0$.

To precise our second question, let us explain what is meant here by the *isoperimetric problem.* In its classical form to solve the isoperimetric problem, one needs to minimize the "surface area" $\mu^+(A)$ when the "volume" $\mu(A) = p \in (0, 1)$ is constant. In other words, one needs to find the *isoperimetric function*

$$I_\mu(p) = \inf_{\mu(A)=p} \mu^+(A), \tag{1.11}$$

where the infimum is taken over all Borel measurable (or, equivalently, closed) sets $A \subset X$ of measure $\mu(A) = p$, and where by our separable and non–atomic assumptions, the value of $I_\mu(p)$ is defined for all $p \in (0, 1)$. The sets A_p, of measure p, which attain the infimum in (1.11) are called *extremal.* Moreover, a set A_p is said to possess the *isoperimetric property,* if for all $h > 0$, A_p minimizes $\mu(A^h)$ among all the sets of measure p. The problem of minimizing $\mu(A^h)$ represents an "integral" version of the isoperimetric problem in (X, d, μ) and in some canonical cases it is equivalent to finding (1.11). These types of connections will be further explained in the sequel.

To further clarify our first two questions, let us see what is the information conveyed by Theorem 1.1 when $\mathcal{L}(f) = ||f - m(f)||_\alpha$, $(1 \le \alpha < +\infty)$ (for $\alpha = +\infty$, the result below continue to hold replacing $|| \cdot ||_\alpha$ by $|| \cdot ||_\infty$ and $(p^\alpha(1-p) + p(1-p)^\alpha)^{1/\alpha}$ by $\max(p, 1-p)$).

Theorem 1.2 *Let $c \ge 0$, then the following are equivalent:*

a) for all $p \in (0, 1)$,

$$I_\mu(p) \ \ge \ c(p^\alpha(1-p) + p(1-p)^\alpha)^{1/\alpha}; \tag{1.12}$$

b) for all μ–integrable, Lipschitz on every ball, function f

$$\int_X |\nabla f(x)| d\mu(x) \ \ge c \ \left(\int_X |f(x) - m(f)|^\alpha d\mu(x) \right)^{1/\alpha}. \tag{1.13}$$

Moreover, the optimal constant is given by

$$c(\alpha) = \inf_{0<p<1} \frac{I_\mu(p)}{(p^\alpha(1-p) + p(1-p)^\alpha)^{1/\alpha}}. \tag{1.14}$$

Indeed, for $p = \mu(A)$, $||\chi_A - p||_\alpha = (p^\alpha(1-p) + p(1-p)^\alpha)^{1/\alpha}$, hence (1.8) together with (1.11) becomes (1.12). Furthermore, the optimal constant $c = c(\alpha)$ in (1.13) has the stated form by the very definition of the isoperimetric function. Thus the geometric information contained in inequalities like (1.13) is exactly (1.12). Moreover, it may occur that (1.12) turns into an equality, and then (1.13) expresses (i.e., is equivalent to) the isoperimetric property of some (extremal) sets. For example, let $X = \mathbf{R}$ with the usual metric, and let μ be the probability distribution such that

$$\mu((-\infty, x]) = \frac{1}{1 + \exp(-x)},$$

for all $x \in \mathbf{R}$ (in probability and statistics, μ is known as the *logistic distribution*). It will be shown (see Section 13) that for the logistic distribution, the intervals $(-\infty, x]$ are extremal in the isoperimetric problem, and that they moreover possess the isoperimetric property in the sense defined above. Hence by (1.11), $I_\mu(p) = p(1-p)$, and for this measure (1.12) turns into an equality with $\alpha = 1$ and $c = 1/2$. Therefore, in such a situation, (1.13) is equivalent to the isoperimetric property of the intervals $(-\infty, x]$. Likewise (see Section 2), for any fixed $\alpha \geq 1$, $c > 0$, there exists a Borel measure μ on the real line for which (1.12) becomes an equality. For instance, if $\alpha = 2, c = 1$, such a measure μ has density $\mu(dx)/dx = (\cos x)/2, |x| \leq \pi/2$. As it will be seen below, the same conclusion holds, with some constant c, for the normalized Lebesgue measure on the 2–sphere. Of course, in general, the isoperimetric function differs from the function on the right–hand side of (1.12). Therefore (1.13) looses part of the geometric content and cannot provide exact information about the extremal sets in the isoperimetric problem.

To further illustrate Theorem 1.2, we apply it to the uniform (i.e., normed Lebesgue) distribution σ_n on the n–sphere $S^n_\rho \subset \mathbf{R}^{n+1}$ of radius $\rho > 0$. In this case, the isoperimetric function I_{σ_n} can be found via the Lévy–Schmidt theorem [Lev, pp.219–222], [Sch], on the extremal property of the balls (caps) on the sphere: $I_{\sigma_n}(p) = \sigma_n^+(A_p)$, where A_p is an arbitrary ball of σ_n–measure p. For the circle ($n = 1$), A_p are just intervals on S^1_ρ of (Lebesgue) length $(2\pi\rho)p$, so by (1.3), $\sigma_1^+(A_p) = 1/(\pi\rho)$, and

$$I_{\sigma_1}(p) = \frac{1}{\pi\rho}, \quad 0 < p < 1, \tag{1.15}$$

is thus constant. As for the best constant in (1.13), one can conclude using (1.14), and as done in Section 7, that:

Proposition 1.3 *For S^1_ρ, the infimum in (1.14) is attained at $p = 1/2$ whenever $1 \leq \alpha \leq 3$, and then $c(\alpha) = 2/(\pi\rho)$, and is attained at another point $p(\alpha)$ when $\alpha > 3$, and then $c(\alpha)$ depends on α decreasing from $2/(\pi\rho)$ to $1/(\pi\rho)$ at infinity.*

In other words, the inequality

$$\int_{S^1_\rho} |\nabla f(x)| \, \frac{dx}{2\pi\rho} \;\; \geq \;\; \frac{2}{\pi\rho} \left(\int_{S^1_\rho} |f(x) - m(f)|^\alpha \frac{dx}{2\pi\rho} \right)^{1/\alpha}, \tag{1.16}$$

holds for all $1 \leq \alpha \leq 3$, with (asymptotic) equality (and up to an additive) for $f = \chi_{A_{1/2}}$. When $\alpha > 3$, the optimal constant is smaller, e.g., $c(4) = 12^{1/4}/(\pi\rho)$, and the extremal functions are of the form $f = \chi_{A_p}$ for some $p \neq 1/2$. For $\alpha = 1$, (1.16) written for the (usual, non–normalized) Lebesgue measure, is already mentioned in [Oss, p.1205] and in [Rot, p.303] (as Feinberg's Wirtinger–type inequality) and for $\alpha = 2$, it is obtained in [Rot, p.303] (it is not stressed there that when written with respect to the normalized measure, the optimal constants coincide). Finally, let us note that since I_{σ_1} is not of the form (1.12), (1.13) does not involve the isoperimetric inequality on the circle whenever $\alpha \geq 1$. On the other hand, Wirtinger's inequality

$$\int_{S^1_1} |\nabla f(x)|^2 dx \;\geq\; \int_{S^1_1} |f(x) - m(f)|^2 dx,$$

which also holds for all Lipschitz functions, is equivalent to the isoperimetric property of the disks on the plane, i.e., to the classical isoperimetric inequality (1.2) for $n = 2$ ([Oss, p.1184]).

For $S^n_\rho, n \geq 2$, and as shown in the present notes, it follows from the Lévy–Schmidt theorem that the isoperimetric function I_{σ_n} can also be written analytically as

$$I_{\sigma_n}(p) \;=\; \frac{s_{n-1}}{s_n \rho^n} (\rho^2 - F_n^{-2}(p))^{\frac{n-1}{2}}, \tag{1.17}$$

where $s_{n-1} = 2\pi^{n/2}/?(n/2)$ is the area of the unit sphere S^{n-1}_1, and where $F^{-1}_n : (0,1) \longrightarrow (-\rho, \rho)$ is the inverse of the distribution function F_n, of density

$$\frac{dF_n(x)}{dx} \;=\; \frac{s_{n-1}}{s_n \rho^{n-1}} (\rho^2 - x^2)^{\frac{n-2}{2}}, \quad |x| < \rho. \tag{1.18}$$

For the 2–sphere, (1.17) and (1.18) give $I_{\sigma_2}(p) = \sqrt{p(1-p)}/\rho$ which coincides for $\alpha = 2$, and $c = 1/\rho$, with the right–hand side of (1.12). Therefore, the Sobolev–type inequality

$$\int_{S^2_\rho} |\nabla f(x)| d\sigma_2(x) \;\geq \frac{1}{\rho} \left(\int_{S^2_\rho} |f(x) - m(f)|^2 d\sigma_2(x) \right)^{1/2}$$

is equivalent to the isoperimetric property of the balls on the 2–sphere.

For $n \geq 3$, the function I_{σ_n} does not have a further analytic expression beyond (1.17) and anyway, it is not of the form (1.12). It is nevertheless possible to find the optimal constant and the extremal functions in (1.13) by solving the analytic problem of minimizing (1.14) for the function I defined by (1.17)–(1.18). To do so, one can first observe that $I_{\sigma_n}(p)$ is asymptotically equivalent to $p^{(n-1)/n}$ as $p \to 0^+$, so $c(\alpha) > 0$, only for $1 \le \alpha \le n/(n-1)$. To find a point p of minimum in (1.14), we then suggest the following useful sufficient condition (see Section 8):

Lemma 1.4 *Let $\alpha \ge 1$, and let I be a non–negative function defined on $(0,1)$ such that:*

(i) I is symmetric about $1/2$;

(ii) $I(0^+) = I(1^-) = 0$;

(iii) I is continuously differentiable on $(0,1)$ and dI^α/dp is convex on $(0,1/2]$.

Then, the function $I^\alpha(p)/(p(1-p))$ attains its infimum at $p = 1/2$. Therefore, for $1 \le \alpha \le 2$, and for functions I_μ satisfying (i), (ii) and (iii) above, the minimum in (1.14) is attained at $p = 1/2$, i.e., the infimum $c(\alpha)$ is equal to $2I(1/2)$.

It turns out that when $I = I_{\sigma_n}$, the isoperimetric function of the n–sphere ($n \ge 2$), the requirements (i)–(iii) of Lemma 1.4 are fulfilled for the whole range of allowed values of α, i.e., for $\alpha \in [1, n/(n-1)]$ (see Section 9). Thus by (1.17),

$$c(\alpha) \;=\; 2I_{\sigma_n}(1/2) \;=\; 2s_{n-1}/(\rho s_n),$$

which does not depend on α. This is somehow unexpected, since for any individual Lipschitz function f, the L_α–norm on the right-hand side of (1.13) is an increasing function of α, and the left side of (1.13) does not depend on α either. This discussion can be summarized in:

Proposition 1.5 *Let $n \ge 2$ and $1 \le \alpha \le n/(n-1)$. Then, for any Lipschitz function f on S^n_ρ,*

$$\int_{S^n_\rho} |\nabla f(x)| d\sigma_n(x) \ge c_n \left(\int_{S^n_\rho} |f(x) - m(f)|^\alpha d\sigma_n(x) \right)^{1/\alpha}, \tag{1.19}$$

where the constant $c_n = 2s_{n-1}/(\rho s_n)$ is optimal. For $\alpha > n/(n-1)$, there is no positive constant satisfying (1.19).

For the extremal value $\alpha = n/(n-1)$, let us rewrite (1.19) with respect to the Lebesgue measure μ_n on the sphere S^n_ρ. For any Lipschitz function f with $m(f) = \int_{S^n_\rho} f d\sigma_n = 0$, and if $\|f\|_{\alpha,\mu_n}$ denotes the L_α–norm with respect to μ_n, we have

$$\|f\|_{\alpha,\mu_n} \;\le\; K_n \|\nabla f\|_{1,\mu_n}, \tag{1.20}$$

where the optimal constant $K_n = s_n^{(n-1)/n}/(2s_{n-1})$ does not depend on ρ. For smooth f with $m(f)$ arbitrary, rewrite (1.20) as $\|f\|_{\alpha,\mu_n} \le K_n \|\nabla f\|_{1,\mu_n} + \rho^n s_n |m(f)|$, which implies

$$\|f\|_{\alpha,\mu_n} \;\le\; K_n\|\nabla f\|_{1,\mu_n} + \|f\|_{1,\mu_n}. \tag{1.21}$$

However, K_n is now suboptimal. The optimal constant in front of $||\nabla f||_1$ in (1.21) is known, due to Aubin [Aub, p.50], and given by $K(n,1) = 1/(n\omega_n^{1/n}) = n^{(1-n)/n}/s_{n-1}^{1/n}$. Curiously, $1/K(n,1)$ is the optimal constant in (1.1) for the Lebesgue measure in $\mathbf{R}^n$. On the other hand, it is also clear that the optimal inequality of Aubin, applied to $f - m(f)$ can only give (1.19) with a suboptimal constant.

A remark on the cases of equality in (1.19) when $n \ge 2$, $1 \le \alpha \le n/(n-1)$. Equality can only be asymptotic for a sequence of smooth functions converging, up to an additive constant, to the indicator function $\chi_{A_{1/2}}$ of the half–sphere $A_{1/2}$ and 0. Note also that among all the functions f with $m(f) = 0$, only the functions $f = const(\chi_A - 1/2)$ (where A has measure $1/2$) are such that $||f||_\alpha$ does not depend on α.

It is also worthwhile to exploit Theorem 1.1 for other types of functionals. One such functional of interest to both probabilists (in connection with the αth mean) and to geometers (in connection with the first eigenvalue problem, see Yau [Yau], Li [Li]) is

$$|||f|||_\alpha \;=\; \inf_{a\in\mathbf{R}} \left(\int_X |f(x)-a|^\alpha d\mu(x)\right)^{1/\alpha}.$$

Since $2|||f|||_\alpha \ge (\int_X |f-m(f)|^\alpha d\mu)^{1/\alpha} \;\ge\; |||f|||_\alpha$, (1.13) always implies

$$\int_X |\nabla f(x)| d\mu(x) \;\ge\; d|||f|||_\alpha, \tag{1.22}$$

where $d \ge c(\alpha)$ and where $c(\alpha)$ is given by (1.14). One might wonder whether or not in (1.22), the constant can be sharpened, i.e., is $d > c(\alpha)$ possible? In many interesting cases, including the n–sphere ($n \ge 2$), the answer is negative. Indeed, by Theorem 1.1 the inequality (1.22), for all bounded Lipschitz functions on (X, d, μ), is equivalent to

$$\mu^+(A) \;\ge\; d|||\chi_A|||_\alpha, \tag{1.23}$$

for all Borel measurable $A \subset X$. Now, if $p = \mu(A)$, a simple computation shows that for $\alpha > 1$, $|||\chi_A|||_\alpha^\alpha \;=\; p(1-p)/(p^{\frac{1}{\alpha-1}} + (1-p)^{\frac{1}{\alpha-1}})^{\alpha-1}$, while for $\alpha = 1$, $|||\chi_A|||_1 = \min(p, 1-p)$. From (1.23), to find the optimal constant $d(\alpha)$ in (1.22) we can now appeal to:

Lemma 1.6 *Let $1 \le \alpha \le 2$, let the function I satisfy the hypotheses of Lemma 1.4, and let*

$$d(1) \;=\; \inf_{0<p<1} \frac{I(p)}{\min(p, 1-p)}, \qquad \alpha = 1,$$

$$d(\alpha)^\alpha \;=\; \inf_{0<p<1} \frac{I(p)^\alpha}{p(1-p)} \left(p^{\frac{1}{\alpha-1}} + (1-p)^{\frac{1}{\alpha-1}}\right)^{\alpha-1}, \qquad 1 < \alpha \le 2.$$

Then, $d(1) = d(\alpha) = 2I(1/2)$.

In other words, the infima above are attained at $p = 1/2$ and for $I = I_\mu$, satisfying (i), (ii) and (iii) of Lemma 1.4,

$$d(\alpha) \;=\; c(\alpha) \;=\; 2I_\mu(1/2),\; 1 \le \alpha \le 2. \tag{1.24}$$

Thus, for such isoperimetric functions I_μ and for $1 \le \alpha \le 2$, (1.13) is stronger than (1.22) when these inequalities are written with optimal constants. Of course, for $\alpha = 2$, the converse statement is true. Applying Lemma 1.6 to the sphere S_ρ^n, we have

Proposition 1.7 *Let $n \ge 2$ and $1 \le \alpha \le n/(n-1)$. Then, for any Lipschitz function f on S_ρ^n,*

$$\int_{S_\rho^n} |\nabla f(x)| d\sigma_n(x) \ge d_n \inf_{a\in\mathbf{R}} \left(\int_{S_\rho^n} |f(x) - a|^\alpha d\sigma_n(x) \right)^{1/\alpha}, \tag{1.25}$$

where the constant $d_n = 2s_{n-1}/(\rho s_n)$ is optimal. For $\alpha > n/(n-1)$, there is no positive constant satisfying (1.25).

Again, the case of the circle S_ρ^1 is somehow different since the isoperimetric function is constant. The optimal constant is independent of α for all values of α, this contrasts Proposition 1.3.

Proposition 1.8 *For any Lipschitz function f on S_ρ^1, and for $1 \le \alpha \le +\infty$,*

$$\int_{S_\rho^1} |\nabla f(x)| d\sigma_1(x) \ge \frac{2}{\pi\rho} \inf_{a\in\mathbf{R}} \left(\int_{S_\rho^1} |f(x) - a|^\alpha d\sigma_1(x) \right)^{1/\alpha}, \tag{1.26}$$

where the constant $2/(\pi\rho)$ is optimal.

Let us further comment on Theorem 1.1 and its range of applicability. First, note that $2^{(1-\alpha)}\min(p, 1-p) \le p(1-p)/(p^{\frac{1}{\alpha-1}} + (1-p)^{\frac{1}{\alpha-1}})^{\alpha-1} \le \min(p, 1-p), \alpha \ge 1$. Hence, Theorem 1.1 applied, for example, to the functional $|||\cdot|||_{n/(n-1)}$ on an n–dimensional, $n \ge 2$, compact Riemannian manifold (without boundary) M^n recovers some well known results (see [Li, p.452]) and [Yau, p.499]). The case of the n–sphere is doubly important. First, by the Lévy–Gromov isoperimetric inequality ([Gro]), the isoperimetric function of S_ρ^n is a lower bound for the isoperimetric function of classes of M^n. Indeed, let $R(M^n)$ denote the infimum, over all the unit tangent vectors of M^n, of the Ricci tensor and let I_{M^n} be the isoperimetric function of the manifold (with respect to the normalized Riemannian measure). If $R(M^n) \ge (n-1)\kappa > 0$, then for all $p \in (0,1)$, $I_{M^n}(p) \ge I_{\sigma_n,\sqrt{\kappa^{-1}}}(p)$, where $I_{\sigma_n,\sqrt{\kappa^{-1}}}$ is the isoperimetric function of the n–sphere of radius $1/\sqrt{\kappa}$. Thus, for such manifolds, the Sobolev constants appearing in (1.13) and (1.22), $1 \le \alpha \le n/(n-1)$, are bounded below by $2s_{n-1}\sqrt{\kappa}/s_n = \sqrt{2/\pi}\sqrt{\kappa}?((n+1)/2)/?(n/2)$. We also note that the inequalities (1.19) and (1.25) (for $\alpha = n/n-1$) are the strongest in the hierarchy of Sobolev inequalities in that they respectively imply ($n > p$, n integer),

$$\left(\int_{S_\rho^n} |\nabla f(x)|^p d\sigma_n(x)\right)^{1/p} \geq c\left(\int_{S_\rho^n} |f(x) - m(f)|^{\frac{np}{n-p}} d\sigma_n(x)\right)^{(n-p)/np},$$

$$\left(\int_{S_\rho^n} |\nabla f(x)|^p d\sigma_n(x)\right)^{1/p} \geq d \inf_{a\in\mathbf{R}}\left(\int_{S_\rho^n} |f(x) - a|^{\frac{np}{n-p}} d\sigma_n(x)\right)^{(n-p)/np},$$

where $c = d/2$ and d are now suboptimal constants (see [Li, p.453] for this last inequality in case $p = 2$ and $d = (n-2)c_n/2(n-1)$ and where c_n is as in Proposition 1.5).

A second important reason for particularly considering the n–sphere stems from a classical result usually attributed to Poincaré. Roughly, this result states that for n large, γ_n the standard Gaussian measure on $\mathbf{R}^n$, of density $(2\pi)^{-n/2}\exp(-|x|^2/2)$, $x \in \mathbf{R}^n$, is somehow almost concentrated and uniformly distributed on the sphere S_ρ^{n-1} of radius $\rho = \sqrt{n-1}$ (in particular, as n goes to infinity, I_{σ_n} converges pointwise to I_{γ_1} while ρ depends on n as above). With the help of the Lévy–Schmidt theorem, Sudakov and Tsirel'son ([Sud–Tsi]) as well as Borell ([Bor]) showed that the half–spaces are extremal in the isoperimetric problem for γ_n (see also Ehrhard [Ehr] for a proof not relying on the Lévy–Schmidt theorem). The isoperimetric function for the Gaussian measure is thus given by $I_{\gamma_n}(p) = \varphi(\Phi^{-1}(p))$, where φ is the density of the standard Gaussian measure on $\mathbf{R}$, and where Φ^{-1} is the inverse of its distribution function. It can easily be checked that for $\mu = \gamma_n$, the best constants in (1.13) and in (1.22) are non–zero only when $\alpha = 1$. Hence, using respectively Lemma 1.4 and Lemma 1.6 we have, $c(1) = 2I_{\gamma_1}(1/2) = \sqrt{2/\pi} = d(1)$. Thus, denoting by $M(f)$ a median of f with respect to γ_n, we have:

Proposition 1.9 *For any Lipschitz function f on $\mathbf{R}^n$, $n \geq 1$,*

$$\int_{\mathbf{R}^n} |\nabla f(x)| d\gamma_n(x) \;\geq\; \sqrt{\frac{2}{\pi}}\int_{\mathbf{R}^n} |f(x) - m(f)| d\gamma_n(x), \tag{1.27}$$

$$\int_{\mathbf{R}^n} |\nabla f(x)| d\gamma_n(x) \;\geq\; \sqrt{\frac{2}{\pi}}\int_{\mathbf{R}^n} |f(x) - M(f)| d\gamma_n(x), \tag{1.28}$$

where the constant $\sqrt{2/\pi}$ is optimal.

The inequality (1.27) is due to Pisier [Pis, p.178] (with a different method), and it has also been independently rederived using isoperimetric methods by Ledoux [Led2]. In fact, (1.19) and (1.25) respectively become (1.27) and (1.28) as $n \to \infty$ and $\rho = \sqrt{n-1}$. The optimal constant $\sqrt{2/\pi}$ can also be found as $\lim_{n\to+\infty} 2s_{n-1}/(\sqrt{n-1}s_n)$. A characteristic feature of the Gaussian measure is that the optimal constants in the above proposition are dimension–free, and so (1.27) and (1.28) continue to hold for infinite dimensional Gaussian measures. Moreover, inequalities (with suboptimal constants) where both the gradient and the function are estimated in $L_p(\gamma_n), 1 < p < +\infty$, instead of $L_1(\gamma_n)$ also follow from (1.27) and (1.28).

According to Theorem 1.1, (1.27) and (1.28) are respectively equivalent to $I_{\gamma_n}(p) \geq \sqrt{2/\pi}2p(1-p)$ and to $I_{\gamma_n}(p) \geq \sqrt{2/\pi}\min(p, 1-p)$. These two estimates only indicate the exponential character of the tails $1 - \gamma_n(A^h)$ as $h \to +\infty$, while the true tails have at most the Gaussian rate of decrease $\varphi(h)/(\sqrt{2\pi}h)$ (provided $\gamma_n(A) = 1/2$). So, there might exist a space more preferable than L_1, e.g., an Orlicz space, in which Sobolev–type inequalities would say more on the isoperimetric problem. Ledoux [Led1] proved an analog of (1.27) with an existential constant for the Orlicz space L_N whose norm is generated by the function $N(x) = |x|\sqrt{2\log(1+|x|)}$. By Theorem 1.1, this gives a lower estimate on I_{γ_n} which is, up to constant, asymptotically equivalent to I_{γ_n} as $p \to 0^+$. Pellicia and Talenti [Pel–Tal] improved this result by modifying N so that it satisfies $2p(1-p)N(1/\varphi(\Phi^{-1}(p))) = 1$, $0 < p < 1$, N is linear on $[0, \sqrt{2\pi}]$, and showed that for this N and all smooth f,

$$\|f - m(f)\|_N \leq \int_{\mathbf{R}^n} |\nabla f| d\gamma_n,$$

with asymptotic equality for some sequence of smooth functions. For indicator functions $f = \chi_A$, this inequality becomes $\gamma_n^+(A) \geq \|\chi_A - \gamma_n(A)\|_N$ which coincides with the isoperimetric inequality $\gamma_n^+(A) \geq I_{\gamma_n}(\gamma_n(A))$ in case $\gamma_n(A) = 1/2$ and differs from it (and, therefore, is weaker) when $\gamma_n(A) \neq 1/2$. So, one might wonder whether or not there exists an Orlicz space $L_N(\mathbf{R}^n, \gamma_n)$ for which an inequality as above becomes the isoperimetric inequality for all $p = \gamma_n(A)$. This is further explained now, and a necessary and sufficient condition for the equivalence of a Sobolev–type inequality and of isoperimetry is presented.

Variational problems and optimal Orlicz spaces

We would now like to have another look at Theorem 1.1 for Orlicz spaces. This look corresponds to a probabilistic point of view according to which the variational problem of minimizing the value of $\int_X |\nabla f| d\mu$ should preferably be solved in terms of F_f the distribution function of f with respect to μ, i.e., $F_f(t) = \mu\{f \leq t\}, t \in \mathbf{R}$. From this, other types of estimations which depend on f only via F_f (like the moments of f), may be of interest. A first reduction of the problem (via a co–area inequality obtained in Section 3) leads to the estimate

$$\int_X |\nabla f(x)| d\mu(x) \ \geq \ J = \int_{a(f)}^{b(f)} I_\mu(1 - F_f(t)) dt, \tag{1.29}$$

where f is Lipschitz on every ball in X, $a(f) = \operatorname{essinf} f$ and $b(f) = \operatorname{esssup} f$. This inequality has not yet lost any geometric information, in that it implies the isoperimetric inequality $\mu^+(A) \geq I_\mu(p)$. This is easily seen by applying (1.29) to a sequence of Lipschitz functions converging to the indicator function χ_A (it should also be noted here that (1.29) already appears when $\mu = \gamma_n$ in Ledoux [Led1]). The functional J on the right–hand side of (1.29) depends only on F_f. Thus, instead of dealing with the notion of gradient, one can study this functional just assuming that f is a measurable function. So, to get a Sobolev–type inequality, for the Orlicz functional $\mathcal{L}(f) = \|f - m(f)\|_N$, one needs to minimize the L^1–norm of the modulus of gradient or (equivalently!) minimize J, under the conditions

$$\int_X f d\mu = a, \qquad \int_X N(f) d\mu = b, \tag{1.30}$$

with $a = 0$ and $b = 1$. In general, the extremal functions which minimize J under (1.30), take at most three values whenever N is continuous. A remarkable fact explaining Theorem 1.1 for Orlicz spaces is that, for N *convex*, the extremal functions take at most *two* values. Moreover, this last statement can be generalized to functions f satisfying boundary conditions, i.e.,

$$k_1 \leq f \leq k_2, \tag{1.31}$$

where in general $-\infty \leq k_1 < k_2 \leq +\infty$. The extremal functions which minimize J under (1.30) and (1.31) take at most four values, two of which are k_1 and k_2 if they are finite, and at most three values one of which is k_1 (in case $k_1 > -\infty$ and $k_2 = +\infty$) or k_2 (in case $k_2 < +\infty$ and $k_1 = -\infty$). The restrictions (1.31) appear naturally in many inequalities such as

$$\|f\|_N \leq c_1 \int_X |\nabla f| d\mu + c_2 \int_X |f| d\mu, \tag{1.32}$$

a partial case of which is the inequality (1.21). Note that in order to find all the optimal (c_1, c_2) in (1.32), we cannot apply Theorem 1.1 since $\mathcal{L}(f) = \|f\|_N - c_2 \int_X |f| d\mu$ is not of the form (1.5). A possible approach to obtaining (1.32) is firstly to note that only non–negative functions need to be considered in the minimizing problem and secondly to note that

$$\int_X |\nabla f| d\mu,$$

can be minimized, using (1.29) by minimizing J under the conditions (1.30)–(1.31), with arbitrary $a \geq 0$, $b > N(a)$, putting also $k_1 = 0, k_2 = +\infty$. However, for our purposes, we do not consider the variational problem with boundary conditions here, although the proof of Theorem 5.1 where we study the infimum of J carries over without essential changes to the boundary conditions case.

Returning to our main interest (the case without boundary conditions), we combine the above observations with (1.29) in the following statement:

Let N be a non–linear (not of the form $N(x) = cx + d$) convex function on the real line $\mathbf{R}$. Fix $p \in (0, 1)$, define $x = x_p(a, b)$ as (the only) positive solution of

$$pN(a + qx) + qN(a - px) = b, \tag{1.33}$$

and let $N^*(a, b) = \inf_{0<p<1} I_\mu(p) x_p(a, b)$.

Theorem 1.10 *For any $b > N(a)$,*

$$\inf \int_X |\nabla f| d\mu \;=\; N^*(a,b), \tag{1.34}$$

where the infimum is taken over all μ-integrable Lipschitz (or Lipschitz on every ball) functions f on (X,d,μ) satisfying (1.30).

Roughly speaking, to minimize the L_1–norm of the gradient under (1.30), one need only consider functions taking two values. Note that no assumption is made on the function I_μ. Also, it is clear that setting $a = 0$, $N(x) = |x|^\alpha$, (1.33) gives $x_p(0,b) = \left(b/(p^\alpha(1-p) + p(1-p)^\alpha)\right)^{1/\alpha}$. Hence,

$$N^*(0,b) = \inf_{0<p<1} \frac{I_\mu(p)}{\left(p^\alpha(1-p) + p(1-p)^\alpha\right)^{1/\alpha}} b^{1/\alpha},$$

and Theorem 1.10 recovers Theorem 1.2.

Let us now return to the case where N is a Young function. Without loss of generality, putting $c = 1$ in Theorem 1.1, we see that by (1.10), $x = 1/\|\chi_A - p\|_N$, where $\mu(A) = p \in (0,1)$, is the only positive solution to (1.33) for $a = 0$ and $b = 1$. Therefore, the Sobolev-type inequality

$$\int_X |\nabla f| d\mu \;\geq\; \|f - m(f)\|_N, \tag{1.35}$$

for bounded Lipschitz functions, which by Theorem 1.1 (as well as by Theorem 1.10) is equivalent to the inequality $\mu^+(A) \geq \|\chi_A - p\|_N$, coincides with the (exact) isoperimetric inequality $\mu^+(A) \geq I_\mu(p)$ if and only if

$$\|\chi_A - p\|_N = I_\mu(p), \tag{1.36}$$

for all $p \in (0,1)$, i.e., if and only if

$$pN\left(\frac{1-p}{I_\mu(p)}\right) + (1-p)N\left(\frac{p}{I_\mu(p)}\right) = 1. \tag{1.37}$$

In this case and only in this case is (1.35) equivalent to the solution of the isoperimetric problem. We are now left with the question of finding a Young function N satisfying (1.36) or equivalently (1.37). Recall that $W(X,\mu)$ denotes the space (equipped with the norm $||f||_W = \int_X |\nabla f| d\mu$) of all μ–integrable functions f which are Lipschitz on every ball in X, and such that $m(f) = \int_X f d\mu = 0$. The following completely characterizes the existence of such Young functions.

Theorem 1.11 *There exists a Young function N satisfying (1.36) or, equivalently, (1.37) if and only if the isoperimetric function $I = I_\mu$ possesses the following properties:*

1) $I(p) > 0$, for all $p \in (0,1)$; $I(0^+) = I(1^-) = 0$;

2) $I(p) = I(1-p)$, for all $p \in (0,1)$;

3) the function $p(1-p)/I(p)$ is concave on $(0,1)$.

In this case, any Orlicz space $L_M(X,\mu)$, containing $W(X,\mu)$ as an embedded space, contains $L_N(X,\mu)$ as well.

If they exist, there are a lot of functions N satisfying (1.37). For example, when $I(p) = p(1-p)$, all the N satisfying (1.37) are described by: $N(x) = (1-2a) + a|x|$, $|x| \geq 1$, where $a \in [0,1)$ is an arbitrary parameter, and where the behavior of N on $[-1,1]$ can also be chosen arbitrarily (as long as N remains a Young function). For the 2–sphere of unit radius, where $I(p) = \sqrt{p(1-p)}$, one can choose $N(x) = x^2, x \in \mathbf{R}$ or (see Remark 10.2)

$$N(x) = \begin{cases} |x| & \text{if } |x| \leq 1 \\ 1 - |x| + x^2 & \text{if } |x| \geq 1. \end{cases}$$

In any case, given a function I with the properties 1)–3), all the Young functions N satisfying (1.37) generate equivalent Banach spaces. One can therefore say that there is unique minimal Orlicz space containing the Sobolev space W.

We now apply this general statement to the uniform distribution on the sphere and to the standard Gaussian measure on $\mathbf{R}^n$. As a rule, the properties 1)–2) are easily verified but, often, 3) is not so simple.

Theorem 1.12 *The isoperimetric function corresponding to the uniform distribution on the n–sphere ($n \geq 2$) and the one corresponding to the standard Gaussian measure on $\mathbf{R}^n$, satisfy the conditions 1)–3) of Theorem 1.11.*

For the n–sphere, ($n \geq 2$), all the Orlicz spaces L_N with N satisfying (1.37) are equivalent (via Proposition 1.5) to the Lebesgue space $L_{n/(n-1)}$, which is thus minimal in the sense described above. As already noted, for the circle the isoperimetric function is constant, so 1) fails. Therefore, the extremal property of the intervals on the circle cannot be expressed via a Sobolev–type inequality (when the gradient is estimated in the L_1–norm). The same conclusion is true for the two–sided exponential distribution μ, of density $(\exp -|x|)/2, x \in \mathbf{R}$, in which case, $I_\mu(p) = \min(p, 1-p)$, so 1)–2) are fulfilled but 3) fails.

Isoperimetry and Sobolev Inequalities on the real line.

To finish these notes, our setting is $\mathbf{R}$ equipped with a non–atomic Borel probability measure. In that context, we study isoperimetry for a class of "regular" μ. By "regular", we mean those measures whose distribution function F is concentrated on an interval (finite or not) $(a_F = \inf\{F > 0\}, b_F = \sup\{F < 1\})$ and such that F is absolutely continuous with positive and continuous density f. For those probability measures, we give necessary and sufficient conditions under which the intervals $(-\infty, x]$ are extremal in the isoperimetric problem. In this case, these intervals turn out to also be extremal in the "integral" problem, i.e., they possess the isoperimetric property. The isoperimetric function for such μ is of the form $I_\mu(p) = f(F^{-1}(p))$, where F^{-1} is the inverse of F, and in fact, any such μ is the Lipschitz image of the double exponential distribution. Symmetric (around a point) log–concave measures satisfy these sufficient conditions. This implies, in particular, that if μ is a symmetric log–concave measure, then

$$\int_{\mathbf{R}} |f - m(f)| d\mu \leq c(\mu) \int_{\mathbf{R}} |f'| d\mu,$$

for some finite, positive constant $c(\mu)$ and all integrable smooth functions f.

