Diffusion equations and entropy inequalities

Giuseppe Toscani *

Abstract. The present notes contain the material presented at the *XLI* Summer School on Mathematical Physics (Ravello - September 05 - 10, 2016). The notes aim in introducing a physical way to look at well-known functional inequalities in sharp form. Among them, the classical Young's inequality and its converse in sharp form, Brascamp-Lieb type inequalities, Babenko's inequality and Prékopa-Leindler inequality as well as Shannon's entropy power inequality. These classical inequalities can be obtained by looking at the monotonicity in time of convex functionals (playing the role of entropies) evaluated on solutions to linear and nonlinear diffusion equations. These arguments have been investigated by the author in a number of recent papers [31] [49] [76] [82] [83] [84], [85] [86] [87], where the interested reader can find more details.

*Department of Mathematics, University of Pavia, via Ferrata 1, 27100 Pavia, Italy. giuseppe.toscani@unipv.it

Contents

1 Introduction

In his famous papers on the mathematical theory of communication [77], considered a foundation of modern information theory, Shannon introduced the entropy power inequality. This inequality is very easy to describe. The main concept in it is the notion of entropy.

Given a random vector X in $\mathbb{R}^d$, $d \geq 1$ with probability density $f(x)$, let

$$H(X) = H(f) = -\int_{\mathbb{R}^d} f(x) \log f(x)\, dx \tag{1.1}$$

denote its entropy functional (or Shannon's entropy). The entropy power is then defined by

$$N(X) = N(f) = \exp\left(\frac{2}{d} H(X)\right). \tag{1.2}$$

The entropy power is built to be linear at Gaussian random vectors. Indeed, let $Z_\sigma = \mathcal{N}(0, \sigma I_d)$ denote the d-dimensional Gaussian random vector having mean vector 0 and covariance matrix σI_d, where I_d is the identity matrix. The probability density of Z_σ equals

$$M_\sigma(x) = \frac{1}{(2\pi\sigma)^{d/2}} \exp\left(-\frac{|x|^2}{2\sigma}\right), \tag{1.3}$$

and $N(Z_\sigma) = 2\pi e \sigma$.

Shannon's entropy power inequality (EPI) gives a lower bound on the entropy power of the sum of two independent random variables X, Y in $\mathbb{R}^d$ with densities

$$N(X+Y) \geq N(X) + N(Y), \tag{1.4}$$

with equality if and only X and Y are Gaussian random vectors with proportional covariance matrices.

In [77], Shannon gave a semi-formal proof of (1.4), using a variational argument (cf [77], Appendix 6). The first rigorous proof of (1.4) was obtained ten years later by Stam [78] in one dimension. Then, Stam's proof was simplified and extended to dimension $d > 1$ by Blachman. It was 1965, and many other proofs and extensions followed [40, 55, 56, 75, 84, 92].

In 1985 Costa [40] proposed a stronger version of EPI (1.4), valid for the case in which $Y = Z_t$, a Gaussian random vector independent of X. In this case

$$N(X + Z_t) \geq (1-t) N(X) + t N(X + Z_1), \qquad 0 \leq t \leq 1 \tag{1.5}$$

or, equivalently, $N(X + Z_t)$, is concave in t, i.e.

$$\frac{d^2}{dt^2} N(X + Z_t) \leq 0. \tag{1.6}$$

Note that equality to zero in (1.6) holds if and only if X is a Gaussian random variable, $X = N(0, \sigma I_d)$. In this case, considering that Z_σ and Z_t are independent each other, and Gaussian densities are stable under convolution, $N(Z_\sigma + Z_t) = N(Z_{\sigma+t}) = \sigma + t$, which implies

$$\frac{d^2}{dt^2} N(Z_\sigma + Z_t) = 0. \tag{1.7}$$

The novelty in Stam's proof [78] was to use the heat equation and its solution to connect Shannon entropy (1.1) with Fisher information. For a given random vector X in $\mathbb{R}^d$ with a smooth density $f(x)$, its Fisher information is defined as

$$I(X) = I(f) = \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)}\, dx. \tag{1.8}$$

The main reason is that Fisher information, which is of *quadratic* nature, is more treatable to give bounds on $I(X+Y)$, where X and Y are independent random vectors. Indeed, a reasonably simple and direct proof allows to prove Blachman–Stam inequality [20, 46, 78]. This inequality takes the name exactly from the original papers in which Stam and Blachman gave a rigorous proof of entropy power inequality. It gives a lower bound on the reciprocal of Fisher information of the sum of independent random vectors with (smooth) densities

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}, \tag{1.9}$$

still with equality if and only X and Y are Gaussian random vectors with proportional covariance matrices.

Hence, the heat equation started to be used as a powerful instrument to obtain mathematical inequalities in sharp form in the years between the late fifties to mid sixties. In addition to the pioneering paper by Stam [78], an interesting application of this idea can be found in a paper by Linnik [63]. Stam [78] was motivated by the finding of a rigorous proof of Shannon's entropy power inequality [77], while Linnik [63] used the information measures of Shannon and Fisher in a proof of the central limit theorem of probability theory. However, at the same time in which Blachman [20] presented his proof, still by resorting massively to properties of the linear diffusion equation, similar computations were presented by McKean [69], in connection with the problem of convergence to equilibrium for Kac's caricature of a Maxwell gas. Motivated by proving that for some kinetic system the subsequent derivatives of entropy alternate in sign, he evaluated these derivatives in correspondence to the one-dimensional solution of the heat equation up to the third one, by obtaining sharp inequalities between the second and the first derivative (the Fisher information). It is now clear that these results were very close to obtain the logarithmic Sobolev inequality, proven by Gross ten years later [53], and to show the concavity property of entropy power, proven by Costa twenty years later [40]. Indeed, McKean ideas have been used by Villani to give an alternative proof of the concavity property [91], and by the present author to obtain an improved version of the logarithmic Sobolev inequality [82].

The huge potentialities of the use of the heat equation to prove inequalities have been rediscovered in more recent times by Carlen, Lieb and Loss [29], that first introduced a Lyapunov functional of solutions to the heat equation which allows to prove Young's inequality and its converse for functions of one variable. Later on, Bennett Carbery Christ and Tao [18] were able to extend the result in [29] to general functions. Other very closely-related works can be found in papers of Bennett and Bez [16], Borell [27], Barthe and Cordero-Erausquin [11] and Barthe-Huet [12]. In particular, Young's inequality and its converse have been proven by Bennett and Bez [16] by showing that a suitable functional of the convolution of powers to the solution to the heat the heat equation exhibits monotonicity properties.

As often happens, however, the seminal ideas of Stam [20, 78] remained confined within the framework of information theory, where, however, functional inequalities gained a lot of interest, in reason of their connections with properties of Shannon's and Rényi's entropies [47]. A notable exception to this confinement is a recent paper by Gardner [52], that clarifies the relationship between the Brunn-Minkowski inequality and other inequalities in geometry and analysis. In [52], clear connections between the entropy power inequality of information theory and Young's inequality and others are described in details, together with an exhaustive list of references.

As far as the classical Young's inequality is concerned, the original proof of the sharp form is due to Beckner [14] and Brascamp and Lieb [28]. In [28] Brascamp and Lieb also proved the sharp form of Young inequality also in the so-called reverse case. A different proof of this sharp reverse Young inequality was subsequently done by Barthe [10]. In their recent paper, Young's inequality has been seen in a different light by Bennett and Bez [16] (cf. also [15, 18, 29]). In this paper, Young's inequality is derived by looking at the monotonicity properties of a suitable functional of the convolution of powers to the solution to the heat equation. In this respect, the arguments of [16] are close to the ones presented in systematic form in [84].

The connections of the sharp form of Young's inequality with the Prékopa–Leindler inequality has been enlightened by Brascamp and Lieb [28]. Then, the connection of Young's inequality with Shannon's entropy power inequality has been noticed by Lieb [61].

Most proofs in these papers are based on properties of the solution to the heat equation. Indeed, it is now clear that diffusion equations, linear and nonlinear, constitute a useful tool to obtain inequalities in sharp form [84, 49]. In these notes, we aim in giving a self-contained presentation of this topic.

2 Learning from kinetic theory

2.1 The radiative transfer and Jensen's inequality

The Boltzmann equation is the most famous kinetic model, both due to its important current applications, and for historical reasons. For this model, Boltzmann [26] proved his celebrated H-theorem about the increase of entropy, which represented the first analytical proof ever of the second principle of thermodynamics. In reason of binary collisions between molecules, the gas density is shown to relax towards the Maxwellian equilibrium, a Gaussian density of type (1.3). Equilibria are indeed important in kinetic theory, as the following example shows.

Consider the radiative transfer equation [36, 34]. In the three-dimensional physical space it reads

$$\left(\frac{\partial f}{\partial t} + v \cdot \nabla_x f\right)(x, v, t) = \sigma \left(\int_\Omega f(x, w, t)\, dw - \mathfrak{m}(\Omega) f(x, v, t)\right). \tag{2.1}$$

The unknown is the non-negative specific radiation intensity $f(x, v, t)$ which depends on position $x \in \mathbb{R}^3$, velocity $v \in \Omega$ and time $t \in \mathbb{R}_+$. In general, Ω is a convex subset of $\mathbb{R}^3$ of finite measure $\mathfrak{m}(\Omega)$. The right-hand of equation (2.1) describes the relaxation process, which in this case is a linear operator. Dropping the dependence on the spatial variable x, we can investigate the relaxation process, here driven by the equation

$$\frac{\partial f(v, t)}{\partial t} = \sigma \left(\int_\Omega f(w, t)\, dw - \mathfrak{m}(\Omega) f(v, t)\right). \tag{2.2}$$

Taking the integral over Ω on both sides of (2.2) shows that the mass of the solution is preserved in time,

$$\int_\Omega f(w, t)\, dw = \int_\Omega f_0(w)\, dw, \tag{2.3}$$

so that the relaxation operator can be equivalently written as

$$\frac{\partial f(v, t)}{\partial t} = \sigma \left(\int_\Omega f_0(w)\, dw - \mathfrak{m}(\Omega) f(v, t)\right). \tag{2.4}$$

Thanks to (2.4), it is then immediate to obtain the stationary solution, which appears to be the unique constant function in Ω of mass equal to the mass of the initial distribution

$$f_\infty(v) = \frac{1}{\mathfrak{m}(\Omega)} \int_\Omega f_0(w)\, dw.$$

Equation (2.2) can be easily solved to give, for $v \in \Omega$,

$$f(v, t) = f_0(v) e^{-\bar{\sigma} t} + \left(1 - e^{-\bar{\sigma} t}\right) \frac{1}{\mathfrak{m}(\Omega)} \int_\Omega f_0(v)\, dv. \tag{2.5}$$

where we set

$$\bar{\sigma} = \mathfrak{m}(\Omega)\sigma.$$

Formula (2.5) shows convergence of the solution towards the steady state as time tends to infinity.

It is remarkable that the relaxation process produces the monotonicity of convex functionals of the solution. Given a convex (smooth) function Φ, let $H_\Phi(f)$ denote the Lyapunov functional

$$H_\Phi(f) = \int_\Omega \Phi(f(v))\, dv. \tag{2.6}$$

To evaluate the time evolution of H_Φ let us make use of the radiative transfer equation (2.2). We obtain

$$\begin{aligned} \frac{dH_\phi(f(t))}{dt} &= \int_\Omega \Phi'(f(v,t))\frac{\partial f(v,t)}{\partial t}\, dv \\ &= \sigma \int_\Omega \Phi'(f(v,t)) \left[\int_\Omega f(w,t)\, dw - \mathfrak{m}(\Omega) f(v,t)\right] dv \\ &= \sigma \int_\Omega dw \int_\Omega dv\ \Phi'(f(v,t))\ [f(w,t) - f(v,t)] = -I_\Phi(f(t)). \end{aligned}$$

Exchanging v and w (respectively w and v) in the double integral, we can write I_Φ in the equivalent form

$$\begin{aligned} I_\Phi(f(t)) &= \sigma \int_\Omega dw \int_\Omega dv\ \Phi'(f(v,t))\ [f(v,t) - f(w,t)] \\ &= \sigma \int_\Omega dv \int_\Omega dw\ \Phi'(f(w,t))\ [f(w,t) - f(v,t)]\,. \end{aligned}$$

Using both expressions for I_Φ, we finally obtain

$$I_\Phi(f(t)) = \frac{\sigma}{2} \int_\Omega dw \int_\Omega dv\ \ [\Phi'(f(v,t)) - \Phi'(f(w,t))]\, [f(v,t) - f(w,t)] \geq 0.$$

In fact, the integrand is non-negative as a consequence of the convexity of Φ. Moreover, $I_\Phi(f(t)) = 0$ if and only if $f(v,t)$ is constant on Ω.

Since $I_\Phi \geq 0$,

$$H_\Phi(f(t)) \geq H_\Phi(f_\infty).$$

In particular,

$$\int_\Omega \Phi(f_0(v))\, dv \geq \int_\Omega \Phi(f_\infty(v))\, dv = \int_\Omega \Phi\left(\frac{1}{\mathfrak{m}(\Omega)} \int_\Omega f_0(w)\, dw\right) dv.$$

Since $f_\infty(v)$ is constant on Ω, while $f_0(v)$ is arbitrary, we proved that, given a convex function Φ,

$$\Phi\left(\frac{1}{\mathfrak{m}(\Omega)}\int_\Omega f_0(w)\,dw\right) \le \frac{1}{\mathfrak{m}(\Omega)}\int_\Omega \Phi(f_0(v))\,dv. \tag{2.7}$$

The study of the time evolution of the Lyapunov functional (2.6) along the solution to the radiative transfer equation gives a new physical interpretation of Jensen's inequality [50].

This property is crucial. While theoretical inequality have a universal validity, some of them also present a physical nature, in that they are deeply linked to some relaxation process. This is well understood in kinetic theory, where Shannon's entropy is linked to the Boltzmann and related kinetic equations [80, 81].

2.2 The BGK model and relative entropies

Our second example is concerned with the Bhatnagar-Gross-Krook (BGK) model of gas dynamics, which is a simplified version of the Boltzmann equation [19]. The initial value problem for the space homogeneous version describes relaxation of the probability density $f(x,t)$ with $x \in \mathbb{R}^d$, towards the Gaussian density (1.3)

$$\frac{\partial f(x,t)}{\partial t} = \mu\left(M_\sigma(x) - f(x,t)\right), \tag{2.8}$$

where μ is a relaxation parameter. By fixing $f(x,t=0) = \varphi(x)$, where $\varphi(x)$ is a probability density with a certain number of moments bounded, it is immediate to recover the solution to (2.8), which results to be a linear combination of the initial and of the Gaussian densities

$$f(x,t) = \varphi(x)e^{-\mu t} + M_\sigma(x)\left(1 - e^{-\mu t}\right). \tag{2.9}$$

Hence, in particular, $f(x,t)$ is a probability density for all $t \ge 0$. Rewriting (2.9) as

$$f(x,t) - M_\sigma(x) = \left(\varphi(x) - M_\sigma(x)\right)\, e^{-\mu t} \tag{2.10}$$

then shows convergence of the solution to M_σ in all norms $\|\cdot\|$ such that the difference $\|\varphi - M_\sigma\|$ is bounded.

Let $\Phi(r)$, $r \ge 0$ be a convex function such that $\Phi'(1) = 0$. Then, for any $t > 0$

$$\frac{d}{dt}\int_{\mathbb{R}^d} \Phi\left(\frac{f(x,t)}{M_\sigma(x)}\right) M_\sigma(x)\,dx = \int_{\mathbb{R}^d} \Phi'\left(\frac{f(x,t)}{M_\sigma(x)}\right) \frac{\partial f(x,t)}{\partial t}\,dx =$$

$$\mu \int_{\mathbb{R}^d} \Phi'\left(\frac{f(x,t)}{M_\sigma(x)}\right) \left(M_\sigma(x) - f(x,t)\right)\,dx =$$

$$-\mu \int_{\mathbb{R}^d} \left[\Phi' \left(\frac{f(x,t)}{M_\sigma(x)} \right) - \Phi'(1) \right] \left(\frac{f(x,t)}{M_\sigma(x)} - 1 \right) M_\sigma(x)\, dx \le 0.$$

Therefore the quantity $\int \Phi(f(t)/M_\sigma)M_\sigma$ is decreasing towards zero, which implies

$$\int_{\mathbb{R}^d} \Phi \left(\frac{\varphi(x)}{M_\sigma(x)} \right) M_\sigma(x)\, dx \ge 0 \tag{2.11}$$

Equation (2.8) is a particular case of a relaxation process towards a general probability density $g(x) > 0$, given by

$$\frac{\partial f(x,t)}{\partial t} = \mu \left(g(x) - f(x,t) \right), \tag{2.12}$$

with solution

$$f(x,t) = \varphi(x) e^{-\mu t} + g(x) \left(1 - e^{-\mu t} \right). \tag{2.13}$$

Repeating step-by-step the previous reasoning, one concludes with the inequality

$$\int_{\mathbb{R}^d} \Phi \left(\frac{f(x)}{g(x)} \right) g(x)\, dx \ge 0, \tag{2.14}$$

valid for any pair of probability densities, and convex function Φ satisfying $\Phi'(1) = 0$. The quantity in (2.14) is usually known with the name of *relative* Φ-*entropy* [1]. The choice $\Phi(r) = r \log r - r + 1$ gives rise to the relative Shannon's entropy

$$H(f|g) = \int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{g(x)}\, dx. \tag{2.15}$$

Let $g = M_\sigma$. Then, if $\int |x|^2 f(x) \le \int |x|^2 M_\sigma$ one has

$$\int_{\mathbb{R}^d} f(x) \log M_\sigma\, dx = -\frac{d}{2} \log 2\pi\sigma \int_{\mathbb{R}^d} f(x)\, dx - \frac{1}{2\sigma} \int_{\mathbb{R}^d} |x|^2 f(x)\, dx \ge$$

$$-\frac{d}{2} \log 2\pi\sigma \int_{\mathbb{R}^d} M_\sigma(x)\, dx - \frac{1}{2\sigma} \int_{\mathbb{R}^d} |x|^2 M_\sigma(x)\, dx.$$

Under this condition

$$0 \le H(f|M_\sigma) = \int_{\mathbb{R}^d} f(x) \log f(x)\, dx - \int_{\mathbb{R}^d} f(x) \log M_\sigma(x)\, dx \le$$

$$\int_{\mathbb{R}^d} f(x) \log f(x)\, dx - \int_{\mathbb{R}^d} M_\sigma(x) \log M_\sigma(x)\, dx,$$

or, what is the same

$$H(f) \le H(M_\sigma) \tag{2.16}$$

whenever

$$\int_{\mathbb{R}^d} |x|^2 f(x)\, dx \le \int_{\mathbb{R}^d} |x|^2 M_\sigma(x)\, dx.$$

This result is known under the name of Gibbs lemma. We will be back on this result later on Section 5.2.1.

2.3 Fokker-Planck equation and Gaussian densities

The Fokker-Planck equation is a partial differential equation describing the time evolution of a density function $f(v,t)$, where $v \in \mathbb{R}^d$, $d \geq 1$ and $t \geq 0$, departing from a nonnegative initial density $\varphi(v)$. The standard assumptions on $\varphi(v)$ is that it possesses finite mass ρ, mean velocity u and temperature θ, where for any given density $g(v)$

$$\rho = \int_{\mathbb{R}^d} g(v)\, dv \tag{2.17}$$

is the mass density,

$$u = \frac{1}{\rho} \int_{\mathbb{R}^d} v g(v)\, dv \tag{2.18}$$

is the mean velocity, and θ is the temperature defined by

$$\theta = \frac{1}{n\rho} \int_{\mathbb{R}^d} |v-u|^2 g(v)\, dv. \tag{2.19}$$

The Fokker-Planck equation appears in many different contexts. It was originally derived for the distribution function of a Brownian particle in a fluid [35], and is applicable in a more general form to a plasma [37]. In normal form, the Fokker-Planck equation reads

$$\frac{\partial f}{\partial t} = J_{FP}(f)(v,t) = \sigma \Delta f(v,t) + \nabla \cdot (v f(v,t)). \tag{2.20}$$

The Fokker-Planck operator J_{FP} has the usual conservation property of mass, and $\int logf\, J_{FP}(f)\, dv < 0$, which guarantees the increasing in time of Shannon's entropy (1.1). If the initial condition $\varphi(v)$ is a probability density, the equilibrium solution is the Maxwellian density (1.3). Indeed, the stationary solution (of mass 1) of equation (2.20) satisfies the equation in divergence form

$$J_{FP}(f)(v) = \nabla \cdot (\sigma \nabla f(v) + v f(v)) = 0.$$

Since

$$\sigma \nabla f(v) + v f(v) = f(v) \nabla \left(\sigma \log f(v) - \frac{|v|^2}{2} \right),$$

on the set $f(v) > 0$

$$\log f(v) - \frac{|v|^2}{2\sigma} = 0,$$

which implies $f(v) = c M_\sigma(v)$, with c constant. Then $c = 1$ fixes the mass equal to one.

Given a probability density $h(x)$, $x \in \mathbb{R}^d$, we define its *Fourier transform* $\widehat{f}(\xi)$, $\xi \in \mathbb{R}^d$ by

$$\hat{h}(\xi) = \int_{\mathbb{R}^d} e^{-i\xi \cdot x} h(x)\, dx.$$

In terms of Fourier transform, the Fokker-Planck equation reads

$$\frac{\partial \hat{f}}{\partial t} = \widehat{J}_{FP}(\hat{f})(\xi, t) = -\sigma|\xi|^2 \hat{f}(\xi, t) - \xi \cdot \nabla \hat{f}(\xi, t). \tag{2.21}$$

Thus the Fourier transform of the stationary solution (of mass 1) of equation (2.21) satisfies

$$\widehat{J}_{FP}(\hat{f})(\xi) = -\sigma|\xi|^2 \hat{f}(\xi) - \xi \cdot \nabla \hat{f}(\xi) = 0,$$

Since

$$\sigma|\xi|^2 \hat{f}(\xi) + \xi \cdot \nabla \hat{f}(\xi) = \hat{f}(\xi)\, \xi \cdot \left(\sigma\xi + \nabla \log |\hat{f}(\xi)| \right),$$

on the set $\xi\, \hat{f}(\xi) \neq 0$

$$\log |\hat{f}(\xi)| - \frac{\sigma|\xi|^2}{2} = 0,$$

which implies, by fixing the mass of $f(v)$ equal to one

$$\hat{f}(\xi) = \widehat{M_\sigma}(\xi) = \exp\left\{ -\sigma \frac{|\xi|^2}{2} \right\}. \tag{2.22}$$

Hence, by means of the study of the stationary solution of the Fokker–Planck equation, we can conclude that the Gaussian density (1.3) has an explicit Fourier transform, given by (2.22).

3 Some facts about the linear diffusion equation

3.1 Scaling and convergence

We begin by recalling some properties of the solution to the heat equation in $\mathbb{R}^d$, $d \geq 1$

$$\frac{\partial u(x,t)}{\partial t} = \kappa \Delta u(x,t), \tag{3.1}$$

where $\kappa > 0$ is the (constant) diffusion coefficient. In the rest of the paper, for the sake of simplicity we will assume that the initial datum is a probability density function $f(x)$, so that $f(x) \geq 0$, and

$$\int_{\mathbb{R}^d} f(x)\, dx = 1. \tag{3.2}$$

This assumption will not affect the generality of the results that follow. The solution to equation (3.1) is given by the function $u(x,t) = f * M_{2\kappa t}(x)$, convolution of the initial datum with the fundamental solution $M_{2\kappa t}$, where $M_\sigma(x)$, for $\sigma > 0$, denotes the Gaussian density in $\mathbb{R}^d$ of mean 0 and variance $d\sigma$

The easiest way to recover the solution to the initial value problem for equation (3.1) is to resort to Fourier transform. In fact, by passing to Fourier transform in (3.1) we obtain

$$\frac{\partial \hat{u}(\xi,t)}{\partial t} = -\kappa|\xi|^2\hat{u}(\xi,t), \tag{3.3}$$

that can be easily integrated by separation of variables to give, for every $t \geq 0$

$$\hat{u}(\xi,t) = \hat{f}(\xi)\exp\{-\kappa|\xi|^2 t\}. \tag{3.4}$$

Hence, owing to the result of Section 2.3, the Fourier transform of the solution is found to coincide with the product of the Fourier transforms of the initial datum $\hat{f}$ and the the Fourier transform $\widehat{M_{2\kappa t}}$ of the Gaussian density (1.3) of mean zero and variance $2\kappa dt$, usually known as fundamental or source-type solution, namely the convolution product specified above. Hence, in terms of random vectors, the solution $u(x,t)$ to the heat equation (3.1) coincides with the density function of the sum of two independent random vectors $X + Z_{2\kappa t}$, where X has density $f(x)$ and $Z_{2\kappa t}$ is the Gaussian random vector with density given by (1.3), with $\sigma = 2\kappa t$.

It is important to remark that formula (3.4) admits a simple generalization. Even if somewhat unusual, we will briefly introduce it, since it is closely related to the original proof of the entropy power inequality given by Stam [78]. Consider the heat equation (3.1) with a time-dependent diffusion coefficient $\kappa(t) = \mu'(t)$, where $\mu(t)$ is an increasing function such that $\mu(0) = 0$ (the classical case corresponds to $\mu(t) = \kappa t$). Then, equation (3.3) can be integrated as well by separation of variables, and the corresponding solution is

$$\hat{u}(\xi,t) = \hat{f}(\xi)\exp\{-\mu(t)|\xi|^2\}. \tag{3.5}$$

Hence, in terms of random vectors, the solution $u(x,t)$ to the heat equation (3.1) with a time-dependent coefficient of diffusion $\kappa(t) = \mu'(t)$, where $\mu(t)$ is increasing from $\mu(0) = 0$, coincides with the density function of the sum of two independent random vectors $X + Z_{2\mu(t)}$, where X has density $f(x)$ and $Z_{2\mu(t)}$ is the Gaussian random vector with density given by (1.3), with $\sigma = 2\mu(t)$. We do not insist more on this standard result.

Maybe not so well-known is that we can make use of formula (3.4) to control the large–time behaviour of the solution to equation (3.1) in various norms. To this aim, let us consider a family of metrics that has been introduced in the paper [51] to study the trend to equilibrium of solutions to the space homogeneous Boltzmann equation for Maxwell molecules, and subsequently applied to a variety of problems related to kinetic models of Maxwell type. For a more detailed description, we address the interested reader to the lecture notes [33].