Then, to better understand the possible behavior of I_μ, we look for its analytic expression when μ is unimodal (log–concave, not necessarily symmetric, measures are unimodal). An analytic expression for I_μ can help to find the optimal (as previously defined) Orlicz space. However, it is possible (as shown in Section 14) to find the optimal constant c in (1.9) for $\mathcal{L}(f) = \|f - m(f)\|_N$ without knowing I_μ. To this end, we find the "minimal" (for the pointwise order) weight w for which the Sobolev inequality

$$\|f - m(f)\|_N \leq \int_{\mathbf{R}} |f'(x)| w(x) dx,$$

holds for all smooth functions f on $\mathbf{R}$. The paper concludes by presenting a simple induction process which allow to extend some of these inequalities to product probability measures on $\mathbf{R}^n$.

Organization of the paper.

The paper is divided into several sections.

Section 2. We discuss here two approaches to the definition of the isoperimetric problem. In particular, we look for an equivalent "integral" form $\mu(A^h) \geq R_h(\mu(A))$ for the isoperimetric inequality $\mu^+(A) \geq I_\mu(A)$.

Section 3. A co–area inequality is proved from which Theorem 1.1 follows.

Sections 4,5,6. Theorem 1.10 is proved. Although this statement is a partial case of Theorem 1.1 when N is a convex function, and although its proof is much longer than that of Theorem 1.1, the arguments developed there are important for probabilistic extensions. In Section 4 the equivalence of (1.29) and of the isoperimetric inequality is stated. In Sections 5 and 6 the functional J is minimized under (1.30) (Section 6 only deals with the discrete version of Theorem 1.10).

Section 7. Proposition 1.3 and 1.8 are proved. In fact, the behavior of the function $I_\alpha(p) = ((p^\alpha(1-p) + p(1-p)^\alpha))^{1/\alpha}$ is studied.

Section 8. A partial case of Theorem 1.2 is studied, Lemma 1.4 and 1.6 are proved.

Section 9. The isoperimetric function of the n–sphere, $n \geq 2$ is found and Proposition 1.5 and 1.7 are proved.

Section 10. Theorem 1.11 is proved.

Section 11 and 12. Theorem 1.12 is respectively proved for the sphere and for Gauss space.

Section 13, 14 and 15. Isoperimetry and Sobolev type inequalities are studied on $\mathbf{R}$. Some extensions of these inequalities from $\mathbf{R}$ to $\mathbf{R}^n$ are presented.

2 Some connections between differential and integral forms of isoperimetric inequalities

Again, let (X, d, μ) be a metric space with a separable and non–atomic Borel probability measure μ. Therefore for any $p \in (0,1)$, there exist closed sets $A \subset X$ of μ-measure p, so that the function

$$I_\mu(p) \ = \ \inf_{\mu(A)=p} \mu^+(A), \tag{2.1}$$

where the infimum is taken over all Borel measurable sets of measure p, is well-defined on $(0,1)$. This infimum can also be taken over all closed sets of measure p. Indeed, for arbitrary $A \subset X$, $\bigcap_{h>0} A^h = \bar{A}$, where $\bar{A}$ is the closure of A. Hence if $\mu(\bar{A}) - \mu(A) > 0$,

$$\frac{\mu(A^h) - \mu(A)}{h} \ \geq \ \frac{\mu(\bar{A}) - \mu(A)}{h} \ \to \ +\infty,$$

as $h \to 0^+$, thus $\mu^+(A) = +\infty$ and (2.1) can be taken over all sets A such that $\mu(\bar{A}) = \mu(A) = p$. But for such sets, and taking into account the identity $A^h = \bar{A}^h$ (valid for any $A \subset X$), we have $\mu(A^h) - \mu(A) = \mu(\bar{A}^h) - \mu(\bar{A})$. Therefore $\mu^+(A) = \mu^+(\bar{A})$, and only closed sets of measure p need to be taken in (2.1). In fact, this also applies to all the inequalities of the form

$$\mu^+(A) \ \geq \ I(\mu(A)), \tag{2.2}$$

and, in particular, when I is the isoperimetric function I_μ.

In this section, we look for an inequality of the type

$$\mu(A^h) \ \geq \ R_h(\mu(A)), \tag{2.3}$$

for some function R_h, which would be equivalent to (2.2). For example, $p = \mu(A)$, the inequalities

$$\mu^+(A) \geq 2cp(1-p), \quad \text{and } \mu^+(A) \geq c(p(1-p))^{1/2}, \tag{2.4}$$

which are just (1.12) for $\alpha = 1$ and $\alpha = 2$, (c is a positive constant), turn out to be respectively equivalent, to

$$\mu(A^h) \ \geq \ \frac{p}{p + (1-p)\exp(-2ch)}, \ h > 0, \tag{2.5}$$

and

$$\mu(A^h) \ \geq \ \frac{1}{2}(1 - \cos(ch)) + p\cos(ch) + (p(1-p))^{1/2}\sin(ch), \tag{2.6}$$

$0 < ch \leq \pi/2 - \arcsin(2p-1)$.

Now, let I be a positive and continuous function on $(0,1)$, and let

$$a_I = -\int_0^{1/2} \frac{dp}{I(p)}, \quad b_I = \int_{1/2}^0 \frac{dp}{I(p)},$$

where $-\infty \leq a_I < 0 < b_I \leq +\infty$. Let also F_I be the unique function from $\mathbf{R}$ to [0,1] such that

(i) F_I is continuous and non–decreasing on $\mathbf{R}$;

(ii) $F_I(a_I^+) = 0$, $F_I(0) = 1/2$, $F_I(b_I^-) = 1$;

(iii) F_I has a positive continuous derivative f_I on (a_I, b_I) and for all $p \in (0,1)$,

$$f_I(F_I^{-1}(p)) = I(p), \tag{2.7}$$

where $F_I^{-1} : (0.1) \longrightarrow (a_I, b_I)$, is the inverse of F_I restricted to (a_I, b_I). Note that F_I can be defined on (a_I, b_I) as the inverse of the function

$$F_I^{-1}(p) = \int_{1/2}^{p} \frac{dt}{I(t)}, \quad 0 < p < 1. \tag{2.8}$$

With the above notation, we now have:

Theorem 2.1 *Let I be a positive continuous function on $(0,1)$, then the following statements are equivalent:*

(a) For all $h > 0$, and for all Borel measurable A with $0 < \mu(A) < 1$,

$$\mu(A^h) \;\geq\; F_I(F_I^{-1}(\mu(A)) + h). \tag{2.9}$$

(b) For all Borel measurable A with $0 < \mu(A) < 1$,

$$\mu^+(A) \;\geq\; I(\mu(A)). \tag{2.10}$$

(c) For any $h > 0$, let the h-neighbourhood of any open ball $D(x,r) \subset X$ be a ball. The inequality (2.10) is satisfied for all sets A, with $0 < \mu(A) < 1$, which are finite unions of open balls in X.

Before presenting the proof of this theorem, let us provide some comments and examples. If $I = I_\mu$, is known one can, with the help of (2.9), estimate the best function in (2.3), i.e., the function

$$R_h(p) = \inf_{\mu(A)=p} \mu(A^h). \tag{2.11}$$

Moreover, if the extremal sets A_p in the isoperimetric problem (2.1) exist and possess the property

$$\mu(A_p^h) = F_I(F_I^{-1}(\mu(A)) + h), \tag{2.12}$$

for all $h > 0$, then these sets minimize the infimum in (2.11), thus providing the solution to the "integral" isoperimetric problem (2.11). The property (2.12) is therefore sufficient to pass from the original problem (2.1) to the "integral" one (2.11).

Here are some examples of extremal sets which satisfy (2.12): the balls on the sphere with respect to the uniform distribution; the half–spaces for the standard Gaussian measure; the intervals of the form $(-\infty, x]$ for an arbitrary symmetric log–concave distribution on the real line (see Section 13). On the other hand, the extremal sets in the integral problem (2.11) for the exponential distribution on $[0, +\infty)$, of density $\exp(-x), x \geq 0$, do not satisfy (2.12). They depend on h and are either intervals of the form $[0, a]$ or intervals of the form $[b, +\infty)$. In this last case, $I_\mu(p) = \min(p, 1-p)$ coincides with the isoperimetric function for the two–sided exponential law of density $(\exp -|x|)/2, x \in \mathbf{R}$ (again, see Section 13). It is easily checked that the value $R_h(p)$, defined by (2.11) is greater than the right–hand side of (2.9). Note also that the function I_μ does not change when the metric d is replaced by an equivalent one, but $R_h(p)$ is essentially determined by the metric.

Let us see how to apply the above theorem to the function $I(p) = 2cp(1-p)$. From (2.8), we get $F_I^{-1}(p) = (\log{(p/1-p)})/2c$, i.e.,

$$F_I(x) = \frac{1}{1 + \exp(-2cx)}, x \in \mathbf{R},$$

and finally,

$$F_I(F_I^{-1}(p) + h) = \frac{p}{p + (1-p)\exp(-2ch)}, \;\; h > 0.$$

This proves the equivalence between (2.5) and the first inequality in (2.4). For the other example, when $I(p) = c(p(1-p))^{1/2}$, we then have

$$cF_I^{-1}(p) = \int_{1/2}^{p} \frac{dt}{(t(1-t))^{1/2}} = \arcsin(2p-1).$$

Therefore $F_I(x) = (1 + \sin(cx))/2$, $|cx| \leq \pi/2$, and

$$\begin{aligned} F_I(F_I^{-1}(p) + h) &= \frac{1 + \sin(\arcsin(2p-1) + ch)}{2} \\ &= \frac{1 + (2p-1)\cos(ch) + \cos(\arcsin(2p-1))\sin(ch)}{2}, \end{aligned}$$

which clearly coincides with the right–hand side of (2.6). The condition $0 \leq ch \leq \pi/2 - \arcsin(2p-1)$ corresponds to $F_I^{-1}(p) + h \leq b_I = \pi/2$, otherwise $F_I(F_I^{-1}(p) + h) = 1$. When $\alpha \neq 1, 2$, an equivalent form for (1.12) is of more complicated nature.

We now formulate the equivalence of part b) and c) of Theorem (2.1) separately to make further use of it in our next remark.

Corollary 2.2 *For any $h > 0$, let the h–neighbourhood of any open ball $D(x, r) \subset X$ be a ball. If (2.10) is satisfied for the sets A, with $0 < \mu(A) < 1$, which are finite* *unions of open balls in X, then it is satisfied for all Borel measurable sets $A \subset X$, with* $0 < \mu(A) < 1$.

Example 2.3 Let $X = S^1_\rho \subset \mathbf{R}^2$ be the circle of radius ρ in the plane and let $\mu = \sigma_1$ be the uniform distribution on S^1_ρ, i.e., $\mu = \mu_1/(2\pi\rho)$, where μ_1 is the Lebesgue measure on the circle. Then for any set $A \subset X$, with $0 < \mu(A) < 1$, representable as a finite union of disjoint open intervals D_k, $1 \le k \le N$, we have that $2\pi\rho\mu^+(A)$ is the number of points of the D_k which do not belong to the closure of other intervals. Anyhow, $2\pi\rho\mu^+(A) \ge 2$, with equality for $N = 1$. Therefore, by Corollary 2.2, this last inequality remains true for any Borel measurable $A \subset X$, with $0 < \mu(A) < 1$, and we finally obtain that, for all $p \in (0,1)$,

$$I_\mu(p) = \frac{1}{\pi\rho}. \tag{2.13}$$

Using (2.8) as well as the equivalence of (2.9) and (2.10), (2.13) can also be written as:

$$\mu(A^h) \ge \min\left(\mu(A) + \frac{h}{\pi\rho}, 1\right), \tag{2.14}$$

which is valid for any Borel measurable $A \subset X$, with $0 < \mu(A) < 1$. (2.14) is a one-dimensional (and, of course, trivial) case of the Lévy–Schmidt theorem on the isoperimetric property of the balls on the sphere. Note also that, in this case, the inequality $\mu^+(A) \ge I_\mu(\mu(A))$ fails if $\mu(A) = 1$ or if $A = \emptyset$. On the other hand, (2.14) fails only for the empty set.

Proof of Theorem 2.1. The proof of the equivalence of a) and b) does not differ from that of a) and c), so we simply prove the latter. Trivially from (2.7), a) implies c) and one needs only to prove the converse implication. Given $h \ge 0$, $0 < p < 1$, set

$$R_h(p) = F(F^{-1}(p) + h), \tag{2.15}$$

where $F = F_I$, and set also $R_h(0) = 0$, $R_h(1) = 1$. Then, R_h forms a family of non–decreasing continuous functions on $[0,1]$ with the following semi–group property: for all $h,\ h' \ge 0$:

$$R_{h+h'}(p) = R_h(R_{h'}(p)), \tag{2.16}$$

for all $p \in [0,1]$. Indeed, (2.16) is trivial for $p = 0$. Then let $p > 0$, and $R_{h'}(p) < 1$ (hence $p < 1$), and by (2.15), $R_{h'}(p) = F(F^{-1}(p) + h')$. Since $0 < R_{h'}(p) < 1$, we then have

$$R_h(R_{h'}(p)) = F(F^{-1}(R_{h'}(p) + h) = F(F^{-1}(p) + h + h') = R_{h+h'}(p).$$

If $R_{h'}(p) = 1$, then $R_h(R_{h'}(p)) = R_h(1) = 1$, and also $1 \ge R_{h+h'}(p) \ge R_{h'}(p) = 1$, since R_h is a non–decreasing function of h. Thus again, $R_{h+h'}(p) = R_{h'}(p) = 1$ and (2.16) is established.

Now we need to show that, for all Borel measurable sets $A \subset X$,

$$\mu(A^h) \ge R_h(\mu(A)). \tag{2.17}$$

First, we slightly modify (2.17) by introducing a parameter $\sigma > 1$ and defining

$$R_h^\sigma(p) = F\left(F^{-1}(p) + \frac{h}{\sigma}\right),\ 0 \le p \le 1,\ h \ge 0.$$

As for $\sigma = 1$, the family R_h^σ satisfies (2.16). First, we prove that for finite unions A of open balls in X and for all $h > 0$,

$$\mu(A^h) \ge R_h^\sigma(\mu(A)),\ \sigma > 1. \tag{2.18}$$

Then, letting $\sigma \to 1$ we will obtain (2.17) for the same sets. Fix such a set A of measure $0 < \mu(A) < 1$, and put

$$\Delta = \{h > 0 : (2.18) \text{ is true for all } h' \in (0, h]\}.$$

Note that the function $h \longrightarrow R_h^\sigma(\mu(A))$ is continuous on $[0, +\infty)$, and that the function $h \longrightarrow \mu(A^h)$ is left continuous on $(0, +\infty)$. Therefore, to prove that $\Delta = (0, +\infty)$, it suffices to show that

i) $\epsilon \in \Delta$, for $\epsilon > 0$ small enough;

ii) If $h \in \Delta$, then $h + \epsilon \in \Delta$, for $\epsilon > 0$ small enough.

For $\epsilon > 0$ small enough, and by the definition of μ^+,

$$\mu(A^\epsilon) \ge \mu(A) + \mu^+(A)\epsilon + \mathrm{o}(\epsilon). \tag{2.19}$$

On the other hand, the Taylor expansion of $R_h^\sigma(p)$ at $h = 0$ gives

$$\begin{aligned} R_\epsilon^\sigma(\mu(A)) &= \mu(A) + f(F^{-1}(\mu(A))\frac{\epsilon}{\sigma} + \mathrm{o}(\epsilon) \\ &= \mu(A) + I(\mu(A))\frac{\epsilon}{\sigma} + \mathrm{o}(\epsilon), \end{aligned} \tag{2.20}$$

where $f = f_I$. By the assumption (2.10), $\mu^+(A) \ge I(\mu(A))$, and comparing (2.19) and (2.20), we get (2.18) for $h > 0$ small enough, i.e., we proved i). Suppose now that $h \in \Delta$. If $\mu(A^h) > R_h^\sigma(\mu(A))$, then this inequality remains true for all $h + \epsilon$ with ϵ small enough since the function $h \longrightarrow \mu(A^h)$ is non–decreasing, and since the function $h \longrightarrow R_h^\sigma(\mu(A))$ is continuous. In the other possible case, i.e., when $\mu(A^h) = R_h^\sigma(\mu(A))$, put $B = A^h$, and note that $A^{h+\epsilon} = B^\epsilon$ for all $\epsilon > 0$. If $\mu(B^\epsilon) = 1$, for all $\epsilon > 0$, then $h \in \Delta$ automatically. Suppose now that $\mu(B^{\epsilon'}) < 1$ for some $\epsilon' > 0$, and let $\epsilon \in (0, \epsilon')$. In particular, $0 < \mu(B) < 1$, and since by assumption, $A = D_1 \cup \cdots \cup D_n$ is the union of the balls D_i, then $A^h = D_1^h \cup \ldots \cup D_n^h$ is also an union of balls. Therefore, (2.10) can be applied to B and $\mu^+(B) \ge I(\mu(B))$. Again, writing (2.19) and (2.20) for B one gets

$$\begin{aligned} \mu(B^\epsilon) &\ge \mu(B) + \mu^+(B)\epsilon + \mathrm{o}(\epsilon), \\ R_\epsilon^\sigma(\mu(B)) &= \mu(B) + f(F^{-1}(\mu(B))\frac{\epsilon}{\sigma} + \mathrm{o}(\epsilon) \\ &= \mu(B) + I(\mu(B))\frac{\epsilon}{\sigma} + \mathrm{o}(\epsilon), \end{aligned}$$

and thus concludes that $\mu(B^\epsilon) \geq R^\sigma_\epsilon(\mu(B))$, for all $\epsilon > 0$ small enough. It remains to note that

$$R^\sigma_{h+\epsilon}(\mu(A)) = R^\sigma_\epsilon(R^\sigma_h(\mu(A))) = R^\sigma_\epsilon(\mu(A^h)) = R^\sigma_\epsilon(\mu(B)) \leq \mu(B^\epsilon) = \mu(A^{h+\epsilon}).$$

Therefore $h+\epsilon \in \Delta$ for all $\epsilon > 0$ small enough. Thus, (2.18) and therefore (2.17) are true for any set A, with $0 < \mu(A) < 1$, and which is a finite union of open balls. If $\mu(A) = 0$ or 1, (2.17) is automatically true.

If A is an arbitrary open set in X, then since μ is separable, there exists a sequence of open balls $D_i \subset A, i \geq 1$, such that $\mu(A_n) \longrightarrow \mu(A)$ as $n \to \infty$, where $A_n = D_1 \cup \cdots \cup D_n$. Since (2.17) is valid for A_n, it extends to A and (2.17) extends to all open sets. Now, let $K \subset X$ be closed. The set K^ϵ is open, hence for all $h > 0$, $\mu((K^\epsilon)^h) = \mu(K^{h+\epsilon}) \geq R_h(\mu(K^\epsilon))$. Letting $\epsilon \to 0^+$, and since $\bigcap_{\epsilon>0} A^\epsilon = \bar{A}$, we get $\mu\left(\overline{K^h}\right) \geq R_h(\mu(K))$, for all $h > 0$. But for all $h' < h$, $\overline{K^{h'}} \subset K^h$ and therefore $\mu(K^h) \geq \mu\left(\overline{K^{h'}}\right) \geq R_{h'}(\mu(K))$. Letting $h' \to h^-$, we obtain $\mu(K^h) \geq R_h(\mu(K))$. Finally, for an arbitrary Borel measurable set A, there exists a sequence of closed sets $K_n \subset A$ such that $\mu(K_n) \to \mu(A)$ as $n \to \infty$, hence for all $h > 0$,

$$\mu(A^h) \geq \mu(K^h_n) \geq R_h(\mu(K_n)) \to R_h(\mu(A)),$$

as $n \to \infty$. Theorem 2.1 is thus proved.

3 Proof of Theorem 1.1

Lemma 3.1 (Co-area Inequality) *Let f be a Lipschitz function on X, then*

$$\int_X |\nabla f(x)| d\mu(x) \ \geq \ \int_{-\infty}^{+\infty} \mu^+\{x \in X : f(x) > t\} dt. \tag{3.1}$$

Remark 3.2 If f is Lipschitz on every ball in X, the function $|\nabla f|$ is Borel measurable and finite. Indeed,

$$|\nabla f(x)| \ = \ \lim_{n\to\infty} \sup_{d(x,y)<1/n} \frac{|f(x) - f(y)|}{d(x,y)}$$

is Borel measurable as the monotone limit of a sequence of lower semi–continuous functions. Finiteness follows from the Lipschitz property.

Remark 3.3 Let $A \subset X$ be Borel measurable, and let r take only rational values. Whenever $h > 0$, $\cup_{0<r<h} A^r = A^h$, hence for any $\epsilon > 0$,

$$\inf_{0<h<\epsilon} \frac{\mu(A^h) - \mu(A)}{h} = \inf_{0<r<\epsilon} \frac{\mu(A^r) - \mu(A)}{r}. \tag{3.2}$$

Therefore,

$$\liminf_{r\to 0^+} \frac{\mu(A^r) - \mu(A)}{r} = \mu^+(A).$$

Thus, for any non–increasing family of Borel measurable sets A_t, $t \in \mathbf{R}$, the function $t \longrightarrow \mu^+(A_t)$ is Borel measurable (on the real line), and so is the integrand on the right hand side of (3.1).

Remark 3.4 Since $\mu^+(\emptyset) = \mu^+(X) = 0$, the second integral in (3.1) is in fact taken over the interval $(a(f), b(f))$, where $a(f) = \text{ess inf } f$, and where $b(f) = \text{ess sup } f$.

Remark 3.5 It should be noted that the proof of Lemma 3.1 does not require any assumption on the Borel probability measure μ (not even that μ is non–atomic). Equality in (3.1) requires some additional properties of μ, such as non–singularity. In fact, let $X = \mathbf{R}$ with its usual metric, let μ be an arbitrary Borel probability measure on $\mathbf{R}$ and let ν denote the absolutely continuous (with respect to the Lebesgue measure) part of μ. If $f(x) = x$, then $p(t) = \mu^+\{x \in X : f(x) > t\}$ is a Radon–Nikodym derivative (with respect to the Lebesgue measure) of ν, and (3.1) becomes

$$\nu(\mathbf{R}) \leq 1.$$

Therefore, and for $X = \mathbf{R}$, equality in (3.1) requires that $\mu = \nu$, i.e., that μ is absolutely continuous. As well known, the "usual" coarea formula tells us that this property is also sufficient.

To prove Theorem 1.1, we will also need the following result which is an immediate consequence of Lemma 3.1.

Corollary 3.6 *Let f be a Lipschitz function on X, such that $\mu\{x \in X : f(x) = 0\} = 0$. Then,*

$$\int_X |\nabla f(x)| d\mu(x) \geq \int_{-\infty}^{0} \mu^+\{x \in X : f(x) < t\} dt + \int_0^{+\infty} \mu^+\{x \in X : f(x) > t\} dt.$$

Proof. It is enough to apply Lemma 3.1 to f^+ and f^- and to also note that $|\nabla f| = |\nabla f^+| + |\nabla f^-|$ on the open set $\{x \in X : f(x) \neq 0\}$.

Proof of Lemma 3.1. First, let us assume that f is bounded. Then, without loss of generality one may assume that $f \geq 0$, since the left and the right hand side of (3.1) remain unchanged if a constant is added to f. Since f is Lipschitz on X,

$$|f(x) - f(y)| \leq c d(x, y), \tag{3.3}$$

for some $c > 0$ and all $x, y \in X$. Then, let

$$f_h(x) = \sup_{d(x,y)<h} f(y),$$

where $h > 0$, and let $A_t = \{x \in X : f(x) > t\}$. Then, for all $t \in \mathbf{R}$ and $h > 0$, the set $\{x \in X : f_h(x) > t\} = \{x \in X : f(x) > t\}^h = A_t^h$ is open as the open h–neighbourhood of A_t. Therefore f_h is lower semi–continuous and in addition,

$$\int_X f_h d\mu \;=\; \int_0^{+\infty} \mu\{x \in X : f_h(x) > t\} dt \;=\; \int_0^{+\infty} \mu(A_t^h) dt.$$

Since $\int_X f d\mu = \int_0^{+\infty} \mu(A_t) dt$, we have

$$\int_X \frac{f_h - f}{h} d\mu \;=\; \int_0^{+\infty} \frac{\mu(A_t^h) - \mu(A_t)}{h} dt. \tag{3.4}$$

From (3.3), $f_h(x) - f(x) \leq ch$, for all $x \in X$ and $h > 0$, hence the integrand on the left hand side of (3.1) is bounded. Therefore, using (3.4), the Lebesgue dominated convergence theorem and Fatou's Lemma (via property (3.2)) and noting that

$$\limsup_{h \to 0^+} \frac{f_h(x) - f(x)}{h} = \limsup_{y \to x} \frac{f(y) - f(x)}{d(x, y)} \leq |\nabla f(x)|,$$

we get

$$\begin{aligned}\int_X |\nabla f| d\mu &= \int_X \limsup_{h\to 0^+} \frac{f_h - f}{h} d\mu \geq \limsup_{h\to 0^+} \int_X \frac{f_h - f}{h} d\mu \\ &\geq \liminf_{h\to 0^+} \int_X \frac{f_h - f}{h} d\mu = \liminf_{h\to 0^+} \int_0^{+\infty} \frac{\mu(A_t^h) - \mu(A_t)}{h} dt \\ &\geq \int_0^{+\infty} \liminf_{h\to 0^+} \frac{\mu(A_t^h) - \mu(A_t)}{h} dt = \int_0^{+\infty} \mu^+(A_t) dt.\end{aligned}$$

Thus, (3.1) is established for f Lipschitz and bounded. Let now f be an arbitrary Lipschitz function. Let a_n be an increasing sequence of positive numbers such that $\lim_{n\to+\infty} a_n = +\infty$, and such that the sets $D_n = \{x \in X : |f(x)| = a_n\}$ have $\mu-$measure 0, for all n. Let $A_n = \{x \in X : |f(x)| < a_n\}$, and define the function

$$f_n(x) = \begin{cases} f(x), & \text{if } |f(x)| < a_n \\ a_n, & \text{if } f(x) \geq a_n \\ -a_n, & \text{if } f(x) \leq -a_n \end{cases}$$

That is, $f_n(x) = \max\{-a_n, \min\{a_n, f(x)\}\}$, so f_n is also a Lipschitz function (of Lipschitz constant at most $\max(c, 1)$) and thus one can apply (3.1) to f_n. Since on A_n, which is open, $f = f_n$, we have $|\nabla f_n(x)| = |\nabla f(x)|$, for any $x \in A_n$. Now, the sets

$$B_n = \{x \in X : f(x) < -a_n\}, \quad C_n = \{x \in X : f(x) > a_n\}$$

are also open, and f_n is constant on both B_n and C_n, so $|\nabla f_n| = 0$ on $B_n \cup C_n$. In addition, $D_n = X \setminus (A_n \cup B_n \cup C_n)$ has $\mu-$measure 0, consequently taking into account Remark 3.4, (3.1) can be written as

$$\int_{A_n} |\nabla f(x)| d\mu(x) \;\geq\; \int_{-a_n}^{a_n} \mu^+\{x \in X : f(x) > t\} dt. \tag{3.5}$$

Finally, applying Tonelli's monotone convergence theorem, we get (3.1) from (3.5).

Proof of Theorem 1.1.

$a) \Rightarrow b)$. Without loss of generality, let $c = 1$. We only need to consider the case of single pair sets $G = \{(g_1, g_2)\}$. Indeed, if

$$\mathcal{L}_{(g_1,g_2)}(f) = \int_X (f^+ g_1 + f^- g_2) d\mu \leq \int_X |\nabla f| d\mu \tag{3.6}$$

follows from $\mu^+(A) \geq \max(\mathcal{L}_{(g_1,g_2)}(\chi_A), \mathcal{L}_{(g_1,g_2)}(-\chi_A))$, then taking the supremum in (3.6) over all $(g_1, g_2) \in G$ will give $\mathcal{L}(f) \leq \int_X |\nabla f| d\mu$, under the condition $\mu^+(A) \geq \max(\mathcal{L}(\chi_A), \mathcal{L}(-\chi_A))$. Now, let $\mathcal{L} = \mathcal{L}_{(g_1,g_2)}$ and assume that

$$\mu^+(A) \geq \max(\mathcal{L}(\chi_A), \mathcal{L}(-\chi_A)), \tag{3.7}$$

for any Borel measurable $A \subset X$. First assume that $\mu\{x \in X : f(x) = 0\} = 0$, then by Corollary 3.6, and by (3.7), we have putting $A_t = \{f > t\}$ and $B_t = \{f < t\}$:

$$\begin{aligned}
\int_X |\nabla f(x)| d\mu(x) &\geq \int_0^{+\infty} \mathcal{L}(\chi_{A_t}) dt + \int_{-\infty}^0 \mathcal{L}(-\chi_{B_t}) dt \\
&= \int_0^{+\infty} \int_X \chi_{A_t}(x) g_1(x) d\mu(x) dt + \int_{-\infty}^0 \int_X \chi_{B_t}(x) g_2(x) d\mu(x) dt \\
&= \int_X \int_0^{+\infty} \chi_{A_t}(x) g_1(x) dt d\mu(x) + \int_X \int_{-\infty}^0 \chi_{B_t}(x) g_2(x) dt d\mu(x) \\
&= \int_X \chi_{A_0}(x) f(x) g_1(x) d\mu(x) + \int_X \chi_{B_0}(x) f(x) g_2(x) d\mu(x) \\
&= \int_X f^+ g_1 d\mu + \int_X f^- g_2 d\mu = \mathcal{L}(f).
\end{aligned}$$

Hence, (3.6) is proved when $\mu\{x \in X : f(x) = 0\} = 0$, note also that we have applied Fubini's Theorem to change the order of integrations. This is valid since using the integrability of $f^+ g_1$ and of $f^- g_2$, we have $\int_X \int_0^{+\infty} \chi_{A_t}(x)|g_1(x)| dt d\mu(x) = \int_X f^+ |g_1| d\mu < +\infty$, and $\int_X \int_{-\infty}^0 \chi_{B_t}(x)|g_2(x)| dt d\mu(x) = \int_X f^+ |g_2| d\mu < +\infty$.

Let us now show how to get rid of the extra assumption $\mu\{x \in X : f(x) = 0\} = 0$. Let $C = \{a \in X : \mu(f = a) > 0\}$, then C is at most countable and by the previous arguments, for any $a \notin C$,

$$\mathcal{L}(f - a) \leq \int_X |\nabla(f - a)| d\mu = \int_X |\nabla f| d\mu, \tag{3.8}$$

whenever $(f - a)^+ g_1$ and $(f - a)^- g_2$ are μ–integrable which is always true by the integrability of $f^+ g_1, f^- g_2, g_1$ and g_2. In addition for this same last reason,

$$\mathcal{L}(f - a) = \int_X (f - a)^+ g_1 d\mu + \int_X (f - a)^- g_2 d\mu \longrightarrow \mathcal{L}(f),$$

as $a \to 0$. Therefore, (3.8) holds for $a = 0$ and b) is established.

b) $\Rightarrow$ *a)*. Again, and without loss of generality, let $c = 1$. By (1.9), for any bounded Lipschitz function $f \geq 0$ on X, and for all $(g_1, g_2) \in G$,

$$\int_X |\nabla f| d\mu \geq \int_X f g_1 d\mu, \quad \int_X |\nabla f| d\mu = \int_X |\nabla(-f)| d\mu \geq \int_X f g_2 d\mu. \tag{3.9}$$

Now we approximate sets A by Lipschitz functions f. Let $A \subset X$ be a closed set such that $0 < \mu(A) < 1$. For any $\epsilon > 0$, there exists a Lipschitz function f^ϵ on X with values in $[0, 1]$, of Lipschitz constant at most $1/\epsilon$ such that $f^\epsilon = 1$ on some open neighbourhood of A and $f^\epsilon = 0$ on $X \setminus A^\epsilon$. One may choose, for example,

$$f^\epsilon(x) = \max\left(1 - \frac{1}{\epsilon} d(x, A^{\frac{\epsilon}{2}}), 0\right),$$

where for B a non–empty subset of X, $d(x,B)=\inf\{d(x,b):b\in B\}$. By (1.4), $|\nabla f^\epsilon|\leq 1/\epsilon$ everywhere (since the function $d(x,B)$ has Lipschitz constant at most 1 when B is chosen as above), and $|\nabla f^\epsilon|=0$ on $X\setminus\overline{A^\epsilon}$. Hence,

$$\int_X|\nabla f^\epsilon|d\mu \;\leq\; \frac{\mu(\overline{A^\epsilon})-\mu(A)}{\epsilon}.$$

Taking into account that for all $\epsilon'<\epsilon$, $\overline{A^{\epsilon'}}\subset A^\epsilon$, we have

$$\mu^+(A) \;=\; \liminf_{\epsilon\to 0^+}\frac{\mu(\overline{A^\epsilon})-\mu(A)}{\epsilon} \;\geq\; \liminf_{\epsilon\to 0^+}\int_X|\nabla f^\epsilon|d\mu. \tag{3.10}$$

On the other hand, since A is closed, f^ϵ converges pointwise, as $\epsilon\to 0^+$, to the indicator function χ_A, hence whenever g is μ-integrable, $\int_X f^\epsilon g d\mu\to\int_X\chi_A g d\mu$, as $\epsilon\to 0^+$. So,

$$\lim_{\epsilon\to 0^+}\int_X f^\epsilon g_1 d\mu \;=\; \int_X\chi_A g_1 d\mu,\quad \lim_{\epsilon\to 0^+}\int_X f^\epsilon g_2 d\mu \;=\; \int_X\chi_A g_2 d\mu. \tag{3.11}$$

Now (3.10), (3.9) and (3.11) yield

$$\mu^+(A)\geq\int_X\chi_A g_1 d\mu, \tag{3.12}$$

$$\mu^+(A)\geq\int_X\chi_A g_2 d\mu, \tag{3.13}$$

whenever $(g_1,g_2)\in G$. Taking the supremum over all $(g_1,g_2)\in G$ in respectively (3.12) and (3.13) gives $\mu^+(A)\geq\mathcal{L}(\chi_A)$ and $\mu^+(A)\geq\mathcal{L}(-\chi_A)$. Thus, a) follows for all closed sets $A\subset X$, with $0<\mu(A)<1$. If A is Borel measurable but not closed, then two cases occur. Either $\mu(\overline{A})>\mu(A)$, then as noted in Section 2, $\mu^+(A)=+\infty$, hence there is nothing to prove. Or, $\mu(\overline{A})=\mu(A)$, then $\mathcal{L}(\chi_A)=\mathcal{L}(\chi_{\overline{A}}),\mathcal{L}(-\chi_A)=\mathcal{L}(-\chi_{\overline{A}})$, and $\mu^+(\overline{A})=\mu^+(A)$, as again noted in the previous section.