Given $s > 0$ and two random vectors X_1, X_2 in $\mathbb{R}^d$ with probability distributions

f_1 and f_2, their Fourier based distance $d_s(X_1, X_2)$ is given by the quantity

$$d_s(X_1, X_2) = d_s(f_1, f_2) := \sup_{\xi \in \mathbb{R}^d} \frac{\left|\widehat{f}_1(\xi) - \widehat{f}_2(\xi)\right|}{|\xi|^s}.$$

The distance is finite, provided that X_1 and X_2 have the same moments up to order $[s]$, where, if $s \notin \mathbb{N}$, $[s]$ denotes the entire part of s, or up to order $s-1$ if $s \in \mathbb{N}$. Moreover d_s is an ideal metric. Its main properties are the following

1. Let X_1, X_2, X_3, with X_3 independent of the pair X_1, X_2 be random vectors with probability distributions f_1, f_2, f_3. Then

$$d_s(X_1 + X_3, X_2 + X_3) = d_s(f_1 * f_3, f_2 * f_3) \le d_s(f_1, f_2) = d_s(X_1, X_2);$$

2. Define for a given nonnegative constant a the mass-preserving dilation in $\mathbb{R}^d$

$$f_a(x) = \frac{1}{a^d} f\left(\frac{v}{a}\right). \tag{3.6}$$

   Then, given two random vectors X_1, X_2 with probability distributions f_1 and f_2, for any nonnegative constant a

$$d_s(aX_1, aX_2) = d_s(f_{1,a}, f_{2,a}) \le a^s\, d_s(f_1, f_2) = a^s\, d_s(X_1, X_2).$$

Dilations of functions will be considered often in this work. In particular, the property of invariance under dilation will be essential in the proof of many results. The following definition clarifies this concept.

Definition 1. For a given probability density $f(x)$, $x \in \mathbb{R}^d$, let $G(f)$ denote a functional acting on f. We will say that G is *invariant under dilation* if, for any positive constant a

$$G(f_a) = G(f), \tag{3.7}$$

where f_a is defined as in (3.6).

Define, for $t > 0$

$$U(x,t) = \left(\sqrt{1 + 2\kappa t}\right)^d u(x\sqrt{1 + 2\kappa t}, t). \tag{3.8}$$

Then, $U(x,t)$ is a mass-preserving time-dependent dilation of the solution $u(x,t)$, so that

$$\int_{\mathbb{R}^d} U(x,t)\,dx = \int_{\mathbb{R}^d} u(x,t)\,dx = 1. \tag{3.9}$$

Then formula (3.4) implies that, for all $t \geq 0$

$$\widehat{U}(\xi, t) = \hat{u}\left(\frac{\xi}{\sqrt{1+2\kappa t}}, t\right) = \hat{f}\left(\frac{\xi}{\sqrt{1+2\kappa t}}\right) \exp\left\{-\frac{|\xi|^2}{2}\frac{2\kappa t}{1+2\kappa t}\right\}. \tag{3.10}$$

On the other hand, if $M(x) = M_1(x)$ denotes the Gaussian density defined in (1.3) corresponding to $\sigma = 1$, for each $t > 0$ we have the identity

$$\widehat{M}(\xi) = \exp\left\{-\frac{|\xi|^2}{2}\right\} = \widehat{M}\left(\frac{\xi}{\sqrt{1+2\kappa t}}\right) \exp\left\{-\frac{|\xi|^2}{2}\frac{2\kappa t}{1+2\kappa t}\right\}. \tag{3.11}$$

Therefore

$$|\widehat{U}(\xi, t) - \widehat{M}(\xi)| = \left|\hat{f}\left(\frac{\xi}{\sqrt{1+2\kappa t}}\right) - \widehat{M}\left(\frac{\xi}{\sqrt{1+2\kappa t}}\right)\right| \exp\left\{-\frac{|\xi|^2}{2}\frac{2\kappa t}{1+2\kappa t}\right\}.$$

Let us suppose now that the distance $d_s(f, M)$ between the initial datum and the Gaussian M is bounded for some $s > 0$. We remark that the distance is certainly bounded for $s \leq 1$ since both f and M are probability densities. In this case, for any positive constant a and for any $\xi \neq 0$, property 2 of the Fourier based distance implies

$$\left|\hat{f}(a\xi) - \widehat{M}(a\xi)\right| = \frac{\left|\hat{f}(a\xi) - \widehat{M}(a\xi)\right|}{|\xi|^s} \cdot |\xi|^s \leq d_s(f, M)\, a^s |\xi|^s. \tag{3.12}$$

If now $a = 1/\sqrt{1+2\kappa t}$ we obtain the bound

$$|\widehat{U}(\xi, t) - \widehat{M}(\xi)| \leq \left(\frac{1}{\sqrt{1+2\kappa t}}\right)^s d_s(f, M)\, |\xi|^s \exp\left\{-\frac{|\xi|^2}{2}\frac{2\kappa t}{1+2\kappa t}\right\}. \tag{3.13}$$

The bound (3.13) can be directly applied to evaluate convergence of $U(x, t)$ to $M(x)$ as $t \to \infty$ in various norms. As main example, let us consider the L^2 norm. In Fourier variables, for any given $t > 0$

$$\int_{\mathbb{R}^d} |\widehat{U}(\xi, t) - \widehat{M}(\xi)|^2\, d\xi \leq d_s(f, M)^2 \int_{\mathbb{R}^d} \left(\frac{|\xi|}{\sqrt{1+2\kappa t}}\right)^{2s} \exp\left\{-|\xi|^2 \frac{2\kappa t}{1+2\kappa t}\right\} d\xi =$$
$$\left(\frac{1}{2\kappa t}\right)^s \left(\frac{1+2\kappa t}{2\kappa t}\right)^{n/2} d_s(f, M)^2 \int_{\mathbb{R}^d} |\eta|^{2s} \exp\left\{-|\eta|^2\right\} d\eta. \tag{3.14}$$

Thus, since the integral

$$F(s) = \int_{\mathbb{R}^d} |\eta|^{2s} \exp\left\{-|\eta|^2\right\} d\eta \tag{3.15}$$

is finite for eaxch $s > 0$, we have convergence of $U(x,t)$ to $M(x)$ in $L^2(\mathbb{R}^d)$ at the rate $1/t^s$. Note that the same bound holds true (with a different constant) if we are looking for a bound, by formula (3.13), of the quantity

$$\int_{\mathbb{R}^d} |\xi|^{2p} |\widehat{U}(\xi,t) - \widehat{M}(\xi)|^2 \, d\xi.$$

In conclusion we prove

Proposition 2. *Let the initial datum f for the heat equation (3.1) be a probability density such that, for some $s \geq 1$ the Fourier based distance $d_s(f,M)$ is finite. Then, there is convergence of the scaled solution $U(x,t)$ defined by (7.36) to the Gaussian $M(x)$ at the rate $C_{s,p}/t^s$ in any Sobolev space H^p, with $p \geq 0$. The constant $C_{s,p}$ depends on $d_s(f,M)$, κ and $\Gamma(s+p)$.*

Remark 3. It is remarkable that the rate of convergence in any Sobolev space H^p, with $p \geq 0$ depends only on the values of the moments of the initial datum f. More moments of the initial datum coincide with moments of the same order of the Gaussian density M, more rapid is convergence to equilibrium. Hence, the rate of convergence does not depend on the regularity of the initial datum. This fact has been noticed in [54], in connection with the central limit theorem of probability theory (cf. also [13] for recent results and references).

A second important remark is concerned with a particular property of the fundamental solutions to the heat equation (3.1). Since the fundamental solutions are Gaussian probability densities, they are closed under the operation of convolution [59], namely

$$M_{\sigma_1} * M_{\sigma_2}(x) = M_{\sigma_1+\sigma_2}(x).$$

Hence, if we consider at time $t > 0$ the convolution of n powers of the fundamental solutions of heat equations with diffusion coefficients κ_j, $j = 1, 2, \dots, n$, we obtain

$$M_{2\kappa_1 t}^{\alpha_1} * M_{2\kappa_2 t}^{\alpha_2} * \cdots * M_{2\kappa_n t}^{\alpha_n} =$$

$$\prod_{j=1}^{n} (4\pi\kappa_j t)^{-\alpha_j d/2} \left(4\pi \frac{\kappa_j}{\alpha_j} t\right)^{d/2} M_{2t\kappa_1/\alpha_1} * M_{2t\kappa_2/\alpha_2} * \cdots * M_{2t\kappa_n/\alpha_n} =$$

$$\prod_{j=1}^{n} (4\pi\kappa_j t)^{-\alpha_j d/2} \left(4\pi \frac{\kappa_j}{\alpha_j} t\right)^{d/2} M_{2\Sigma t},$$

where

$$\Sigma = \sum_{j=1}^{n} \frac{\kappa_j}{\alpha_j}.$$

In the expression above the time-dependent quantity in front of the exponential is given by

$$\phi(t) = t^{-\frac{d}{2}\sum_{j=1}^{n}\alpha_j + \frac{1}{2}(n-1)}.$$

Therefore, if the exponents α_j are such that

$$\sum_{j=1}^{n} \alpha_j = n - 1, \tag{3.16}$$

independently of the values of the diffusion coefficients κ_j, $\phi(t) = 1$, and

$$M_{2\kappa_1 t}^{\alpha_1} * M_{2\kappa_2 t}^{\alpha_2} * \cdots * M_{2\kappa_n t}^{\alpha_n} = \Sigma_1 \exp\left\{-|x|^2/4\Sigma t\right\},$$

where Σ_1 denotes the constant

$$\Sigma_1 = \left(\frac{\kappa_j}{\alpha_j}\right)^{n/2} \Sigma^{-d/2} \prod_{j=1}^{n} (\kappa_j)^{-\alpha_j d/2}.$$

Consequently, independently of the values of the diffusion coefficients κ_j, if the exponents α_j satisfy condition (3.16), for every $x \in \mathbb{R}^d$

$$\frac{d}{dt} M_{2\kappa_1 t}^{\alpha_1} * M_{2\kappa_2 t}^{\alpha_2} * \cdots * M_{2\kappa_n t}^{\alpha_n} \ge 0. \tag{3.17}$$

This property is obviously restricted to a set of positive constants α_j satisfying (6.1). Note moreover that for a single fundamental solution condition (6.1) implies $\alpha_1 = 0$, so that property (3.17) becomes trivial.

Let us finally recall the evolution equation for a power of the solution to the heat equation. If $\alpha > 0$ is a positive constant, and $u(x,t)$ solves (3.1), then $u^\alpha(x,t)$ solves

$$\frac{\partial u^\alpha(x,t)}{\partial t} = \kappa \left[\Delta u^\alpha(x,t) + \alpha(1-\alpha) u^\alpha(x,t) |\nabla \log u(x,t)|^2\right]. \tag{3.18}$$

Equation (3.18) is particularly adapted to work with convolutions of powers. Note that equation (3.18) connects in a natural way dual exponents. In fact, if $\alpha = 1/p$, with $p > 1$, equation (3.18) takes the form

$$\frac{\partial u^{1/p}(x,t)}{\partial t} = \kappa \left[\Delta u^{1/p}(x,t) + \frac{1}{pp'} u^{1/p}(x,t) |\nabla \log u(x,t)|^2\right],$$

where $1/p + 1/p' = 1$.

3.2 Convex and other functionals

Let $f(x)$ be a probability density on $\mathbb{R}^d$, and let $u(x,t) = f * M_{2\kappa t}$ be the solution to the heat equation (3.1). Let $\Phi(r)$ be a convex function, and let G denote the convex functional

$$G(f) = \int_{\mathbb{R}^d} \Phi(f(x))\, dx. \tag{3.19}$$

Then, since M_σ, $\sigma > 0$, is a probability density function, Jensen's inequality implies

$$\Phi(f * M_\sigma)(x) = \Phi\left(\int_{\mathbb{R}^d} f(x-y) M_\sigma(y)\, dy\right) \le \int_{\mathbb{R}^d} \Phi(f(x-y)) M_\sigma(y)\, dy,$$

so that

$$G(f * M_\sigma) = \int_{\mathbb{R}^d} \Phi(f * M_\sigma)(x)\, dx \le \int_{\mathbb{R}^d} \Phi(f)(x)\, dx = G(f). \tag{3.20}$$

Owing to the same argument, we obtain that $G(f * M_\sigma) \le G(M_\sigma)$. Moreover, since for $\delta < \sigma$ $M_\sigma = M_\delta * M_{\sigma-\delta}$, $G(u(x,t))$ is a decreasing function of t.

Shannon's entropy is obtained by choosing $\Phi(r) = -r \log r$, which is concave. Therefore, Shannon's entropy is increasing along the solution to the heat equation. In addition, since $H(M_\sigma) = \frac{n}{2}\log(2\pi\sigma e)$, at time t

$$H(u(t)) \ge H(M_{2\kappa t}) = \frac{n}{2}\log(4\pi\kappa t e).$$

This shows that $H(u(t))$ diverges with time.

The precise growth of Shannon's entropy along the solution to the heat equation is given by the so-called DeBruijn's identity [69, 42]).

$$\frac{d}{dt} H(u(t)) = \kappa I(u(t)) = \kappa \int_{\{u(t)>0\}} \frac{|\nabla u(x,t)|^2}{u(x,t)}\, dx > 0. \tag{3.21}$$

In (3.21) $I(u(t))$ defines as usual the Fisher information of the probability density $u(x,t)$. The proof of equality (3.21) is left as exercise. It follows simply integrating by parts, and using the smoothness of the solution of the heat equation. Details can be found in [69, 80].

Also, Fisher information $I(u(t))$ is decreasing in time. Let us prove it for $d = 1$. Indeed, for any given $\sigma > 0$ Cauchy–Schwarz inequality implies

$$\int_{\mathbb{R}} f'(x-y) M_\sigma(y)\, dy \le \left[\int_{\mathbb{R}} \frac{f'(x-y)^2}{f(x-y)} M_\sigma(y)\, dy\right]^{1/2} \left[\int_{\mathbb{R}} f(x-y) M_\sigma(y)\, dy\right]^{1/2}.$$

Hence

$$\frac{1}{u(x,t)}\left(\frac{\partial u(x,t)}{\partial x}\right)^2 \le \int_{\mathbb{R}} \frac{f'(x-y)^2}{f(x-y)} M_{2\kappa t}(y)\, dy,$$

and integration over $\mathbb{R}$ leads at once to $I(u(t)) \le I(f)$. Using once more that for $\delta < \sigma$ $M_\sigma = M_\delta * M_{\sigma-\delta}$, we conclude that $I(u(x,t))$ is a decreasing function of t. Exchanging the roles of f and M_σ we also have that $I(u(t)) \le I(M_\sigma)$. The same properties hold in dimension $d > 1$. The precise decay of Fisher information can be evaluated as well. We will postpone it to Section 5.2.