We are thus just left with the case $\mu(A)=0$ and $\mu(A)=1$. If $\mu(A)=0$, then by definition $\mathcal{L}(\chi_A)=0$ and (1.8) holds since $\mu^+(A)\geq 0=\mathcal{L}(\chi_A)$. Let now $\mu(A)=1$, then applying (1.9) to $f_1=1$ and $f_2=-1$, we obtain

$$0=\mathcal{L}(f_1)=\sup_{(g_1,g_2)\in G}\int_X g_1 d\mu,\quad 0=\mathcal{L}(f_2)=\sup_{(g_1,g_2)\in G}\int_X g_2 d\mu.$$

Hence, $\int_X g_1 d\mu\;\leq\;0,\int_X g_2 d\mu\;\leq\;0$ and therefore, $\mu^+(A)\;\geq\;0\;\geq\;\mathcal{L}(\chi_A),\mu^+(A)\;\geq\;0\;\geq\;\mathcal{L}(-\chi_A)$. The proof of Theorem 1.1 is complete.

4 The isoperimetric problem as a relation between the distribution of a function and its derivative

Let f be a function on (X,d) which is Lipschitz on every ball, and let $F_f(t) = \mu\{x \in X : f(x) \le t\}, t \in \mathbf{R}$, be its distribution function with respect to the measure μ.

Theorem 4.1 *Let I be a non–negative continuous function on $(0,1)$. If for all Borel sets $A \subset X$, with $0 < \mu(A) < 1$,*

(4.1) $$\mu^+(A) \ge I(\mu(A)),$$

then for any function f which is Lipschitz on every ball in X,

(4.2) $$\int_X |\nabla f(x)| d\mu(x) \ \ge \ \int_{a(f)}^{b(f)} I(1 - F_f(t)) dt,$$

where $a(f) = \text{ess inf } f$ and $b(f) = \text{ess sup } f$. Conversely, if (4.2) holds for all bounded Lipschitz functions, then (4.1) also holds for all Borel sets $A \subset X$, with $0 < \mu(A) < 1$.

Remark 4.2 *If $I(0^+) = I(1^-) = 0$, then (4.2) takes the form*

$$\int_X |\nabla f(x)| d\mu(x) \ \ge \ \int_{-\infty}^{+\infty} I(1 - F_f(t)) dt.$$

Proof. For f bounded and Lipschitz, Lemma 3.1 as well as (4.1) imply (4.2). For f bounded and Lipschitz on ever ball, a truncation argument can be used to prove the result. Let

$$T_r(x) = \begin{cases} 1, & \text{if } d(a,x) < r \\ r + 1 - d(a,x), & \text{if } r \le d(a,x) \le r+1 \\ 0, & \text{if } d(a,x) > r+1, \end{cases}$$

where $r > 0$, $x \in X$, and where a is a fixed point in X. Clearly, the function T_r is Lipschitz, of Lipschitz constant at most 1. Let $f_r(x) = f(x)T_r(x)$, then f_r is a bounded Lipschitz function of Lipschitz constant at most $d_r = f(a) + rC_r$, where C_r is the Lipschitz constant of f on the open ball $D(a,r)$. In addition, since the sets $D(a,r)$ and $\{x \in X : d(a,x) > r+1\}$ are open, we have by (1.4):

$$|\nabla f_r(x)| = \begin{cases} |\nabla f_r(x)|, & \text{if } d(a,x) < r \\ 0, & \text{if } d(a,x) > r+1. \end{cases}$$

Let $r < d(a,x) < r+1$, $r < d(a,y) < r+1$. Then,

$$\begin{aligned} f_r(y) - f_r(x) &= f(y)(r+1-d(a,y)) - f(x)(r+1-d(a,x)) \\ &= (f(y)-f(x))(r+1-d(a,y)) + f(x)(d(a,x)-d(a,y)). \end{aligned}$$

Therefore,

$$|f_r(y) - f_r(x)| \le |f(y)-f(x)| + |f(x)|d(x,y),$$

hence

$$|\nabla f_r(x)| = \limsup_{y\to x} \frac{|f_r(y)-f_r(x)|}{d(x,y)} \le |\nabla f(x)| + |f(x)|.$$

Note that, $\mu\{x : d(a,x) = r\} + \mu\{x : d(a,x) = r+1\} = 0$, for all $r > 0$, except maybe, for countably many r. So, not taking such values of r, we have

$$\int_X |\nabla f_r| d\mu \le \int_X |\nabla f| d\mu + \int_X |f(x)| \chi_{\{r<d(a,x)<r+1\}}(x) d\mu(x). \tag{4.3}$$

Now, let $u_r = 1 - F_r$, where F_r is the distribution function of f_r with respect to μ. When $r \to +\infty$, f_r converges to f pointwise and therefore, u_r converges to u weakly. That is, $u_r(t) \longrightarrow u(t)$, and by the continuity of I, $I(u_r(t)) \longrightarrow I(u(t))$, for every point of continuity of u. Consequently, since u is non–increasing, this convergence takes place for all t except countably many t. In addition, $a(f_r) \longrightarrow a(f)$, $b(f_r) \longrightarrow b(f)$ as $r \to +\infty$. Again, applying Fatou's Lemma to the right hand side of (4.2) with f_r and noting that, since f is bounded, the last integral in (4.3) tends to zero as $r \to +\infty$, we finally obtain

$$\begin{aligned} \int_{a(f)}^{b(f)} I(1-F_f(t))dt &= \int_{-\infty}^{+\infty} I(u(t))\chi_{(a(f),b(f))}(t)dt \\ &= \int_{-\infty}^{+\infty} \liminf_{r\to+\infty} I(u_r(t))\chi_{(a(f_r),b(f_r))}(t)dt \\ &\le \liminf_{r\to+\infty} \int_{-\infty}^{+\infty} I(u_r(t))\chi_{(a(f_r),b(f_r))}(t)dt \\ &\le \liminf_{r\to+\infty} \int_X |\nabla f_r| d\mu \\ &\le \int_X |\nabla f| d\mu. \end{aligned}$$

The inequality (4.2) is thus proved for f bounded and Lipschitz on every ball. If f is unbounded, one can use a truncation argument similar to the one used in the proof of Lemma 3.1. Let

$$f_n(x) = \max\{-a_n, \min\{a_n, f(x)\}\},$$

where a_n is an increasing sequence such that $a_n \to +\infty$, and $\mu\{x \in X : |f(x)| = a_n\} = 0$. Clearly, f_n is Lipschitz on every ball, and since $|f_n| \le a_n$, one can apply (4.2) to f_n. Finally, letting $n \to \infty$, (4.2) for such f follows by applying Tonelli's monotone convergence theorem to the left hand side of (4.2) and Fatou's lemma to the right hand side of (4.2).

This proves the direct part of the theorem, and it just remains to prove the converse. Let $A \subset X$ be a closed set such that $0 < \mu(A) < 1$. As in the proof of Theorem 1.1, taking the family of Lipschitz functions f^ϵ, $\epsilon > 0$, which approximate the indicator function $f = \chi_A$. We have that

$$\int_X |\nabla f^\epsilon| d\mu \ \leq \ \frac{\mu(\overline{A^\epsilon}) - \mu(A)}{\epsilon}. \tag{4.4}$$

Since f^ϵ converges pointwise to f, as $\epsilon \to 0^+$, F_ϵ the distribution function (with respect to μ) of f^ϵ converges weakly to the distribution function F of f. In other words, $F_\epsilon(t) \longrightarrow F(t)$ as $\epsilon \to 0$, for all t except at $t = 0$ and $t = 1$ where F is discontinuous. So, the continuity of I and once more Fatou's lemma give

$$\liminf_{\epsilon \longrightarrow 0} \int_0^1 I(1 - F_\epsilon(t))dt \geq \int_0^1 I(1 - F(t))dt = I(\mu(A)). \tag{4.5}$$

Note that, for all $\epsilon > 0$, ess inf $f^\epsilon \geq$ ess inf $f = 0$, ess inf $f^\epsilon \to 0$ as $\epsilon \to 0^+$, and ess sup $f^\epsilon =$ ess sup $f = 1$. So, we get from (4.2), (4.4) and (4.5), taking into account that for all $\epsilon' < \epsilon$, $\overline{A^{\epsilon'}} \subset A^\epsilon$:

$$\begin{aligned}
I(\mu(A)) \ &\leq \ \liminf_{\epsilon \to 0^+} \int_X |\nabla f^\epsilon| d\mu \\
&\leq \ \liminf_{\epsilon \to 0^+} \frac{\mu(\overline{A^\epsilon}) - \mu(A)}{\epsilon} \\
&= \ \liminf_{\epsilon \to 0^+} \frac{\mu(A^\epsilon) - \mu(A)}{\epsilon} \\
&\leq \ \mu^+(A).
\end{aligned}$$

At noted at the beginning of Section 2 (see (2.2), the inequality $\mu^+(A) \geq I(\mu(A))$ extends to all Borel sets A of measure $0 < \mu(A) < 1$. This completes the proof of Theorem 4.1.

5 A variational problem

In order to minimize the value of

$$\int_X |\nabla f| d\mu, \tag{5.1}$$

under the conditions

$$\int_X f d\mu = a, \ \int_X N(f) d\mu = b, \tag{5.2}$$

where a and b are fixed constants and where $N : \mathbf{R} \longrightarrow \mathbf{R}$ is an arbitrary convex function, the inequality (4.2) will be used. The first restriction in (5.2) requires that f should be μ–integrable. In this case, the second integral in (5.2) always exists, being finite or not, because $\int_X \max(N(f), 0) d\mu = \int_X \max(-N(f), 0) d\mu = +\infty$ cannot occur for N convex and f integrable. Thus, the integrability of f is the only restriction in our problem.

If the function N is linear, i.e., of the form $N(x) = cx + d$, then $b = ca + d$, the infimum of (5.1) is zero and is attained for the constant function $f = a$, a.e.μ. Otherwise (N not linear), by Jensen's inequality, b can be any real number such that $b \geq N(a)$. If $b = N(a)$, only the constant function $f = a$, a.e., satisfies (5.2), and the integral in (5.1) is again zero.

Let $b > N(a)$, and let the function f take only two values, say, $a + qx$ and $a - px$, $x \in \mathbf{R}$, with respective μ–probabilities p and q, $(0 < p < 1, q = 1 - p)$. Then, the first integral in (5.2) is equal to a and the second integral which is equal to

$$N_{p,a}(x) = pN(a + qx) + qN(a - px), \tag{5.3}$$

is a convex function of the real variable x. This function is non–increasing on $(-\infty, 0]$, non–decreasing on $[0, +\infty)$, and $N_{p,a}(0) = N(a)$. Moreover, since N is not linear,

$$N_{p,a}(+\infty) = N_{p,a}(-\infty) = +\infty.$$

Therefore, for any $p \in (0,1), b > N(a)$, the equation $N_{p,a}(x) = b$ admits only one positive solution $x = x_p(a,b)$. Furthermore, when p and a are fixed, $x_p(a,b)$ is an increasing, concave function of the variable $b > N(a)$. Thus, for any $b > N(a)$, there exist Borel measurable functions satisfying (5.2). We will show below that, in fact, there exist Lipschitz functions on X satisfying (5.2).

Given $b > N(a)$, let us now define

$$N_I^*(a,b) = \inf_{0<p<1} \left(I(p) x_p(a,b) \right), \tag{5.4}$$

where I is a non–negative function on $(0,1)$. In our main case of interest, when $I = I_\mu$ is the isoperimetric for (X, d, μ), we simply write N^* instead of $N_{I_\mu}^*$.

For the function f described above (taking only two values), the expression inside the infimum in (5.4) is easily seen to be exactly

$$\int_{a(f)}^{b(f)} I(1 - F_f(t))dt, \tag{5.5}$$

where F_f denotes the distribution function of f with respect to the measure μ, and where $a(f) = \mathrm{essinf} f$, $b(f) = \mathrm{esssup} f$. Thus, $N_I^*(a, b)$ is the infimum of such integrals over all functions which take two values and satisfy the conditions (5.2) (such infimum might only be attained asymptotically). Note also that, if $I(0^+) = I(1^-) = 0$, then the integral (5.5) can be extended to the whole real line.

To prove Theorem 1.10, we first need to prove the following statement.

Theorem 5.1 *Let I be a non–negative, continuous function on $(0,1)$, such that the function $I(p)/p$ is non–increasing on $(0,1)$. Then,*

$$\inf \int_{a(f)}^{b(f)} I(1 - F_f(t))dt \ = \ N_I^*(a, b), \tag{5.6}$$

where the infimum is taken over all μ-integrable functions f on X, which satisfy (5.2).

Before proving Theorem 5.1 and Theorem 1.10, we need some preparatory results. Then Theorem 5.1 and Theorem 1.10 are proved, assuming that Theorem 5.1 has already been established for all probability measures $F = F_f$ with finite support, i.e., for the class of functions f which take only finitely many values. The discrete version of Theorem 5.1, which is of interest for different applications and is, in particular, a basic key to comprehend Theorem 1.10, is considered separately in the next section. It should be stressed here that Theorem 5.1, which is used to prove Theorem 1.10, will *not* be applied to the isoperimetric function I_μ which might not satisfy the conditions of Theorem 5.1, but rather to the function

$$I(p) = N^*(a, b)/x_p(a, b)$$

which satisfy these conditions, serves as a lower estimate for I_μ and generates the same function $N_I^* = N^*$. Thus, Theorems 4.1 applied, via lemma 3.1, to the function I and Theorem 5.1 will give the estimate

$$\int_X |\nabla f| d\mu \ \geq \ N^*(a, b).$$

This inequality can become asymptotic equality for Lipschitz functions approximating indicator functions of sets of measure p. Note that one needs only consider in Theorem 5.1 (as well as in in the variational problem (5.1)–(5.2)) the case $a = 0$, because one can replace N by the function $x \to N(x + a)$ which is also convex. However, to fulfill the suggested strategy of proof, it will be essential to state some properties of N_I^* as a function of the two variables a and b.

Let f be a bounded, Borel measurable function on X such that $\int_X f d\mu = 0$, $f \neq 0$ a.e.μ and let a be a real number. Generalizing (5.3), we introduce the function

$$N_{f,a}(x) = \int_X N(xf + a)d\mu = \int_{-\infty}^{+\infty} N(xt + a)dF_f(t) \tag{5.7}$$

of the real variable x. Clearly, if the function f takes only two values q and $-p$ with respective μ–probabilities p and q, $(0 < p < 1, q = 1 - p)$, then $N_{f,a} = N_{p,a}$. Note also that $N_{f,a}(0) = N(a)$. Let us now point out the following facts:

Lemma 5.2 *The function $N_{f,a}$ is convex, non–increasing on $(-\infty, 0]$, non–decreasing on $[0, +\infty)$, with also $N_{f,a}(-\infty) = N_{f,a}(+\infty) = +\infty$. Therefore, for any $b > N(a)$, the equation $N_{f,a}(x) = b$ admits only one positive solution $x = x_f(a, b)$. Moreover, $x_f(a, b)$ is an increasing, concave function of the variable b on $(N(a), +\infty)$.*

Lemma 5.3 *The function x_f is concave, hence continuous, on the open convex set $\{(a, b) \in \mathbf{R^2} : b > N(a)\}$. Moreover, if a sequence $\{f_n\}$ of uniformly bounded functions with μ–mean zero converges to f a.e., then $x_{f_n}(a, b) \to x_f(a, b)$ for all $b > N(a)$.*

By Lemma 5.3, N_I^* is concave as the infimum over $p \in (0, 1)$ of the concave functions $I(p)x_p$. Therefore we also separately state:

Corollary 5.4 *For any non–negative function I defined on $(0, 1)$, the function N_I^* is concave, hence continuous, as a function of two variables, in the region $b > N(a)$.*

Corollary 5.4 will be used to prove Theorem 5.1, while the last statement of Lemma 5.3 will be used to prove Theorem 1.10. In addition, we also point out the following corollary which asserts that our minimizing problem is non–vacuous.

Corollary 5.5 *All the bounded functions f satisfying (5.2), are of the form $f = xg + a$, where g is an arbitrary bounded, non–constant, Borel measurable function with μ–mean zero, and where x is the unique value $x_g(a, b)$.*

We finally state a last lemma which is used in the proof of the variational theorem and in proving the condition 3) of Theorem 1.11.

Lemma 5.6 *Given $b > N(a)$, the function $I(p) = 1/x_p(a, b)$ is continuous, and $I(p)/p$ is decreasing on $(0, 1)$.*

Note that the function x_p is continuous by the last statement of Lemma 5.3 since it is a particular case of function x_f.

Proof of Lemma 5.2. Let M_C be the space of all signed measures concentrated on some compact interval $[-C, C]$. Then $M_C^1 \subset M_C$, the family of all probability distributions F concentrated $[-C, C]$ is a compact symplex (for the topology of weak convergence in M_C), and the delta–measures δ_t, $|t| \le C$, are the extremal points of this symplex. Therefore the elements of V_C, the intersection of the extremal points of M_C^1 with the hyperspace $\{F \in M_C : \int_{-C}^{C} t dF(t) = 0\}$, lie on the one–dimensional edges of M_C^1, i.e., take the form $p\delta_t + q\delta_s$, where $p \in [0,1], q = 1-p, pt + qs = 0, \; -C \le t, s \le C$. Replacing t and s by qt and $-pt$, and applying Choquet's theorem, one can represent an arbitrary $F \in V_C$ as a mixture

$$F = \iint_{R_C} (p\delta_{qt} + q\delta_{-pt}) \; d\pi(p,t), \tag{5.8}$$

where π is a probability measure concentrated on the set $R_C = \{(p,t) : 0 \le p \le 1, 0 \le t \le C/\max(p,q)\}$ of all allowed values of (p,t). Of course, V_C is not a symplex, i.e., the measure π in (5.8) is not unique.

Let π_f be a mixing measure for the distribution F_f of the function f and let $C = \operatorname{esssup}|f|$. Then, from the representation (5.8) and using (5.3) and (5.7), we obtain

$$N_{f,a}(x) = \iint_{R_C} N_{p,a}(xt) \; d\pi_f(p,t). \tag{5.9}$$

As itself a mixture of the functions $N_{p,a}$, the function $N_{f,a}$ inherits many of their properties. Clearly, it is convex, it is non–increasing on $(-\infty, 0]$, non–decreasing on $[0, +\infty)$, and $N_{f,a}(0) = N(a)$. Moreover,

$$N_{f,a}(+\infty) = N_{f,a}(-\infty) = +\infty.$$

Indeed, if this last claim were false and say that $N_{f,a}(+\infty) < +\infty$, then one would have $N_{f,a}(x) = N_{f,a}(0)$ for all $x \ge 0$. But, this last statement is possible if and only if the measure π_f is concentrated on the line $t = 0$. This means, according to (5.8), that $F_f = \delta_0$, i.e., that $f = 0$, a.e., and this contradicts the assumption made on f. Thus, for any $b > N(a)$, there exists only one positive solution $x = x_f(a,b)$ to the equation $N_{f,a}(x) = b$, and as a function of b, x_f is increasing and concave on the interval $(N(a), +\infty)$. This proves Lemma 5.2.

Proof of Lemma 5.3. Since x_f is an increasing function of b, we obtain that for all $x \ge 0, b > N(a)$,

$$\int_X N(xf + a) d\mu \;\le b \implies x \;\le x_f(a,b).$$

We use this property to establish the concavity of x_f. Let $b_1 > N(a_1), b_2 > N(a_2), 0 \le \alpha \le 1$. Put $x_1 = x_{f,a}(a_1, b_1)$, $x_2 = x_{f,a}(a_2, b_2)$, so that

$$\int_X N(x_1 f + a_1) d\mu = b_1, \text{and } \int_X N(x_2 f + a_2) d\mu = b_2.$$

By the convexity of N,

$$\alpha b_1 + (1-\alpha) b_2 \;\geq\; \int_X N((\alpha x_1 + (1-\alpha)x_2) f + (\alpha a_1 + (1-\alpha) a_2)) d\mu.$$

Hence, by the property mentioned above,

$$\alpha x_1 + (1-\alpha) x_2 \;\leq\; x_f(\alpha a_1 + (1-\alpha) a_2, \alpha b_1 + (1-\alpha) b_2),$$

which is exactly concavity.

It remains to establish the last statement in Lemma 5.3. Put $x_n = x_{f_n}(a,b)$, then

$$\int_X N(x_n f_n + a) d\mu = b. \tag{5.10}$$

There always exists a subsequence $\{x_{n_k}\}$ converging to some $x \in [0, +\infty]$. If x is finite, then by the Lebesgue dominated convergence theorem, one can take the limit in (5.10) as $n \to \infty$, and have $\int_X N(xf + a) = b$. Necessarily $x > 0$, since if $x = 0$ then $b = N(a)$ which contradicts the assumption $b > N(a)$. This implies that $x = x_f(a,b)$ and proves the statement when x is finite. To prove that indeed x is finite, it is sufficient to show that for any sequence $x_n \to +\infty$,

$$\int_X N(x_n f_n + a) d\mu \to +\infty, \tag{5.11}$$

as $n \to \infty$. For simplicity (recalling that it is possible to replace N by the function $x \to N(x + a)$), one needs only to consider the case $a = 0$. Put also $N_{f_n,0} = N_{f_n}$, $\pi_{f_n} = \pi_n$. Again using the representation (5.9), one can write the integral (5.11) as

$$N_{f_n}(x) = \iint\limits_{R_C} N_p(xt)\; d\pi_n(p,t), \tag{5.12}$$

with $x = x_n$. To prove (5.11), it suffices to estimate (from below) all the functions N_{f_n}, with n large enough, by a function which is unbounded, and non–decreasing on $[0, +\infty)$. To this end, for a given $\epsilon \in (0, 1/2)$, introduce the set

$$A_\epsilon \;=\; \{(p,t) \in R_C : \epsilon < p < 1 - \epsilon,\; t > 2C\epsilon\}.$$

Then, the integral (5.12) can be estimated by

$$N_{f_n}(x) \;\geq\; T_\epsilon(2C\epsilon x)\pi_n(A_\epsilon),$$

where $T_\epsilon(x) \;=\; \inf_{\epsilon < p < 1-\epsilon} N_p(x)$. Thus it suffices to show that for ϵ small enough,

$$T_\epsilon(x) \;\longrightarrow\; +\infty, \tag{5.13}$$

as $x \to +\infty$, and that $\liminf_{n \to \infty} \pi_n(A_\epsilon) > 0$.

To prove that $\liminf_{n\to\infty} \pi_n(A_\epsilon) > 0$, let us write down the representation (5.8) for the distribution function F_n of f_n, and for the interval $[-2C\epsilon, 2C\epsilon]$:

$$(5.14)\quad F_n([-2C\epsilon, 2C\epsilon]) \;=\; \iint_{R_C} p\delta_{qt}([-2C\epsilon, 2C\epsilon]) + q\delta_{-pt}([-2C\epsilon, 2C\epsilon])\, d\pi_n(p,t).$$

Then, consider the three inequalities which define the complement $R_C \setminus A_\epsilon$, noting that for all $(p,t) \in R_C$, we have $0 \le t \le C/\max(p,q) \le 2C$:

(i) If $p \le \epsilon$, then $pt \le 2C\epsilon$, hence the second term $q\delta_{-pt}([-2C\epsilon, 2C\epsilon]) = q \ge 1-\epsilon$.

(ii) Analogously, for $p \ge 1-\epsilon$, the first term $p\delta_{qt}([-2C\epsilon, 2C\epsilon]) = p \ge 1-\epsilon$. In both cases (i) and (ii), the integrand in (5.14) is greater or equal to $1-\epsilon$.

(iii) In a similar way, if $t \le 2C\epsilon$, then the integrand is equal to 1.

Consequently, (5.14) implies

$$F_n([-2C\epsilon, 2C\epsilon]) \;\ge\; (1-\epsilon)\pi_n(R_C \setminus A_\epsilon),$$

or, in other words,

$$\pi_n(A_\epsilon) \;\ge\; 1 - \frac{1}{1-\epsilon} F_n([-2C\epsilon, 2C\epsilon]).$$

The functions f_n converge to f a.e., hence the sequence F_n converges weakly to the distribution function F of f. In particular, for all $\epsilon > 0$, $\limsup_{n\to\infty} F_n([-2C\epsilon, 2C\epsilon]) \le F([-2C\epsilon, 2C\epsilon])$. This gives

$$\liminf_{n\to\infty} \pi_n(A_\epsilon) \;\ge\; 1 - \frac{1}{1-\epsilon} F([-2C\epsilon, 2C\epsilon]).$$

The right–hand side of this inequality is positive for all ϵ small enough because its limit as $\epsilon \to 0^+$, is equal to $1 - F(\{0\}) = \mu\{f \ne 0\} > 0$, since by assumption f is not a.e. 0.

To prove (5.13), introduce the Radon–Nikodym derivative (the density) N' of the function N. It is defined a.e., but can be chosen to be non–decreasing with possibly, $N(x) = N(0)$ on some interval (x_0, x_1), $-\infty \le x_0 \le 0 \le x_1 \le +\infty$. Let x_0 be maximally small and let x_1 be maximally large. Then, for all $x > x_1$, $N'(x) > 0$, and for all $x < x_0$, $N'(x) < 0$. Note that having simultaneously, $x_0 = -\infty$ and $x_1 = +\infty$ is impossible, since N is not constant. By (5.3), the derivative of N_p can be chosen to be equal to

$$(5.15)\qquad N_p'(x) = pq(N'(qx) - N'(-px)).$$

From (5.15), for all $x \ge 0$, and for all $p \in [\epsilon, 1-\epsilon]$, we obtain

$$N_p'(x) \ge \epsilon(1-\epsilon)(N'(\epsilon x) - N'(-\epsilon x)),$$

$$N_p'(x) \le \frac{(N'((1-\epsilon)x) - N'((1-\epsilon)x))}{4}.$$

These inequalities show that, on any compact interval within $(d_\epsilon, +\infty)$, where $d_\epsilon = \min(x_1/\epsilon, -x_0/\epsilon)$, the value of the Lipschitz norm of the function N_p is bounded. Hence, the function T_ϵ is Lipschitz on such intervals. Therefore, T_ϵ is absolutely continuous, non–decreasing on $(d_\epsilon, +\infty)$, and its Radon–Nikodym derivative clearly satisfies the same inequalities. In particular, for all $x > d_\epsilon$,

$$T'_\epsilon(x) \geq \epsilon(1-\epsilon)(N'(\epsilon x) - N'(-\epsilon x)).$$

The right–hand side of this last inequality is a non–negative, non–decreasing function of x and, in fact, it is positive for x large enough, because N is not linear. This proves (5.13).

Proof of Lemma 5.6. One can assume that $a = 0, N(a) = 0$, so $b > 0$. Put $y(p) = px_p(a, b)$. Then one needs to show that the function y is increasing on $(0, 1)$. By the convexity of N, the function $T(x) = N(x)/x$ is non–increasing on $(-\infty, 0)$ and non–decreasing on $(0, +\infty)$. From (5.3),

$$pN(qx_p) + qN(-px_p) = pqx_p(T(qx_p) - T(-px_p)) = qy(p)\left(T\left(\frac{qy(p)}{p}\right) - T(-y(p))\right) = b,$$

where $q = 1 - p$. Let us now assume that for some $p_1 < p_2$, we have $y(p_1) \geq y(p_2)$. Let $q_1 = 1 - p_1$, $q_2 = 1 - p_2$. Then, $q_1 > q_2$, $q_1/p_1 > q_2/p_2$ and therefore,

$$T\left(\frac{q_1 y(p_1)}{p_1}\right) \;\geq\; T\left(\frac{q_2 y(p_2)}{p_2}\right), \quad T(-y(p_1)) \;\leq\; T(-y(p_2)),$$

since $b > 0$, $T(qy(p)/p) - T(-y(p)) > 0$, for all $p \in (0, 1)$. Thus, we finally have

$$q_1 y(p_1)(T(q_1 y(p_1)/p_1) - T(-y(p_1))) \;>\; q_2 y(p_2)(T(q_2 y(p_2)/p_2) - T(-y(p_2))),$$

i.e., $b > b$!

Proof of Theorem 5.1 (reduction to the discrete case).

Let us assume that for all F with finite support, and for any continuous, non–negative function I, such that $I(p)/p$ is non–increasing on $(0, 1)$,

$$\int_{a(F)}^{b(F)} I(1 - F(t))dt \geq N_I^*(a, b), \tag{5.16}$$

where $F(t) = F((-\infty, t])$ is the ditribution function of the measure F, $a(F) = \inf\{t \in \mathbf{R} : F(t) > 0\}$, $b(F) = \sup\{t \in \mathbf{R} : F(t) < 1\}$, $\int_{\mathbf{R}} t dF(t) = a$, $\int_{\mathbf{R}} N(t) dF(t) = b$.

Step 1. We first consider the case where the function I is bounded. Assume that F is concentrated on $[-C, C]$ (more precisely that $[a(F), b(F)] \subset [-C, C]$). The probability measures with finite support form a dense (for the weak convergence topology) set in M_C^1. Hence, for any F in M_C^1 there exists a sequence F_n of measures in M_C^1, with finite support, converging to F weakly, i.e., such that $F_n(t) \to F(t)$ at all points t, where F is continuous. Therefore, $a(F_n) \to a(F)$, $b(F_n) \to b(F)$, and

$$\int_{a(F_n)}^{b(F_n)} I(1 - F_n(t))dt \longrightarrow \int_{a(F)}^{b(F)} I(1 - F(t))dt,$$

by the Lebesgue dominated convergence theorem, because I is bounded, and $I(1 - F_n(t)) \to I(1 - F(t))$ at all t except at possibly countably many values. In addition,

$$\int_{-C}^{C} x dF_n(x) \to a, \quad \int_{-C}^{C} N(x) dF_n(x) \to b, \ (n \to +\infty).$$

Thus, one can take the limit on both sides of (5.16) (with F there replaced by F_n). Using the continuity of N_I^* this gives (5.16) for F compactly supported and (recall) I bounded.

In order to extend (5.16) to an arbitrary F, a truncation argument is again used. Namely, let f be μ–integrable and satisfy (5.2). We need to prove (5.16) for $F = F_f$. If f is unbounded, define the functions

$$f_n(x) = \begin{cases} -n, & \text{if } f(x) < -n \\ f(x), & \text{if } -n \le f(x) \le n \\ n, & \text{if } f(x) > n. \end{cases}$$

The distribution function F_n of f_n, is bounded, and $a(F) \le a(F_n) \le b(F_n) \le b(F)$, hence from (5.16):

$$\begin{aligned} N_I^*(a_n, b_n) &\le \int_{a(F_n)}^{b(F_n)} I(1 - F_n(t))dt \\ &= \int_{a(F_n)}^{b(F_n)} I(1 - F(t))dt \\ &\le \int_{a(F)}^{b(F)} I(1 - F(t))dt, \end{aligned} \tag{5.17}$$

where $a_n = \int_X f_n d\mu$, and where

$$b_n = \int_X N(f_n) d\mu = N(-n)F(-n^-) + N(n)(1 - F(n)) + \int_{-n}^{n} N(t) dF(t).$$

By the integrability of f, a_n converges to $a = \int_X f d\mu$, and since $b = \int_X N(f) d\mu$ is finite, b_n converges to b. In addition, by the continuity of N_I^* (see Corollary 5.4),

$$N_I^*(a_n, b_n) \longrightarrow N_I^*(a, b), \tag{5.18}$$

as $n \to +\infty$. Now, (5.18) and (5.17) imply (5.16).

Step 2. It remains to remove the boundedness assumption on I and to prove (5.16) for an arbitrary continuous I. Define the function $I_n = \min(I, n)$. Clearly, $I \geq I_n$, moreover I_n is bounded, non–negative, continuous, and $I_n(p)/p$ is non–increasing on $(0,1)$. Therefore, by Step 1,

$$\int_{a(F)}^{b(F)} I(1 - F(t))dt \geq N^*_{I_n}(a, b),$$

and one needs only to show that $N^*_{I_n} \longrightarrow N^*_I$, as $n \to \infty$. First from (5.4), it is clear that if I and J are two non–negative functions on $(0,1)$,

$$N^*_{\min(I,J)} = \min(N^*_I, N^*_J).$$

Thus taking $J(p) = n$, we obtain

$$N^*_{I_n} = \min(N^*_I, cn),$$

where $c = \inf_{0<p<1} x_p(a, b)$. The above expression converges pointwise to N^*_I if and only if $c > 0$. But $c = 0$ is impossible. Indeed, if for some sequence $p_n \in (0,1)$ converging to $p \in [0,1]$, $x_n = x_{p_n}(a, b) \to 0$, then we would have

$$b = N_{p_n}(x_n) = p_n N(a + (1 - p_n)x_n) + (1 - p_n)N(a - p_n x_n) \ \to \ N(a),$$

i.e., we would have $b = N(a)$ which contradicts the assumption $b > N(a)$. Thus, the reduction of Theorem 5.1 to the discrete case has been achieved.

Proof of Theorem 1.10.

Denote by $Z(a, b)$, $b > N(a)$, the infimum of (5.1) under the conditions (5.2). Combining Theorems 4.1 and 5.1, we immediately have $Z(a, b) \geq N^*_I(a, b)$, whenever the function $I \leq I_\mu$ is non–negative, continuous, and $I(p)/p$ is non–increasing on $(0,1)$. Now, take $I(p) = N^*(a, b)/x_p(a, b)$. Since

$$N^*(a, b) = \inf_{0<p<1} \left(I_\mu(p) x_p(a, b)\right),$$

we have $I(p) \leq I_\mu(p)$, for all $p \in (0,1)$. By Lemma 5.6, I is continuous, and $I(p)/p$ is non–increasing on $(0,1)$, so I satisfies the conditions of Theorems 4.1 and 5.1. Furthemore, according to (5.4),

$$N^*_I(a, b) = \inf_{0<p<1} \left(I(p) x_p(a, b)\right) = N^*(a, b).$$

One thus concludes that $Z(a, b) \geq N^*(a, b)$, and only the reversed inequality needs to be proved. Let $0 < p < 1$, and let $A \subset X$ be a closed set of the measure $\mu(A) = p$. There exists a sequence $\epsilon_n \to 0^+$ such that

$$\frac{\mu(\overline{A^{\epsilon_n}}) - \mu(A)}{\epsilon_n} \longrightarrow \mu^+(A), \tag{5.19}$$

as $n \to \infty$. One can then take a sequence g_n of Lipschitz functions with values in $[0,1]$, of Lipschitz constant at most $1/\epsilon_n$, such that $g_n = 1$ on some open neighbourhood of A and $g_n = 0$ on $X \setminus A^{\epsilon_n}$. Thus, g_n converges everywhere to the indicator function $g = \chi_A$ of the set A, and according to the definition of the modulus of gradient,

$$\int_X |\nabla g_n| d\mu \le \frac{\mu(\overline{A^{\epsilon_n}}) - \mu(A)}{\epsilon_n}. \tag{5.20}$$

Since the sequence $f_n = g_n - a_n$, where $a_n = \int_X g_n d\mu$, converges to the function $f = g - p$, which takes the value $q = 1-p$ with μ–probability p and the value $-p$ with μ–probability q, by Lemma 5.3 we have

$$x_{f_n}(a,b) \longrightarrow x_f(a,b) = x_p(a,b), \tag{5.21}$$

as $n \to \infty$. Then, recall that the sequence $x_n = x_{f_n}(a,b)$ corresponds to the condition $\int_X N(x_n f_n + a) = b$, i.e., the functions $x_n f_n + a$ satisfy (5.2). Hence using (5.20), for all n,

$$\begin{aligned} Z(a,b) &\le \int_X |\nabla(x_n f_n + a)| d\mu \\ &= x_n \int_X |\nabla f_n| d\mu \\ &= x_n \int_X |\nabla g_n| d\mu \\ &\le x_n \frac{\mu(\overline{A^{\epsilon_n}}) - \mu(A)}{\epsilon_n}. \end{aligned}$$

By (5.19) and (5.21), this last expression converges, as $n \to \infty$, to $x_p(a,b)\mu^+(A)$. Taking the infimum over all possible A, one obtains $Z(a,b) \le x_p(a,b) I_\mu(p)$, for all $p \in (0,1)$. Finally, taking the infimum over all $p \in (0,1)$, on the right hand side of this last inequality yields $Z(a,b) \le N^*(a,b)$. This finishes the proof of the theorem.