3.3 Two examples

3.3.1 Maximizing entropy under constraints

An example will clarify the importance of functionals which are invariant under dilation invariance to get inequalities. Given a solution to the heat equation (3.1), with initial datum the probability density $f(x)$ with finite Shannon's entropy, we showed in Section 3.2 that the time derivative of its Shannon's entropy $H(u(t))$ is non-negative and it converges to infinity as time goes to infinity. This time behavior is consequence of the fact that Shannon's entropy is a functional that is not dilation invariant

$$H(u_a) = H(u) - d\log a. \tag{3.22}$$

It is easily checked that the second moment of a probability density function scales according to

$$E(u_a) = \int_{\mathbb{R}^d} |x|^2 u_a(x)\,dx = \frac{1}{a^2}E(u). \tag{3.23}$$

Hence, if the initial probability density in the heat equation has bounded second moment, a dilation invariant functional of $u(x,t)$ is obtained by coupling Shannon's entropy of $u(x,t)$ with the logarithm of the second moment of $u(x,t)$

$$\Gamma(t) = \Gamma(u(t)) = H(u(t)) - \frac{d}{2}\log E(u(t)). \tag{3.24}$$

Following [84], let us compute the time derivative of $\Gamma(t)$. We obtain

$$\frac{d}{dt}\Gamma(t) = \kappa\left(I(u(t)) - \frac{d^2}{E(u(t))}\right), \tag{3.25}$$

which is a direct consequence of DeBruijn's identity (3.21), and of the time evolution of the second moment of the solution to the heat equation,

$$\frac{d}{dt}E(u(t)) = \frac{d}{dt}\int_{\mathbb{R}^d} |x|^2 u(x,t)\,dx = 2\kappa d\int_{\mathbb{R}^d} u(x,t)\,dx = 2\kappa d.$$

The right-hand side of (3.25) is nonnegative. This can be easily shown by an argument which is often used in this type of proofs, and goes back probably to Mckean [69]. It holds

$$0 \le \int_{\mathbb{R}^d}\left|\frac{\nabla u(x,t)}{u(x,t)} + x\frac{d}{E(u(x,t))}\right|^2 u(x,t)\,dx =$$

$$I(u(t)) + \frac{d^2}{E(u(t))^2}\int_{\mathbb{R}^d} |x|^2 u(x,t)\,dx + 2\frac{d}{E(u(t))}\int_{\mathbb{R}^d} x\cdot\nabla u(x,t)\,dx =$$

$$I(u(t)) + \frac{d^2}{E(u(t))} - 2\frac{d^2}{E(u(t))} = I(u(t)) - \frac{d^2}{E(u(t))}. \tag{3.26}$$

Note that, since $u(x,t)$ is the (smooth) solution to the heat equation (3.1), equality to zero in (3.26) holds if and only if

$$\frac{\nabla u(x)}{u(x)} + x\frac{d}{E(u(x))} = 0$$

for all $x \in \mathbb{R}^d$. This condition can be rewritten as

$$\nabla\left(\log u(x) + \frac{d}{E(u(x))}\frac{x^2}{2}\right) = 0 \tag{3.27}$$

which identifies the probability density $u(x,t)$ as a Gaussian density in $\mathbb{R}^d$. By (8.17), this also shows that, among all densities with finite variance, Fisher's information takes its minimum value in correspondence to the Gaussian density (1.3), where $\sigma = E(u)/d$.

Thus, unless the initial value f is a Gaussian density, the functional $\Gamma(t)$ is monotone increasing, and it will reach its (eventual) maximum value as time $t\to\infty$. The computation of the limit value uses in a substantial way the dilation invariance of Γ. In fact, at each time $t>0$, the value of $\Gamma(t)$ does not change if we substitute $u(x,t)$ with $U(x,t)$ defined by (7.36).

Thanks to Proposition 2,

$$\lim_{t\to\infty} U(x,t) = M(x)$$

in any Sobolev space $H^p(\mathbb{R}^d)$. Therefore, passing to the limit one obtains

$$\Gamma(0) = H(f) - \frac{d}{2}\log E(f) \le H(M) - \frac{d}{2}\log E(M) = \frac{d}{2}\log\frac{2\pi e}{d}. \tag{3.28}$$

This inequality holds for all probability density functions with bounded second moment, and does not require that the second moment of f equals the second moment of the Gaussian density.

This relatively simple example contains the main ingredients we will use to obtain inequalities in sharp form: monotonicity of a functional of solutions to the heat equation coupled with the dilation invariance property.

3.3.2 Hölder's inequality revisited

We show now that Hölder's inequality can be viewed as a consequence of the time monotonicity of a suitable Lyapunov functional of the solution to the heat equation [16, 29]. For the sake of simplicity, we will present the proof in dimension $d = 1$. The corresponding higher-dimensional inequality can be deduced as well by making use of standard properties of the Gaussian function.

Hölder's inequality for integrals states that, if $p, q > 1$ are such that $1/p+1/q = 1$

$$\int_{\mathbb{R}} |f(x)g(x)|\, dx \leq \left(\int_{\mathbb{R}} |f(x)|^p \, dx \right)^{1/p} \left(\int_{\mathbb{R}} |g(x)|^q \, dx \right)^{1/q}. \tag{3.29}$$

Moreover, there is equality in (3.29) if and only f and g are such that there exist positive real numbers a and b such that $af^p(x) = bg^q(x)$ almost everywhere. Hölder's inequality can be proven in many ways, for example resorting to Young's inequality for constants, which states that, if $1/p + 1/q = 1$

$$cd \leq \frac{c^p}{p} + \frac{d^q}{q}, \tag{3.30}$$

for all nonnegative c and d, where equality is achieved if and only if $c^p = d^q$.

Without loss of generality, one can assume that the functions f, g in (3.29) are nonnegative. A different way to achieve inequality (3.29) is contained into the following

Theorem 4. *Let $\Phi(u(t), v(t))$ be the functional*

$$\Phi(t) = \Phi(u(t), v(t)) = \int_{\mathbb{R}} u(x,t)^{1/p} v(x,t)^{1/q} \, dx, \tag{3.31}$$

where $1/p+1/q = 1$, and $u(x,t)$ and $v(x,t)$, $t > 0$, are solutions to the heat equation (3.1) corresponding to the initial values $u(x) \in L^1(\mathbb{R})$ (respectively $v(x) \in L^1(\mathbb{R})$). Then $\Phi(u(t), v(t))$ is increasing in time from

$$\Phi(u(t=0), v(t=0)) = \int_{\mathbb{R}} u(x)^{1/p} v(x)^{1/q} \, dx,$$

to

$$\lim_{t\to\infty} \Phi(u(t), v(t)) = \left(\int_{\mathbb{R}} u(x)\, dx \right)^{1/p} \left(\int_{\mathbb{R}} v(x)\, dx \right)^{1/q}.$$

Proof. Without loss of generality, let us fix $\kappa = 1$ in equation (3.1). It is immediate to verify that the functional $\Phi(u(t), v(t))$ is invariant with respect to dilation (3.6) (applied to both u and v with the same constant a). Moreover, the condition

$u(x), v(x) \in L^1(\mathbb{R})$ is enough to ensure that $\Phi(u(t), v(t)) \in L^1(\mathbb{R})$ at any time $t \geq 0$. Indeed, inequality (3.30) implies

$$u(x,t)^{1/p} v(x,t)^{1/q} \leq \frac{1}{p} u(x,t) + \frac{1}{q} v(x,t),$$

where, since $u(x,t)$ and $v(x,t)$ are solution to the heat equation,

$$\int_{\mathbb{R}} u(x,t)\, dx = \int_{\mathbb{R}} u(x)\, dx, \qquad \int_{\mathbb{R}} v(x,t)\, dx = \int_{\mathbb{R}} v(x)\, dx.$$

Let us first proceed in a formal way. However, by resorting to the smoothness properties of the solution to the heat equation, all mathematical details can be rigorously justified.

Let us evaluate the time derivative of $\Phi(t)$. It holds

$$\frac{d}{dt}\Phi(t) = \int_{\mathbb{R}} \left[(u(x,t)^{1/p})_t v(x,t)^{1/q} + u(x,t)^{1/p} (v(x,t)^{1/q})_t \right] dx =$$

$$\int_{\mathbb{R}} \left[\frac{1}{p} u^{1/p-1} v^{1/q} u_{xx} + \frac{1}{q} u^{1/p} v^{1/q-1} v_{xx} \right] dx =$$

$$\int_{\mathbb{R}} \left[\frac{1}{p} u^{-1/q} v^{1/q} u_{xx} + \frac{1}{q} u^{1/p} v^{-1/p} v_{xx} \right] dx.$$

Integrating by parts we end up with

$$\frac{d}{dt}\Phi(u(t), v(t)) = \frac{1}{pq} \int_{\mathbb{R}} u^{1/p} v^{1/q} \left[\left(\frac{u_x}{u} \right)^2 - 2 \frac{u_x}{u} \frac{v_x}{v} + \left(\frac{v_x}{v} \right)^2 \right] dx =$$

$$\frac{1}{pq} \int_{\mathbb{R}} u^{1/p}(x,t) v^{1/q}(x,t) \left(\frac{u_x(x,t)}{u(x,t)} - \frac{v_x(x,t)}{v(x,t)} \right)^2 dx \geq 0.$$

Hence the functional $\Phi(u(t), v(t))$ is increasing in time. Note that the time derivative of the functional is equal to zero if and only if, for every $t > 0$

$$\frac{u_x(x,t)}{u(x,t)} - \frac{v_x(x,t)}{v(x,t)} = 0$$

for all points $x \in \mathbb{R}$. This condition can be rewritten as

$$\frac{d}{dx} \log \frac{u(x,t)}{v(x,t)} = 0.$$

Consequently $\Phi'(t) = 0$ if and only if

$$u(x,t) = c\, v(x,t) \tag{3.32}$$

for some positive constant c. Thus, unless condition (3.32) is verified almost everywhere at time $t = 0$, the functional $\Phi(t)$ is monotone increasing, and it will reach its eventual maximum value as time $t \to \infty$. Once again, the computation of the limit value uses in a substantial way the scaling invariance of Φ. In fact, at each time $t > 0$, the value of $\Phi(u(t), v(t))$ does not change if we substitute $u(x,t)$ and $v(x,t)$ with $U(x,t)$ and $V(x,t)$ defined by

$$U(x,t) = \sqrt{1+2t}\, u(x\sqrt{1+2t}, t)$$
$$V(x,t) = \sqrt{1+2t}\, v(x\sqrt{1+2t}, t).$$

By proposition 2

$$\lim_{t\to\infty} U(x,t) = M(x)\int_{\mathbb{R}} u(x)\, dx \quad \lim_{t\to\infty} V(x,t) = M(x)\int_{\mathbb{R}} v(x)\, dx,$$

Therefore, passing to the limit one obtains

$$\lim_{t\to\infty}\Phi(u(t),v(t)) = \lim_{t\to\infty}\Phi(U(t),V(t)) =$$
$$\left(\int_{\mathbb{R}} u(x)\, dx\right)^{1/p}\left(\int_{\mathbb{R}} v(x)\, dx\right)^{1/q}\int_{\mathbb{R}} M_1(x)^{1/p}M_1(x)^{1/q}\, dx =$$
$$\left(\int_{\mathbb{R}} u(x)\, dx\right)^{1/p}\left(\int_{\mathbb{R}} v(x)\, dx\right)^{1/q}\int_{\mathbb{R}} M_1(x)\, dx =$$
$$\left(\int_{\mathbb{R}} u(x)\, dx\right)^{1/p}\left(\int_{\mathbb{R}} v(x)\, dx\right)^{1/q}.$$

Since

$$\lim_{t\to 0^+}\Phi(u(t),v(t)) = \int_{\mathbb{R}} u(x)^{1/p}v(x)^{1/q}\, dx,$$

the monotonicity of the functional $\Phi(t)$ implies the inequality

$$\int_{\mathbb{R}} u(x)^{1/p}v(x)^{1/q}\, dx \le \left(\int_{\mathbb{R}} u(x)\, dx\right)^{1/p}\left(\int_{\mathbb{R}} v(x)\, dx\right)^{1/q},$$

with equality if and only if (3.32) is verified at time $t = 0$, that is

$$u(x) = cv(x), \tag{3.33}$$

for some positive constant c. Setting $f = u^{1/p}$ and $g = v^{1/q}$ proves both Hölder inequality (3.29) and the equality cases. □

Despite its apparent complexity, this way of proof is based on a solid physical argument, namely the monotonicity in time of a Lyapunov functional of the solution to the heat equation. This gives a clear indication that many inequalities reflect the physical principle of the tendency of a system to move towards the state of maximum entropy.

4 Blachman–Stam inequality

Blachman-Stam inequality is concerned with the behavior of the Fisher information with respect to convolutions [78, 20]. Historically, it was the key argument to prove Shannon's entropy power inequality. It is very instructive to give a proof, since it makes evident that the quadratic nature of Fisher's information plays a fundamental role in proofs. In what follow, we present a natural extension of the original Blachman-Stam inequality that applies to any power of convolution. We will present a one-dimensional proof of this inequality, that the interested reader can without essential difficulties extend to any dimension $d > 1$.

4.1 Fisher information bounds

To start with, we need the following definition. Given two (smooth) probability densities f and g on $\mathbb{R}$, let us define

$$I(f|g)(x) = \int_{\mathbb{R}} \frac{(f_x(x-y))^2}{f(x-y)} g(y)\, dy. \tag{4.1}$$

It is evident that the Fisher information of the density f coincides with the integral of $I(f|g)(x)$ for any choice of the probability density g

$$I(f) = \int_{\mathbb{R}} I(f|g)(x)\, dx. \tag{4.2}$$

Then, the following Lemma holds.

Lemma 5. *Let $f(x)$ and $g(x)$ be probability density functions such that both $I(f|g)$ and $I(g|f)$ are well defined. Then, for all positive constants a, b and $r > 0$*

$$\left(a^2 + b^2 + 2abr\right) \int_{\mathbb{R}} (f * g)^{r-2} \left((f * g)_x\right)^2\, dx \leq$$

$$a^2 \int_{\mathbb{R}} (f * g)^{r-1} I(f|g)\, dx + b^2 \int_{\mathbb{R}} (f * g)^{r-1} I(g|f)\, dx. \tag{4.3}$$

Moreover, there is equality in (1.8) if and only if, for any positive constant c and constants γ_1, γ_2, f and g are Gaussian densities, $f(x) = M_{ca}(x - \gamma_1)$ and $g(x) = M_{cb}(x - \gamma_2)$.

Proof. The proof follows along the same lines of the proof of inequality (1.9), given by Blachman [20]. First of all, to easily justify computations, let us prove the lemma by considering smooth functions f and g. This can be easily done by considering, for some $t > 0$, $f * M_{2t}$ and $g * M_{2t}$, solutions to the heat equation (3.1). Then the proof for general f and g will follow owing to the convexity properties of $I(f|g)$ and $I(g|f)$ [64]. Let

$$k(x) = f * g(x).$$

Then, for any pair of positive constants a, b the derivative of the convolution can be written in the following form

$$k'(x) = \frac{a}{a+b}\int_{\mathbb{R}} f'(x-y)g(y)\,dy + \frac{b}{a+b}\int_{\mathbb{R}} f(x-y)g'(y)\,dy.$$

Therefore

$$(a+b)\frac{k'(x)}{k(x)} = a\int_{\mathbb{R}} \frac{f'(x-y)}{f(x-y)}\frac{f(x-y)g(y)}{k(x)}\,dy + b\int_{\mathbb{R}} \frac{g'(y)}{g(y)}\frac{f(x-y)g(y)}{k(x)}\,dy =$$

$$\int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)}\right)\,d\mu_x(y),$$

where we denoted

$$d\mu_x(y) = \frac{f(x-y)g(y)}{k(x)}\,dy.$$

Note that, for every $x \in \mathbb{R}$, $d\mu_x$ is a unit measure on $\mathbb{R}$. Consequently, by Jensen's inequality

$$(a+b)^2\left[\frac{k'(x)}{k(x)}\right]^2 = \left[\int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)}\right)\,d\mu_x(y)\right]^2 \leq$$

$$\int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)}\right)^2\,d\mu_x(y). \tag{4.4}$$

Hence, for every constant $r > 0$

$$(a+b)^2\int_{\mathbb{R}} k^r(x)\,\left[\frac{k'(x)}{k(x)}\right]^2\,dx \leq$$

$$\int_{\mathbb{R}} k^r(x)\int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)}\right)^2\,\frac{f(x-y)g(y)}{k(x)}\,dy\,dx =$$

$$\int_{\mathbb{R}} k^{r-1}(x)\,\left[a^2\int_{\mathbb{R}} \frac{(f'(x-y))^2}{f(x-y)}g(y)\,dy + b^2\int_{\mathbb{R}} \frac{(g'(y))^2}{g(y)}f(x-y)\,dy\right]\,dx+$$

$$2ab\int_{\mathbb{R}} k^{r-1}(x)\int_{\mathbb{R}} f'(x-y)g'(y)\,dy\,dx.$$

On the other hand,

$$\int_{\mathbb{R}} f'(x-y)g'(y)\,dy = k''(x),$$

so that

$$\int_{\mathbb{R}} k^{r-1}(x) \int_{\mathbb{R}} f'(x-y)g'(y)\,dy\,dx =$$

$$\int_{\mathbb{R}} k^{r-1}(x)k''(x)\,dx = -(r-1)\int_{\mathbb{R}} k^{r-2}(x)(k'(x))^2\,dx.$$

This concludes the proof of the lemma. The cases of equality are easily found resorting to the following argument. Equality follows if, after application of Jensen's inequality, there is equality in (4.4). On the other hand, for any convex function φ and unit measure $d\mu$ on the set Ω, equality in Jensen's inequality

$$\varphi(\int_{\Omega} f\,d\mu) \le \int_{\Omega} \varphi(f)\,d\mu$$

holds true if and only if f is constant, so that

$$f = \int_{\Omega} f\,d\mu.$$

In our case, this means that there is equality if and only if the function

$$a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)}$$

does not depend on y. If this is the case, taking the derivative with respect to y, and using the identity

$$\frac{d}{dy}\left(\frac{f'(x-y)}{f(x-y)}\right) = -\frac{d}{dx}\left(\frac{f'(x-y)}{f(x-y)}\right),$$

we conclude that f and g have to satisfy

$$a\frac{d^2}{dx^2}\log f(x-y) = b\frac{d^2}{dy^2}\log g(y). \tag{4.5}$$

Note that (4.5) can be verified if and only if the functions on both sides are constant. Thus, there is equality if and only if

$$\log f(x) = b_1x^2 + c_1x + d_1, \quad \log g(x) = b_2x^2 + c_2x + d_2. \tag{4.6}$$

By coupling (4.6) with (4.5), we obtain that there is equality in (1.8) if and only if f and g are gaussian densities, of variances ca and cb, respectively, for any given positive constant c.

□

If $r = 1$ using (4.2) into inequality (4.3) gives the classical result by Blachman and Stam about Fisher information of convolution. Let a, b be positive constants. Then

$$(a^2 + b^2)I(f * g) \leq a^2 I(f) + b^2 I(g). \tag{4.7}$$

The classical Blachman–Stam inequality easily follows from (4.7). Indeed, inequality (4.7) is equivalent to

$$I(f * g) \leq \gamma^2 I(f) + (1-\gamma)^2 I(g),$$

where $0 \leq \gamma = a/(a+b) \leq 1$. Optimizing over γ implies inequality (1.9).

Remark 6. *Note that inequality (4.3) continues to hold even if the functions f and g are not density functions. Indeed, inequality (4.3) is invariant respect to the substitution of f with Af, and g with Bg, for any pair of positive constants A, B.*

5 Shannon's entropy power inequality

5.1 The proof of Stam and Blachman

The original proof of the entropy power inequality (1.4), given by Stam [78] and concluded by Blachman [20] makes an essential use both of the solution to the heat equation, and of the inequality (4.7) proven in the previous section. We write this proof by using the notations of Section 3. Let us fix the dimension $d = 1$. Let X and Y be independent random variables with probability densities f and g, and denote by $f(t)$ and $g(t)$ the densities of $X + Z_{\nu(t)}$ (respectively $X + Z_{\mu(t)}$), namely the convolutions of f and g with Gaussian probability densities having variances $\nu(t)$ and $\mu(t)$ respectively, where both $\nu(t)$ and $\mu(t)$ are increasing in time from $\nu(0) = 0$ (respectively $\mu(0) = 0$). We suppose that the Gaussian variables $Z_{\nu(t)}$ and $Z_{\nu(t)}$ are independent from each other. Thanks to formula (3.5) $f(t)$ and $g(t)$ are recognized as solutions to the heat equation with a time-dependent coefficient, and initial data f and g.

Now, consider the functional

$$\mathcal{V}(t) = \frac{\exp\{2H(f(t))\} + \exp\{2H(g(t))\}}{\exp\{2H(r(t))\}}, \tag{5.1}$$

with $r(t) = f(t) * g(t)$. It is evident that this functional is invariant under dilation. Moreover

$$\mathcal{V}(0) = \frac{N(f) + N(g)}{N(r)}.$$

Let us differentiate with respect to time. We obtain

$$\frac{d\mathcal{V}(t)}{dt} \exp\{2H(r(t))\} = \nu'(t)I(f(t))\exp\{2H(f(t))\} + \mu'(t)I(g(t))\exp\{2H(g(t))\} -$$

$$I(r(t))\,[\nu'(t)+\mu'(t)]\,[\exp\{2H(f(t))\}+\exp\{2H(g(t))\}]\,.$$

Let us choose the functions $\nu(t)$ and $\mu(t)$ so that

$$\nu'(t)=\exp\{2H(f(t))\},\quad \mu'(t)=\exp\{2H(g(t))\}.$$

For this choice of $\nu(t)$ and $\mu(t)$, applying inequality (4.7) with

$$a=\exp\{2H(f(t))\},\quad b=\exp\{2H(g(t))\},$$

it follows

$$\frac{d\mathcal{V}(t)}{dt}\geq 0.$$

By Lemma 5 the derivative is equal to zero if and only if both $f(t)$ and $g(t)$ are Gaussian densities. But then f and g have to be Gaussian and the derivative is identically zero for all t. In other words $\mathcal{V}(t)$ is either strictly increasing or a constant. Since $\mathcal{V}(t)$ is continuous from the right in $t=0$, we have

$$\mathcal{V}(0)=\frac{N(f)+N(g)}{N(r)}\leq \lim_{t\to\infty}\mathcal{V}(t).$$

To conclude the proof it remains to show that the limit is equal to 1. Here I report exactly the concluding argument of Stam [78], using the same words.

It is clear that

$$\lim_{t\to\infty}\nu(t)=\lim_{t\to\infty}\mu(t)=\infty.$$

The fact that $\lim_{t\to\infty}\mathcal{V}(t)$ exists and is equal to 1 can be proved easily, making use of the fact that $f(t)$, $g(t)$ and $r(t)$ *become more and more Gaussians.*

Remark 7. While we know that the result is correct, the argument *become more and more Gaussians* is not rigorous and the proof of the validity of the entropy power inequality can not be claimed on this basis. Some years later, Blachman completed the proof by resorting to a variant of the property of invariance under dilation given in Definition 1.

In order to prove that $\lim_{t\to\infty}\mathcal{V}(t)=1$, Blachman used the scaling property of Shannon's entropy we gave in (3.22). Let us define

$$F(x,t)=\sqrt{\nu(t)}f(x\sqrt{\nu(t)},t);\quad G(x,t)=\sqrt{\mu(t)}f(x\sqrt{\mu(t)},t)$$

and

$$R(x,t)=\sqrt{\nu(t)+\mu(t)}r(x\sqrt{\nu(t)+\mu(t)},t).$$

Then, by (3.22)

$$H(f(t))=H(F(t))+\log\sqrt{\nu(t)},$$

which gives

$$N(f(t)) = \nu(t)N(F(t)),$$

Same expression for $g(t)$ and $r(t)$. On the other hand, even if proved it when $\nu(t) = \kappa t$, it is clear that Proposition 2 applies any time the function $\nu(t)$ tends to infinity as time goes to infinity, like in the present case. Hence

$$\lim_{t\to\infty} N(F(t)) = 2\pi e,$$

and the same limit value is found for $N(G(t))$ and $N(R(t))$. This implies

$$\lim_{t\to\infty} \mathcal{V}(t) = \lim_{t\to\infty} \frac{\nu(t)N(F(t)) + \mu(t)N(G(t))}{(\nu(t) + \mu(t))N(R(t))} = 1.$$

This concludes the proof.

Remark 8. A careful reading of the arguments used in the proof indicates that we can use the same arguments in dimension $d > 1$. This has been done by Blachman in [20]. We will postpone to Section 6.4 a different proof of the entropy power inequality, fully based on a functional which is invariant under dilation, that simplifies noticeably the original proof.

5.2 The concavity of entropy power

The entropy power inequality proven in the previous section has a lot of interesting consequences. Among others, one of these consequences has been noticed by Costa [40], who named it emphconcavity property of entropy power.