6 The discrete version of Theorem 5.1

Let the probability distribution F on the real line $\mathbf{R}$ have finite support, say, $\{t_1, \dots, t_n\}$, $t_1 \geq \dots \geq t_n$, $n \geq 2$. Hence, $F = p_1\delta_{t_1} + \dots + p_n\delta_{t_n}$, where as usual δ_t denotes the unit mass at the point $t \in \mathbf{R}$, and where $p_i > 0, p_1 + \dots + p_n = 1$. Then, the conditions (5.2) take the form

$$p_1 t_1 + \dots + p_n t_n = a, \quad p_1 N(t_1) + \dots + p_n N(t_n) = b, \tag{6.1}$$

The integral (5.5)

$$\int_{a(F)}^{b(F)} I(1 - F(t))dt, \tag{6.2}$$

where $a(F) = \inf\{t : F(t) > 0\}$, $b(F) = \sup\{t : F(t) < 1\}$, and where $F(t) = F((-\infty, t])$ denotes the distribution function associated with the measure F, becomes

$$G(t_1, \dots, t_n) = \sum_{i=1}^{n-1} c_i(t_i - t_{i+1}), \tag{6.3}$$

where $c_i = I(p_1 + \dots + p_i)$. To complete the proof of Theorem 5.1, it remains to show that among all the discrete measures F, satisfying (6.1), the infimum of (6.2) is attained (possibly asymptotically) within the family of measures with only two atoms. So, we fix the values $p_i > 0, p_1 + \dots + p_n = 1$, and minimize the functional G on the $(n-2)$–dimensional set

$$C^+(a,b) \ = \ C(a,b) \cap \Delta_+,$$

where $C(a,b) \subset \mathbf{R}^n$ denotes the hypersurface defined by (6.1), and where $\Delta_+ = \{t = (t_1, \dots, t_n) \in \mathbf{R}^n : t_1 \geq \dots \geq t_n\}$. By Lemma 5.2, the set $C^+(a,b)$ is not empty, whenever $b > N(a)$. Recall that the function I, defining the coefficients c_i, is assumed to be such that $I(p)/p$ is non–increasing on $(0,1)$. Of course, the continuity property of I does not matter in the discrete case. In this section, we extend I to $(0,1]$ by putting $I(1) = 0$ ($I(1^-) > 0$ is possible). Now, in the discrete case, Theorem 5.1 takes the following form.

Lemma 6.1 *Let $b > N(a)$. On $C^+(a,b)$, the functional G in (6.3) attains its minimum at a point $(t_1, \dots, t_n)$ such that only two of its coordinates t_i are distinct.*

Using an induction over n, we can reduce the above statement to the three–dimensional case where a proposition a little bit more general is proved:

Lemma 6.2 *Let $p_1, p_2, p_3 > 0$, $p_1 + p_2 + p_3 \leq 1$, $c_i = I(p_1 + \dots + p_i)$, $i = 1, 2, 3$. For* *any $u, a, b \in \mathbf{R}$, and if it is not empty, let*

$$C_u^+(a,b) = \{(t_1,t_2,t_3) : \sum_1^3 p_i t_i = a,\ \sum_1^3 p_i N(t_i) = b,\ t_1 \geq t_2 \geq t_3 \geq u\}.$$

Then on $C_u^+(a,b)$, the functional

$$G(t_1,t_2,t_3) \ = \ c_1(t_1-t_2) + c_2(t_2-t_3) + c_3(t_3-u)$$

attains its minimum at a point (t_1,t_2,t_3) such that $t_1 = u$, or $t_2 = t_1$, or $t_3 = t_2$.

First, by induction, we show how Lemma 6.1 follows from Lemma 6.2. Let $n = 3$, since $p_1 + p_2 + p_3 = 1$ and $I(1) = 0$, we have $c_3 = 0$, and therefore the functional G from Lemma 6.2 is of the form $G(t_1,t_2,t_3) = c_1(t_1-t_2) + c_2(t_2-t_3)$, i.e., it coincides with (6.3). Letting $u \to -\infty$ in Lemma 6.2, we get the statement of Lemma 6.1 for $n = 3$. Now, let $n > 3$, and assume that the statement of Lemma 6.1 is valid for all dimensions lower or equal to $n-1$, and for all admissible functions I. Let $(t_1,\cdots,t_n) \in C^+(a,b)$. We fix the values $t_4,\ldots,t_n$, but t_1,t_2 and t_3 vary. Since the sums

$$p_4t_4 + \cdots + p_nt_n = a', \quad p_4N(t_4) + \cdots + p_nN(t_n) = b',$$

are also fixed, the first three variables can vary arbitrarily under the conditions

$$p_1t_1 + p_2t_2 + p_3t_3 = a - a', \quad p_1N(t_1) + p_2N(t_2) + p_3N(t_3) = b - b',$$

and $t_1 \geq t_2 \geq t_3 \geq u = t_4$. This means that the triple (t_1,t_2,t_3) belongs to the set $C_u^+(a'',b'')$ from Lemma 6.2 with $a'' = a - a'$, $b'' = b - b'$.

Therefore, one can apply Lemma 6.2 according to which the functional G, as a function of $(t_1,t_2,t_3) \in C_u^+(a'',b'')$, attains its minimum at a point (t_1,t_2,t_3) such that $t_1 = u$, or $t_2 = t_1$, or $t_3 = t_2$. In all these cases, we decrease the number of different coordinates of the vector $(t_1,\cdots,t_n)$ and again we need to minimize G under the additional restriction $t_i = t_{i+1}$ for some $i = 1,\cdots,n-1$. But, when $t_i = t_{i+1}$, we obtain the original $(n-1)$–dimensional problem since the conditions (6.1) as well as the functional (6.3) remain of the same type, and since a and b do not change. Thus, one can use the induction assumption and Lemma 6.1 is proved.

The proof of Lemma 6.2 will occupy the rest of this section. For the reader's convenience we also present two pictures illustrating our minimizing problem.

Proof of Lemma 6.2. Without loss of generality, we assume that $u = 0$ since otherwise N can be replaced by a shifted version $x \longrightarrow N(x-u)$. Now change notations and set $x = t_1, y = t_2, z = t_3, p = p_1, q = p_2, r = p_3$. So, we fix the values $p,q,r > 0$, $p+q+r \leq 1$, and minimize the functional

(6.4) $$G = c_1(x-y) + c_2(y-z) + c_3 z$$

on the curve $C_0^+(a,b)$ defined by the conditions (6.1) and the restrictions $x \geq y \geq z \geq 0$. From (6.1),

(6.5) $$z = \frac{a - px - qy}{r},$$

therefore we can treat the problem of minimizing G as a problem in the plane $\mathbf{R}^2$. So redefine $C_0^+(a,b)$ as a curve in the plane:

$$C_0^+(a,b) \;=\; \{(x,y) \in \mathbf{R}^2 : x \geq y \geq z \geq 0,\; pN(x) + qN(y) + rN(z) = b\},$$

where z is always understood to be as in (6.5). Necessarily, $a \geq 0$, since otherwise $C_0^+(a,b)$ would be empty. If $a = 0$, $C_0^+(a,b) = \{(0,0)\}$, and there is nothing to prove. Thus, one can assume in the following that $a > 0$.

The inequalities $x \geq y \geq z$ define in the plane, the sector $Sec(a)$ with vertex P which is the point of intersection of the lines $x = y$ and $y = z$. Note also that

$$y = z \iff y = \frac{a - px - qy}{r} \iff y = \frac{a - px}{q + r}.$$

Thus, the line $y = z$ has the equation $y = (a - px)/(q + r)$, and the vertex P has coordinates $(a/(p+q+r), a/(p+q+r))$.

Furthemore, the line $z = 0$ (i.e., $px + qy = a$) intersects the line $x = y$ at a point Q of coordinates $(a/(p+q), a/(p+q))$ and intersects the line $y = z$ at a point R of coordinates $(a/p, 0)$.

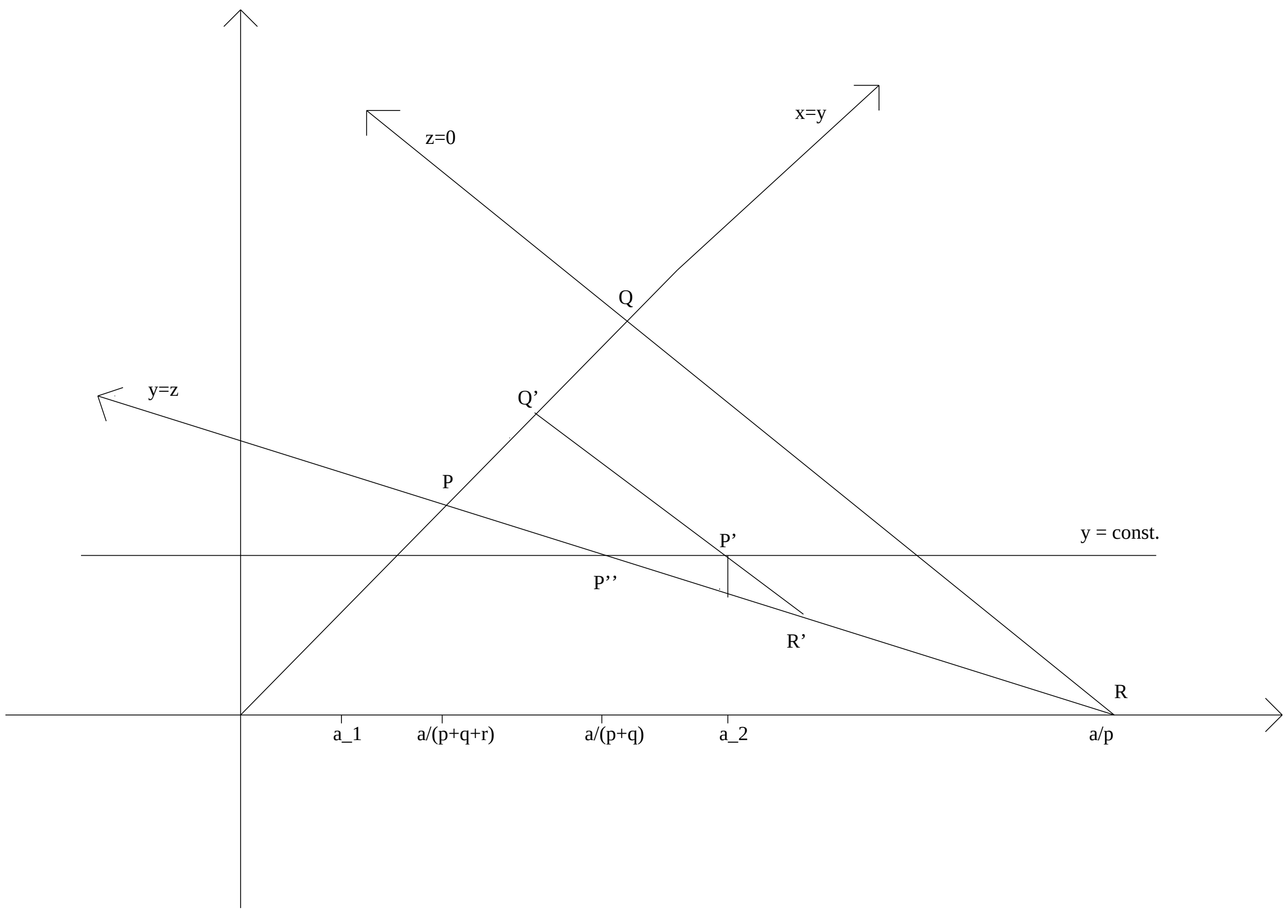
Thus, the restrictions $x \geq y \geq z \geq 0$ determine the triangle $Tri(a)$ with vertices P, Q, and R, and $C_0^+(a,b)$ is the intersection of the curve

$$C(a,b) \;=\; \{(x,y) \in \mathbf{R}^2 : pN(x) + qN(y) + rN(z) = b\}$$

with $Tri(a)$. We thus need to show that a point of minimum of G on $C_0^+(a,b)$ lies on one of the sides of the triangle $Tri(a)$.

Put $a_0 = a/(p+q+r)$. By the convexity of N, there exists a point $b \geq (p+q+r)N(a_0)$, since otherwise $C_0^+(a,b)$ would be empty. Let us first treat the case $b = (p+q+r)N(a_0)$ and present a picture illustrating our minimizing problem in this case.

The case b = (p+q+r)N(a/(p+q+r))



The equality $b = (p+q+r)N(a_0)$, is only possible if there exists a non–degenerate interval containing a_0 and where the function N is linear, i.e., has the form $N(x) = cx+d$. Let $[a_1, a_2] \subset (-\infty, +\infty)$ be the maximal (possibly infinite) interval containing a_0 where the function N is linear. Then, the two conditions $pN(x) + qN(y) + rN(z) = b$, and $px + qy + rz = a$, are equivalent to $x, y, z \in [a_1, a_2]$, so

$$C_0^+(a, b) \;=\; \{(x, y) \in \mathbf{R}^2 : x \geq y \geq z \geq 0,\; x, y, z \in [a_1, a_2]\}.$$

Since $a_0 \in [a_1, a_2]$, and since for all points $(x, y) \in Tri(a)$, one has $x \geq a_0$, $z \leq a_0$, the conditions $x \geq y \geq z \geq 0$ and $x, y, z \in [a_1, a_2]$ reduce to $x \leq a_2, z \geq a_1^+ = \max(a_1, 0)$. Therefore,

$$C_0^+(a, b) \;=\; \{(x, y) \in \mathbf{R}^2 : a_2 \geq x \geq y \geq z \geq a_1^+\}.$$

This set is either a triangle Tri' with vertices P, Q', R' such that Q' and R' respectively lie on the segments $[P, Q]$ and $[P, R]$, the lines (Q, R) and (Q', R') being parallel (this case corresponds to $a/p \leq a_2$), or is the intersection of Tri' with the half–plane $x \leq a_2$. In the first case, all three extremal points of $C_0^+(a,b)$ lie on the sides of $Tri(a)$, and the linear functional G attains its minimum at one of these points. Therefore, a point of minimum of G lies on a side of $Tri(a)$. In the second case, where $a/p > a_2$ (i.e., the point R is on the right of the line $x = a_2$), $C_0^+(a,b)$ has a fourth extremal point P', inside Tri', which lies on the line (Q', R') and has x–coordinate a_2. So one needs show that,

$$G(P') \ \geq \ \min\{G(P), G(Q'), G(R')\}.$$

To prove this, it suffices to see that given y, $G(x,y)$ is a non–decreasing function of x. Then, taking for P'' the intersection of the line $y = const$ containing the point P' with the segment $[P,Q]$ or $[P,R]$, we will have that $G(P'') \leq G(P')$. This means that a point of minimum of G lies on these segments. Let us now see that, indeed, G is monotone in x. From (6.4) and (6.5), and using the definition of the coefficients c_i, we have

$$\begin{aligned} G(x,y) &= c_1(x-y) + c_2(y-z) + c_3\frac{a-px-qy}{r} \\ &= \left(c_1 + (c_2-c_3)\frac{p}{r}\right)x + \left((c_2-c_1) + (c_2-c_3)\frac{q}{r}\right)y + (c_3-c_2)\frac{a}{r}. \end{aligned}$$

So G is a non–decreasing function of x if and only

$$c_1 + (c_2-c_3)\frac{p}{r} \ \geq \ 0.$$

Recalling the definition of the coefficients c_i, this last condition can be rewriten as

$$rI(p) + pI(p+q) \ \geq \ pI(p+q+r).$$

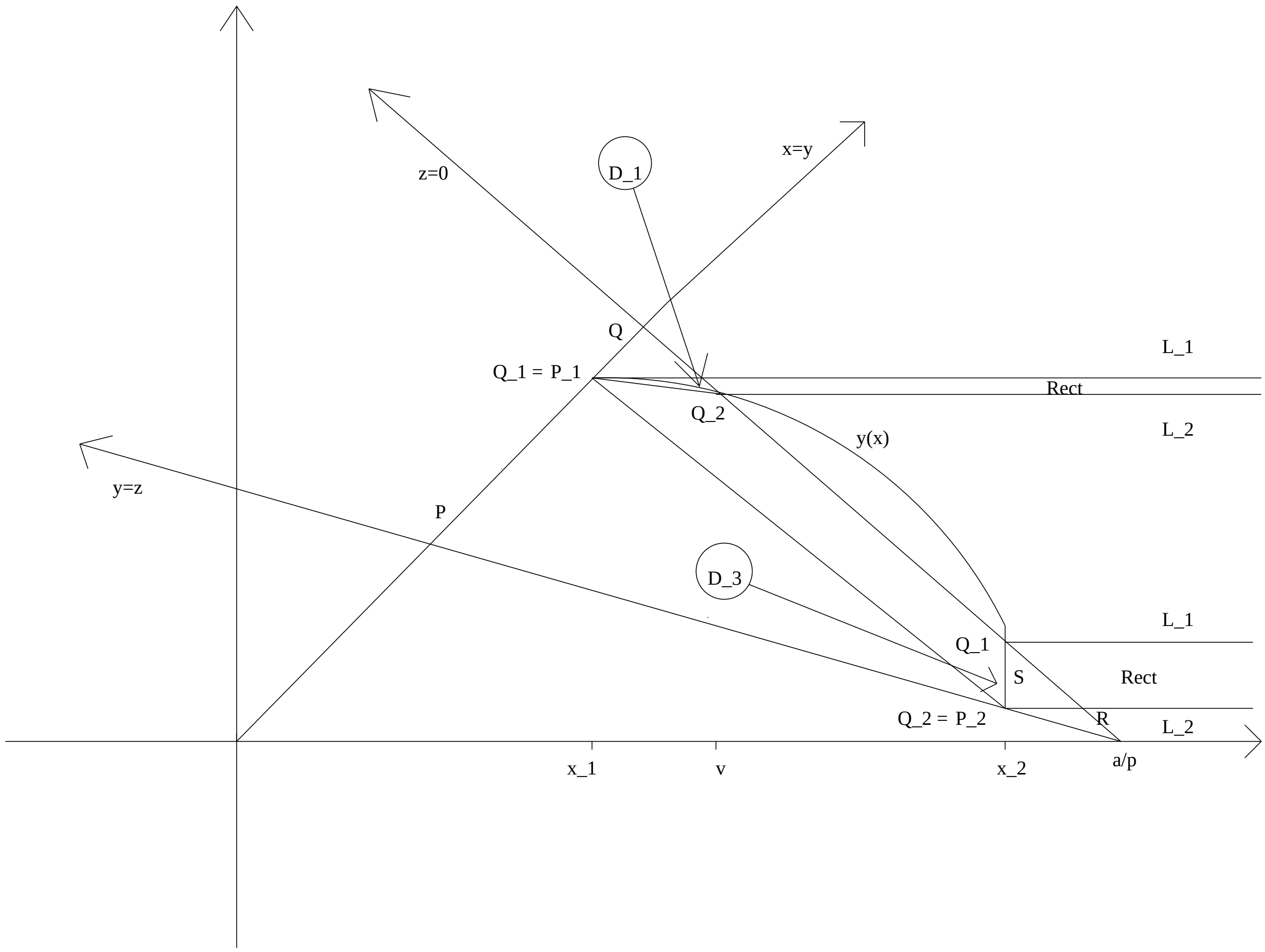
When $p+q+r = 1$, this last inequality is obviously true since $I(1) = 0$. When $p+q+r < 1$, introduce the function $J(p) = I(p)/p$ which by assumption is non–increasing on $(0,1)$. In terms of J, the monotonicity of G is thus equivalent to

$$prJ(p) + p(p+q)J(p+q) \ \geq \ p(p+q+r)J(p+q+r).$$

This is clearly true since $J(p) \geq J(p+q+r)$, $J(p+q) \geq J(p+q+r)$. The case $b = (p+q+r)N(a/(p+q+r))$ has thus been resolved.

We now consider the case $b > (p+q+r)N(a/(p+q+r))$. To briefly describe the main ideas used in minimizing G on $C_0^+(a,b)$, we point out the following steps which are also illustrated with the following picture.

The case b > (p+q+r)N(a/(p+q+r))



Step 1). On the half–plane $x \geq a$ (which contains $Sec(a)$), the curve $C(a,b)$ respectively intersects the line $x = y$ and the line $y = z$ at the (unique) point $P_1(x_1, y_1)$ (hence $y_1 = x_1$) and $P_2(x_2, y_2)$ (hence $y_2 = (a - px_2)/(q + r)$).

Step 2). $a < x_1 < x_2$.

Step 3). There is a non–increasing, concave, continuous function $y = y(x)$, defined on $[x_1, x_2]$, such that the graph of y is situated inside the sector $Sec(a)$, and the curve $C^+(a,b) = C(a,b) \cap Sec(a)$ is representable as

$$C^+(a,b) = \{(x, y(x)) : x_1 \leq x \leq x_2\} \cup \{(x_2, y) : y_2 \leq y \leq y(x_2)\}.$$

Thus, $C^+(a,b)$ is the graph of $y = y(x)$ plus a vertical segment S connecting the points (x_2, y_2) and the right end point $(x_2, y(x_2))$ of that graph. The function y automatically satisfies the equation $pN(x) + qN(y) + rN(z) = b$.

Let now $C_0^+(a,b)$ be the part of $C^+(a,b)$ which is on the left of the line $z=0$, i.e., let

$$C_0^+(a,b) \;=\; \{(x,y(x)) : x_1 \le x \le x_2,\ px+qy(x) \le a\} \cup \{(x_2,y) \in S : px_2+qy \le a\}\,.$$

Step 4). $C_0^+(a,b)$ is made of one or two pieces of $C^+(a,b)$. The case of one piece is only possible if $C_0^+(a,b)=C^+(a,b)$. The case of two pieces is possible only if one of these two pieces is a left piece of $C^+(a,b)$ and the other one is a right piece of $C^+(a,b)$. Thus, the left piece should have the form

$$D_1 \;=\; \{(x,y(x)) : x_1 \le x \le v,\ px+qy(x) \le a\}, \tag{6.6}$$

for some $x_1 \le v < x_2$, and the right piece has either the form

$$D_2 \;=\; \{(x,y(x)) : w \le x \le x_2,\ px+qy(x) \le a\} \cup S \tag{6.7}$$

for some $x_1 < w < x_2$, or the form

$$D_3 \;=\; \{(x_2,y) \in S : px_2+qy \le a\}. \tag{6.8}$$

In all the above cases, the end points of these pieces are on the sides of the triangle $Tri(a)$.

Let us see how to finish the proof of Lemma 6.1, provided the steps 1)–4) have been done. It then remains to show that for any curve D of either type (6.6)–(6.8), with end points Q_1 and Q_2, the functional G attains its minimum on D at either Q_1 or Q_2. Since the function y is concave and non–increasing, the curve D is situated on the right of the segment $[Q_1,Q_2]$. Moreover, D is a subset of the infinite rectangle *Rect* which is deliminated by the line (Q_1,Q_2) and the two lines ℓ_1 and ℓ_2 which are parallel to the x–axis and respectively contain the points Q_1 and Q_2. Because G is a linear functional, its infimum on *Rect* is attained at an extremal point of *Rect*. Two of these points, Q_1 and Q_2, belong to D. The other ones are infinite points on the lines ℓ_1 and ℓ_2. But as noted before, G is a non–decreasing function of x (again, we use the fact that $I(p)/p$ is non–increasing). Hence, Q_1 is the point of minimum of G on ℓ_1, and Q_2 is the point of minimum of G on ℓ_2. So G attains its minimum on *Rect* either at Q_1 or at Q_2. Since $D \subset Rect$, and $Q_1, Q_2 \in D$, these points are also points of minimum of G on the curve D. This completes this part of the proof, and only the statements claimed in 1)–4) need to be established.

Before proving these statements, we would like to give the reader a simple, intuitively clear explanation for some of them. Introduce the function

$$\phi(x,y) \;=\; pN(x) + qN(y) + rN(z), \tag{6.9}$$

where as usual, $z = (a - px - qy)/r$, and assume that N has a continuous, increasing derivative N'. Clearly, ϕ is convex and the curve $C(a,b) = \{(x,y) : \phi(x,y) = b\}$ surrounds the convex set $V_b = \{(x,y) : \phi(x,y) \le b\}$. Furthermore, inside the open sector $Sec_0(a) = \{(x,y) : x > y > z\}$,

$$\frac{\partial \phi}{\partial x} = p(N'(x) - N'(z)) > 0, \quad \frac{\partial \phi}{\partial y} = q(N'(y) - N'(z)) > 0.$$

Therefore, ϕ increases on $Sec_0(a)$ as a function of two variables. Hence, the convexity of $C^+(a,b) = C(a,b) \cap Sec(a)$ is directed to the right. Now, let us fix a point $(x,y) \in C(a,b) \cap Sec_0(a)$ and find the tangent line $y = bx + c$ at this point. One can differentiate the equality $\phi = b$ and get

$$pN'(x) + bqN'(y) - (p + bq)N'(z) = 0,$$

that is

$$b \;=\; -\,\frac{q(N'(y) - N'(z))}{p(N'(x) - N'(z))}.$$

Since $x > y > z$ and since N' is increasing, we obtain that $b < 0$, i.e., y is a decreasing function of x. This function is also concave since $C^+(a,b)$ is a part of the boundary of V_b.

We prove next the above claims in a more careful manner. Since the Steps 1)–4) concern the shape of the set $C_0^+(a,b)$ where G is minimized, one may assume in the following that $p + q + r = 1$. Thus, the assumption on b becomes $b > N(a)$.

Step 1). Let ϕ_1 and ϕ_2 be the restrictions of the function ϕ, defined by (6.9), to the lines $x = y$ and $y = z$ $(y = (a - px)/(q + r))$. In other words, let

$$\phi_1(x) \;=\; \phi(x,x) \;=\; (p+q)N(x) + rN\left(\frac{a - (p+q)x}{r}\right), \tag{6.10}$$

$$\phi_2(x) \;=\; \phi\left(x, \frac{a - px}{q + r}\right) \;=\; pN(x) + (q + r)N\left(\frac{a - px}{q + r}\right). \tag{6.11}$$

As restrictions of a convex function, these functions are convex too. Let N' be a non–decreasing Radon–Nikodym derivative of N. Then, the function

$$\phi_1'(x) \;=\; (p+q)\left(N'(x) - N'\left(\frac{a - (p+q)x}{r}\right)\right)$$

can also serve as a non–decreasing Radon–Nikodym derivative for ϕ_1. Since $x \le (a-(p+q)x)/r \iff x \le a$, we have $\phi_1' \le 0$ on $(-\infty, a)$ and $\phi_1' \ge 0$ on $(a, +\infty)$, i.e., ϕ_1 is non–increasing on $(-\infty, a)$ and non–decreasing on $(a, +\infty)$. It is possible to have $\phi_1 = \phi_1(a)$ (a constant) on some (maximal) interval $[\xi_1, \eta_1] \ni a$, but in view of its convexity, ϕ_1 should increase on $[\eta_1, +\infty)$. In addition, since N is not affine, $\phi_1(+\infty) = \phi_1(-\infty) = +\infty$. Finally, let us note that $\phi_1(a) = N(a)$. Hence, for any $b > N(a)$, there is unique solution $x = x_1$ to the equation $\phi_1(x) = b$ on $(\eta_1, +\infty)$. In particular, $x_1 > a$.

The same type of reasoning can be applied to the function ϕ_2. Again,

$$\phi_2'(x) \;=\; p\left(N'(x) - N'\left(\frac{a-px}{q+r}\right)\right)$$

can serve as a non–decreasing Radon–Nikodym derivative of ϕ_2, and $x \le (a-px)/(q+r) \iff x \le a$. Analogously, ϕ_2 is non–increasing on $(-\infty, a)$ and non–decreasing on $(a, +\infty)$; $\phi_2(+\infty) = \phi_2(-\infty) = +\infty$, $\phi_2(a) = N(a)$. If $[\xi_2, \eta_2] \ni a$ is the maximal interval where $\phi_2 = \phi_2(a)$, then ϕ_2 increases on $(\eta_2, +\infty)$. Therefore, there is unique solution $x = x_2$ to the equation $\phi_2(x) = b$ on $(\eta_2, +\infty)$. In particular, $x_2 > a$.

Step 2). Let

$$T(x) \;=\; \frac{N(x) - N(a)}{x-a} \;=\; \int_0^1 N'(a + t(x-a))dt,$$

which, trivially, is a non–decreasing function on the whole real line. The above integral does not depend on the choice of the Radon–Nikodym derivative N', which is thus always assumed to be non–decreasing, while for $x = a$, one can also set $T(x) = N'(a)$. From (6.10), and since $p + q + r = 1$, we have

$$\begin{aligned}
\phi_1(x) \;&=\; N(a) + (p+q)(x-a)T(x) + r\left(\frac{a-(p+q)x}{r} - a\right) T\left(\frac{a-(p+q)x}{r}\right) \\
&=\; N(a) + (p+q)(x-a)\left(T(x) - T\left(\frac{a-(p+q)x}{r}\right)\right).
\end{aligned}$$

Analogously, from (6.11)

$$\begin{aligned}
\phi_2(x) \;&=\; N(a) + p(x-a)T(x) + (q+r)\left(\frac{a-px}{q+r} - a\right) T\left(\frac{a-px}{q+r}\right) \\
&=\; N(a) + p(x-a)\left(T(x) - T\left(\frac{a-px}{q+r}\right)\right).
\end{aligned}$$

Note that the functions,

$$T_1(x) = T(x) - T\left(\frac{a-(p+q)x}{r}\right), \quad T_2(x) = T(x) - T\left(\frac{a-px}{q+r}\right),$$

are non–negative on $(a, +\infty)$, since

$$(p+q)T_1(x) = \frac{\phi_1(x) - \phi_1(a)}{x-a}, \quad pT_2(x) = \frac{\phi_2(x) - \phi_2(a)}{x-a},$$

and since, as noted on step 1), ϕ_1 and ϕ_2 are non–decreasing, convex functions on $(a, +\infty)$. Moreover, ϕ_1 and ϕ_2 respectively increase on the intervals $(\eta_1, +\infty)$ and $(\eta_2, +\infty)$. Therefore, T_1 and T_2 are positive on these intervals and, in particular, $T_2(x_2) > 0$.

To finish the proof of Step 2), we proceed by contradiction and assume that $x_1 \geq x_2$. Since $x_1, x_2 > a$, then $T(x_1) \geq T(x_2)$. Moreover,

$$\frac{a-(p+q)x_1}{r} \;\leq\; \frac{a-px_2}{q+r} \;<\; a.$$

Indeed, $(a-(p+q)x_1)/r \leq (a-(p+q)x_2)/r \;<\; a$, and $(a-(p+q)x_2)/r \leq (a-px_2)/(q+r) \Longleftrightarrow (q+r)a-(p+q)(q+r)x_2 \leq ra - prx_2 \Longleftrightarrow qa \leq (pq+q^2+qr)x_2 \Longleftrightarrow a \leq (p+1+r)x_2 = (2-q)x_2$ which is true since $x_2 > a$, $q < 1$ (recall that $p, q, r > 0$, $p+q+r = 1$). Since T is non–increasing on $(-\infty, a)$, we thus get

$$T\left(\frac{a-(p+q)x_1}{r}\right) \;\leq\; T\left(\frac{a-px_2}{q+r}\right).$$

But, $p+q > p$, $x_1 - a \geq x_2 - a > 0$, and $T_1(x_1) \geq T_2(x_2) > 0$. Thus, we finally get

$$\begin{aligned} b \;=\; \phi_1(x_1) \;&=\; N(a) + (p+q)(x_1-a)T_1(x_1) \\ &>\; N(a) + p(x_2-a)T_2(x_2) \;=\; \phi_2(x_2) \;=\; b! \end{aligned}$$

Step 3). Existence and uniqueness of the solution to $\phi = b$.

First, we fix $x \geq a$ and show that, above the line $y = z$ (i.e., for $y \geq (a-px)/(q+r)$), the equation $g_x(y) = \phi(x, y) = b$ has a unique solution, $y = y(x)$, when $x < x_2$, and no solution when $x > x_2$. In addition, we need to show that when $x = x_2$, the solution to $g_x(y) = b$ above the line $y = z$ forms a segment $[y_2, y(x_2^-)]$. First, recall that $g_{x_2}(y_2) = b$.

Again using a non–decreasing Radon–Nikodym derivative N', one can construct a non–decreasing Radon–Nikodym derivative for the convex function g_x, differentiating (6.9) with respect to y:

$$g_x'(y) = q\left(N'(y) - N'\left(\frac{a-px-qy}{r}\right)\right).$$

Clearly, since N' is non–decreasing, g_x is non–increasing in the interval $y \le (a-px-qy)/r$, i.e., for $y \le y_0(x) = (a-px)/(q+r)$, and is non–decreasing for $y \ge y_0(x)$. Since N is not linear, we have (and similarly for the functions N_p of which g_x is a particular case) that $g_x(-\infty) = g_x(+\infty) = +\infty$. Therefore, $g_x = g_x(y_0)$ on some maximal interval $[\xi(x), \eta(x)] \ni 0$ (with possibly, $\xi(x) = \eta(x) = y_0(x)$), and g_x increases on $[\eta(x), +\infty)$. Thus, if $b > g_x(y_0(x))$ then, on the interval $[y_0(x), +\infty)$ there is only one solution $y = y(x) > \eta(x)$ to $g_x(y) = b$. If $b = g_x(y_0(x))$, then on the interval $[y_0(x), +\infty)$ the solution to that equation represents the segment $[y_0(x), \eta(x)]$. When $b < g_x(y_0(x))$, there is no solution on the interval $[y_0(x), +\infty)$.

Now observe that

$$g_x(y_0(x)) = pN(x) + (q+r)N\left(\frac{a-px}{q+r}\right) = \phi_2(x),$$

and recall, as shown in the previous steps, that the inequality $\phi_2(x) \le b$ is equivalent to $x \le x_2$, provided $x \ge a$. Therefore,

if $x > x_2$, then $\phi_2(x) > b$, and the equality $g_x(y) = b$ is impossible.

If $x = x_2$, then $\phi_2(x) = b$, and the equality $g_x(y) = b$, provided $y \ge y_2$, is equivalent to $y_2 \le y \le \eta(x_2)$.

If $a \le x < x_2$, then $\phi_2(x) < b$, and the equality $g_x(y) = b$ is attained only at $y = y(x) \ge (a-px)/(q+r)$. Moreover, $y(x) > \eta(x)$. In this case, we also note that since the function g_x increases on $[\eta(x), +\infty)$, and since $g_x(\eta(x)) = g_x(y_0(x)) = \phi_2(x) < b$, for any $x \in [a, x_2)$, we have:

$$g_x(y) \le b \implies y \le y(x), \tag{6.12}$$

whenever $y \in \mathbf{R}$.

Concavity of y. Let $a \le u, v < x_2$ and let $t \in [0,1]$. By the convexity of ϕ,

$$b = tb + (1-t)b = t\phi(u, y(u)) + (1-t)\phi(v, y(v)) \ge \phi(tu + (1-t)v, ty(u) + (1-t)y(v)).$$

Since $tu + (1-t)v \in [a, x_2)$, we get by (6.12) that

$$ty(u) + (1-t)y(v) \;\le\; y(tu + (1-t)v).$$

Thus, y is concave on $[a, x_2)$. In particular, y is continuous on (a, x_2), hence continuous on $[x_1, x_2)$.

Monotonicity of y. We prove here that y is non–increasing on $[a, x_2)$. Given $y \in \mathbf{R}$, the function $x \to \phi(x, y) = pN(x) + qN(y) + rN(z)$ is convex, and its non–decreasing Radon–Nikodym derivative can be chosen to be

$$\frac{\partial \phi}{\partial x} = p(N'(x) - N'(z)).$$

This derivative is non–negative for $x \geq z = (a - px - qy)/r$, i.e., for x such that the points (x, y) are above the line $x = z$, i.e., $y = (a-(p+r)x)/q$. Clearly, this line contains the points $(0, a/q)$ and $P(a, a)$, therefore the half–plane $\{(x, y) : x \geq z\}$ contains the area $H(a) = \{(x, y) : x \geq a, y \geq (a - px)/(q + r)\}$, where the graph $\{(x, y(x)) : a \leq x < x_2\}$ is situated. Thus, when y is fixed and $y \geq (a - px)/(q + r)$, ϕ is non–decreasing with respect to $x \geq a$.

Now, let $a \leq u < v < x_2$. To prove that y is non–increasing on $[a, x_2)$, proceed by contradiction and assume that $y(u) < y(v)$. Since g_u increases on $(\eta(u), +\infty)$ and since $y(u) > \eta(u)$, we have

$$b = g_u(y(u)) < g_u(y(v)) = \phi(u, y(v)) \leq \phi(v, y(v)) = b!$$

We thus proved that inside the area $H(a)$, the equation $\phi(x, y) = b$ has only one solution $y = y(x)$ when $a \leq x < x_2$, has no solution when $x > x_2$, and has the interval $\{x_2\} \times [y_2, \eta(x_2)]$ for solutions when $x = x_2$. Now, the function y is concave and non–increasing on $[a, x_2)$. In particular, it is continuous on (a, x_2), and moreover for all $x \in [a, x_1)$,

$$y(x) \geq y(x_1) = x_1 > x.$$

Hence, when restricted to $[a, x_1)$, the graph of y is outside of the sector $Sec(a)$. On the contrary, for $x \in [x_1, x_2)$,

$$y(x) \leq y(x_1) = x_1 \leq x.$$

Thus, when restricted to $[x_1, x_2)$, the graph of y is inside the sector $Sec(a)$. In addition, it easy to see that only when $\eta(x_2) = y(x_2^-)$, is the graph $\{(x, y(x)) : x_1 \leq x < x_2\}$ plus the segment $\{x_2\} \times [0, \eta(x_2)]$ a part the boundary of the convex set $V_b = \{(x, y) : \phi(x, y) \leq b\}$. Therefore,

$$\begin{aligned} C^+(a, b) &= C(a, b) \cap Ang(a) \\ &= \{(x, y(x)) : x_1 \leq x < x_2\} \cup \{(x_2, y) : y_2 \leq y \leq y(x_2^-)\}. \end{aligned}$$

To complete the proof of this step, it just remains to set $y(x_2) = y(x_2^-)$ so that the function y should be defined on the closed interval $[x_1, x_2]$.

Step 4). Note that, for any continuous, concave function g defined on a segment $[x_1, x_2]$, the set of solutions $\mathcal{S} \subset [x_1, x_2]$ to the inequality $g(x) \leq 0$ has only of one of the following five possible descriptions.

1. $\mathcal{S} = \emptyset$.
2. $\mathcal{S} = [x_1, x_2]$.
3. $\mathcal{S} = [x_1, x_3]$, for some $x_3 \in [x_1, x_2)$.
4. $\mathcal{S} = [x_4, x_2]$, for some $x_4 \in (x_1, x_2]$.
5. $\mathcal{S} = [x_1, x_3] \cup [x_4, x_2]$, for some $x_1 \leq x_3 < x_4 \leq x_2$.

Applying this observation to the function $g(x) = px + qy(x) - a$, one concludes that the above cases correspond to the following possible curves $C_0^+(a,b)$:

1. $C_0^+(a,b)$ is empty. This case is excluded by the assumptions of Lemma 6.2.

2. $C_0^+(a,b) = C^+(a,b)$, i.e., the curve $C^+(a,b)$ is a subset of the triangle $Tri(a)$ and connects the segments $[P,Q]$ and $[P,R]$. Then, necessarily, $P_1(x_1,y_1) \in [P,Q]$, $P_2(x_2,y_2) \in [P,R]$.

3. $C_0^+(a,b)$ is a "left" part (truncation) of $C^+(a,b)$ and connects the segments $[P,Q]$ and $[Q,R]$, with possibly, $C_0^+(a,b) = \{Q\}$.

4. $C_0^+(a,b)$ is a "right" part (truncation) of $C^+(a,b)$ and connects the segments $[Q,R]$ and $[P,R]$, with possibly, $C_0^+(a,b) = \{R\}$.

5. $C_0^+(a,b)$ consists of two disjoint parts of $C^+(a,b)$ which respectively connect $[P,Q]$ with $[Q,R]$ and $[Q,R]$ with $[P,R]$. A middle part of $C^+(a,b)$ is on the right of the segment $[Q,R]$.

In these five cases (except for the first one), $C_0^+(a,b)$ consists of one or two pieces of $C^+(a,b)$ of one of three types (6.6)–(6.8), and the ends of these pieces lie on the sides of the triangle $Tri(a)$. This completes the proof of Lemma 6.2.

7 Proof of Proposition 1.3 and 1.8

We start by stating some elementary properties of the function

$$I_\alpha(p) = (p^\alpha q + pq^\alpha)^{1/\alpha}, \quad 0 \le p \le 1, \; q = 1 - p. \tag{7.1}$$

When $p \in (0,1)$, $p \neq 1/2$, is fixed, $I_\alpha(p)$ is an increasing function of α. This can easily be seen from the identity

$$I_\alpha = (\mathbf{E}|\xi - \mathbf{E}\xi|^\alpha)^{1/\alpha},$$

where ξ (defined on some probability space) is a Bernoulli random variable with parameter p, i.e., ξ takes the value 1 with probability p and the value 0 with probability q, and where $\mathbf{E}$ is the mathematical expectation. The value $p = 1/2$ is the only one in $(0,1)$ for which $|\xi - \mathbf{E}\xi| = \mathit{const}$ $(= 1/2)$ almost surely. Note also that for all $\alpha \ge 1$, I_α is symmetric around $1/2$, i.e., $I_\alpha(p) = I_\alpha(q)$, and that $I_\alpha(0) = I_\alpha(1) = 0$, $I_\alpha(1/2) = 1/2$. Furthermore, when $\alpha \to +\infty$, $I_\alpha(p)$ converges pointwise on $(0,1)$ to the *convex* function $\max(p,q)$.

To minimize I_μ / I_α, when the isoperimetric function I_μ is constant, and thus to prove Proposition 1.3, we establish the following:

Lemma 7.1 *The function I_α is concave if and only if $1 \le \alpha \le 3$, and it then attains its maximum at $1/2$. For $\alpha \ge 3$, $\max_{0<p<1} I_\alpha(p)$ is an increasing function of α varying from $1/2$ to 1 at infinity.*

For example, if $\alpha = 4$ we have that $p^4 q + pq^4 = pq(1 - 3pq)$ attains its maximum at $pq = 1/6$, i.e., at $p = 1/2 \pm \sqrt{1/12}$, and the maximum of I_4 is equal to $(1/12)^{1/4} > 1/2$.

Proof. We beging by introducing several functions of the variable $p \in [0,1]$, where again $q = 1 - p$.

$$u_\alpha(p) = p^\alpha q + pq^\alpha, \tag{7.2}$$

$$v_\alpha(p) = p^\alpha + q^\alpha, \quad w_\alpha(p) = p^\alpha - q^\alpha, \tag{7.3}$$

$$x = pq. \tag{7.4}$$

Step 1: $2 \le \alpha \le 3$.
We show that u_α is concave for such values of α. Via (7.2), the first and second derivatives of u_α are given by

$$u'_\alpha(p) = \alpha(p^{\alpha-1} q - pq^{\alpha-1}) - (p^\alpha - q^\alpha).$$

$$u''_\alpha(p) = \alpha(\alpha - 1)(p^{\alpha-2} q + pq^{\alpha-2}) - 2\alpha(p^{\alpha-1} + q^{\alpha-1}). \tag{7.5}$$

If $0 \le \alpha \le 1$, then from (7.5), $u''_\alpha(p) \le 0$, for all $p \in (0,1)$, therefore, u_α is concave on the same interval. Rewriting (7.5) as $u''_\alpha(p) = \alpha(\alpha-1)u_{\alpha-2} - 2\alpha v_{\alpha-1}$ and noting that v_α is convex for all $\alpha \ge 1$, we obtain that, for all $\alpha \in [2,3]$, u''_α is concave. Since u''_α is symmetric around $1/2$, it attains its maximum at this point. But,

$$\begin{aligned} u''_\alpha\left(\frac{1}{2}\right) &= \alpha(\alpha-1)2^{-\alpha+2} - 2\alpha 2^{-\alpha+2} \\ &= \alpha(\alpha-3)2^{-\alpha+2} \le 0, \end{aligned} \tag{7.6}$$

Consequently, for all $p \in (0,1)$, $u''_\alpha(p) \le 0$, and u_α is concave and so is $I_\alpha = u_\alpha^{1/\alpha}$ which is the composition of an increasing concave function with a concave one.

Remark. The function u_α is not concave if $1 < \alpha < 2$, since from (7.5), $\lim_{p\to 0^+} u''_\alpha(p) = +\infty$. Thus for such α, the preceding arguments do not work.

Step 2: $1 \le \alpha \le 2$.
It is clear that any function I of the form $I = u^{1/\alpha}$, where $\alpha > 0$ and where u is positive with continuous second derivative, is concave if and only if

$$\alpha u u'' \le (\alpha-1)(u')^2. \tag{7.7}$$

Let us check that the functions u_α satisfy the condition (7.7). First, as direct consequences of (7.2)–(7.4), the following identities are true:

$$u_\alpha = v_\alpha - v_{\alpha+1}, \quad v'_\alpha = \alpha w_{\alpha-1}, \quad w'_\alpha = \alpha v_{\alpha-1}, \tag{7.8}$$

$$u'_\alpha = \alpha w_{\alpha-1} - (\alpha+1)w_\alpha, \tag{7.9}$$

$$u''_\alpha = \alpha(\alpha-1)v_{\alpha-2} - \alpha(\alpha-1)v_{\alpha+1}. \tag{7.10}$$

In addition, for $\alpha \le \beta$,

$$v_\alpha v_\beta = v_{\alpha+\beta} + x^\alpha v_{\beta-\alpha}, \tag{7.11}$$

$$w_\alpha w_\beta = v_{\alpha+\beta} - x^\alpha v_{\beta-\alpha}. \tag{7.12}$$

Using (7.8)–(7.12), we get

$$\begin{aligned} \alpha u_\alpha u''_\alpha &= \alpha^2(\alpha-1)v_\alpha v_{\alpha-2} + \alpha^2(\alpha+1)v_{\alpha+1}v_{\alpha-1} - \alpha^2(\alpha-1)v_{\alpha+1}v_{\alpha-2} - \alpha^2(\alpha+1)v_\alpha v_{\alpha-1} \\ &= \alpha^2(\alpha-1)(v_{2\alpha-2} + x^{\alpha-2}v_2) + \alpha^2(\alpha+1)(v_{2\alpha} + x^{\alpha-1}v_2) \\ &\quad -\alpha^2(\alpha-1)(v_{2\alpha-1} + x^{\alpha-2}v_3) - \alpha^2(\alpha+1)(v_{2\alpha-1} + x^{\alpha-1}v_1) \\ &= \alpha^2(\alpha+1)v_{2\alpha} - 2\alpha^3 v_{2\alpha-1} + \alpha^2(\alpha-1)v_{2\alpha-2} - \alpha^2(\alpha-1)x^{\alpha-2}v_3 \\ &\quad +[\alpha^2(\alpha-1)x^{\alpha-2} + \alpha^2(\alpha+1)x^{\alpha-1}]v_2 - \alpha^2(\alpha+1)x^{\alpha-1}v_1. \end{aligned} \tag{7.13}$$

In a similar way,

$$(\alpha-1)(u'_\alpha)^2 = \alpha^2(\alpha-1)w^2_{\alpha-1} + (\alpha+1)^2(\alpha-1)w^2_\alpha - 2\alpha(\alpha+1)(\alpha-1)w_{\alpha-1}w_\alpha$$

$$= \alpha^2(\alpha-1)(v_{2\alpha-2}-x^{\alpha-1}v_0)+(\alpha+1)^2(\alpha-1)(v_{2\alpha}-x^{\alpha}v_0)-2\alpha(\alpha+1)(\alpha-1)(v_{2\alpha-1}-x^{\alpha-1}v_1)$$
$$= (\alpha+1)^2(\alpha-1)v_{2\alpha} - 2\alpha(\alpha+1)(\alpha-1)v_{2\alpha-1} + \alpha^2(\alpha-1)v_{2\alpha-2}$$
$$+ 2\alpha(\alpha+1)(\alpha-1)x^{\alpha-1}v_1 - [\alpha^2(\alpha-1)x^{\alpha-1} - (\alpha+1)^2(\alpha-1)x^{\alpha}]v_0. \tag{7.14}$$

Now, we need to show that (7.13) is dominated by (7.14). Considering the difference (7.14)–(7.13) and noting that $v_0 = 2, v_1 = 1$, we collect the coefficients of

1) $v_{2\alpha}:\ (\alpha+1)^2(\alpha-1) - \alpha^2(\alpha+1) = -(\alpha+1);$
2) $v_{2\alpha-1}:\ -2\alpha(\alpha+1)(\alpha-1) + 2\alpha^3 = 2\alpha;$
3) $v_{2\alpha-2}:\ \alpha^2(\alpha-1) - \alpha^2(\alpha-1) = 0;$
4) $x^{\alpha-1}:\ -2\alpha^2(\alpha-1) + 2\alpha(\alpha+1)(\alpha-1) + \alpha^2(\alpha+1) = \alpha(\alpha^2+3\alpha-2).$

Therefore, the domination of (7.13) by (7.14) takes the form:

$$-\alpha^2(\alpha-1)x^{\alpha-2}v_3 + [\alpha^2(\alpha-1)x^{\alpha-2} + \alpha^2(\alpha+1)x^{\alpha-1}]v_2 - \alpha(\alpha^2+3\alpha-2)x^{\alpha-1}$$
$$\leq -(\alpha+1)v_{2\alpha} + 2\alpha v_{2\alpha-1}. \tag{7.15}$$

One can simplify the left hand–side of (7.15) with the help of the identities:

$$v_2 = 1-2x; \quad v_3 = 1-3x.$$

Indeed, $v_2 = p^2+q^2 = (p^2+2pq+q^2) - 2pq = 1-2x; v_3 = p^3+q^3 = (p+q)(p^2-pq+q^2) = 1-2x-x$. Now, we get that the left hand side of (7.15) is equal to:

$$\begin{aligned}
- &\quad \alpha^2(\alpha-1)x^{\alpha-2}(1-3x) + [\alpha^2(\alpha-1)x^{\alpha-2} + \alpha^2(\alpha+1)x^{\alpha-1}](1-2x)\\
&\quad -\alpha(\alpha^2+3\alpha-2))x^{\alpha-1}\\
= &\quad -2\alpha^2(\alpha+1)x^{\alpha} + [3\alpha^2(\alpha-1) + \alpha^2(\alpha+1) - 2\alpha^2(\alpha-1) - \alpha(\alpha^2+3\alpha-2)]x^{\alpha-1}\\
&\quad +(-\alpha^2(\alpha-1) + \alpha^2(\alpha-1))x^{\alpha-1}\\
= &\quad -2\alpha^2(\alpha+1)x^{\alpha} + \alpha(\alpha-1)(\alpha-2)x^{\alpha-2}.
\end{aligned}$$

Therefore, (7.15) takes the final form

$$-2\alpha^2(\alpha+1)x^{\alpha} + \alpha(\alpha-1)(\alpha-2)x^{\alpha-1} \ \leq\ -(\alpha+1)v_{2\alpha} + 2\alpha v_{2\alpha-1}. \tag{7.16}$$

Both terms on the left of (7.16) are non–positive for $1 \leq \alpha \leq 2$. Therefore, it suffices to show that

$$(\alpha+1)v_{2\alpha} \ \leq\ 2\alpha v_{2\alpha-1}.$$

This last inequality follows from $\alpha+1 \leq 2\alpha$ and $v_{2\alpha} \leq v_{2\alpha-1}$ (v_α is a decreasing function of α).

Step 3: $\alpha > 3$.
First, we use (7.6) to prove that I_α is not concave for $\alpha > 3$. For such values of α, we have

$u''_\alpha(1/2) > 0$. Hence, by the continuity of the second derivative, this inequality holds in some neighborhood of $p = 1/2$, i.e., the first derivative u'_α increases inside some interval containing the point $1/2$. Since $u'_\alpha(1/2) = 0$, we therefore obtain that $u'_\alpha(p) > 0$, for p close enough (from the right) to $1/2$. Hence, u_α and therefore I_α are increasing in some interval $[1/2, 1/2 + \epsilon]$. But I_α is symmetric around $1/2$ and consequently is not concave on $(0, 1)$. In addition, I_α attains its maximum on $[0, 1]$ at some point $p(\alpha) \neq 0, 1/2, 1$. Hence, for all $3 < \alpha < \beta$, using the monotonicity of the function $\alpha \longrightarrow I_\alpha(p), \alpha \geq 1$, with $p \in (0, 1/2) \cup (1/2, 1)$, we obtain that

$$\max_{0<p<1} I_\alpha(p) = I_\alpha(p(\alpha)) < I_\beta(p(\alpha)) \leq I_\beta(p(\beta)) = \max_{0<p<1} I_\beta(p).$$

Thus, $\max_{0<p<1} I_\alpha(p)$ is an increasing function of $\alpha > 3$. Lemma 7.1 is proved, and Proposition 1.3 follows.

Proof of Proposition 1.8.

Since for $p \in (0, 1), p \neq 1/2$, I_α is an increasing function of α, it is enough to notice that

$$\frac{2pq}{\pi \rho pq} \leq \frac{\left(p^{\frac{\alpha}{\alpha-1}} q + p q^{\frac{\alpha}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}}}{\pi \rho pq} = \frac{\left(p^{\frac{1}{\alpha-1}} + q^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}}}{\pi \rho (pq)^{\frac{1}{\alpha}}} \leq \frac{\max(p, q)}{\pi \rho pq} = \frac{1}{\pi \rho \min(p, q)},$$

and then to take the infimum.

8 A special case of Theorem 1.2

Proofs of Lemma 1.4 and 1.6 are presented in this section. As a consequence, we have a version of Theorem 1.2 which will be applied to the uniform distribution on the n–sphere S^n_ρ with $n \geq 2$. Assume that the isoperimetric function $I = I_\mu$ satisfies the following properties:

(i) I is symmetric around 1/2, i.e., for all $p \in (0,1)$,

$$I(1-p) = I(p).$$

(ii) $I(0^+) = I(1^-) = 0$.

(iii) I is continuously differentiable on $(0,1)$ and dI^α/dp is convex on $(0,1/2]$.

Proposition 8.1 *Let $1 \leq \alpha \leq 2$. Then, under the conditions (i), (ii) and (iii) above, for any μ–integrable, Lipschitz on every ball, function f,*

$$\int_X |\nabla f(x)| d\mu \geq c \left(\int_X |f(x) - m(f)|^\alpha d\mu\right)^{\frac{1}{\alpha}},$$

$$\int_X |\nabla f(x)| d\mu \geq c \inf_{a \in \mathbf{R}} \left(\int_X |f(x) - a|^\alpha d\mu\right)^{\frac{1}{\alpha}},$$

where $c = 2I_\mu(1/2)$ is the optimal constant.

To prove Proposition 8.1 and thus Lemma 1.4 and 1.6, it is in fact enough to only consider a partial case of these results.

Lemma 8.2 *If a non-negative function I defined on $(0,1)$ satisfies the conditions (i), (ii), and (iii) above, with $\alpha = 1$, then the infimum of $I(p)/p(1-p)$ on the interval $(0,1)$ is attained at $p = 1/2$.*

Provided this statement is proved, one can apply it to the function I^α, and obtain that the infimum of $I^\alpha(p)/(p(1-p))$ is attained at $p = 1/2$. Then, the function $w(p) = p^{\alpha-1} + (1-p)^{\alpha-1}$, $1 \leq \alpha \leq 2$, has its maximum at $p = 1/2$ since it is concave, symmetric around 1/2 with also $w(0) = w(1) = 1$. Therefore,

$$\frac{I^\alpha(p)}{p^\alpha(1-p) + p(1-p)^\alpha} = \frac{I^\alpha(p)}{(p(1-p))w(p)}$$

also attains its minimum at $p = 1/2$. Similarly, since the functions $\max(p, 1-p)$ and $(p^{1/(\alpha-1)} + (1-p)^{1/(\alpha-1)})^{\alpha-1}, 1 < \alpha \leq 2$, are convex, symmetric around 1/2 and are equal to one at 0 and at 1, the respective minimum of $I(p)/\min(p, 1-p)$ and of $I^\alpha(p)(p^{1/(\alpha-1)} + (1-p)^{1/(\alpha-1)})^{\alpha-1}/p(1-p)$ is attained at $p = 1/2$.

Proof of lemma 8.2. Let us first extend the function I to the interval $[0,1]$ by setting $I(0) = I(1) = 0$, and let us define the function

$$u(p) = I(p) - 4I\left(\frac{1}{2}\right)p(1-p),\ 0 \le p \le 1.$$

We need to show that $u(p) \ge 0$ for all $p \in (0,1)$. In view of (8.1), it is enough to only consider the case $0 < p \le \frac{1}{2}$. The function u has the following properties:

a) $u(0) = u(1/2) = 0$.

b) The derivative $u'(p) = I'(p) - 4I(1/2)(1-2p)$ is a convex, continuous function on $(0,1/2]$.

c) $u'(1/2) = 0$.

The equality in c) follows from the $I'(1-p) = -I'(p)$ (applied at $1/2$) which itself follows from (i). The property b) follows from (iii): u' is a sum of convex functions.

Formally, $u'(0^+)$ has three possible behaviors.

1) $u'(0^+) < 0$.

By b), u' is non–decreasing on $(0,1/2]$, while by c) $u'(1/2) = 0$. Hence, $u'(p) \le 0$ for all $0 < p < 1/2$ with strict inequality in a neighbourhood of $p = 0$. Therefore, u is non–increasing on $[0,1/2]$ and decreases in a neighborhood of $p = 0$, hence $u(1/2) < u(0)$. This contradicts a), and thus $u'(0^+) \ge 0$.

2) $u'(0^+) = 0$.

An argument as in 1) can also be applied here. This does not lead to a contradiction only when $u' = 0$ on $[0,1/2]$. But, by a) this gives $u = 0$, thus $I(p) = 4I(1/2)p(1-p)$ for all $p \in [0,1/2]$, and for such I there is nothing to prove.

3) $u'(0^+) > 0$.

Since $u'(0^+) > 0$, since $u'(1/2) = 0$, and since u' is convex and continuous on $(0,1/2]$, two formal possibilities have to be considered. First, let $u'(p) \ge 0$, for all $0 < p < 1/2$, and for some $p_0 \in (0,1/2]$, $u'(p) > 0$ for all $0 < p < p_0$, and $u'(p) = 0$ for all $p_0 \le p < 1/2$. Again, we get the contradiction:

$$0 = u\left(\frac{1}{2}\right) = u(p_0) > u(0) = 0.$$

Thus, only the second possibility can take place, i.e., there is a unique $p_0 \in (0,1/2]$ such that $u'(p) > 0$ for all $0 < p < p_0$, $u'(p_0) = 0$, and $u'(p) < 0$ for all $p_0 < p < 1/2$. In addition, $u(0) = u(1/2) = 0$, and $u(p) > 0$ for $p > 0$ small enough by the assumption 3) and by the continuity of u'. Clearly, for functions u possessing these properties, the inequality $u \ge 0$ on $[0,1/2]$ has to hold, since otherwise u' would be zero at two or more points in $(0,1/2)$. This finishes the proof of Lemma 8.2.

9 The uniform distribution on the sphere

We show here how to apply Proposition 8.1 to the uniform distribution σ_n on the sphere $S^n_\rho \subset \mathbf{R}^{n+1}$ of radius $\rho > 0$, and thus to prove Proposition 1.5 and 1.7. For simplicity, we may assume that the center of the sphere is at the origin. Denote by μ_n the Lebesgue measure on S^n_ρ, i.e., $\mu_n = \rho^n s_n \sigma_n$. The isoperimetric property of the balls on S^n_ρ states (see [Lev], [Sch]) that, for any $h > 0$, among all the Borel sets $A \subset S^n_\rho$ of fixed volume $\mu_n(A) = \rho^n s_n p$, $0 < p < 1$, the value of

$$\mu_n(A^h)$$

is minimal if A is an arbitrary ball on the sphere and, in particular, if A is the ball

$$B_n(t) = \{x \in S^n_\rho : x_1 \le t\}, \quad |t| < \rho.$$

Here t is chosen so that

$$F_n(t) = \sigma_n(B_n(t)) = p, \quad |t| < \rho. \tag{9.1}$$

That is, $t = F_n^{-1}(p)$ is the quantile of order p of the distribution function F_n of the random variable $\xi(x) = x_1$ defined on the probability space (S^n_ρ, σ_n), where as usual F_n^{-1} denotes the inverse of F_n. From the very definition of the isoperimetric function, we then have

$$I_{\sigma_n}(p) = \sigma_n^+(B_n(t)). \tag{9.2}$$

To apply Proposition 8.1, we need two elementary results which for the sake of completeness are derived below. From now on, we assume that $n \ge 2$.

Lemma 9.1 *For all $p \in (0,1)$,*

$$I_{\sigma_n}(p) = \frac{s_{n-1}}{s_n \rho^n}(\rho^2 - F_n^{-2}(p))^{\frac{n-1}{2}}. \tag{9.3}$$

Proof. By (1.3) and (9.2),

$$\begin{aligned} I_{\sigma_n}(p) &= \liminf_{h \to 0^+} \frac{\sigma_n(B_n^h(t)) - \sigma_n(B_n(t))}{h} \\ &= \liminf_{h \to 0^+} \frac{\mu_n(B_n^h(t)) - \mu_n(B_n(t))}{\rho^n s_n h}, \end{aligned} \tag{9.4}$$

where $t = F_n^{-1}(p)$. Note that for $t \in (-\rho, \rho)$, the boundary

$$\partial B_n(t) = S_r^{n-1}(t) = \{x \in S^n_\rho : x_1 = t\}$$

is the $(n-1)$ dimensional sphere of radius r such that $r^2+t^2=\rho^2$. Therefore, the lim inf in (9.4) can be replaced by lim which is equal to

$$\frac{\mu_{n-1}(S_r^{n-1})}{\rho^n s_n} = \frac{r^{n-1}s_{n-1}}{\rho^n s_n},$$

and which coincides with the right hand side of (9.3).

Lemma 9.2 *The distribution function F_n is absolutely continuous with density*

$$f_n(t) = \frac{s_{n-1}}{\rho^{n-1}s_n}(\rho^2-t^2)^{\frac{n-2}{2}}, \quad |t|<\rho. \tag{9.5}$$

Proof. First note that the geodesic metric as well as the Euclidean metric on the sphere can be used. For the geodesic metric, the h–neighborhood $B_n^h(t), |t|<\rho$, if it is not the whole sphere, is the ball $B_n(s)$, $t<s<\rho$, where s is defined by

$$\rho\arccos(t) - \rho\arccos(s) = h.$$

Hence,

$$h = \frac{\rho(s-t)}{\sqrt{(1-t^2)}} + \mathrm{O}((s-t)^2) = \frac{\rho(s-t)}{r} + \mathrm{O}((s-t)^2),\ s\to t^-, \tag{9.6}$$

where r is defined as in Lemma 9.1, i.e., $r^2+t^2=\rho^2$. Taking small positive values of h, we get by (9.6) and arguments similar to the one used in Lemma 9.1, that

$$\begin{aligned} F_n(s)-F_n(t) &= \sigma_n(B_n(s)) - \sigma_n(B_n(t)) \\ &= \sigma_n(B_n^h(t)) - \sigma_n(B_n(t)) \\ &= \frac{\mu_{n-1}(S_r^{n-1})h}{\rho^n s_n} + \mathrm{O}(h^2) \\ &= \frac{r^{n-1}s_n h}{\rho^n s_n} + \mathrm{O}(h^2) \\ &= \frac{r^{n-2}s_{n-1}(s-t)}{\rho^{n-1}s_n} + \mathrm{O}((s-t)^2). \end{aligned}$$

Therefore,

$$f_n(t) = \lim_{s\to t^-}\frac{F_n(s)-F_n(t)}{s-t} = \frac{r^{n-2}s_{n-1}}{\rho^{n-1}s_n},$$

which coincides with (9.5).

Let us now check the conditions (i), (ii) and (iii) required in Proposition 8.1. The first two are trivially verified and only (iii) requires some proof. Without loss of generality, let $\rho=1$. Let

$$I(p) = (1-F_\lambda^{-2}(p))^\tau, \tag{9.7}$$

where $\tau \geq 0, \lambda \geq 0$, and where $F_\lambda^{-1} : (0,1) \longrightarrow (-1,1)$ is the inverse of the distribution F_λ of density

$$f_\lambda(t) = d_\lambda(1-t^2)^\lambda, \quad |t| < 1. \tag{9.8}$$

The normalizing constant d_λ corresponds to the condition $F_\lambda(1) = 1$.

Lemma 9.3 *If* $\alpha \leq \frac{\lambda+1}{\tau}$, *then the derivative of* I^α *is convex on the interval* $(0, \frac{1}{2}]$.

Proof. Differentiating (9.7), we have

$$\begin{aligned} -\frac{dI(p)}{dp} &= \frac{2\tau(1-F_\lambda^{-2}(p))^{\tau-1}F_\lambda^{-1}(p)}{f_\lambda(F_\lambda^{-1}(p))} \\ &= 2\tau d_\lambda^{-1} F_\lambda^{-1}(p)(1-F_\lambda^{-2}(p))^{\tau-\lambda-1}. \end{aligned}$$

Hence,

$$-\frac{dI^\alpha(p)}{dp} = 2\alpha\tau d_\lambda^{-1}F_\lambda^{-1}(p)(1-F_\lambda^{-2}(p))^\beta, \tag{9.9}$$

where $\beta = (\alpha-1)\tau + (\tau-\lambda-1) = \alpha\tau - \lambda - 1$. Note that whenever $\beta \leq 0$, the function $x(1-x^2)^\beta$ is increasing and convex on $[0,1]$. From (9.8), F_λ is increasing and concave on $[0,1]$, therefore F_λ^{-1} is increasing and convex on $[1/2, 1]$. Thus, when $\beta \leq 0$, the right hand side of (9.9) is non–decreasing and convex on $[1/2, 1]$, as the composition of two functions with the same properties. In addition, since f_λ is symmetric around 0, $F_\lambda^{-1}(1-p) = -F_\lambda^{-1}(p)$ and the right hand side in (9.9) is odd around $p = 1/2$. Taking into account the minus sign in (9.9), we obtained that dI^α/dp is convex on $(0, 1/2]$ provided $\beta = \alpha\tau - \lambda - 1 \leq 0$, i.e., $\alpha \leq (\lambda+1)/\tau$. This completes the proof.

We can now apply Lemma 9.3 to prove Proposition 1.5 and 1.7. Indeed, applying it to the function I_{σ_n}, i.e., taking $\tau = (n-1)/2, \lambda = (n-2)/2$, we see that the function $dI_{\sigma_n}^\alpha/dp$ $(\alpha \geq 1)$ is convex on $(0, \frac{1}{2}]$ if

$$\alpha \leq \frac{\lambda+1}{\tau} = \frac{n}{n-1}.$$

Note also, that the condition $\alpha \leq 2$ is automatically satisfied since $n \geq 2$. By Proposition 8.1, we thus obtain (1.19) and (1.25) for the measure $\mu = \sigma_n, n \geq 2$, where according to (9.3) the best constant is given by

$$c_n = 2I_{\sigma_n}\left(\frac{1}{2}\right) = \frac{2s_{n-1}}{\rho s_n}.$$

To prove the last parts of Proposition 1.5 and 1.7, we note that near $p = 0$,

$$I_n(p) \asymp p^{\frac{n-1}{n}}.$$

Indeed, it is easily seen that for $-1 \leq x \leq 0$ and $\lambda \geq 0$,

$$\frac{d_\lambda}{\lambda+1}(1+x)^{\lambda+1} \leq F_\lambda(x) \leq 2^\lambda d_\lambda (1+x)^{\lambda+1}.$$

Therefore, setting $x = F_\lambda^{-1}(p)$, we get if $I(p)$ is as in (9.7),

$$d_\lambda \left[1 - \left(\left(\frac{p}{2^\lambda d_\lambda}\right)^{\frac{1}{\lambda+1}} - 1\right)^2\right]^\tau \leq I(p) \leq d_\lambda \left[1 - \left(\left(\frac{p(\lambda+1)}{d_\lambda}\right)^{\frac{1}{\lambda+1}} - 1\right)^2\right]^\tau,$$

from which we conclude that as $p \to 0^+$, $I(p) \asymp p^{\frac{\tau}{\lambda+1}}$. Finally, setting $\tau = (n-1)/2, \lambda = (n-2)/2$, we see that for $\alpha > n/n-1$, zero is the best non–negative constant in Proposition 1.5 and 1.7.