Let X be a random vector with a (smooth) density function $f(x)$, $x \in \mathbb{R}^d$, and let us denote by $f(x,t)$ the solution to the Cauchy problem for the heat equation (3.1) with diffusion constant $\kappa = 1$, posed in the whole space and such that $f(x,t = 0) = f(x)$. As discussed in Section 3, $f(x,t)$ is the density of the random variable $X + Z_{2t}$. Given the entropy power $N(X + Z_{2t})$, where N is defined as in (1.2), we shall now prove that $N(X + Z_{2t})$ is concave with respect to time

$$\frac{d^2}{dt^2} N(X + Z_{2t}) \le 0. \tag{5.2}$$

The proof of concavity then requires to evaluate, for any time $t > 0$, two time derivatives of the entropy power of $f(x,t)$. The first derivative of the entropy power is easily evaluated resorting to DeBruijn's identity (3.21) which connects Shannon's entropy functional with the Fisher information of a random variable with density Using identity (3.21) we get

$$\frac{d}{dt} N(f(t)) = \frac{2}{d} \exp\left\{\frac{2}{d} H(f(t))\right\} \frac{d}{dt} H(f(t)) =$$

$$\frac{2}{d}\exp\left\{\frac{2}{d}H(f(t))\right\}\, I(f(t)).$$

Hence

$$\frac{d^2}{dt^2}N(f(t)) = \frac{2}{d}\frac{d}{dt}\left[\exp\left\{\frac{2}{d}H(f(t))\right\}\, I(f(t))\right].$$

Let us set

$$\Upsilon(f) = \exp\left\{\frac{2}{d}H(f)\right\}\, I(f). \tag{5.3}$$

Then, the concavity of entropy power can be rephrased as the decreasing in time property of the functional $\Upsilon(f(t))$ along the solution to the heat equation. If

$$-J(f(t)) = \frac{dI(f(t))}{dt}, \tag{5.4}$$

denotes the derivative of Fisher information along the solution to the heat equation, we obtain

$$\frac{d}{dt}\Upsilon(f(t)) = \exp\left\{\frac{2}{d}H(f(t))\right\}\left(\frac{dI(f(t))}{dt} + \frac{2}{d}I(f(t))^2\right) =$$

$$\exp\left\{\frac{2}{d}H(f(t))\right\}\left(-J(f(t)) + \frac{2}{d}I(f(t))^2\right).$$

Hence, $\Upsilon(f(t))$ is non increasing if and only if

$$J(f(t)) \ge \frac{2}{d}I(f(t))^2. \tag{5.5}$$

It is interesting to remark that, aiming in proving the old conjecture that subsequent derivatives of Boltzmann's H-functional, evaluated on the solution to heat equation, alternate in sign, the functional $J(f(t))$ was first considered by McKean [69]. Indeed, in one space dimension, inequality (5.5) is essentially due to him. Let us repeat his highlighting idea. In the one dimensional case one has

$$I(f) = \int_{\mathbb{R}} \frac{f'(x)^2}{f(x)}\, dx,$$

while

$$J(f) = 2\left(\int_{\mathbb{R}} \frac{f''(x)^2}{f(x)}\, dx - \frac{1}{3}\int_{\mathbb{R}} \frac{f'(x)^4}{f(x)^3}\, dx\right). \tag{5.6}$$

McKean observed that $J(f)$ is positive. In fact, resorting to integration by parts, $J(f)$ can be rewritten as

$$J(f) = 2\int_{\mathbb{R}} \left(\frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2}\right)^2 f(x)\, dx \ge 0. \tag{5.7}$$

Having this formula in mind, consider that, for any constant $\lambda > 0$

$$0 \le 2 \int_{\mathbb{R}} \left(\frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2} + \lambda \right)^2 f(x)\, dx =$$

$$J(f) + 2\lambda^2 + 4\lambda \int_{\mathbb{R}} \left(f''(x) - \frac{f'(x)^2}{f(x)} \right) dx = J(f) + 2\lambda^2 - 4\lambda I(f).$$

Choosing $\lambda = I(f)$ shows (5.5) for $d = 1$.

Note that equality in (5.5) holds if and only if f is a Gaussian density. In fact, the condition

$$\frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2} + \lambda = 0,$$

can be rewritten as

$$\frac{d^2}{dx^2} \log f(x) = -\lambda,$$

which corresponds to

$$\log f(x) = -\lambda x^2 + bx + c. \tag{5.8}$$

Joining condition (5.8) with the fact that $f(x)$ has to be a probability density, we conclude.

The argument of McKean was used by Villani [91] to obtain (5.5) for $d > 1$. In the general d-dimensional situation, Villani proved the formula

$$J(f) = 2 \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \left[\frac{\partial^2}{\partial v_i \partial v_j} \log f \right]^2 f\, dx =$$

$$2 \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \left[\frac{1}{f} \frac{\partial^2}{\partial v_i \partial v_j} - \frac{1}{f^2} \frac{\partial f}{\partial v_i} \frac{\partial f}{\partial v_j} \right]^2 f\, dx. \tag{5.9}$$

By means of (5.9), the nonnegative quantity

$$A(\lambda) = \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} \left[\frac{1}{f} \frac{\partial^2}{\partial v_i \partial v_j} - \frac{1}{f^2} \frac{\partial f}{\partial v_i} \frac{\partial f}{\partial v_j} + \lambda \delta_{ij} \right]^2 f\, dx,$$

with the choice $\lambda = I(f)/d$, allows to recover inequality (5.5) for $d > 1$. This proves the concavity property of entropy power.

To enlighten the consequences of the concavity of entropy power, consider that the functional $\Upsilon(f)$ is invariant under dilation (cf. Definition 1). In fact, Shannon's entropy is such that,

$$H(f_a) = H(f) - d \log a,$$

while Fisher's information satisfies

$$I(f_a) = \int_{\mathbb{R}^d} \frac{|\nabla f_a(x)|^2}{f_a(x)}\, dx = a^2 \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)}\, dx = a^2 I(f).$$

Therefore, for any constant $a > 0$

$$\Upsilon(f(t)) = \Upsilon\left(f_a(t)\right). \tag{5.10}$$

Thanks to Proposition 2 we can identify the long-time behavior of the functional $\Upsilon(f(t))$. Unless the initial value $f(x)$ in the heat equation is a Gaussian function, the functional $\Upsilon(f(t))$ is monotone decreasing, and it will reach its eventual minimum value as time $t \to \infty$. Grace to the invariance under dilation property, at each time $t > 0$, the value of $\Upsilon(f(t))$ does not change if we scale the argument $f(x,t)$ according to (7.36), namely

$$f(x,t) \to F(x,t) = \left(\sqrt{1+2t}\right)^d f(x\sqrt{1+2t}, t),$$

which is such that the initial value $f(x)$ is left unchanged. On the other hand, Proposition 2 implies

$$\lim_{t\to\infty} F(x,t) = M(x)$$

Moreover, the limit value of $\Upsilon(f(t))$ does not change if we consider a dilation of the limit Gaussian function in order to have a variance different from one. Therefore, passing to the limit one obtains, for any $\sigma > 0$, the inequality

$$\Upsilon(f) \geq \Upsilon(M_\sigma),$$

or, what is the same

$$\exp\left\{\frac{2}{d}H(f)\right\} I(f) \geq \exp\left\{\frac{2}{d}H(M_\sigma)\right\} I(M_\sigma). \tag{5.11}$$

5.2.1 The logarithmic Sobolev inequality

Inequality (5.11) has various important consequences. First, let us rewrite it in the form

$$\frac{I(f)}{I(M_\sigma)} \geq \exp\left\{-\frac{2}{d}\left(H(f) - H(M_\sigma)\right)\right\}. \tag{5.12}$$

Since

$$I(M_\sigma) = \frac{d}{\sigma},$$

while

$$H(M_\sigma) = \frac{d}{2}\log 2\pi\sigma + \frac{d}{2},$$

using that $e^{-x} \geq 1 - x$, we obtain from (5.12)

$$\int_{\mathbb{R}^d} f(x) \log f(x)\, dx + d + \frac{d}{2} \log 2\pi\sigma \leq \frac{\sigma}{2} \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)}\, dx. \tag{5.13}$$

Inequality (5.13) is nothing but the logarithmic Sobolev inequality by Gross [53], written in an equivalent form.

Consider now the case in which the probability density $f(x)$ of the random vector X is such that the second moment of X is bounded. Then, for any σ such that

$$\sigma \geq \frac{1}{n} \int_{\mathbb{R}^d} |x|^2 f(x)\, dx,$$

it holds

$$-H(f) + H(M_\sigma) = \int_{\mathbb{R}^d} f(x) \log f(x)\, dx - \int_{\mathbb{R}^d} M_\sigma(x) \log M_\sigma(x)\, dx =$$

$$\int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{M_\sigma(x)}\, dx + \frac{1}{2\sigma} \int_{\mathbb{R}^d} |x|^2 \left(M_\sigma(x) - f(x) \right)\, dx \geq$$

$$\int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{M_\sigma(x)}\, dx.$$

By the Csiszar-Kullback inequality [58]

$$2 \int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{M_\sigma(x)}\, dx \geq \|f - M_\sigma\|_{L^1}^2. \tag{5.14}$$

By expanding the right-hand side of inequality (5.12) up to the second order, we end up with the inequality

$$\frac{\sigma}{2} \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)}\, dx - \int_{\mathbb{R}^d} f(x) \log f(x)\, dx + d + \frac{d}{2} \log 2\pi\sigma \geq \frac{d^2}{8} \|f - M_\sigma\|_{L^1}^4. \tag{5.15}$$

The right-hand side of (5.15) improves the logarithmic Sobolev inequality when the density function involved into inequality (5.12) has bounded second moment, and it is different from a Gaussian density. In this case, it is possible to quantify the positivity of the difference between the right and left sides of (5.12) in terms of the distance of the density $f(x)$ from the manifold of the Gaussian densities, with a precise estimate of this distance in terms of the L^1-norm.

5.2.2 Nash's inequality revisited

A second interesting consequence of the concavity of entropy power is a new proof of Nash's inequality [70]. To this aim, note that the right-hand side of inequality (5.11), thanks to the invariance under dilation property of $\Upsilon(f)$, does not depend of σ. The choice

$$\sigma = \bar{\sigma} = (2\pi e)^{-1}, \tag{5.16}$$

gives

$$I(M_{\bar{\sigma}}) = 2\pi e d,$$

and

$$H(M_{\bar{\sigma}}) = 0.$$

Thus, substituting the value $\sigma = \bar{\sigma}$ in (8.24) we obtain the inequality

$$\exp\left\{\frac{2}{d}H(f)\right\} I(f) \geq 2\pi e d. \tag{5.17}$$

Inequality (5.17) is know in information theory with the name of *Isoperimetric Inequality for Entropies* (cf. [47] for a different proof).

The case in which $f(x) \geq 0$ is a nonnegative function of mass different from 1, leads to a modified inequality. Let us set

$$\mu = \int_{\mathbb{R}^d} f(x)\,dx \neq 1$$

Then, the function $\phi(x) = f(x)/\mu$ is a probability density, which satisfies (5.17). Therefore

$$I(\mu\phi) = \mu\, I(\phi) \geq \mu\, I(M_\sigma) \exp\left\{\frac{2}{d}H(M_\sigma)\right\}\, \exp\left\{-\frac{2}{d}H(\phi)\right\} =$$

$$\mu\, I(M_\sigma) \exp\left\{\frac{2}{d}(H(M_\sigma) - \log\mu)\right\}\, \exp\left\{-\frac{2}{d}(H(\phi) - \log\mu)\right\} =$$

$$\mu\, I(M_\sigma) \exp\left\{\frac{2}{d}\frac{1}{\mu}H(\mu M_\sigma)\right\}\, \exp\left\{-\frac{2}{d}\frac{1}{\mu}H(\mu\phi)\right\}. \tag{5.18}$$

In (5.18) we used the identity

$$H(\mu\phi) = \mu H(\phi) - \mu\log\mu.$$

Setting now $\sigma = \bar{\sigma}$, as given by (5.16), we conclude with the inequality

$$I(f) \geq 2\pi e d\, \|f\|_{L^1} \exp\left\{-\frac{2}{d\|f\|_{L^1}}\left[H(f) - \|f\|_{L^1}\log\|f\|_{L^1}\right]\right\}, \tag{5.19}$$

which clearly holds for any integrable function $f(x) \geq 0$.

Given a probability density function $g(x)$, let us set $f(x) = g^2(x)$. In this case

$$H(f) = H(g^2) = -\int_{\mathbb{R}^d} g^2(x) \log g^2(x)\, dx = -2\int_{\mathbb{R}^d} (g(x) \log g(x))\, g(x)\, dx.$$

Since the function $h(r) = r \log r$ is convex, and $\|g\|_{L^1} = 1$, Jensen's inequality implies

$$-H(g^2) \geq 2\int_{\mathbb{R}^d} g^2(x)\, dx \log \int_{\mathbb{R}^d} g^2(x)\, dx. \tag{5.20}$$

Using (5.20) into (5.19) gives

$$I(g^2) \geq 2\pi ed \int_{\mathbb{R}^d} g^2(x)\, dx\ e^{\frac{2}{d}\log\int_{\mathbb{R}^d} g^2(x)\, dx} = 2\pi ed \left(\int_{\mathbb{R}^d} g^2(x)\, dx\right)^{1+2/d}. \tag{5.21}$$

Using the identity

$$I(g^2) = 4\int_{\mathbb{R}^d} |\nabla g(x)|^2\, dx$$

we obtain from (5.21) the classical Nash's inequality in sharp form

$$\left(\int_{\mathbb{R}^d} g^2(x)\, dx\right)^{1+2/d} \leq \frac{2}{\pi ed}\int_{\mathbb{R}^d} |\nabla g(x)|^2\, dx \tag{5.22}$$

Inequality (5.22) clearly holds for all probability density functions $g(x)$. Note that, if $\|g\|_{L^1} \neq 1$, (5.22) implies

$$\left(\int_{\mathbb{R}^d} g^2(x)\, dx\right)^{1+2/d} \leq \frac{2}{\pi ed}\left(\int_{\mathbb{R}^d} |g(x)\, dx\right)^{4/d} \int_{\mathbb{R}^d} |\nabla g(x)|^2\, dx. \tag{5.23}$$

The constant $2/(\pi ed)$ in (5.23) is sharp.

5.3 Dembo's proof of the concavity property

In a short note [46], Dembo showed that the concavity of entropy power is a direct consequence of Blachman–Stam inequality (1.9). The idea is very simple. If we write the Blachman–Stam inequality for the random vector $X + Z_{2t}$, we obtain

$$\frac{1}{I(X + Z_{2t})} \geq \frac{1}{I(X)} + \frac{1}{I(Z_{2t})} = \frac{1}{I(X)} + \frac{2t}{d}.$$

Indeed, for a Gaussian vector Z_σ, $I(Z_\sigma) = d/\sigma$. Hence, for each $\tau > 0$ we obtain the inequality

$$\frac{1}{\tau}\left[\frac{1}{I(X + Z_{2t+2\tau})} - \frac{1}{I(X + Z_{2t})}\right] \geq \frac{2}{d}$$

Passing to the limit as $\tau \to 0$, we finally obtain

$$\frac{d}{d\tau}\frac{1}{I(X+Z_{2t+2\tau})} = -\frac{1}{I(X+Z_{2t})^2}\frac{dI(X+Z_{2t+2\tau})}{d\tau} = -\frac{1}{I(f(t))^2}J(f(t)) \geq \frac{2}{d},$$

which, grace to definition (5.4) coincides with (5.5).