10 Existence of optimal Orlicz spaces

Recall that a convex function N on the real line is said to be a Young function if it is non–negative, even, and if $N(x) = 0$ only for $x = 0$. Given a measurable space (X, μ) (as usual for us, X is a metric space with metric d and μ is a separable and non–atomic Borel probability measure), such a function N generate the Orlicz space $L_N(X, \mu)$ of μ-measurable real valued functions f equipped with the norm

$$\|f\|_N \;=\; \inf\left\{\lambda > 0 : \int_X N\left(\frac{f(x)}{\lambda}\right) d\mu(x) \le 1\right\}. \tag{10.1}$$

In particular, for any Borel measurable set $A \subset X$ of measure $\mu(A) = p \in (0, 1)$, the value $I(p) = \|\chi_A - p\|_N$ is the only positive one which satisfies

$$pN\left(\frac{q}{I(p)}\right) + qN\left(\frac{p}{I(p)}\right) = 1, \quad q = 1 - p. \tag{10.2}$$

First part of Theorem 1.11 *Given a positive function I on $(0, 1)$, there exists a Young function N satisfying (10.2) for all $p \in (0, 1)$ if and only if*

1) $I(0^+) = I(1^-) = 0$;

2) $I(p) = I(q)$, for all $p \in (0, 1)$;

3) the function $pq/I(p)$ is concave on $(0, 1)$.

Proof. *Necessity.* In (1.33) of Section 1 and (5.3) of Section 5, we introduced the function $x_p = x_p(0, 1)$ of the variable $p \in (0, 1)$ as the positive solution of the equation

$$pN(qx) + qN(px) = 1, \quad q = 1 - p. \tag{10.3}$$

So, $x_p = 1/I(p)$. By (10.3), $x_p = x_q$, and the property 2) follows. By Lemma 5.6, the function

$$y(p) = px_p$$

increases on $(0, 1)$, therefore $\ell = \lim_{p\to 1^-} y(p)$ exists, being finite or not. If $\ell < +\infty$, then since $x_p = x_q$, rewriting (10.3) as

$$pN(y(q)) + qN(y(p)) = 1 \tag{10.4}$$

and letting $p \to 0^+$, we get $N(y(0^+)) = 1$. But then, we obtain by (10.4) that, for all $p \in (0, 1)$,

$$1 > pN(y(0^+)) + qN(y(0^+)) = N(y(0^+)) = 1.$$

Therefore, $\ell = +\infty$, i.e., $I(p) \to 0$, as $p \to 1^-$, and the property 1) is proved. It remains to establish 3).

Introducing the function $z(p) = pq/I(p)$ where as usual $q = 1 - p$, let us rewrite (10.2) in the form

$$pN\left(\frac{z(p)}{p}\right) + qN\left(\frac{z(q)}{q}\right) = 1. \tag{10.5}$$

We now show that (10.5) implies that $z = z(p)$ is a concave function. Since N is a non–negative, convex function on $[0, +\infty)$ and $N(0) = 0$, $N(x) > 0$ for $x > 0$, N can be represented on $(0, +\infty)$ in the form

$$N(x) \;=\; \sup_{(c,d)\in T}(cx - d), \tag{10.6}$$

for some $T \subset (0, +\infty) \times [0, +\infty)$ with the following property: whenever $x > 0$, (10.6) is attained at some $(c, d) \in T$. From (10.5) and (10.6), noting that $z(p) = z(q)$, we have

$$\begin{aligned} pN\left(\frac{z}{p}\right) + qN\left(\frac{z}{q}\right) \;&=\; p \sup_{(c,d)\in T}\left(c\frac{z}{p} - d\right) + q \sup_{(c',d')\in T}\left(c'\frac{z}{q} - d'\right) \\ &=\; \sup_{(c,d)\in T,\ (c',d')\in T}[(c+c')z - (dp + d'q)] \;=\; 1. \end{aligned} \tag{10.7}$$

In particular, for all $(c, d) \in T$ and $(c', d') \in T$, we have $(c + c')z - (dp + d'q) \le 1$, i.e.,

$$z \;\le\; \inf_{(c,d)\in T,\ (c',d')\in T} \frac{1 + dp + d'q}{c + c'}. \tag{10.8}$$

Conversely, since (10.6) is attained at some $(c, d) \in T$, (10.7) and therefore, (10.8) are also attained at some $(c, d), (c', d') \in T$. Hence, the inequality in (10.8) is in fact an equality, and the function z is thus concave as the infimum of a family of affine functions.

To prove the sufficiency part of the result, we first state

Lemma 10.1 *If a positive function I on $(0, 1)$ satisfies the properties 1)–3), then the function $y(p) = p/I(p)$ is increasing on $(0, 1)$.*

Proof. Since z is concave and positive on $(0, 1)$, $z(0^+)$ is finite and non–negative. Let z' be a non–increasing Radon–Nikodym derivative of z. Since $y(p) = z(q)/q$, where $q = 1 - p$, one can write

$$y(p) = \frac{z(0^+)}{q} + \int_0^1 z'(tq)dt.$$

The functions $p \longrightarrow 1/q$ and $p \longrightarrow z'(tq)$ is non–decreasing and so is y as the sum of non–decreasing functions. Moreover, if $z(0^+) > 0$, then, since $p \longrightarrow 1/q$ is increasing on $(0, 1)$, y is also increasing on $(0, 1)$. Let now $z(0^+) = 0$, and thus

$$y(p) = \int_0^1 z'(tq)dt.$$

Assume that $y(p_0) = y(p_1)$, for some $0 < p_1 < p_0 < 1$. Then

$$z'(tq_1) = z'(tq_0),$$

for almost all (with respect to Lebesgue measure) $t \in (0,1)$, where $q_0 = 1-p_0$, $q_1 = 1-p_1$. But then, as easily seen, z' is constant on $(0, q_0)$, hence z is of the form $z(s) = a + bs$, for all $0 < s < q_0$, and since $z(0^+) = 0$, we have $z(s) = bs$, for some $b > 0$ and for all $s \in (0, q_0)$. If such is the case, we get

$$I(s) = \frac{s(1-s)}{z(s)} = \frac{1-s}{b},$$

hence $I(0^+) = 1/b > 0$, which contradicts the assumption 1). Lemma 10.1 is proved.

Sufficiency. From Lemma 10.1, we know that the function $y(p) = p/I(p)$ is increasing on $(0,1)$, and by the assumption 1) that $y(1^-) = +\infty$. Without loss of generality, one can assume that $y(1/2) = 1$. Then,

$$c_0 = y(0^+) < 1,$$

so y is an increasing bijection from $(0, 1/2]$ to $(c_0, 1]$, and from $[1/2, 1)$ to $[1, +\infty)$. Denote by $y^{-1} : (c_0, +\infty) \longrightarrow (0,1)$ its inverse.

We need to find a convex function $N : (0, +\infty) \longrightarrow (0, +\infty)$ satisfying (10.2) and such that $N(0^+) = 0$. Then, N will be extended to the whole real line by putting $N(-x) = N(x)$, for $x > 0$. To define N on $(0, +\infty)$, we first set $N(x) = x$, for $0 < x \leq 1$. For $x > 1$, we define $N(x)$ (in a unique way) in according (10.4). First, for $0 < p \leq 1/2$, from the above discussion we have $N(y(p)) = y(p)$. Therefore (10.4) becomes

$$N(y(q)) \ = \ \frac{1 - qy(p)}{p}, \quad 0 < p \leq 1/2, \ q = 1-p,$$

hence letting $x = y(q)$, we have

$$N(x) \ = \ \frac{1 - y^{-1}(x)y(1 - y^{-1}(x))}{1 - y^{-1}(x)}, \quad x \geq 1. \tag{10.9}$$

Clearly, the function N defined in (10.9) is increasing and continuous on $[1, +\infty)$, and in addition, $N(1) = 1$. So, it just remains to show that N is convex on $[1, +\infty)$, and that the right derivative $N'(1^+) \geq 1$. Then, the extended function N will be convex on $[0, +\infty)$.

To obtain the first requirement, we give instead of (10.9), another representation for N. Note that, by the definition of N, any of the identities (10.2), (10.4) and (10.5) (which are equivalent to one another) are fulfilled. Recall also that, by the assumption

2), $z(p) = z(q)$. For $0 < p \le 1/2$, $z(p)/q = y(p) \le 1$, hence $N(z(p)/q) = z(p)/q$, and the identity (10.5) becomes

$$N\left(\frac{z(p)}{p}\right) = \frac{1}{p} - \frac{z(p)}{p}, \quad 0 < p \le 1/2,$$

or in terms of the function

$$\delta(p) \; = \; \frac{z(p)}{p} \; = \; y(q),$$

we obtain the identity

$$N(\delta(p)) \; = \; \frac{1}{p} \; - \; \delta(p), \quad 0 < p \le 1/2. \tag{10.10}$$

Note that δ decreases on (0,1/2] and satisfies $\delta(1/2) = 1$, $\delta(0^+) = y(1^-) = +\infty$. Let $\delta^{-1} : [1, +\infty) \longrightarrow (0, 1/2]$ be the inverse of δ. Then, putting $x = \delta(p)$, (10.10) can be rewriten as

$$N(x) \; = \; \frac{1}{\delta^{-1}(x)} - x, \quad x \ge 1. \tag{10.11}$$

Therefore, N is convex on the interval $x \ge 1$ if and only if the function $1/\delta^{-1}(x)$ is convex on the same interval. Since this last function is increasing, continuous on $[1, +\infty)$ and satisfies, by (10.11), $1/\delta^{-1}(1) = 2$, $1/\delta^{-1}(+\infty) = \infty$, the convexity of $1/\delta^{-1}$ is equivalent to the concavity of its inverse function R on $[2, +\infty)$. Let us find R. For any $t \ge 2$,

$$\frac{1}{\delta^{-1}(x)} = t \;\Leftrightarrow\; \delta^{-1}(x) = \frac{1}{t} \;\Leftrightarrow\; x = \delta\left(\frac{1}{t}\right) = tz\left(\frac{1}{t}\right).$$

Thus,

$$R(t) = tz\left(\frac{1}{t}\right), \quad t \ge 2.$$

By the assumption 3), z is a concave function, hence can be represented in the form

$$z(p) \; = \; \inf_{(c,d) \in S} (cp + d),$$

for some $S \subset \mathbf{R} \times \mathbf{R}$. Therefore,

$$R(t) = tz\left(\frac{1}{t}\right) = t \inf_{(c,d) \in S} \left(\frac{c}{t} + d\right) = \inf_{(c,d) \in S} (c + dt)$$

is concave as the infimum of a family of affine functions.

To obtain the second requirement, i.e., to show that $N'(1^+) \ge 1$, one needs to check, according to (10.11), that the right derivative

$$\left[\frac{1}{\delta^{-1}(x)}\right]'_{x=1^+} \; \ge \; 2,$$

or, equivalently, that $R'(2^+) \leq 1/2$. As a concave function, z is differentiable at all points p except, maybe, on a countable set $U \subset (0,1)$. For $t > 2$ such that $1/t$ does not belong to U, we have

$$R'(t) = \left[tz\left(\frac{1}{t}\right)\right]' = z\left(\frac{1}{t}\right) - \frac{z'\left(\frac{1}{t}\right)}{t}. \tag{10.12}$$

Letting in (10.12) $t \to 2$, $t > 2$, we get

$$R'(2^+) = z\left(\frac{1}{2}\right) - \frac{z'\left(\left(\frac{1}{2}\right)^-\right)}{2} \leq z\left(\frac{1}{2}\right) = \frac{1}{2},$$

because, by the assumptions 2) and 3), $z'((1/2)^-) \geq 0$. The first part of Theorem 1.11 is thus proved.

Remark 10.2 In general, when $c = 2I(1/2)$ is an arbitrary positive constant (not necessarily 1), a function N satisfying (10.2) is constructed in the same way, by putting:

(i) $N(\frac{1}{c}) = 1$;

(ii) $N(x) = cx$, for all $0 \leq x \leq \frac{1}{c}$;

(iii) $N(x) = \frac{1}{1-q(x)} - cx$, for all $x \geq \frac{1}{c}$, where $q(x) \in [1/2, 1)$ is the only solution of $q = I(q)x$.

The condition (i) necessarily follows from (10.2) when $p = 1/2$, while given (ii), (iii) follows from (10.2) by putting $x = q/I(q)$, $1/2 \leq q < 1$ there.

Now, given a function I satisfying 1)–3), assume that another Young function N_0 is defined by the equality

$$2pqN_0\left(\frac{1}{I(p)}\right) = 1, \quad q = 1 - p,$$

so that $N_0(2/c) = 2$, and N_0 is linear on $[0, 2/c]$, that is as in (ii), $N_0(x) = cx$, for all $0 \leq x \leq 2/c$. As noted in the introductory section, such a function N_0 was studied by Pellicia and Talenti ([Pel–Tal]) when $I(p) = \varphi(\Phi^{-1}(p))$ (the isoperimetric function of the canonical Gaussian measure).

Let us now compare the Orlicz norms (with respect to N and N_0) of the function $\chi_A - p$, where A is a Borel set in X of μ−measure p. In addition to satisfying (i)–(iii), N is also assumed to be strictly convex on $[1/c, +\infty)$ (which is certainly the case of $I(p) = \varphi(\Phi^{-1}(p)))$. Since for any $p \in (0,1)$ and $x \geq 0$, $N(px) \leq pN(x)$ with strict inequality for $x \in (1/c, +\infty)$, we have putting $x = 1/I(p)$, $q = 1 - p$,

$$2pqN_0(x) = 1 = pN(qx) + qN(px) \leq 2pqN(x)$$

with strict inequality for $x > 1/c$. Therefore, $N_0(x) < N(x)$ for all $x \geq \inf_p 1/I(p) = 2/c$. Also, $N_0(x) = cx < N(x)$ for all $x \in (1/c, 2/c]$, by the strict convexity of N on $[1/c, +\infty)$.

Moreover, $N_0(x) = N(x)$ for all $x \in [0, 1/c]$. Now fix $p \in (0,1)$ and put

$$x = \|\chi_A - p\|_N, \quad x_0 = \|\chi_A - p\|_{N_0}.$$

By the very definition of the Orlicz norm, we have $pN(qx)+qN(px) = 1$. Hence $pN_0(qx)+qN_0(px) \le 1$ with strict inequality if and only if $px > 1/c$ or $qx > 1/c$. Since the function $y(p) = px$ is strictly increasing, and $y(1/2) = 1/c$, this is possible if and only if $p > 1/2$ or $q > 1/2$. Therefore, $x_0 < x$ when $p \neq 1/2$, and $x = x_0$ when $p = 1/2$. Thus, one concludes that the inequality

$$\|f - m(f)\|_{N_0} \le \int_{\mathbf{R}^n} |\nabla f| d\gamma_n$$

becomes the isoperimetric inequality for the indicator functions $f = \chi_A$ if and only if the sets A have measure $p = 1/2$, and is weaker otherwise.

Second part of Theorem 1.11 *Let the conditions 1)-3) be fulfilled for the isoperimetric function $I = I_\mu$, and let N be a Young function satisfying (10.2). Then, for any Young function M such that $L_M(X,\mu)$ contains $W(X,\mu)$ as an embedded space, $L_M(X,\mu)$ also contains $L_N(X,\mu)$ as an embedded space.*

We recall that $W(X,\mu)$ denotes the space of μ–integrable functions f, which are Lipschitz on every ball in X, such that $\int_X f d\mu = 0$, equipped with norm

$$\|f\|_W \ = \ \int_X |\nabla f| d\mu.$$

Proof. To prove that $L_N(X,\mu)$ is embedded in $L_M(X,\mu)$, i.e., to prove that for some $c > 0$ and all $f \in L_N(X,\mu)$,

$$\|f\|_M \ \le \ c\|f\|_N,$$

it suffices to find constants $c, d > 0$ (below and above c might denote two different absolute constant) such that

$$M(x) \ \le \ \frac{1}{d} N(cx), \tag{10.13}$$

for all $x > 0$ large enough.

By assumption, for some $c > 0$,

$$\|f\|_M \ \le \ c\|f\|_W, \tag{10.14}$$

for all $f \in W$, where $\|f\|_M$ is defined as in (10.1). By Theorem 1.1 (or, Theorem 1.10), (10.14) implies

$$\|\chi_A - p\|_M \ \le \ cI(p),$$

where $A \subset X$ is an arbitrary Borel set of measure $\mu(A) = p \in (0,1)$. By definition (10.1), this means that for all $p \in (0,1)$,

$$pM\left(\frac{qx_p}{c}\right) + qM\left(\frac{px_p}{c}\right) \ \le \ 1, \tag{10.15}$$

where, as usual, $x_p = 1/I(p), q = 1 - p$. We get from (10.15) that, for all $0 < p \le 1/2$,

$$pM\left(\frac{qx_p}{c}\right) \ \le \ 1.$$

Hence, since $q \ge 1/2$ and since M is increasing on $[0, +\infty)$,

$$x_p \ \le \ 2cM^{-1}\left(\frac{1}{p}\right), \quad 0 < p \le 1/2, \tag{10.16}$$

where M^{-1} is the inverse of M restricted to $[0, +\infty)$.

Let us return to the function N and to the identity (10.4) which is equivalent to (10.2). Since the function $y(p) = px_p$ is strictly increasing on $(0,1)$, we must have $N(c_0) < 1$ where

$$c_0 = c_0(\mu) = \lim_{p \to 0^+} px_p,$$

since otherwise we would have

$$1 = pN(qx_p) + qN(px_p) > pN(c_0) + qN(c_0) = N(c_0) \ge 1$$

Letting in (10.4) $p \to 0^+$, we obtain that for all small enough $p \in (0, 1/2]$,

$$pN(qx_p) \ \le \ d = \frac{1 - N(c_0)}{2} \ > 0.$$

Hence for such p, $pN(x_p/2) \le d$, i.e.,

$$x_p \ \ge \ 2N^{-1}\left(\frac{d}{p}\right), \tag{10.17}$$

where N^{-1} is the inverse of N restricted to $[0, +\infty)$. Comparing (10.16) and (10.17), we have that for all small enough $p \in (0, 1/2]$,

$$N^{-1}\left(\frac{d}{p}\right) \ \le \ cM^{-1}\left(\frac{1}{p}\right), \tag{10.18}$$

and (10.13) follows from (10.18) by putting $y = 1/p$, $x = M^{-1}(y)$ and since N is increasing on $[0, +\infty)$.

11 Proof of Theorem 1.12 (the case of the sphere)

Given two non–negative parameters τ and λ, let

$$(11.1) \qquad I(p) \;=\; (1 - F^{-2}(p))^{\tau}, \quad 0 < p < 1,$$

where $F^{-1} : (0,1) \longrightarrow (-1,1)$ is the inverse of the distribution function F which is concentrated in $(-1,1)$ and has density

$$(11.2) \qquad F'(x) \;=\; d_\lambda (1 - x^2)^{\lambda}, \quad |x| < 1,$$

where d_λ is a normalizing constant. As we saw in Section 9, I_{σ_n} the isoperimetric function corresponding to the uniform distribution on the n–sphere $S_1^n, n \geq 2$, (in the sequel, and for simplicity $\rho = 1$) has up to a constant the form (11.1)–(11.2) with

$$(11.3) \qquad \tau = \frac{n-1}{2}, \quad \lambda = \frac{n-2}{2}.$$

In this section we verify that I_{σ_n} $(n \geq 2)$ satisfies the conditions 1)–3) of Theorem 1.11. The properties 1) and 3) are trivially satisfied, and only 3) requires some proof. Note that the case $n = 2$ was studied in Section 1: $I_{\sigma_2}(p) = \sqrt{pq}$ $(q = 1 - p)$, so the function

$$\frac{pq}{I_{\sigma_2}(p)} \;=\; \sqrt{pq}$$

is clearly concave on (0,1).

Lemma 11.1 *If* $1 \leq \lambda \leq \tau \leq \lambda + 1, \tau \geq 2$, *then the function*

$$z(p) \;=\; \frac{pq}{I(p)}, \quad q = 1 - p,$$

is concave on $(0,1)$.

If τ and λ are of the form (11.3), then the assumptions of Lemma 11.1 are only fulfilled for $n \geq 5$, so this result gives the proof of Theorem 1.12 for the n–sphere only when $n \geq 5$. The cases $n = 3$ and $n = 4$ are treated separately after the proof of Lemma 11.1.

Proof. Let us first give another equivalent wording (with arbitrary τ and λ) of the statement of the lemma. Introduce the function

$$(11.4) \qquad H_\lambda(x) \;=\; \int_0^x (s(1-s))^{\lambda} ds, \quad 0 < x < 1.$$

By (11.2), making the change of variables $t = 2s - 1$, we have

$$F(x) = d_\lambda \int_{-1}^{x} (1-t^2)^{\lambda} dt = d_\lambda 2^{2\lambda+1} \int_0^{\frac{1+x}{2}} (s(1-s))^{\lambda} ds = d_\lambda 2^{2\lambda+1} H_\lambda(\frac{1+x}{2}).$$

Hence, by (11.1) and taking into account the symmetry identity $1 - F(x) = F(-x)$, we get

$$
\begin{aligned}
z(F(x)) &= \frac{F(x)(1-F(x))}{I(F(x))} = \frac{F(x)F(-x)}{(1-x^2)^\tau} \\
&= d_\lambda^2 4^{2\lambda+1-\tau} \frac{H_\lambda(\frac{1+x}{2})H_\lambda(\frac{1-x}{2})}{(\frac{1+x}{2})^\tau(\frac{1-x}{2})^\tau} \\
&= c\,\frac{H_\lambda(y)H_\lambda(1-y)}{y^\tau(1-y)^\tau},
\end{aligned}
\tag{11.5}
$$

where $y = (1+x)/2$, and where c is a constant depending only on τ and λ. Differentiating (11.5) (note that $dy/dx = 1/2$) gives

$$
z'(F(x))F'(x) = \frac{d}{dy}\left[\frac{H_\lambda(y)H_\lambda(1-y)}{y^\tau(1-y)^\tau}\right].
$$

Therefore using (11.2) and since $1 - x^2 = 4y(1-y)$, we obtain the identity

$$
z'(F(x)) = c(y(1-y))^{-\lambda}\,\frac{d}{dy}\left[\frac{H_\lambda(y)H_\lambda(1-y)}{(y(1-y))^\tau}\right],
\tag{11.6}
$$

where again, c depends only on τ and λ. Note that the function $z = z(p)$ is symmetric around $1/2$, $z(0^+) = z(1^-) = 0$, and that it is easy to see that $z'(1/2) = 0$. Therefore, to prove that z is concave on $(0,1)$, it suffices to show the concavity of z on $(0,1/2)$. In other words, it is enough to show that its derivative $z'(p)$ is non–increasing on (0,1/2). Since F is increasing and continuous on $(-1,1)$, this is in turn equivalent to showing that $z'(F(x))$ is non–increasing on $(-1,0)$, i.e., that the right–hand side of (11.6) is non–increasing for $0 < y < 1/2$. Thus, one has:

Lemma 11.2 *The function z is concave on $(0,1)$ if and only if the function*

$$
(x(1-x))^{-\lambda}\frac{d}{dx}\left[\frac{H_\lambda(x)H_\lambda(1-x)}{(x(1-x))^\tau}\right]
$$

is non–increasing on $(0,1/2)$.

We now need some further preparatory work.

Lemma 11.3 *Consider the two functions:*

$$
\frac{H_\lambda(x)H_\lambda(1-x)}{(x(1-x))^{2\lambda+1}}, \quad \frac{H_\lambda(x)H_\lambda(1-x)}{(x(1-x))^{\lambda+1}}.
$$

For $\lambda \geq 1$, the first function decreases on $(0,1/2]$, while for $\lambda \geq 0$, the second one increases on $(0,1/2]$.

Proof. By (11.4), making the change of variables $t = sx$, where $x \in (0,1)$ is fixed, we have

$$H_\lambda(x) \;=\; \int_0^x (t(1-t))^\lambda dt \;=\; x^{\lambda+1}\int_0^1 (s(1-sx))^\lambda ds.$$

Hence,

$$\frac{H_\lambda(x)}{x^{2\lambda+1}} \;=\; \int_0^1 s^\lambda \left(\frac{1}{x} - s\right)^\lambda ds,$$

and therefore

$$\frac{H_\lambda(x)}{x^{2\lambda+1}}\frac{H_\lambda(1-x)}{(1-x)^{2\lambda+1}} \;=\; \int_0^1 t^\lambda \left(\frac{1}{x} - t\right)^\lambda dt \int_0^1 t^\lambda \left(\frac{1}{1-x} - t\right)^\lambda dt \;=\; T(u)T(v),$$

where $u = 1/x$, $v = 1/(1-x)$, $T(u) = \int_0^1 t^\lambda (u-t)^\lambda dt$.

Note that $(u-1)(v-1) = 1$, and that $u = u(x) \geq 2$ is a decreasing function of $x \in (0, 1/2]$. Therefore, replacing u by $u+1$ and v by $v+1$, we get

$$\frac{H_\lambda(x)H_\lambda(1-x)}{(x(1-x))^{2\lambda+1}} \;=\; S(u)S\left(\frac{1}{u}\right),$$

where

$$\begin{aligned} S(u) \;=\; T(u+1) \;&=\; \int_0^1 t^\lambda (u + (1-t))^\lambda dt \\ &=\; \int_0^1 (1-t)^\lambda (u+t)^\lambda dt \\ &=\; \int_0^1 (u+t)^\lambda p(t) dt, \end{aligned}$$

$p(t) = (1-t)^\lambda$. To prove the first part of the lemma, we thus need to show that the function $S(u)S(1/u)$ increases on $[1, +\infty)$. But,

$$\begin{aligned} S(u)S\left(\frac{1}{u}\right) \;&=\; \int_0^1\int_0^1 (u+t)^\lambda \left(\frac{1}{u} + s\right)^\lambda p(t)p(s)dtds \\ &=\; 2\int\int_{0<t<s<1} \left[(u+t)^\lambda \left(\frac{1}{u} + s\right)^\lambda + (u+s)^\lambda \left(\frac{1}{u} + t\right)^\lambda\right] p(t)p(s)dtds. \end{aligned}$$

Consequently, it suffices to show that the function

$$f(u) \;=\; (u+t)^\lambda \left(\frac{1}{u} + s\right)^\lambda + (u+s)^\lambda \left(\frac{1}{u} + t\right)^\lambda$$

which is within the square brackets in the integral, is an increasing function of $u \geq 1$, for any fixed $0 < t < s < 1$. Rewriting $f(u)$ as

$$f(u) \;=\; \left((1+ts) + us + \frac{t}{u}\right)^\lambda + \left((1+ts) + ut + \frac{s}{u}\right)^\lambda,$$

we obtain that its derivative

$$\frac{\lambda}{f'(u)} \;=\; \left((1+ts)+us+\frac{t}{u}\right)^{\lambda-1}\left(s-\frac{t}{u^2}\right)+\left((1+ts)+ut+\frac{s}{u}\right)^{\lambda-1}\left(t-\frac{s}{u^2}\right) \;>\; 0$$

if and only if

$$\left[\frac{(1+ts)+us+\frac{t}{u}}{(1+ts)+ut+\frac{s}{u}}\right]^{\lambda-1} \;>\; -\frac{t-\frac{s}{u^2}}{s-\frac{t}{u^2}}. \tag{11.7}$$

Now, it is easy to see that the right–hand side of (11.7) is strictly less than one and that the expression in the square brackets on the left–hand side of (1.17) is strictly greater than one, when $u > 1$. Indeed, first note that $s - t/u^2 > 0$, because $s > t$. Therefore,

$$-\frac{t-\frac{s}{u^2}}{s-\frac{t}{u^2}} < 1 \;\Leftrightarrow\; -t+\frac{s}{u^2} < s-\frac{t}{u^2} \;\Leftrightarrow\; \frac{s+t}{u^2} < s+t$$

which is, of course, true. Analogously,

$$\frac{(1+ts)+us+\frac{t}{u}}{(1+ts)+ut+\frac{s}{u}} > 1 \Leftrightarrow (1+ts)+us+\frac{t}{u} > (1+ts)+ut+\frac{s}{u} \Leftrightarrow u(s-t) > \frac{s-t}{u}$$

which is also true when $u > 1$. Since $\lambda - 1 \geq 0$, the first part of Lemma 11.3 follows.

To establish the second part of the lemma, let

$$h_\lambda(x) \;=\; \frac{H_\lambda(x)}{x^{\lambda+1}} \;=\; \frac{1}{x^{\lambda+1}}\int_0^x t^\lambda(1-t)^\lambda dt \;=\; \int_0^1 s^\lambda(1-sx)^\lambda ds.$$

A differentiation and an integration by parts give

$$\begin{aligned} h'_\lambda(x) &= -\lambda\int_0^1 s^{\lambda+1}(1-sx)^{\lambda-1}ds \\ &= \frac{1}{x}\int_0^1 s^{\lambda+1}d(1-sx)^\lambda \\ &= \frac{(1-x)^\lambda}{x} \;-\; \frac{\lambda+1}{x}\int_0^1 s^\lambda(1-sx)^\lambda ds. \end{aligned}$$

Therefore, h_λ satifies the following differential equation:

$$h'_\lambda(x) \;=\; \frac{(1-x)^\lambda}{x} \;-\; \frac{\lambda+1}{x}h_\lambda(x). \tag{11.8}$$

By (11.8), the derivative of the second function in Lemma 11.3 is

$$\begin{aligned} & (h_\lambda(x)h_\lambda(1-x))' \\ =\; & h'_\lambda(x)h_\lambda(1-x) - h_\lambda(x)h'_\lambda(1-x) \\ =\; & \left[\frac{(1-x)^\lambda}{x} - \frac{\lambda+1}{x}h_\lambda(x)\right]h_\lambda(1-x) \;-\; \left[\frac{x^\lambda}{1-x} - \frac{\lambda+1}{1-x}h_\lambda(1-x)\right]h_\lambda(x) \\ =\; & \frac{(1-x)^\lambda}{x}h_\lambda(1-x) \;-\; \frac{x^\lambda}{1-x}h_\lambda(x) \;+\; \frac{(\lambda+1)(1-2x)}{x(1-x)}h_\lambda(x)h_\lambda(1-x). \end{aligned}$$

The third term in this last expression is positive, since $1-2x>0$. Hence, to prove that $(h_\lambda(x)h_\lambda(1-x))'>0$, it is enough to show that

$$\frac{(1-x)^\lambda}{x}h_\lambda(1-x) \ \geq \ \frac{x^\lambda}{1-x}h_\lambda(x). \tag{11.9}$$

Multiplying (11.9) by $x(1-x)$, leads to the inequality $(1-x)^{\lambda+1}h_\lambda(1-x)\geq x^{\lambda+1}h_\lambda(x)$, which is equivalent to $H_\lambda(1-x)\geq H_\lambda(x)$. Now, this last inequality holds true since H_λ increases on $(0,1)$ and since by assumption, $0<x\leq 1/2$. Lemma 11.3 is thus proved.

Continuation of the proof of Lemma 11.1.

According to Lemma 11.2, we need show that

$$(x(1-x))^{-\lambda}\frac{d}{dx}\left[\frac{H_\lambda(x)H_\lambda(1-x)}{(x(1-x))^\tau}\right]$$

is non–increasing on $(0,1/2)$. Let $\kappa=x(1-x)$, so that $d\kappa=(1-2x)dx$, and let $\alpha=\lambda-\tau+1$. By assumption, $0\leq\alpha\leq 1$ (note that, in the case of the sphere, i.e., when λ and τ are defined by (11.3), we have $\alpha=1/2$). For $\lambda\geq 0$, by Lemma 11.3, the function

$$V(x) \ = \ \frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^{\lambda+1}}$$

increases on $(0,1/2]$. Rewriting the function in the square brackets above as

$$\frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^\tau} \ = \ \kappa^\alpha V(x),$$

we have for its derivative:

$$\frac{d}{dx}\left[\frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^\tau}\right] \ = \ \alpha\kappa^{\alpha-1}V(x)(1-2x)+\kappa^\alpha V'(x).$$

Therefore,

$$\begin{aligned}(x(1-x))^{-\lambda}\frac{d}{dx}\left[\frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^\tau}\right] \ &= \ \alpha\frac{V(x)}{\kappa^{\lambda+1-\alpha}}(1-2x)+\frac{V'(x)}{\kappa^{\lambda-\alpha}} \\ &= \ \alpha\frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^{2\lambda+2-\alpha}}(1-2x)+\frac{V'(x)}{\kappa^{\lambda-\alpha}}.\end{aligned}$$

For $\lambda\geq 1$, and by Lemma 11.3, the first term in this last expression rewriten as

$$\alpha\ \frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^{2\lambda+1}}\ \frac{1}{\kappa^{1-\alpha}}\ (1-2x),$$

is the product of decreasing functions, and so is a decreasing function on $(0, 1/2]$. To study the second term, we first find the derivative of V. Noting that $H'_\lambda(x) = \kappa^\lambda$, we get:

$$V'(x) \;=\; \frac{H_\lambda(1-x) - H_\lambda(x)}{\kappa} \;-\; (\lambda+1)\frac{H_\lambda(x)H_\lambda(1-x)}{\kappa^\lambda}.$$

From this one concludes that the function

$$\frac{V'(x)}{\kappa} \;=\; \frac{H_\lambda(1-x) - H_\lambda(x)}{\kappa^2} \;-\; (\lambda+1)V(x)$$

decreases as the difference of a decreasing function and, as already proved in Lemma 11.3, of an increasing function. Therefore,

$$\frac{V'(x)}{\kappa^{\lambda-\alpha}} \;=\; \frac{V'(x)}{\kappa}\,\frac{1}{\kappa^{\lambda-\alpha-1}}$$

decreases since by assumption,

$$\lambda - \alpha - 1 \;=\; \tau - 2 \geq 0.$$

Lemma 11.1 is proved.

Lemma 11.4 (the case $n = 4$) *The function $pq/I_{\sigma_4}(p)$, $(q = 1-p)$, is concave on $(0, 1)$.*

Proof. When $n = 4$, $\tau = n - 1/2 = 3/2$, and $\lambda = n - 2/2 = 1$. According to Lemma 11.2, one should verify that

$$(x(1-x))^{-1}\frac{d}{dx}\left[\frac{H_1(x)H_1(1-x)}{(x(1-x))^{3/2}}\right]$$

is non–increasing on $(0, 1/2)$. As above, we set $\kappa = x(1-x)$. By (11.4),

$$H_1(x) \;=\; \int_0^x t(1-t)dt \;=\; \frac{1}{6}x^2(3-2x),$$

hence,

$$36\ H_1(x)H_1(1-x) \;=\; \kappa^2(3-2x)(1+2x) \;=\; \kappa^2(3+4\kappa).$$

Therefore,

$$36\frac{H_1(x)H_1(1-x)}{\kappa^{3/2}} \;=\; 3\kappa^{\frac{1}{2}} + 4\kappa^{\frac{3}{2}},$$

so

$$36\ \frac{d}{dx}\left[\ \frac{H_1(x)H_1(1-x)}{\kappa^{3/2}}\right] \;=\; \left(\frac{3}{2}\kappa^{\frac{-1}{2}} + 6\kappa^{\frac{1}{2}}\right)(1-2x),$$

and

(11.10) $$36\kappa^{-1}\frac{d}{dx}\left[\frac{H_1(x)H_1(1-x)}{\kappa^{3/2}}\right] \;=\; \left(\frac{3}{2}\kappa^{\frac{-3}{2}} + 6\kappa^{\frac{-1}{2}}\right)(1-2x).$$

The right hand side of (11.10) is the product of two non–negative, non–increasing functions on $(0,1/2)$ and Lemma 11.4 is proved.

Lemma 11.5 (the case $n=3$) *The function* $pq/I_{\sigma_3}(p)$, $(q=1-p)$, *is concave on* $(0,1)$.

Proof. When $n=3$, $\tau=n-1/2=1$, $\lambda=n-2/2=1/2$, and we set $H(x)=H_{1/2}(x)$. Again, using Lemma 11.2, we just need to verify that

$$(x(1-x))^{-1/2}\frac{d}{dx}\left[\frac{H(x)H(1-x)}{(x(1-x))}\right]$$

is non–increasing on $(0,1/2)$. By (11.4),

$$H(x) \;=\; \int_0^x \sqrt{t(1-t)}dt. \tag{11.11}$$

Noting that $H'(x)=\kappa^{1/2}$, (where again, $\kappa=x(1-x)$) we have

$$(H(x)H(1-x)\;\kappa^{-1})' \;=\; (H(1-x)-H(x))\kappa^{-1/2} \;-\; H(x)H(1-x)\kappa^{-2}(1-2x),$$

hence

$$g(x)=(\kappa)^{-1/2}\frac{d}{dx}\left[\frac{H(x)H(1-x)}{\kappa}\right] \;=\; (H(1-x)-H(x))\kappa^{-1}-H(x)H(1-x)\kappa^{-5/2}(1-2x),$$

is non–increasing on $(0,1/2)$, if its derivative is non–positive, i.e., if

$$\begin{aligned} g'(x) \;&=\; -2\kappa^{-1/2} \;-\; (H(1-x)-H(x))\kappa^{-2}(1-2x) \\ &\quad -[-2H(x)H(1-x)\kappa^{-5/2}+(H(1-x)-H(x))\kappa^{-2}(1-2x) \\ &\quad +\frac{5}{2}H(x)H(1-x)\kappa^{-7/2}(1-2x)] \;\le\; 0. \end{aligned} \tag{11.12}$$

Multiplying (11.12) by $\kappa^{7/2}$, we obtain the inequality

$$\begin{aligned} 2H(x)H(1-x)\kappa \;\;\le\;\; & 2\kappa^3+2(H(1-x)-H(x))\kappa^{3/2}(1-2x) \\ & +\frac{5}{2}H(x)H(1-x)(1-2x). \end{aligned}$$

Since the middle term on the above right–hand side is positive for $x\in(0,1/2)$, it suffices to show that

$$2H(x)H(1-x)\kappa \;\le\; 2\kappa^3+\frac{5}{2}H(x)H(1-x)(1-2x),$$

i.e., that

$$H(x)H(1-x)(4\kappa-5(1-2x)) \;\le\; 4\kappa^3. \tag{11.13}$$

Now from (11.11) we immediately have, noting that $\sqrt{t(1-t)}$ increases on $(0,1/2)$ and that its maximum on $(0,1)$ is $1/2$, that for all $x \in (0,1/2)$,

$$H(x) \;\le\; x\kappa^{1/2}, \quad H(1-x) \;\le\; \frac{(1-x)}{2}.$$

Therefore, $H(x)H(1-x) \le \kappa^{3/2}/2 \le \kappa/4$. Applying this to (11.13), it is enough to see that

$$4\kappa - 5(1-2x) \;\le\; 16\kappa^2. \tag{11.14}$$

Changing variables ($\kappa = x - x^2 = s/4$, $0 < s < 1$), we have $1 - 2x = \sqrt{1-s}$, and (11.14) takes the form

$$s - 5\sqrt{1-s} \le s^2,$$

i.e., the form $s\sqrt{1-s} \le 5$. Again, this last inequality is trivially true, and Lemma 11.5 is proved.

12 Proof of Theorem 1.12 (the Gaussian case)

The half–spaces $A_p = \{x \in \mathbf{R}^n : x_1 \le t\}$ are extremal in the isoperimetric problem for the standard Gaussian measure γ_n (see [Sud–Tsi] and [Bor]), i.e., the value $\gamma_n(A^h)$ is minimal among all Borel sets $A \subset \mathbf{R}^n$ of measure $\gamma_n(A) \ge p$, if $A = A_p$. The value of t is chosen so that $\gamma_n(A_p) = \gamma_1((-\infty, t]) \equiv \Phi(t) = p$, i.e., $t = \Phi^{-1}(p)$ is the inverse of the distribution function Φ of the standard univariate Gaussian density $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Since,

$$\{x \in \mathbf{R}^n : x_1 \le t\}^h = \{x \in \mathbf{R}^n : x_1 \le t + h\},$$

the isoperimetric function corresponding to γ_n has the form

$$I_{\gamma_n}(p) = \gamma_n^+(A_p) = \varphi(\Phi^{-1}(p)).$$

Note that I_{γ_n} does not depend on the dimension, so it can simply be denoted by I_γ. The properties 1) and 2) of Theorem 1.11 are trivially true for I_γ. To complete the proof of Theorem 1.12, it remains to state

Lemma 12.1 *The function*

$$z(p) \;=\; \frac{pq}{I_\gamma(p)}, \quad q = 1 - p,$$

is concave on $(0,1)$.

Indirect proof. It is easy to check by (1.17)–(1.18) (or (9.3) and (9.5)) that the sequence $I_{\sigma_n}(p)$ of isoperimetric functions corresponding to the uniform distributions on the n–spheres of radius $\rho = \sqrt{n}$ converges pointwise to $I_\gamma(p)$, as $n \to \infty$. Therefore, by Lemma 11.1, pq/I_γ is concave as a limit of concave functions.

Direct proof. Clearly, z is concave if it is concave on $(0, 1/2)$, i.e., if

$$g(x) \;=\; z'(\Phi(x)) \;=\; \frac{(1-2\Phi(x))\varphi(x) + \Phi(x)(1-\Phi(x))x}{\varphi^2(x)}$$

does not increase on $(-\infty, 0)$, i.e., if $g'(x) \le 0$. After differentiating and with the help of the identity $\varphi'(x) = -x\varphi(x)$, this last inequality takes the form

$$u(x) \;=\; (1+2x^2)\Phi(x)(1-\Phi(x)) - \varphi^2(x) + 2x(1-2\Phi(x))\varphi(x) \;\le\; 0.$$

Another differentiation gives

$$u'(x) \;=\; 4x\Phi(x)(1-\Phi(x)) + 4x\varphi(x)(1-\varphi(x)) + 3\varphi(x)(1-2\Phi(x)).$$

Let $v(x) = 4x(1-\varphi(x)) + 3(1-2\Phi(x))$, so that

$$u'(x) \;=\; 4x\Phi(x)(1-\Phi(x)) + \varphi(x)v(x). \tag{12.1}$$

Then,

$$v'(x) = 4 + \varphi(x)(x^2 - 10), \tag{12.2}$$

and $v''(x) = x(12-x^2)\varphi(x) \geq 0$, for $0 \leq x \leq \sqrt{12}$. Therefore, v is convex on $[0, \sqrt{12}]$. In addition, $v(0) = 0$,

$$v'(0) = 4 - 10\varphi(0) = \frac{\sqrt{32\pi} - 10}{\sqrt{2\pi}} > 0,$$

hence $v' > 0$ on $[0, \sqrt{12}]$. By (12.2), $v' > 0$ on $[\sqrt{10}, +\infty)$, thus since v' is even, $v' > 0$ on the whole real line. Since v increases, and $v(0) = 0$, we have $v(x) < 0$ on $(-\infty, 0)$, therefore by (12.1), $u' < 0$ on $(-\infty, 0)$. Consequently, in order to prove that $u \leq 0$ on $(-\infty, 0)$, it is enough to check that $u(-\infty) \leq 0$. To prove this, noting that the middle term in the definition of u is negative, it is in turn enough to show that

$$-x\Phi(x)(1-\Phi(x)) \;\leq\; (1-2\Phi(x))\varphi(x),$$

for all $x < 0$ with $|x|$ large enough. But this follows from the well–known asymptotic expansion

$$\frac{-x\Phi(x)}{\varphi(x)} \;=\; 1 - \frac{1}{x^2} + \mathrm{O}\left(\frac{1}{x^4}\right), \quad x \to -\infty.$$

The direct proof of Lemma 12.1 is obtained.

13 The isoperimetric problem on the real line

In this section we study the isoperimetric problem for a class of "regular" μ. To any Borel measure μ on the real line is associated its distribution function $F(x) = \mu((-\infty, x])$, $x \in \mathbf{R}$. Denote by $\mathcal{F}$ the family of those measure (with distribution functions F) which are concentrated on some (finite or not) interval (a_F, b_F) $(a_F = \inf\{F > 0\}, b_F = \sup\{F < 1\})$ where F is absolutely continuous with continuous, and positive density $f = F'$. Such functions F are strictly increasing on (a_F, b_F), and we introduce the continuous, positive function

$$J_\mu(p) \;=\; f(F^{-1}(p)), \quad 0 < p < 1,$$

where $F^{-1} : (0,1) \longrightarrow (a_F, b_F)$ is the inverse of F restricted to (a_F, b_F). We extend J_μ to $[0,1]$ by putting

$$J_\mu(0) \;=\; J_\mu(1) \;=\; 0.$$

Note that, under a shift transformation, we obtain the measure $\nu(A) = \mu(A+h)$ and that $J_\nu = J_\mu$. So, in the sequel, one can think of μ in terms of J_μ, up to the shift parameter h. The map $\mu \longrightarrow J_\mu$ is a bijection from $\mathcal{F}$ onto the family of all continuous, positive functions on $(0,1)$. If 0 is the median of μ, i.e., if $F(0) = 1/2$, then F is expressed via J_μ as the inverse of the function

$$F^{-1}(p) \;=\; \int_{1/2}^{p} \frac{dt}{J_\mu(t)}, \quad 0 < p < 1, \tag{13.1}$$

where $a_F = F^{-1}(0^+)$, and $b_F = F^{-1}(1^-)$. The isoperimetric function I_μ can be found via J_μ as follows:

Proposition 13.1 *Let $F \in \mathcal{F}$. Then, for any $p \in (0,1)$,*

$$I_\mu(p) \;=\; \inf \sum_{k=1}^{n} \left(J_\mu(p_{2k-1}) + J_\mu(p_{2k})\right), \tag{13.2}$$

where the infimum is taken over all possible $0 \le p_1 < p_2 < \ldots < p_{2n-1} < p_{2n} \le 1$ such that $\sum_{k=1}^{n}(p_{2k} - p_{2k-1}) = p$.

Proof. Put $f(x) = 0$ for $x \in (-\infty, a_F] \cup [b_F, +\infty)$, so that f is continuous on $(-\infty, +\infty)$ except, possibly, at $x = a_F$ or $x = b_F$. Denote by $\mathcal{T}$ the family of all sets $A \subset \mathbf{R}$ which are finite unions $A = \cup_{i=1}^{n} \Delta_i$ of open (finite or not) intervals $\Delta_i = (a_i, b_i)$ with disjoint boundaries. Clearly, for an open (finite or not) interval $\Delta = (a,b)$ with $p_1 = F(a)$ and $p_2 = F(b)$, we have

$$\mu^+(\Delta) \;=\; f(a^-) + f(b^+) \;=\; J_\mu(p_1) + J_\mu(p_2).$$

Therefore, (13.2) just says that in the definition of the isoperimetric function

$$I_\mu(p) \;=\; \inf \mu^+(A), \tag{13.3}$$

the infimum can be taken over all $A \in \mathcal{T}$ of measure p. In fact, this property holds true for any non–atomic μ defined the real line without assuming that $F \in \mathcal{F}$. With the help of this statement, we now claim that an analogous property also holds true for the integral isoperimetric problem, namely, if

$$R_h(p) \;=\; \inf \mu(A^h), \quad h > 0,\; 0 < p < 1, \tag{13.4}$$

where A runs over all Borel measurable $A \subset \mathbf{R}$ of measure p, then the infimum in (13.4) can be restricted to the class $\mathcal{T}$.

To prove this claim, we represent the h–neighbourhood of A,

$$A^h \;=\; \cup_{n\geq 1}(c_n, d_n),$$

as the union of (at most) countably many disjoint open intervals (c_n, d_n). Then, let $a_n = c_n + h$, $b_n = d_n - h$. Clearly, $a_n \leq b_n$, $dist((a_i,b_i),(a_j,b_j)) \geq 2h$ $(i \neq j)$, hence any finite interval $(-c, c)$ contains only finitely many of the (c_n, d_n). In addition, if we take $B = \cup_{n\geq 1}(a_n, b_n)$, then

$$A \subset \cup_{n\geq 1}[a_n, b_n], \text{ and } A^h = B^h.$$

Therefore, using the continuity of F, $\mu(B) \geq p$, $\mu(A^h) = \mu(B^h)$. Now, let

$$B(c) = B \cup (-\infty, -c) \cup (c, +\infty), \quad c > 0,$$

where as noted above, $B(c) \in \mathcal{T}$. In addition,

$$\mu(B(c)) \geq p, \quad \mu(B^h(c)) \longrightarrow \mu(B^h) = \mu(A^h),$$

as $c \to +\infty$. Hence, for any $\epsilon > 0$, there exists $B_1 \in \mathcal{T}$ ($B_1 = B(c)$ with c large enough) such that $\mu(B_1) \geq p$, $\mu(B_1^h) \leq \mu(A^h) + \epsilon$. Decreasing the length of the intervals whose union is B_1, one gets a set $B_2 \subset B_1$ such that $\mu(B_2) = p$, and we also have that $B_2 \in \mathcal{T}$, and that $\mu(B_2^h) \leq \mu(A^h) + \epsilon$. Thus, the infimum in (13.4) can be restricted to the class $\mathcal{T}$.

To complete the proof of the proposition, let us return to the abstract triple (X, d, μ). Let $\mathcal{B}(X)$ be the Borel sets in X, and let $\mathcal{R} \subset \mathcal{B}(X)$ contains sets of any measure. Let

$$R_h^*(p) \;=\; \inf \mu(A^h), \quad I^*(p) \;=\; \inf \mu^+(A), \quad (h > 0,\; 0 < p < 1),$$

where the infima are taken over all $A \in \mathcal{R}$. In particular, when $\mathcal{R} = \mathcal{B}(X)$, we have $R_h^*(p) = R_h(p)$, and $I^*(p) = I_\mu(p)$. Then, it directly follows from above that

$$I^*(p) \;=\; \liminf_{h \to 0^+} \frac{R_h^*(p) - p}{h}. \tag{13.5}$$

We therefore have the following statement: *if for all $h > 0$,*

$$R_h^*(p) = R_h(p), \quad then \quad I^*(p) = I_\mu(p).$$

Now, on the real line, taking $\mathcal{R} = \mathcal{T}$ completes the proof.

Remark 13.2 In a similar way, one can substitute the real variable h in (13.5) with a rational variable r. An important consequence of this possible substitution, is the fact that the isoperimetric function I_μ is always Borel measurable on $(0,1)$ as the limit of a sequence of measurable functions (note that $R_h(p)$ is a non–decreasing function of p).

Remark 13.3 It can also be proved (as in Proposition 13.1, and this will be used in the proof of Theorem 14.1), that if λ is a non–atomic finite Borel probability measure on the real line, then

$$\inf_{\mu(A)=p\in(0,1)} \lambda^+(A) = \inf_{A\in\mathcal{T},\mu(A)=p\in(0,1)} \lambda^+(A)$$

Let us return to the real line $\mathbf{R}$ equipped with the measure μ whose distribution function F belongs to the class $\mathcal{F}$. As shown above, the infimum (13.3) can be taken over all $A \in \mathcal{T}$ of μ-measure p. Note that the interior of $\mathbf{R} \setminus A$ belongs to $\mathcal{T}$, and that

$$\mu(\mathbf{R} \setminus A) = 1 - p, \qquad \mu^+(\mathbf{R} \setminus A) = p.$$

Therefore, the isoperimetric function I_μ is symmetric around $1/2$, i.e., for all $p \in (0,1)$,

$$I_\mu(1-p) \;=\; I_\mu(p).$$

Since

$$I_\mu(p) \;\leq\; J_\mu(p), \tag{13.6}$$

we then have that

$$I_\mu(p) \;\leq\; \min\{J_\mu(p), J_\mu(1-p)\}. \tag{13.7}$$

A natural question arising from (13.6) and (13.7) is: "does there exist necessary and sufficient conditions, in terms of F or of J_μ, to have $I_\mu = J_\mu$?" Equivalently, "when are the intervals $(-\infty, x]$ extremal in the isoperimetric problem (13.3)?" In turn this last statement is equivalent to the extremality of these same intervals in the "integral" problem (13.4) (i.e., to the isoperimetric property of the intervals $(-\infty, x]$).

Proposition 13.4 *Let $F \in \mathcal{F}$. The following properties are equivalent:*

a) for any $p \in (0,1)$, the infimum (13.3) is attained at the interval $(-\infty, x]$, where $x = F^{-1}(p)$, and then $I_\mu = J_\mu$;

b) for any $p \in (0,1)$ and $h > 0$, the infimum (13.4) is attained at the interval $(-\infty, x]$, where $x = F^{-1}(p)$, and then $R_h(p) = F(F^{-1}(p) + h)$;

c) the measure μ is symmetric around its median, i.e., J_μ is symmetric around $1/2$, and for all $p, q > 0$ such that $p + q < 1$,

$$J_\mu(p+q) \ \leq \ J_\mu(p) + J_\mu(q). \tag{13.8}$$

Before coming to the proof of the above equivalences, let us state two corollaries. For the first, one can easily verify using (13.1), that μ is log–concave if and only if J_μ is concave on $(0,1)$. But concave functions clearly satisfy (13.8).

Corollary 13.5 *If the measure μ is log-concave, i.e, if the function $\log f$ is concave on (a_F, b_F), and if μ is symmetric around its median, then μ satisfies the above conditions.*

Corollary 13.6 *Let the measure μ with $F \in \mathcal{F}$ satisfies one of the above conditions a), b), or c), then there exists a positive constant c such that,*

$$J_\mu(p) \ \geq \ c \min\{p, 1-p\}, \tag{13.9}$$

whenever $p \in (0,1)$, that is for all $x \in (a_F, b_F)$,

$$\min\{F(x), 1-F(x)\} \ \leq \ \frac{1}{c} f(x).$$

Equivalently, the increasing map $U : \mathbf{R} \longrightarrow (a_F, b_F)$, which transforms the two–sided exponential distribution ν, of density $f_\nu(x) = d\nu(x)/dx = (\exp -|x|)/2$, $x \in \mathbf{R}$, into μ (i.e., $\nu U^{-1} = \mu$), is a Lipschitz function, of Lipschitz constant at most $1/c$. In particular, the "tails" $F(-h)$, $1 - F(h)$ have at least an exponential rate of decrease (as $h \to +\infty$).

Proof of Corollary 13.6. We first note that since J_μ is continuous, (13.8) extends to countable families, that is, for any sequence $p_n > 0$ such that $\sum_n p_n < 1$,

$$J_\mu(\sum_n p_n) \ \leq \ \sum_n J_\mu(p_n). \tag{13.10}$$

Since J_μ is symmetric, positive and continuous, it suffices to show that

$$c \ = \ \liminf_{p \to 0^+} J_\mu(p)/p \ > \ 0.$$

From this, (13.9) follows with, perhaps, a smaller constant. To prove that this lim inf is positive, assume the contrary. Then, for any $c > 0$, the set $S(c)$ of points $p \in (0,1)$ with $J_\mu(p) \leq cp$ is at least countable, and moreover, $0 \in \overline{S(c)}$. Hence, for any $p \in (0,1)$, there exists a sequence $p_n \in S(c)$ such that $\sum_n p_n = p$. By (13.10), we then have $J_\mu(p) \le cp$. Since $c > 0$ can be arbitrarily small, this gives $J_\mu = 0$, which is impossible. To complete the proof of the corollary, let F_ν be the distribution function of the double exponential ν, i.e., let

$$F_\nu(x) = \begin{cases} \frac{\exp x}{2}, & \text{if } x < 0 \\ 1 - \frac{\exp(-x)}{2}, & \text{if } x \ge 0. \end{cases}$$

Then, by a direct application of the definition,

$$J_\nu(p) \ = \ f_\nu(F_\nu^{-1}(p)) \ = \ \min\{p, 1-p\}.$$

Since $U(x) = F^{-1}(F_\nu(x))$, we obtain after differentiation:

$$U'(F_\nu^{-1}(p)) \ = \ \frac{J_\nu(p)}{J_\mu(p)}.$$

Therefore, U is Lipschitz, of Lipschitz constant at most $1/c$, if and only if $J_\mu(p) \ge cJ_\nu(p)$, for all $0 < p < 1$. This coincides with (13.9), and the corollary is proved.

Proof of Proposition 13.4. That a) and b) are equivalent immediately follows from Theorem 2.1 applied to $I = J_\mu$. Assume now that the property c) is fulfilled. Then, for any $0 < p < q < 1$, we have $p + (1-q) < 1$, hence

$$J_\mu(p) + J_\mu(q) \ = \ J_\mu(p) + J_\mu(1-q) \ \ge \ J_\mu(p + (1-q)) \ = \ J_\mu(q-p).$$

These inequalities remain true also when $p = 0$ and/or $q = 1$, so for any $0 \le p < q \le 1$,

$$J_\mu(p) + J_\mu(q) \ \ge \ J_\mu(q-p).$$

Applying this last inequality to (13.2) and using (13.8) with the notations of Proposition 13.1, we get

$$\sum_{k=1}^n \left(J_\mu(p_{2k-1}) + J_\mu(p_{2k})\right) \ \ge \ \sum_{k=1}^n \ J_\mu(p_{2k} - p_{2k-1}) \ \ge \ J_\mu\Big(\sum_{k=1}^n (p_{2k} - p_{2k-1})\Big) \ = \ J_\mu(p).$$

Therefore, by Proposition 13.1, $I_\mu(p) \ge J_\mu(p)$. Together with (13.7), this gives $I_\mu(p) = J_\mu(p)$, i.e., (13.2) is attained when $n = 2$, $p_1 = 0$, $p_2 = p$. This corresponds to the extremal case $A = (-\infty, x]$ with $x = F^{-1}(p)$ in (13.3). Thus, c) implies a).

Conversely, assume that a) is true. Let $A = (-\infty, x]$, and $B = [y, +\infty)$ be intervals of μ–measure p, i.e., $x = F^{-1}(p)$, and $y = F^{-1}(1-p)$. By assumption, $\mu^+(A) = f(x) \le \mu^+(B) = f(y)$, that is $J_\mu(p) \le J_\mu(1-p)$. Replacing p by $1-p$, we get the opposite inequality, and combining these two inequalities gives $J_\mu(p) = J_\mu(1-p)$. Consider now in (13.2) the case $n = 1$. Replacing there (since a) is assumed), I_μ by J_μ, leads to

$$J_\mu(p) \ \le \ J_\mu(p_1) + J_\mu(p_2) \ = J_\mu(p_1) + J_\mu(1-p_2),$$

whenever $0 \le p_1 < p_2 \le 1$, $p = p_2 - p_1$. Putting $p' = p_1$, $q' = 1 - p_2$, so $p' + q' = p$, turns this last inequality into (13.9), and finishes the proof of the proposition.

Let us try to say something more about the isoperimetric function I_μ for non–log–concave measures. As noted before, I_μ should be symmetric around 1/2. In addition, by Proposition 13.1, we have that if a symmetric (around 1/2) function I on $(0,1)$, satisfies (13.8) and is majorized by J_μ, then $I_\mu \ge I$. This might naturally inspire the conjecture that the isoperimetric function I_μ is maximal among all the functions I which satisfy the condition c) in Proposition 13.4 and which are majorized by J_μ. This is not so. To see that, we consider the isoperimetric problem for unimodal distributions.

Assume $F \in \mathcal{F}$. We say that the measure μ (or its distribution function F) is unimodal, if for some $x_0 \in [a_F, b_F]$, the density f is non–decreasing on the interval (a_F, x_0), and is non–increasing on the interval (x_0, b_F). An equivalent wording for unimodality is: for some $p_0 \in [0,1]$, the function J_μ is non–decreasing on the interval $(0, p_0)$, and non–increasing on the interval $(p_0, 1)$. In this definition, f is allowed to be monotone, say, non–increasing on (a_F, b_F). In this case, $x_0 = a_F$, and $p_0 = 0$, so the left intervals (a_F, x_0) and $(0, p_0)$ are empty. This is the case, for example, of the standard (one-sided) exponential distribution μ of density $f(x) = \exp(-x)$, $x > 0$, for which $J_\mu(p) = 1 - p$, whenever $0 < p < 1$.

Note that the class of unimodal distributions is not larger than the class defined in Proposition 13.3 c). Take for example, the measure μ with $J_\mu(p) = \max\{p, 1-p\}$, and, for simplicity, $x_0 = 0$. From (13.1), we find that its density is: $f(x) = (\exp|x|)/2$, $|x| < \log 2$. So, f decreases on $(-\log 2, 0)$ and increases on $(0, \log 2)$, hence μ is not unimodal. On the other hand, for this measure, (13.8) is fulfilled.

Proposition 13.7 *Let $F \in \mathcal{F}$ be unimodal. Then, for any $p \in (0,1)$, the infimum (13.3) is attained either at an interval (finite or not) $A = (a,b)$ of measure p, or at the complement $A = \mathbf{R} \setminus (a,b)$ of an interval of measure $1-p$. Therefore, for all $p \in (0,1)$,*

$$I_\mu(p) \ = \ \inf\,(J_\mu(p_1) + J_\mu(p_2)), \tag{13.11}$$

where the infimum is taken over all possible $0 \le p_1 < p_2 \le 1$ such that $p_2 - p_1 = p$ and such that $p_2 - p_1 = 1 - p$.

We give the proof of this proposition at the end of the section and provide now some important partial cases where (13.11) simplifies (the proofs of these partial results are also given at the end of the section). Again, let $F \in \mathcal{F}$ be the distribution function of a unimodal measure μ. As usual, f denotes a continuous density of μ. Note also that log–concave distributions are unimodal.

Corollary 13.8 *If the measure μ is log–concave, or if f is monotone on (a_F, b_F), then for all $p \in (0,1)$,*

$$I_\mu(p) \;=\; \min\{J_\mu(p), J_\mu(1-p)\}. \tag{13.12}$$

In other words, the infimum (13.3) is attained either at the interval $(-\infty, x]$, or at the interval $[x, +\infty)$ of measure p (in either case).

Thus, for symmetric log–concave μ, (13.12) recovers the statement of Corollary 13.5. For general (not necessarily symmetric) log–concave μ, we get equality in (13.7).

Remark 13.9 It can be shown that if μ satisfies either one of the assumptions of Corollary 13.8, then the last statement of Corollary 13.8 can be extended to the "integral" isoperimetric problem: *the infimum (13.4) is attained either at the interval $(-\infty, x]$, or at the interval $[x, +\infty)$ of measure p (in either case).*

Corollary 13.10 *Let $F \in \mathcal{F}$ be unimodal and let also assume that the measure μ is symmetric around a point x_0 (which should be its median and mode simultaneously). Then, for all $p \in (0,1)$,*

$$I_\mu(p) \;=\; \min\left\{J_\mu(p), 2J_\mu\left(\frac{\min(p, 1-p)}{2}\right)\right\}. \tag{13.13}$$

This means that the infimum (13.3) is attained either at $(-\infty, x]$, or at $(x_0 - h, x_0 + h)$, or at $(-\infty, x_0 - g) \cup (x_0 + g, +\infty)$, of measure p (in all the cases).

Examples and comments

1) As already noted, for the two–sided exponential distribution ν, of density $f_\nu(x) = (\exp -|x|)/2$, $x \in \mathbf{R}$, we have $J_\nu(p) = \min\{p, 1-p\}$. Since ν is log–concave and symmetric around 0, by Corollary 13.5, $I_\nu(p) = J_\nu(p)$. For the standard (one–sided) exponential distribution of density $f(x) = \exp(-x)$, $x > 0$, we have as also noted before, $J_\mu(p) = 1-p$, $0 < p < 1$. By Corollary 13.8,

$$I_\mu(p) \;=\; \min\{J_\mu(p), J_\mu(1-p)\} \;=\; \min\{p, 1-p\}.$$

Therefore, the one–sided and the two–sided exponential distributions have the same isoperimetric function. On the other hand, the solutions to the integral isoperimetric problem differ for these distributions. Indeed, one may apply Proposition 13.3 b) to ν and get: for any $h > 0$, $0 < p < 1$,

$$R_{h,\nu}(p) \;=\; \inf_{\nu(A)=p} \nu(A^h) \;=\; F_\nu(F_\nu^{-1}(p) + h).$$

A direct calculation of the right–hand side above gives, putting also $\alpha = \exp(-h)$:

$$R_{h,\nu}(p) \;=\; \begin{cases} p/\alpha, & \text{if } p \le \alpha/2 \\ 1-\alpha/(4p), & \text{if } \alpha/2 \le p \le 1/2 \\ 1-\alpha(1-p), & \text{if } p \ge 1/2 \end{cases}$$

Applying Remark 13.9 to the measure μ of density $f(x) = \exp(-x)$, $x > 0$

$$R_{h,\mu}(p) \;=\; \inf_{\mu(A)=p} \mu(A^h) \;=\; \min\{F_\mu(F_\mu^{-1}(p)+h), 1-F_\mu(F_\mu^{-1}(p)-h)\}.$$

Another calculation (setting also $\alpha = \exp(-h)$, gives

$$R_{h,\mu}(p) \;=\; \begin{cases} p/\alpha, & \text{if } p \le \alpha/2 \\ 1-\alpha(1-p), & \text{if } p \ge \alpha/2 \end{cases}$$

Hence, $R_{h,\nu}(p) \le R_{h,\mu}(p)$ with strict inequality when $\alpha/2 < p < 1/2$.

2) Another interesting example is provided by a probability measure μ of density

$$f(x) \;=\; \frac{1}{(|x|+2)^2}.$$

Clearly, μ is unimodal and symmetric around 0, so one can apply Corollary 13.10 in order to find the isoperimetric function I_μ. The distribution function of μ is $F(x) = (x+1)/(x+2)$, for $x \ge 0$, and an easy calculation shows that $J_\mu(p) = (1-p)^2$, for $p \ge 1/2$. Since J_μ is symmetric around $1/2$, we get

$$J_\mu(p) \;=\; \min\{p^2, (1-p)^2\}, \quad 0 < p < 1.$$

Applying (13.13), we finally find

$$I_\mu(p) \;=\; \frac{1}{2}\min\{p^2, (1-p)^2\} \;=\; \frac{1}{2}J_\mu(p).$$

Note that, whenever $0 < p < 1$, the interval $(-\infty, x]$, of measure p, is not an extremal set in (13.3). Note also that there does not exist a positive function I which would be majorized by J_μ and would satisfy the condition c) in Proposition 13.4: otherwise, by Corollary 13.6, $1-F(x)$ would decrease exponentially to zero at infinity.

Proof of Proposition 13.7. Set $J = J_\mu$, and for $\Delta = [p,q)$ define $J(\Delta) = J(p) + J(q)$, $0 \le p < q \le 1$, while for $p = q$ put $J(\emptyset) = 0$. Then, (13.2) can be written as

$$I_\mu(p) \;=\; \inf \sum_{k=1}^{n} J(\Delta_k), \tag{13.14}$$

where the infimum is taken over finite unions of disjoint intervals $\Delta_k \subset [0,1)$, $1 \le k \le n$, of total length p. Note that since I_μ is symmetric around $1/2$, one can also take in (13.14) sets of total length $1-p$.

By assumption, for some $p_0 \in [0,1]$, the function J is non–decreasing on the interval $[0,p_0]$, and non–increasing on the interval $[p_0,1]$. Take all the intervals Δ_k, $k \in V \subset \{1,\dots,n\}$, which are situated on the left of p_0. Then, by the above,

$$\sum_{k\in V} J(\Delta_k) \;\ge\; J(p_1), \tag{13.15}$$

where p_1 is the total length of the Δ_k, $k \in V$. In the same way, if we take all the Δ_k, $k \in W \subset \{1,\dots,n\}$, which are situated on the right of p_0, and denote by p_2 their total lengths, then

$$\sum_{k\in W} J(\Delta_k) \;\ge\; J(1-p_2). \tag{13.16}$$

Therefore, all the Δ_k, $k \in V$ (resp., $k \in W$), can be substituted in (13.14) by a single interval $[0,p_1)$ (resp., $[1-p_2,1)$). Note that one of the sets V or W is empty when $p_0 = 0$, or $p_0 = 1$. In addition, at most one of the Δ_k covers the point p_0. Therefore, to minimize (13.14), one needs only consider unions of three disjoint intervals, more precisely:

$$\Delta_1 = [0,p_1), \quad \Delta_2 = [1-p_2,1), \quad \Delta_3 = [p_3,p_4),$$

where in general, $0 \le p_1 < p_3 \le p_0 \le p_4 < 1-p_2 \le 1$, $p_1 + p_2 + (p_4 - p_3) = p$ (and/or $= 1-p$.)

The middle interval Δ_3 can be excluded from our considerations by putting $p_3 = p_4$ (this is explained in a short while). Then (excluding Δ_3), the right side of (13.14) becomes

$$J(\Delta_1) + J(\Delta_2) = J(p_1) + J(1-p_2),$$

under the assumption $p_1 + p_2 = p$ (and/or $= 1-p$), i.e., under the assumption $(1-p_2) - p_1 = 1-p$ (and/or $= p$), and $p_1 \le p_0 \le 1-p_2$. This gives the right–hand side of (13.11) taking into account the following remark: In (13.11), the case $0 < p_1 < p_2 < 1$, where also p_1 and p_2 lie on the same side of p_0 is non–extremal. Indeed, let $0 < p_1 < p_2 \le p_0$ (of course, $p_2 - p_1 = p$ (or, $= 1-p$)). Then, since J is non–decreasing on $[0,p_0]$, we have

$$J(p_1) + J(p_2) \;\ge J(p_2 - p_1).$$

Thus, the pair $(0, p_2 - p_1)$ is "better" than the pair (p_1,p_2) which is thus non–extremal.