The same idea can be applied starting from the entropy power inequality applied to the random vector $X + Z_{2t}$. In this case, since by definition $N(Z_\sigma) = 2\pi e\sigma$,

$$N(X + Z_{2t}) \geq N(X) + N(Z_{2t}) = N(X) + 2\pi e 2t.$$

Hence for each $\tau > 0$ we have

$$\frac{1}{\tau}\left[N(X+Z_{2t+2\tau}) - N(X+Z_{2t})\right] \geq 4\pi e.$$

Passing to the limit as $\tau \to 0$, we finally obtain

$$\frac{d}{d\tau}N(X+Z_{2t+2\tau}) = N(X+Z_{2t})I(X+Z_{2t})\frac{2}{d} \geq 4\pi e.$$

Thus, we obtain the isoperimetric inequality for entropies we obtained in (5.17).

Remark 9. It is important to outline that this idea allows to prove the isoperimetric inequality without resorting to second-order derivatives in time of Shannon's entropy evaluated on the solution to the heat equation. In other words, inequality (5.17) and its consequences (including the logarithmic Sobolev inequality) follows from the entropy power inequality. On the other hand, the entropy power inequality is a consequence of Blachman–Stam inequality, which is self-contained. Hence, the logarithmic Sobolev inequality can be proven without resorting to the entropy-entropy method used in [1].

6 Inequalities for convolutions

The purpose of this section is to present various results concerned with the monotonicity in time of the convolution of powers of solutions to the heat equation. The main reason behind this investigation is that many functional inequalities can be viewed as the consequence of the tendency of various Lyapunov functionals defined in terms of powers of the solution to the heat equation to reach their extremal values as time tends to infinity. The discovery of a Lyapunov functional which allows to prove Young inequality and its converse [16], is only one of the possible application of this idea (cf. also [82, 83, 84] for a connection of these results with information theory). While the inequalities are not new, and some of the results we present have been obtained before, what is new is the approach to the problem, which takes into account the information-theoretical meaning of inequalities for convolutions, and consequently allows to obtain clean and relatively simple new proofs.

The prototype of these monotonous in time convolutions is as follows. Let n be an integer, and let α_j, $j = 1, \dots, n$, be positive real numbers such that

$$\sum_{j=1}^{n} \alpha_j = n - 1. \tag{6.1}$$

Let $f_j(x)$, $j = 1, \dots, n$, be non-negative functions on $\mathbb{R}^d$, $d \geq 1$, such that $f_j \in L^1(\mathbb{R}^d)$. For any given j, $j = 1, \dots, n$, we denote by $u_j(x,t)$ the solution to the heat equation (3.1) with the diffusion coefficients κ_j

$$\frac{\partial u_j(x,t)}{\partial t} = \kappa_j \Delta u_j(x,t),$$

such that

$$\lim_{t\to 0^+} u_j(x,t) = f_j(x).$$

We consider the n-th convolution

$$w(x,t) = u_1^{\alpha_1} * u_2^{\alpha_2} * \cdots * u_n^{\alpha_n}(x,t). \tag{6.2}$$

where, as usual the n-th convolution of the functions $g_j(x) \in L^1(\mathbb{R}^d)$, $j = 1, 2, \dots, n$ reads

$$g_1 * g_2 * \cdots * g_n(x) =$$
$$\int_{(\mathbb{R}^d)^{n-1}} g_1(x - x_1) \cdots g_{n-1}(x_{n-2} - x_{n-1}) g_n(x_{n-1})\, dx_1 dx_2 \cdots dx_{n-1}.$$

Then, a natural question arises. Can we fix the diffusion coefficients in the heat equation in such a way that $w(x,t)$ behaves monotonically in time? Note that the choice of condition (6.1) is forced by the fact that we want that the monotonicity of $w(x,t)$, $t > 0$ has to hold at least if $u_j(x,t)$ is the fundamental solution to the heat equation, $j = 1, 2, \dots, n$. In this case, in fact, computations are explicit, and, provided condition (6.1) is satisfied, $w(x,t)$ is increasing in time independently of the choice of the diffusion coefficients (cf. Section 3, where this property has been explicitly obtained). In the general case, however, the monotonicity in time of the n-th convolution can be proven under more restrictive assumptions both on the numbers α_j, and only for a unique choice of the diffusion coefficients κ_j (cf. Lemma 10).

The interest in the monotonicity of the convolution of powers of solutions to the heat equation is linked to its consequences. Indeed, the discovery of the monotonicity of $w(x,t)$ for a special choice of the diffusion coefficients translates immediately to the proof of an inequality for convolutions in sharp form. Let n be an integer, and let

p_j, $j = 1, \dots, n$, be real numbers such that $1 \le p_j \le +\infty$ and $\sum_{j=1}^n p_j^{-1} = n-1$. Let $f_j(x)$, $j = 1, \dots, n$, be functions on $\mathbb{R}^d$, $d \ge 1$, such that $f_j \in L^{p_j}(\mathbb{R}^d)$. In Theorem 13 we will show that the monotonicity of $w(x,t)$ implies the following inequality for convolutions:

$$\sup_x |f_1 * f_2 * \cdots * f_n| \le \prod_{j=1}^n C_{p_j}^d \|f_j\|_{p_j}. \tag{6.3}$$

In (6.3), the constant C_p which defines the sharp constant is given by

$$C_p^2 = \frac{p^{1/p}}{p'^{1/p'}}, \tag{6.4}$$

where primes always denote dual exponents, $1/p + 1/p' = 1$. Also, the expression of the best constant in (6.3), in the case in which the functions f_j are probability density functions, is obtained by assuming that the functions f_j are suitable Gaussian densities [62]. This expression naturally appears in this monotonicity approach by considering that for large times the solution to the heat equation behaves as the self-similar Gaussian profile.

Note that inequality (6.3) implies

$$\left| \int f_1(x_1) f_2(x_1 - x_2) \cdots f_n(x_{n-1})\, dx_1 dx_2 \cdots dx_{n-1} \right| \le \prod_{j=1}^n C_{p_j}^d \|f_j\|_{p_j}, \tag{6.5}$$

which is a particular case of the general inequalities obtained by Brascamp and Lieb [28], which are nowadays known as the Brascamp–Lieb inequalities.

Inequality (6.3) is closely related to the monotonicity property of the functional given by L^∞-norm of the n-th convolution $w(x,t)$. Naturally one could ask if a similar property holds for the L^r-norm of $w(x,t)$, where $r \ge 0$. Also in this case, the monotonicity in time can be proven under suitable assumptions both on the numbers α_j, and only for a unique choice of the diffusion coefficients κ_j. The study of the monotonicity in time of $\|w(t)\|_r$ is connected with the classical Young's inequality in sharp form ($r > 1$), or with its reverse form ($r < 1$).

Last, the limiting cases $r \to 1$ and $r \to 0$ lead to the monotonicity in time of Shannon's entropy and of the Renyi entropy of order 0 [42]. The monotonicity here leads to the entropy power inequality of Shannon [77], and to the Prékopa–Leindler inequality [60, 72, 73], respectively.

Therefore, all these well-known functional inequalities can be seen into a unified framework, as consequences of the monotonicity in time of the n-convolution of powers of solutions to the heat equation.

6.1 The monotonicity of convolutions

Let n be an integer, and let p_j, $j = 1, \dots, n$, be real numbers such that

$$1 \le p_j \le +\infty; \quad \sum_{j=1}^{n} \frac{1}{p_j} = n - 1. \tag{6.6}$$

Let $f_j(x)$, $j = 1, \dots, n$, be non-negative functions on $\mathbb{R}^d$, $d \ge 1$, such that $f_j \in L^1(\mathbb{R}^d)$. For any given j, $j = 1, \dots, n$, we denote by $u_j(x,t)$ the solution to the heat equation (3.1) with the diffusion coefficients κ_j

$$\frac{\partial u_j(x,t)}{\partial t} = \kappa_j \Delta u_j(x,t), \tag{6.7}$$

such that

$$\lim_{t \to 0^+} u_j(x,t) = f_j(x). \tag{6.8}$$

The following Lemma shows that there is a (unique) choice of the diffusion coefficients in the heat equation such that $w(x,t)$ behaves monotonically in time.

Lemma 10. *Let $w(x,t)$ be the n-th convolution*

$$w(x,t) = u_1^{1/p_1} * u_2^{1/p_2} * \cdots * u_n^{1/p_n}(x,t) \tag{6.9}$$

where the functions $u_j(x,t)$, $j = 1, 2, \dots, n$, are solutions to the heat equation corresponding to the initial values $0 \le f_j(x) \in L^1(\mathbb{R}^d)$. Then, if for each j the exponents p_j satisfy conditions (6.6) and the diffusion coefficients are given by $\kappa_j = (p_j p_j')^{-1}$, $w(x,t)$ is monotonically increasing in time from

$$w(x, t = 0) = f_1^{1/p_1} * f_2^{1/p_2} * \cdots * f_n^{1/p_n}(x).$$

Moreover, $w(x,t)$ remains constant in time if and only if $f_j(x)$, $j = 1, 2, \dots, n$, is a multiple of a Gaussian density of variance $d\kappa_j$.

Proof. For the sake of simplicity, we will prove the Lemma for $d = 1$. As the proof shows, however, analogous computations can be done in higher dimension.

Since $\sum_{j=1}^{n} p_j^{-1} = n - 1$, Hölder inequality implies that

$$\left| \int f_1(x_1)^{1/p_1} \cdots f_n(x_{n-1})^{1/p_n} \, dx_1 dx_2 \cdots dx_{n-1} \right| \le \prod_{j=1}^{n} \left(\int_{\mathbb{R}} |f_j(x)| \, dx \right)^{1/p_j}.$$

Hence, also

$$f_1^{1/p_1} * f_2^{1/p_2} * \cdots * f_n^{1/p_n}(x) \le \prod_{j=1}^{n} \left(\int_{\mathbb{R}} |f_j(x)| \, dx \right)^{1/p_j}, \tag{6.10}$$

and, since the right-hand side of (6.10) depends only on the L^1-norms of the functions, which are preserved by the heat equation, the function $w(x,t)$ is bounded for all subsequent times $t>0$. Also, using basic considerations on the heat equation, it is sufficient to prove the increasing property of $w(t)$ for very smooth initial data f_j, $j=1,2,\dots,n$, with fast decay at infinity. This will enable us to have the *central limit property*. In order not to worry about derivatives of logarithms, which will often appear in the proof, we may also impose that $|\frac{d}{dx}\log f_j(x)|\le C(1+|x|^2)$ for some positive constant C. The general case will follow by density [64].

For a given $x\in\mathbb{R}$, let us evaluate the time derivative of the n-th convolution $w(x,t)$. We obtain

$$\frac{\partial w(x,t)}{\partial t}=\left(\sum_{j=1}^n \kappa_j\right)\frac{\partial^2 w(x,t)}{\partial x^2}+\sum_{j=1}^n\frac{\kappa_j}{p_jp_j'}R_j(x,t), \tag{6.11}$$

where, for $j=1,2,\dots,n$

$$R_j(x)=\int u_1(x-x_1)^{1/p_1}\cdots u_n(x_{n-1})^{1/p_n}\left|\frac{\partial\log u_j}{\partial x}(x_{j-1}-x_j)\right|^2 dx_1\cdots dx_{n-1} \tag{6.12}$$

Indeed,

$$\frac{\partial w}{\partial t}=\frac{\partial u_1^{1/p_1}}{\partial t}*u_2^{1/p_2}*\cdots*u_n^{1/p_n}+u_1^{1/p_1}*\frac{\partial u_2^{1/p_2}}{\partial t}*\cdots*u_n^{1/p_n}+\dots$$
$$+u_1^{1/p_1}*u_2^{1/p_2}*\cdots*\frac{\partial u_n^{1/p_n}}{\partial t},$$

and the time derivative of each term on the right-hand side can be evaluated by considering that the functions $u_j(x,t)$, $j=1,2,\dots,n$ satisfy the heat equation (6.7) (with diffusion coefficients κ_j, $j=1,2,\dots,n$). Hence

$$\frac{\partial u_1^{1/p_1}}{\partial t}*u_2^{1/p_2}*\cdots*u_n^{1/p_n}=\kappa_1\frac{\partial^2 u_1^{1/p_1}}{\partial x^2}*u_2^{1/p_2}*\cdots*u_n^{1/p_n}+$$
$$\frac{\kappa_1}{p_1p_1'}\left(\left|\frac{\partial\log u_1}{\partial x}\right|^2u_1^{1/p_1}\right)*u_2^{1/p_2}*\cdots*u_n^{1/p_n}=$$
$$\kappa_1\frac{\partial^2 w}{\partial x^2}+\frac{\kappa_1}{p_1p_1'}\left(\left|\frac{\partial\log u_1}{\partial x}\right|^2u_1^{1/p_1}\right)*u_2^{1/p_2}*\cdots*u_n^{1/p_n}. \tag{6.13}$$

An analogous formula holds for the other indexes $j\ge 2$. Note that in (6.13) we used the convolution property

$$\frac{\partial^2}{\partial x^2}f*g(x)=\int f''(x-y)g(y)\,dy=\int f'(x-y)g'(y)\,dy=\int f(x-y)g''(y)\,dy. \tag{6.14}$$

By property (6.14), it holds that, for each pair of indexes (i, j) with $i, j = 1, 2, \dots, n$, with $x_0 = x$ and $x_n = 0$

$$(f_1 * f_2 * \cdots * f_n)'' =$$

$$\int f_1(x - x_1) \dots f_n(x_{n-1})(\log f_i(x_{i-1} - x_i))'(\log f_j(x_{j-1} - x_j))'\, dx_1 \dots dx_{n-1}.$$