It remains to explain why the middle interval Δ_3 can be excluded from our considerations. Let $0 \le p_1 < p_3 \le p_0 \le p_4 < 1-p_2 \le 1$ and $p_3 < p_4$. Then, take the complementary intervals $\Delta_4 = [p_1,p_3)$ and $\Delta_5 = [p_4, 1-p_2)$. Δ_4 is situated on the left of p_0, while Δ_5 is on the right of p_0, their total length is $1-p$ (and/or p) with, finally,

$$J(\Delta_4) + J(\Delta_5) = J(\Delta_1) + J(\Delta_2) + J(\Delta_3). \tag{13.17}$$

By (13.15) and (13.16), these two intervals can be respectively replaced by $[0, p_3 - p_1)$ and by $[1 - p_2 - p_4, 1)$, so the left, hence the right hand–side of (13.17) can be decreased. Thus, we decrease the number of intervals (from three to two), and therefore complete the proof.

Proof of Corollary 13.8. The function $g_p(t) = J_\mu(t) + J_\mu(t+p)$, $0 \le t \le 1-p$, attains its minimum at one of end points $t = 0$ or $t = 1 - p$, because g_p is concave when μ is log–concave (in which case J_μ is concave), and g_p is monotone when f is monotone (in which case J_μ is monotone).

Proof of Corollary 13.10. It suffices to consider the case $p \le 1/2$. If $t \ge 1/2$, then since J_μ is non–increasing on $[1/2, 1]$, g_p attains its minimum on $[1/2, 1-p]$ at $t = 1-p$, and that minimum is equal to $J_\mu(1-p) = J_\mu(p)$. In the same way, g_p attains its minimum on $[0, 1/2 - p]$ at $t = 0$, and that minimum is equal to $J_\mu(p)$. In case $1/2 - p \le t \le 1/2$, we use another representation for g_p, namely,

$$g_p(t) = J_\mu(t) + J_\mu(1 - (t+p)).$$

Clearly, g_p attains its minimum on the middle interval at t where $t = 1 - (t+p)$, i.e., $t = (1-p)/2$, and the minimum is equal to $2J_\mu((1-p)/2)$. Since by (13.11), $I_\mu(p) = \min\{\inf_t g_p(t), \inf_t g_{1-p}(t)\}$, we obtain (13.13).

14 Isoperimetry and Sobolev–type inequalities on the real line

Here we return to Theorem 1.1 in order to better understand the situation on the real line. Again, given a non–atomic probability measure μ on $\mathbf{R}$, and a Young function N, we estimate the Orlicz–norm of $g - m(g)$, where g is a "smooth" function defined on $\mathbf{R}$, with μ–mean $m(g) = \int_{\mathbf{R}} g d\mu$, in terms of the first power of their derivative g' :

$$\|g - m(g)\|_N \ \le \ c_\mu(N) \int_{-\infty}^{+\infty} |g'| d\mu. \tag{14.1}$$

If one wishes to find an "optimal" N (which satisfies (1.36) when (14.1) becomes equivalent to the isoperimetric problem for μ), then it is necessary (at least formally) to find the isoperimetric function I_μ and the results of the previous section can be used. Moreover, the optimal constant c in (14.1), is given by:

$$c_\mu(N) \ = \ \sup_{0<p<1} \frac{I_N(p)}{I_\mu(p)}, \tag{14.2}$$

where $I_N(p) = \|\chi_A - p\|_N$, and where $A \subset \mathbf{R}$ has μ–measure p, i.e., $x_p = 1/I_N(p)$ is the positive solution of

$$pN((1-p)x_p) + (1-p)N(px_p) = 1. \tag{14.3}$$

Recall that such functions I_N are completely characterized by the conditions 1)–3) of Theorem 1.11. When $N(x) = |x|^\alpha$, $1 \le \alpha < +\infty$, the function I_N is simply I_α and (14.2) is just (1.14). Note also that I_N is determined by N but not by μ.

According to (14.2), in order to find $c_\mu(N)$, we have to solve the isoperimetric problem for μ so as to find I_μ. In fact, as explained below, for μ absolutely continuous there is no need to do so, since for such probability measures the isoperimetric function I_μ in (14.2) can be replaced by the function $J_\mu(p) = f(F^{-1}(p))$. Moreover, (14.1) can be essentially improved if one wishes to find the minimal "weight" w such that

$$\|g - m(g)\|_N \ \le \ \int_{-\infty}^{+\infty} |g'(x)| w(x) dx. \tag{14.4}$$

Below, and as usual, $F(x) = \mu((-\infty, x])$ is the distribution function of μ which is continuous on $\mathbf{R}$, since μ is.

Theorem 14.1 *For any locally Lipschitz function g,*

$$\|g - m(g)\|_N \ \le \ \int_{-\infty}^{+\infty} |g'(x)| I_N(F(x)) dx. \tag{14.5}$$

More precisely: if the right–hand side of (14.5) is finite, then $g \in L_N(\mathbf{R}, \mu)$, hence g is μ–integrable, and (14.5) holds. Moreover, the function $w(x) = I_N(F(x))$ is minimal (for the pointwise order and up to a set of Lebesgue measure zero) among all the locally integrable (with respect to the Lebesgue measure) functions w which satisfy (14.4), for all locally Lipschitz g.

Before proceeding to the proof of this theorem, let us present some examples and consequences.

Example 14.2 When $N(x) = |x|^\alpha$, $1 \le \alpha < +\infty$, (14.5) takes the form

$$\left(\int_{-\infty}^{+\infty} |g(x) - m(g)|^\alpha dF(x)\right)^{1/\alpha} \ \le \ \int_{-\infty}^{+\infty} |g'(x)| I_\alpha(F(x)) dx. \tag{14.6}$$

In particular, for $\alpha = 1$ and $\alpha = 2$, we respectively have $I_1(p) = 2p(1-p)$, $I_2(p) = \sqrt{p(1-p)}$, and (14.6) becomes

$$\int_{-\infty}^{+\infty} |g(x) - m(g)| dF(x) \ \le \ 2\int_{-\infty}^{+\infty} |g'(x)| F(x)(1 - F(x)) dx. \tag{14.7}$$

$$\sqrt{\int_{-\infty}^{+\infty} |g(x) - m(g)|^2 dF(x)} \ \le \ \int_{-\infty}^{+\infty} |g'(x)| \sqrt{F(x)(1 - F(x))} dx. \tag{14.8}$$

When $\alpha = +\infty$, then $I_\infty(p) = \max(p, 1-p)$, and we get

$$\|g - m(g)\|_\infty \le \int_{-\infty}^{+\infty} |g'(x)| \max(F(x), (1 - F(x))) dx.$$

Theorem 14.1 also allows to find the optimal constant in (14.1) $c_\mu(N)$ provided μ is absolutely continuous with respect to the Lebesgue measure. Indeed,

Proposition 14.3 *Assume that μ is absolutely continuous with density f, and that (14.1) is satisfied. Then,*

$$c_\mu(N) \ = \ \operatorname*{ess\,sup}_{x \in \mathbf{R}} \frac{I_N(F(x))}{f(x)}. \tag{14.9}$$

Therefore, $c_\mu(N) < +\infty$ if and only if

$$d_\mu(N) \ = \ \operatorname*{ess\,inf}_{x \in \mathbf{R}} f(x) N^{-1}\left(\frac{1}{F(x)(1 - F(x))}\right) \ > \ 0, \tag{14.10}$$

where $N^{-1} : [0, +\infty) \longrightarrow [0, +\infty)$ is the inverse of N restricted to $[0, +\infty)$.

It is reasonable to give another wording for the above statement by comparing μ to some canonical distribution as this was done in the last part of Corollary 13.6. Here we assume for a while that f is continuous and positive on an interval (a_F, b_F) (finite or not) where μ is concentrated. Thus, in the terminology of Section 13, $F \in \mathcal{F}$.

Any Young function N corresponds to a (symmetric around 0) probability measure μ_N, defined via the identity $J_{\mu_N} = I_N$. Its distribution function F_N and density f_N can tentatively be found via (13.1). As already observed in Section 2, when $N(x) = |x|$, we have

$$F_N(x) = \frac{1}{1+\exp(-2x)}, \quad x \in \mathbf{R}.$$

Likewise when $N(x) = |x|^2$, we have

$$F_N(x) = \frac{(1+\sin x)}{2}, \quad |x| < \pi/2.$$

A big difference between these two important examples is that the second distribution has compact support. In general, put

$$b_N \;=\; \int_{1/2}^{1} \frac{1}{I_N(t)} dt, \quad a_N = -b_N,$$

so that μ_N is concentrated on (a_N, b_N). With these notations, and if μ is concentrated on (a_F, b_F), we state:

Proposition 14.4 *Let the probability measure μ have distribution function $F \in \mathcal{F}$. Then, $c_\mu(N) < +\infty$, if and only if the increasing map $U : (a_N, b_N) \longrightarrow (a_F, b_F)$, which transforms μ_N into μ, is Lipschitz, of Lipschitz constant at most $c_\mu(N)$.*

From this proposition, and if $F \in \mathcal{F}$, we also conclude:

Corollary 14.5 *Suppose that $b_N < +\infty$, that is*

$$\int_2^{+\infty} \frac{N^{-1}(t)}{t^2} dt \;<\; +\infty.$$

If $c_\mu(N) < +\infty$, then μ has compact support.

In particular, the probability distributions μ with $F \in \mathcal{F}$ and for which (14.1) holds true for the Lebesgue spaces $L_\alpha(\mathbf{R}, \mu)$, $\alpha > 1$, and for a finite constant $c_\alpha(N)$, are concentrated on finite intervals.

Remark 14.6 Of course, in Proposition 14.4, the characterizing property of $c_\mu(N) < +\infty$, can also be expressed via any probability distributions λ, different but equivalent to μ_N, in that:

$$0 < \inf_{0<p<1} \frac{J_\lambda(p)}{I_N(p)} \leq \sup_{0<p<1} \frac{J_\lambda(p)}{I_N(p)} < +\infty.$$

For example, when $N(x) = |x|$, μ_N can be replaced in the statement of the proposition (changing also the Lipschitz constant) by the two–sided exponential distribution ν for which $J_\nu(p) = \min\{p, 1-p\}$.

Proof of Theorem 14.1

Step 1: In this step, (14.5) is proved for all g bounded and Lipschitz, provided also that

$$\int_{-\infty}^{+\infty} I_N(F(x))dx \ < \ +\infty.$$

By this assumption, the measure ν of density $d\nu(x)/dx = I_N(F(x))$ is finite. Therefore, by Theorem 1.1 with $\mathcal{L}(g) = \|g - m(g)\|_N$, the inequality (14.5) holds for all g bounded and Lipschitz, if and only if, for any Borel measurable set $A \subset \mathbf{R}$, with $\mu(A) = p$,

$$\nu^+(A) \ \geq \ \|\chi_A - p\|_N = I_N(p). \tag{14.11}$$

According to Remark 13.3, to prove this, it suffices to check (14.11) for the sets A which are finite unions of open intervals $\Delta_1, \dots, \Delta_n$ and whose boundaries are disjoint. Note that the function I_N possesses the following property: for any $p, q \geq 0$, such that $p+q \leq 1$,

$$I_N(p+q) \ \leq I_N(p) + I_N(q). \tag{14.12}$$

Indeed, taking disjoint sets A and B of respective measure p and q, we have, by the very definition of I_N:

$$I_N(p+q) = \|\chi_{A\cup B} - (p+q)\|_N \leq \|\chi_A - p\|_N + \|\chi_B - q\|_N = I_N(p) + I_N(q).$$

Now, assume for a while that (14.11) has already been shown for open intervals. Then, we will have for the sets A described above, applying (14.12) to a finite sum of $p_i = \mu(\Delta_i)$ such that $\sum_{i=1}^n p_i = p$:

$$\nu^+(A) = \sum_{i=1}^n \nu^+(\Delta_i) \geq \sum_{i=1}^n I_N(p_i) \geq I_N(\sum_{i=1}^n p_i) = \|\chi_A - p\|_N = I_N(p).$$

So, (14.11) holds for all A, provided it is true for the intervals $A = (a,b)$, $-\infty \leq a < b \leq +\infty$. For such intervals A,

$$\nu^+(A) = I_N(F(a)) + I_N(F(b)).$$

Using $I_N(1-p) = I_N(p)$ and again (14.12), we finally get (note that $\mu(A) = F(b) - F(a) = p$):

$$\begin{aligned} I_N(F(a)) + I_N(F(b)) &= I_N(F(a)) + I_N(1 - F(b)) \\ &\geq I_N(1 - (F(b) - F(a))) \\ &= I_N(F(b) - F(a)) \\ &= I_N(p). \end{aligned}$$

Step 2: In this step, (14.5) is proved for all Lipschitz g which are constant outside a finite interval, say $[a,b]$, and no assumption on the finiteness of ν is made.

Take a sequence μ_n of Borel probability measures on $\mathbf{R}$ whith compact supports, which converge weakly to μ, i.e., such that for any bounded, continuous function h,

$$\int_{\mathbf{R}} h d\mu_n \longrightarrow \int_{\mathbf{R}} h d\mu,$$

as $n \to \infty$. Then, applying the results of step 1 to g and μ_n:

$$\|g - \int_{\mathbf{R}} g d\mu_n\|_{L_N(\mu_n)} \ \leq \ \lambda_n = \int_{-\infty}^{+\infty} |g'(x)| I_N(F_n(x)) dx, \tag{14.13}$$

where F_n is the distribution function of μ_n. Since μ is non–atomic, F is continuous and F_n converges to F pointwise. Hence $I(F_n(x)) \longrightarrow I(F(x))$ for all $x \in \mathbf{R}$, as $n \to \infty$. Since I_N is bounded on $[0,1]$, $|g'|$ is bounded by its Lipschitz seminorm, and since the right integral in (14.3) is taken over $[a,b]$, the Lebesgue dominated convergence theorem gives

$$\lambda_n \longrightarrow \lambda = \int_{-\infty}^{+\infty} |g'(x)| I_N(F(x)) dx.$$

Let us now rewrite (14.13) in the form

$$\int_{\mathbf{R}} N\left(\frac{g - m_n}{\lambda_n}\right) d\mu_n \ \leq \ 1, \tag{14.14}$$

where $m_n = \int_{\mathbf{R}} g d\mu_n$. Since g is bounded and continuous, $m_n \to m(g)$, as $n \to \infty$. In order to take the limit in (14.14), we use the following property of weak convergence (see Billingsley [Bil, Theorem 5.5]): if $\mu_n \to \mu$ weakly and if h_n is a uniformly bounded sequence of continuous functions such that $h_n(x_n) \to h(x)$ whenever $x_n \to x$, where h is a continuous function, then

$$\int_{\mathbf{R}} h_n d\mu_n \longrightarrow \int_{\mathbf{R}} h d\mu,$$

as $n \to \infty$. Applying this result to $h_n = N((g - m_n)/\lambda_n)$, $h = N((g - m(g))/\lambda)$, we get (assuming that $\lambda > 0$):

$$\int_{\mathbf{R}} N\left(\frac{g - m(g)}{\lambda}\right) d\mu \ \leq \ 1,$$

that is

$$\|g - m(g)\|_{L_N(\mu)} \ \leq \ \lambda.$$

This last inequality coincides with (14.5). The case $\lambda = 0$ is trivial: g becomes constant on (a_F, b_F), and the left–hand side in (14.5) is zero.

Step 3: In this step, (14.5) is extended to all g locally Lipschitz. Before verifying this claim, we first note that if the right–hand side of (14.5) is finite, then g is μ–integrable and moreover, $g \in L_N(\mathbf{R}, \mu)$. Indeed, if g is locally Lipschitz, define the functions

$$g_n(x) = \begin{cases} g(x), & \text{if } a_n \le x \le b_n \\ g(b_n), & \text{if } x \ge b_n \\ g(a_n), & \text{if } x \le a_n \end{cases}$$

where $a_n \to -\infty$, $b_n \to +\infty$, as $n \to \infty$. Clearly, g_n is Lipschitz and constant outside of $[a_n, b_n]$, so applying the result of step 2 and writing (14.5) for g_n we get:

$$\|g_n - \int_{\mathbf{R}} g_n d\mu\|_N \ \le \ \int_{a_n}^{b_n} |g'(x)| I_N(F(x)) dx,$$

and therefore,

$$\|g_n - \int_{\mathbf{R}} g_n d\mu\|_N \ \le \ \lambda = \int_{-\infty}^{+\infty} |g'(x)| I_N(F(x)) dx. \tag{14.15}$$

Assume now that $\lambda < +\infty$. The space $L_1(\mu)$ is the largest of all the Orlicz spaces, moreover there exists a constant $A = A(N)$ such that, for all $g \in L_N(\mu)$,

$$\|g\|_1 \ \le \ A\|g\|_N.$$

Hence, (14.15) implies

$$\|g_n - m(g_n)\|_1 \ \le \ A\lambda.$$

Therefore, estimating $|g_n(x) - g_n(y)|$ via $|g_n(x) - m(g_n)| + |g_n(y) - m(g_n)|$, we obtain

$$\int_{\mathbf{R}} \int_{\mathbf{R}} |g_n(x) - g_n(y)| d\mu(x) d\mu(y) \le 2A\lambda. \tag{14.16}$$

Applying Fatou's lemma to the left–hand side of (14.16) gives

$$\int_{\mathbf{R}} \int_{\mathbf{R}} |g(x) - g(y)| d\mu(x) d\mu(y) \le 2A\lambda,$$

and g is μ–integrable. Now,

$$m(g_n) = \int_{a_n}^{b_n} g(x) d\mu(x) + g(a_n)\mu((-\infty, a_n]) + g(b_n)\mu([b_n, +\infty)). \tag{14.17}$$

Since g is μ–integrable, the first term on the right–hand side of (14.17) converges to $m(g)$. The second and the third terms will also converge to 0, if a_n and b_n are chosen in an appropriate way. The existence of such appropriate sequences follows from

$$\inf_{x>b} |g(x)| \mu([b, +\infty)) \ \le \ \int_b^{+\infty} |g(x)| d\mu(x) \to 0 \quad (b \to +\infty),$$

$$\inf_{x<a} |g(x)| \mu((-\infty, a]) \ \le \ \int_{-\infty}^{a} |g(x)| d\mu(x) \to 0 \quad (a \to -\infty).$$

Thus, $m(g_n) \to m(g)$. Let us rewrite (14.15), assuming again that $\lambda > 0$:

$$\int_{\mathbf{R}} N\left(\frac{g_n - m(g_n)}{\lambda}\right) d\mu \ \le \ 1. \tag{14.18}$$

Applying, once more, Fatou's lemma to (14.18), we finally get

$$\int_{\mathbf{R}} N\left(\frac{g - m(g)}{\lambda}\right) d\mu \ \le \ 1,$$

that is, $g - m(g) \in L_N(\mu)$, and moreover, $\|g - m(g)\|_N \le \lambda$. This coincides with (14.5). For $\lambda = 0$, the result is trivially true and Step 3 is complete.

Step 4: In this step, we prove the minimality (for the pointwise order and up to a set of Lebesgue measure zero) of $I_N(F(x))$ among all the locally integrable (with respect to the Lebesgue measure) functions w which satisfy (14.4) for all locally Lipschitz g.

Approximating the indicator function $g = \chi_{(-\infty,x)}$, $x \in \mathbf{R}$, by Lipschitz functions, (14.4) yields

$$I_N(F(x)) = \|g - m(g)\|_N \ \le \ w_1(x), \tag{14.19}$$

where

$$w_1(x) = \liminf_{\epsilon \to 0^+} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} w(t)dt.$$

Since $w_1(x) = w(x)$ almost everywhere, the result follows.

Proof of Proposition 14.3. The identity (14.9) follows directly from the statement on the minimal weight in Theorem 14.1. Then, from (14.3), one can easily obtain two–sided estimates for x_p and show that, for any $p \in (0,1)$, $q = 1 - p$,

$$N^{-1}\left(\frac{1}{2pq}\right) \ \le \ x_p \ \le \ 2N^{-1}\left(\frac{1}{2pq}\right). \tag{14.20}$$

Indeed, assuming that $0 < p \le 1/2$, so that $2q \ge 1$, it follows from (14.3) that $pN(qx_p) \le 1$, hence

$$x_p \le \frac{1}{q} N^{-1}\left(\frac{1}{p}\right) \le 2N^{-1}\left(\frac{1}{2pq}\right),$$

since N^{-1} is concave. On the other hand, let $T(x) = N(x)/x$, $x > 0$. Since N is convex, T is non–decreasing. Again by (14.3),

$$1 = pqx_p(T(qx_p) + T(px_p)) \le 2pqx_pT(x_p) = 2pqN(x_p),$$

and the left inequality in (14.20) follows. Since $I_N(p) = 1/x_p$, (14.10) is equivalent to (14.9), and Proposition 14.3 is proved.

Proof of Proposition 14.4. It suffices to note that

$$U'(F^{-1}(p)) = \frac{I_N(p)}{J_\mu(p)}.$$

Proof of Corollary 14.5. Since N^{-1} is concave and increasing, by (14.20), $N^{-1}(1/2pq)$ behaves, up to a constant, like $N^{-1}(1/p)$, as $p \to 0^+$. Therefore, b_N is finite if and only if the integral

$$\int_0^{1/2} N^{-1}\left(\frac{1}{p}\right) dp$$

is finite. The change of variables $t = 1/p$ and Proposition 14.4 finish the proof, since the Lipschitz image of a compactly supported measure is also compactly supported.

15 Extensions of Sobolev–type inequalities to product measures on $\mathbf{R}^n$

For the space $L^1(\mu)$, the inequality (14.1) is easily extends, by induction on the dimension n, to product measures $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ whose marginals distributions are absolutely continuous. Let f_i be a density (with respect to the Lebesgue measure) of μ_i, again we are looking for conditions on μ for the existence of a constant c such that

$$\int_{\mathbf{R}^n} |g - m(g)| d\mu \le c \int_{\mathbf{R}^n} |\nabla g| d\mu, \tag{15.1}$$

for all locally Lipschitz functions f on $\mathbf{R}^n$. Note that $|\nabla g|$ depends on the metric d on $\mathbf{R}^n$. Of course, the metric dependence is only important in finding the optimal c in (15.1). Below, we find the optimal constant in case d is the ℓ^1–metric in $\mathbf{R}^n$, i.e., $d_1(x, y) = \sum_{i=1}^n |x_i - y_i|$.

Let us denote by $F_i(x) = \mu_i((-\infty, x])$ the distribution function of μ_i, and let also

$$c(\mu) = \max_{1 \le i \le n} \operatorname{ess\,sup}_{x \in \mathbf{R}} \frac{2F_i(x)(1 - F_i(x))}{f_i(x)}. \tag{15.2}$$

Proposition 15.1 *The inequality (15.1) is satisfied for some $c < +\infty$ and all bounded Lipschitz functions if and only if $c(\mu) < +\infty$. If it is so, then (15.1) holds for all locally Lipschitz function g in the following sense: if the right hand side of (15.1) is finite then g is μ–integrable, and moreover (15.1) holds. In addition, for the ℓ^1–metric in $\mathbf{R}^n$, $c(\mu)$ is optimal.*

Proof. Taking in (15.1) bounded Lipschitz functions $g(x) = g(x_i)$ which only depend on the ith variable, (15.1) reduces to (14.1) in $L_1(\mu_i)$. Therefore by Proposition 14.3, the condition

$$c_i = \operatorname{ess\,sup}_{x \in \mathbf{R}} \frac{2F_i(x)(1 - F_i(x))}{f_i(x)} < +\infty,$$

is necessary and in addition $c \ge \max_{1 \le i \le n} c_i$. Let now assume that $c(\mu) < +\infty$. We prove the result by induction on the dimension. By Theorem 14.1 and Proposition 14.3, the result holds for $n = 1$. Then, let g be a bounded Lipschitz function on $\mathbf{R}^n$ and assume that (15.1) is true for the dimension $n - 1, n \ge 2$. Fixing $x_n \in \mathbf{R}$ and applying (15.1) to the function $h(x_1, \cdots, x_{n-1}) = g(x_1, \cdots, x_{n-1}, x_n)$ and to the measure $\nu = \mu_1 \otimes \cdots \otimes \mu_{n-1}$ we get:

$$\int_{\mathbf{R}^{n-1}} |h| d\nu \le \left| \int_{\mathbf{R}^{n-1}} h d\nu \right| + \max_{1 \le i \le n-1} c_i \int_{\mathbf{R}^{n-1}} |\nabla h| d\nu. \tag{15.3}$$

Note that the function h is Lipschitz, hence differentiable almost everywhere (with respect to the Lebesgue measure). So,

$$|\nabla h| = \left|\frac{\partial g}{\partial x_1}\right| + \cdots + \left|\frac{\partial g}{\partial x_{n-1}}\right|,$$

for almost all $(x_1, \cdots, x_{n-1}) \in \mathbf{R}^{n-1}$. Let us now introduce the function $\varphi(x_n) = \int_{\mathbf{R}^{n-1}} h d\nu$. Clearly, φ is Lipschitz, so again, one can apply (14.1) for $L_1(\mathbf{R}, \mu_n)$ and get:

$$\int_{\mathbf{R}} |\varphi| d\mu_n \le \left|\int_{\mathbf{R}} \varphi d\mu_n\right| + c_n \int_{\mathbf{R}} |\varphi'| d\mu_n. \tag{15.4}$$

Integrating (15.3) with respect to μ_n over $\mathbf{R}$, using (15.4) and noting that $\int_{\mathbf{R}} \varphi d\mu_n = \int_{\mathbf{R}^n} g d\mu$, we have by Fubini's Theorem

$$\int_{\mathbf{R}^n} |g| d\mu \le \left|\int_{\mathbf{R}^n} g d\mu\right| + c_n \int_{\mathbf{R}} |\varphi'| d\mu_n + \max_{1\le i\le n-1} c_i \int_{\mathbf{R}^n} \sum_{1\le i\le n-1} \left|\frac{\partial g}{\partial x_i}\right| d\mu. \tag{15.5}$$

It now remains to note that $\varphi'(x_n) = \int_{\mathbf{R}^{n-1}} \frac{\partial g}{\partial x_n} d\nu$, hence

$$|\varphi'(x_n)| \le \int_{\mathbf{R}^{n-1}} \left|\frac{\partial g}{\partial x_n}\right| d\nu,$$

and

$$\int_{\mathbf{R}} |\varphi'(x_n)| d\mu_n(x_n) \le \int_{\mathbf{R}^n} \left|\frac{\partial g}{\partial x_n}\right| d\mu.$$

Since $c_n \le c(\mu)$, $\max_{1\le i\le n-1} c_i \le c(\mu)$, (15.5) gives (15.1) for all bounded Lipschitz g with $c = c(\mu)$. Then, a truncation argument extends (15.1) to all locally Lipschitz g, as stated in the proposition.

Remark 15.2 In a particular case, Proposition 15.1 gives the solution to the isoperimetric problem when $\mathbf{R}^n$ is equipped with the supremum distance $d_\infty(x, y) = \max_{1\le i\le n} |x_i - y_i|$. Indeed, let μ be the n–th power of the logistic distribution $F(x) = 1/(1+\exp(-x))$, $x \in \mathbf{R}$ of density f. Clearly, f is such that $f(x) = F(x)(1 - F(x))$. Hence, by Proposition 15.1, the inequality (15.1) is satisfied with $c = 2$. Thus, by Theorem 1.2 for $\alpha = 1$, (15.1) is equivalent to $\mu^+(A) \ge p(1-p), p = \mu(A)$. But, by Theorem 2.1, this is in turn equivalent to the "integral" inequality $\mu(A^h) \ge p/(p + (1 - p)\exp(-h))$ (see the equivalence of (2.4) and (2.5)) where the enlargement A^h is taken with respect to the d_∞ metric, i.e., $A^h = A + hB_\infty$, where B_∞ is the ℓ_∞ unit ball in $\mathbf{R}^n$. We thus have:

Corollary 15.3 *Let μ be the n–th power of the logisitic distribution. For any $h > 0$, the minimal value of $\mu(A^h)$, when $\mu(A) = p$ is fixed, is attained at a standard half–space $\{x \in \mathbf{R}^n : x_1 \le const.\}$ of measure p.*

Remark 15.4 Proposition 15.1 implies that

$$\int_{\mathbf{R}^n} |g - m(g)| d\mu \leq c(\mu)\sqrt{n} \int_{\mathbf{R}^n} \sqrt{\sum_{i=1}^{n} \left| \frac{\partial g}{\partial x_i} \right|^2} d\mu, \tag{15.6}$$

where now, the constant $c(\mu)\sqrt{n}$ is suboptimal. For the ℓ^1–metric, the optimal constant does not tend to infinity with n. One might thus wonder, even when $\mu_i = \mu_1$ for all i, whether or not, when the gradient is estimated in the Euclidian metric, the optimal constant tends to infinity with n.

Let us rewrite (15.6) with a constant K and the product measure $\mu^n = \mu \otimes \cdots \otimes \mu$:

$$\int_{\mathbf{R}^n} |g - m(g)| d\mu^n \leq 2K \int_{\mathbf{R}^n} \sqrt{\sum_{i=1}^{n} \left| \frac{\partial g}{\partial x_i} \right|^2} d\mu^n, \tag{15.7}$$

where $m(g) = \int_{\mathbf{R}^n} g d\mu^n$. As we know (Theorem 1.2), (15.7) is equivalent to

$$(\mu^n)^+(A) \geq \frac{1}{K} p(1-p), \;\; \mu^n(A) = p. \tag{15.8}$$

In turn, this is equivalent (see (2.5)) to

$$\mu^n(A^h) \geq \frac{p}{p + (1-p)\exp(-h/K)} = R_{\frac{h}{K}}(p), \tag{15.9}$$

where $h > 0$ and where A^h is the Euclidian h–neighboorhood of A. Let us now suppose that μ has a continuous positive density f on (a, b) where μ is concentrated. Then, by Proposition 14.4 and Remark 14.6, (15.1) for $n = 1$, is equivalent to: μ is a Lipschitz image of the two–sided exponential distribution ν. It is thus natural to conjecture that this last requirement on μ (which is necessary) is also a sufficient condition for the validity of (15.7), i.e., of (15.9), in the n–dimensional case and for some constant $K = K(\mu)$ independent of n. It is easy to see that if (15.9) holds for some measure μ, then it holds for all its Lipschitz images μU^{-1}. Therefore the above conjecture is equivalent to (15.9) for $\mu = \nu$. So only one distribution (the double exponential) needs to be consider to solve this problem. To date, the closest form to (15.9) is an inequality due to Talagrand [Tal1, p.95] (see also [Tal2, Chap.2]):

$$\nu^n(A + hB_1 + \sqrt{h}B_2) \geq F_\nu \left(F_\nu^{-1}(p) + \frac{h}{K} \right), \tag{15.10}$$

where K is some universal constant and where B_1 and B_2 are respectively the ℓ_1 and ℓ_2 unit balls in $\mathbf{R}^n$. In Example 1 of Section 13, an expression for the right–hand side of (15.10) was found. On the other hand, the right–hand side of (15.8) is equivalent to $K \min(p, 1-p)$ which is (up to a constant) the isoperimetric function of the one (or two) sided exponential distribution. Hence, the right hand side of (15.10) is equivalent (uniformly in p and h) to $R_{h/K}(p)$, where $R_h(p) = F_\mu(F_\mu^{-1}(p) + h)$, where $F_\mu(x) = 1/(1 + \exp(-x)), x \in \mathbf{R}$ is the logistic distribution. Finally, note that (15.10) is stronger than (15.9) for h large. Unfortunately, for h small (important in estimating the isoperimetric function), (15.10) does not imply (15.9) and in fact becomes weaker. Thus, the question of the existence of necessary and sufficient conditions for the validity of (15.7) with some constant independent of n remains open.

References

[Aub] Aubin, T (1982) Non–linear Analysis on manifolds. Monge Ampère equations. Springer–Verlag, Berlin.

[Bil] Billingsley, P. (1968) Convergence of probability measures. John Wiley, New York.

[Bor] Borell, C. (1975) The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30**, 207-216.

[Bur–Zal] Burago, Y. D., and V. A. Zalgaller (1988). Geometric inequalities. Springer–Verlag, Berlin.

[Ehr] Ehrhard, A. (1983). Symétrisation dans l' espace de Gauss. *Math. Scand.* **53**, 281–301.

[Fed] Federer, H. (1969) *Geometric measure theory.* Springer–Verlag, Berlin.

[Fed–Fle] Federer, H. and W. H. Fleming (1960). Normal and Integral currents. *Ann. Math.* **72**, 458-520.

[Gro] Gromov, M. Paul Lévy's isoperimetric inequality. preprint IHES (1980)

[Kan–Aki] Kantorovich, L.V.K. and Akilov, G.P. (1982) Functional analysis. Pergamon Press, Oxford.

[Led1] Ledoux, M. (1988) Isopérimetrie et inégalités de Sobolev logarithmiques gaussiennes. *C.R. Acad. Sci. Paris* **306**, 79–82.

[Led2] Ledoux, M. (1994) Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. Preprint.

[Lev] Lévy, P. (1951) Problèmes concrets d'analyse fonctionnelle. Gauthier–Villars, Paris. (see also pp 271–273 of Leçons d'analyse fonctionnelle, Gauthier–Villars, Paris (1919)).

[Li] Li, P. (1980) On the Sobolev constant and the p–spectrum of a compact Riemannian manifold. *Ann. scient. Éc. Norm. Sup.* **13**, 451–469.

[Ma1] Maz'ya, V. G. (1960) Classes of domains and imbedding theorems for function spaces. *Dokl. Akad. Nauk. SSSR* **133**, 527-530.

[Ma2] Maz'ya, V. G. (1985) Sobolev Spaces, Springer–Verlag, Berlin.

[Oss] Osserman, R. (1978) The isoperimetric inequality. *Bull. Amer. Math. Soc.* **84**, 1182-1238.

[Pis] Pisier, G. (1986) Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varenna (Italy) 1985. Lecture Notes in Math 1206, 167–241 Springer–Verlag.

[Rot] Rothaus, O. S. (1985) Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities. *J. Func. Anal.* **64**, 296–313.

[Sch] Schmidt, E. (1948) Die Brunn–Minkowskische Ungleichung und Ihr Spiegelbild sowie die isoperimetrische Eingenschaft der Kugel in der euklidischen und nicheuklidischen Geometrie 1. *Math. Nachr.* **1**, 81-157.

[Sud–Tsi] Sudakov, V. N., and B.S. Tsirelson (1974) Extremal properties of half–spaces for spherically invariant measures. *Notes Sci. Seminars Leningrad Branch of Math. Inst.* **41**, 14-24 (English Trans. (1978) J. Soviet. Math. 9, 9–18).

[Tal1] Talagrand, M. (1991) A new isoperimetric inequality and the concentration of measure phenomenon. Israel Seminar (GAFA), Lecture Notes in Math 1469, 94–124 Springer–Verlag.

[Tal2] Talagrand, M. (1994) Concentration of measures and isoperimetric inequalities in product spaces. Preprint.

[Pel–Tal] E. Pellicia and G. Talenti (1993) A proof of a logarithmic Sobolev inequality. *Calc. Var.* **1**, 237–242.

[Yau] Yau, S.-T. (1975) Isoperimetric constants and the first eigenvalue of a compact Riemannian manifold. *Ann. scient. Éc. Norm. Sup.* **8**, 487–507.