Hence, if we take a set of positive constants $a_{i,j}$'s, $i, j = 1, 2, \dots, n$, such that $\sum_{i \neq j} a_{i,j} = 1$, we can express the second derivative of a convolution as

$$(f_1 * f_2 * \cdots * f_n)'' = \sum_{i \neq j} a_{i,j} \int f_1(x - x_1) \dots f_n(x_{n-1}) \cdot$$

$$\cdot (\log f_i(x_{i-1} - x_i))'(\log f_j(x_{j-1} - x_j))'\, dx_1 \dots dx_{n-1}.$$

This shows that, for any set of positive values $a_{i,j}$ such that $\sum_{i \neq j} a_{i,j} = 1$, with $x_0 = x$ and $x_n = 0$, one has

$$\frac{\partial^2 w}{\partial x^2} = \sum_{i \neq j} \frac{a_{i,j}}{p_i p_j} \int u_1^{1/p_1}(x - x_1) \dots u_n^{1/p_n}(x_{n-1}) \cdot$$

$$\cdot (\log u_i(x_{i-1} - x_i))'(\log u_j(x_{j-1} - x_j))'\, dx_1 \dots dx_{n-1}. \tag{6.15}$$

Finally, by setting, for $j = 1, 2, \dots, n$

$$L_j = (\log u_j(x_{j-1} - x_j))', \tag{6.16}$$

we can rewrite (6.11) in the following way:

$$\frac{\partial w(x,t)}{\partial t} = \int u_1^{1/p_1}(x - x_1) \dots u_n^{1/p_n}(x_{n-1}) \cdot$$

$$\left(\sum_{j=1}^{n} \frac{\kappa_j}{p_j p_j'} L_j^2 + \sum_{l=1}^{n} \kappa_l \sum_{i \neq j} \frac{a_{i,j}}{p_i p_j} L_i L_j \right) dx_1 \dots dx_{n-1}. \tag{6.17}$$

The sign of the time derivative of $w(x,t)$ depends on the quantity

$$\mathfrak{L}(u_1, \cdots u_n) = \sum_{j=1}^{n} \frac{\kappa_j}{p_j p_j'} L_j^2 + \sum_{l=1}^{n} \kappa_l \sum_{i \neq j} \frac{a_{i,j}}{p_i p_j} L_i L_j. \tag{6.18}$$

Let us set the coefficient of diffusion $\kappa_j = (p_j p_j')^{-1}$, and define $Q_j = L_j / p_j$, for $j = 1, 2, \dots n$. Then

$$\mathfrak{L} = \sum_{j=1}^{n} \left(\frac{1}{p_j'} \right)^2 Q_j^2 + \sum_{l=1}^{n} \frac{1}{p_l p_l'} \sum_{i \neq j} a_{i,j} Q_i Q_j. \tag{6.19}$$

Now, recall that

$$\sum_{j=1}^{n} \frac{1}{p_j} = n - 1$$

implies that, for all $j = 1, 2, \dots n$

$$\frac{1}{p_j} = \sum_{i \neq j} \frac{1}{p'_i}.$$

Consequently

$$\sum_{l=1}^{n} \frac{1}{p_l p'_l} = \sum_{i \neq j} \frac{1}{p'_i p'_j}.$$

Therefore

$$\mathfrak{L} = \sum_{j=1}^{n} \left(\frac{1}{p'_j} \right)^2 Q_j^2 + \sum_{i \neq j} \frac{1}{p'_i p'_j} \sum_{i \neq j} a_{i,j} Q_i Q_j. \tag{6.20}$$

If we now choose, for $i \neq j$

$$a_{i,j} = \frac{(p'_i p'_j)^{-1}}{\sum_{i \neq j} (p'_i p'_j)^{-1}}, \tag{6.21}$$

which is such that $\sum_{i \neq j} a_{i,j} = 1$, we end up with

$$\mathfrak{L} = \sum_{j=1}^{n} \left(\frac{1}{p'_j} \right)^2 Q_j^2 + \sum_{i \neq j} \frac{1}{p'_i p'_j} Q_i Q_j = \left(\sum_{j=1}^{n} \frac{Q_j}{p'_j} \right)^2 \geq 0. \tag{6.22}$$

The previous argument shows that the time derivative of $w(x,t)$ can be made non-negative by suitably choosing the diffusion coefficients κ_j, $j = 1, 2, \dots n$.

Recalling the definition of Q_j (respectively L_j), equality to zero in (6.22) holds if and only if

$$\frac{1}{p_1 p'_1} (\log u_1(x - x_1))' + \sum_{j=2}^{n-1} \frac{1}{p_j p'_j} (\log u_j(x_{j-1} - x_j))' + \frac{1}{p_n p'_n} (\log u_n(x_{n-1}))' = 0. \tag{6.23}$$

As each variable x_i appears as argument of a pair of functions only, it holds that, for every $j = 1, 2, \dots, n-1$

$$\frac{1}{p_j p'_j} \frac{\partial}{\partial x_j} (\log u_j(x_{j-1} - x_j))' + \frac{1}{p_{j+1} p'_{j+1}} \frac{\partial}{\partial x_j} (\log u_{j+1}(x_j - x_{j+1}))' = 0. \tag{6.24}$$

In (6.24) we set $x_0 = x$ and $x_n = 0$. On the other hand, since

$$(\log u_j(x_{j-1} - x_j))' = \frac{\partial}{\partial x_{j-1}} \log u_j(x_{j-1} - x_j) = -\frac{\partial}{\partial x_j} \log u_j(x_{j-1} - x_j),$$

equation (6.24) coincides with

$$\frac{1}{p_j p_j'}\frac{\partial^2}{\partial x_{j-1}^2}\log u_j(x_{j-1}-x_j) = \frac{1}{p_{j+1}p_{j+1}'}\frac{\partial^2}{\partial x_j^2}\log u_{j+1}(x_j - x_{j+1}). \tag{6.25}$$

Note that (6.25) can be verified if and only if the functions on both sides are constant. Thus, there is equality in (6.25) if and only if

$$\log u_j(x) = c\kappa_j x^2 + c_1 x + d_1, \quad \log u_{j+1}(x) = c\kappa_j x^2 + c_2 x + d_2.$$

The integrability of the function u_j then implies that the constant c has to be negative, $c = -C$, where $C > 0$. Hence, there is equality in (6.25) if and only if u_j and u_{j+1} are multiple of Gaussian densities, of variances $C(p_j p_j')^{-1}$ and $C(p_{j+1}p_{j+1}')^{-1}$, respectively, for any given positive constant C. Therefore, equality in (6.22) holds if and only if each function $u_j(x)$, $j = 1, 2, \ldots, n$ is a multiple of a Gaussian density of variance $C(p_j p_j')^{-1}$.
Finally, with this choice of the diffusion coefficients, for every $x \in \mathbb{R}$ and $t_1 < t_2$,

$$u_1^{1/p_1} * u_2^{1/p_2} * \cdots * u_n^{1/p_n}(x, t_1) < u_1^{1/p_1} * u_2^{1/p_2} * \cdots * u_n^{1/p_n}(x, t_2), \tag{6.26}$$

unless all initial data are multiple of Gaussian densities with the right variances. Clearly, (6.26) is equivalent to say that the n-th convolution $w(x, t)$ is monotone increasing. An identical proof holds in higher dimension. This concludes the proof of the Lemma. □

Remark 11. The result of Lemma 10 remains true if each diffusion coefficient k_j is multiplied by a positive constant D. In this case, equality holds if the functions f_j are Gaussian functions with variances Ddk_j.

Remark 12. As already specified in the introduction, our quantity $w(x, t)$ is related to a particular geometric Brascamp–Lieb inequality. Results concerning more general Brascamp–Lieb inequalities that are related to Lemma 10 have been obtained by Bennett, Carbery, Christ and Tao in [18]. This clearly indicates that the proof of Lemma 10 presented here could be extended to cover more general situations.

Lemma 10 has important consequences. Indeed, let us introduce the functional

$$\Psi(t) = \sup_x w(x, t) = \sup_x u_1^{1/p_1} * u_2^{1/p_2} * \cdots * u_n^{1/p_n}(x, t). \tag{6.27}$$

It is a simple exercise to verify that, in view of conditions (6.6) on the constants p_j, the functional $\Psi(t)$ is dilation invariant. In reason of this property we prove:

Theorem 13. *Let $\Psi(t)$ be the functional (6.27), where the functions $u_j(x, t)$, $j = 1, 2, \ldots, n$, are solutions to the heat equation corresponding to the initial values $0 \le f_j(x) \in L^1(\mathbb{R}^d)$, $d \ge 1$. Then, if for each j the exponents p_j satisfy conditions (6.6)* *and the diffusion coefficients are given by $\kappa_j = (p_j p_j')^{-1}$, or by a multiple of them, $\Psi(t)$ is increasing in time from*

$$\Psi(0) = \sup_x f_1^{1/p_1} * f_2^{1/p_2} * \cdots * f_n^{1/p_n}(x)$$

to the limit value

$$\lim_{t\to\infty} \Psi(t) = \prod_{j=1}^n C_{p_j}^d \left(\int_{\mathbb{R}^d} |f_j(x)|\, dx \right)^{1/p_j}. \tag{6.28}$$

The constants C_{p_j} in (6.28) are defined as in (6.4).

Moreover, $\Psi(0) = \lim_{t\to\infty} \Psi(t)$ if and only if $f_j(x)$, $j = 1, 2, \ldots, n$, is a multiple of a Gaussian density of variance $cd\kappa_j$, with $c > 0$.

Proof. Thanks to Lemma 10 we know that the functional $\Psi(t)$ is monotonically increasing from $\Psi(t = 0)$, unless the initial densities are Gaussian functions with the right variances. To conclude the proof, it remains to show that the functional $\Psi(t)$ converges towards the limit value (6.28) as time converges to infinity. The computation of the limit value uses in a substantial way the scaling invariance of Ψ. In fact, thanks to the dilation invariance, at each time $t > 0$, the value of $\Psi(t)$ does not change if we scale each function $u_j(x)$, $j = 1, 2, \ldots, n$, according to (7.36)

$$u_j(x,t) \to U_j(x,t) = \left(\sqrt{1+2t}\right)^d u(x\sqrt{1+2t}, t).$$

Let us choose, as in Lemma 10 smooth initial values rapidly decreasing at infinity. Then, Proposition 2 implies that, for $j = 1, 2, \ldots, n$

$$\lim_{t\to\infty} U_j(x,t) = M_{\kappa_j}(x) \int_{\mathbb{R}^d} f_j(x)\, dx \tag{6.29}$$

at least in $L^1(\mathbb{R}^d)$. Therefore, passing to the limit one obtains

$$\lim_{t\to\infty} \Psi(t) = \prod_{j=1}^n \left(\int_{\mathbb{R}^d} |f_j(x)|\, dx \right)^{1/p_j} \sup_x M_{\kappa_1}^{1/p_1} * M_{\kappa_2}^{1/p_2} * \cdots * M_{\kappa_n}^{1/p_n}(x). \tag{6.30}$$

Owing to the identity

$$M_{\kappa_j}^{1/p_j}(x) = C_{p_j}^d (2\pi)^{(2p_j'/d)^{-1}} M_{1/p_j'}, \tag{6.31}$$

and recalling that $\sum_{j=1}^n (p_j')^{-1} = 1$, we obtain

$$M_{\kappa_1}^{1/p_1} * M_{\kappa_2}^{1/p_2} * \cdots * M_{\kappa_n}^{1/p_n}(x) =$$

$$(2\pi)^{-d/2} \prod_{j=1}^{n} C_{p_j}^d M_1(x) = \prod_{j=1}^{n} C_{p_j}^d \exp\{-|x|^2/2\}.$$

This implies (6.28), and concludes the proof of the inequality for well chosen initial data. Then, the general case of the theorem follows by standard density arguments. □

Remark 14. Theorem 13 is related to the monotonicity in time of a dilation invariant functional whose components are solutions to the heat equation. Therefore, the main importance of the theorem is to highlight the existence of a new Lyapunov functional related to the heat equation. This result, however, can be rephrased to give a new proof of known inequalities in sharp form. Let us set, in Theorem 13

$$g_j(x) = f_j(x)^{1/p_j},$$

for $j = 1, 2, \dots, n$. Then, it holds

$$\sup_x g_1 * g_2 * \cdots * g_n(x) \le \prod_{j=1}^{n} C_{p_j}^d \prod_{j=1}^{n} \|g_j\|_{p_j}. \tag{6.32}$$

Moreover, since

$$\sup_x g_1 * g_2 * \cdots * g_n(x) \ge \int g_1(-x_1) g_2(x_1 - x_2) \dots g_n(x_{n-1})\, dx_1 \dots dx_{n-1},$$

inequality (6.32) implies, under the same conditions on the constants p_j,

$$\int g_1(x_1) g_2(x_1 - x_2) \dots g_n(x_{n-1})\, dx_1 \dots dx_{n-1} \le \prod_{j=1}^{n} C_{p_j}^d \prod_{j=1}^{n} \|g_j\|_{p_j}. \tag{6.33}$$

Inequality (6.33) is a particular case of the inequalities obtained by Brascamp and Lieb [28] by a different method.

Remark 15. Clearly, the proof of Theorem 13 still holds when $n = 2$. In this case, however, the diffusion coefficients κ_j, $j = 1, 2$ coincide. In fact, when $n = 2$, the condition (6.6) reduces to

$$1 \le p_j \le +\infty; \quad \frac{1}{p_1} + \frac{1}{p_2} = 1,$$

so that p_1 and p_2 are dual exponents. Consequently $p_1' = p_2$ and $p_2' = p_1$, which imply $\kappa_1 = \kappa_2 = \kappa = (p_1 p_2)^{-1}$. But in this case the definition (6.4) of the constant C_p implies $C_{p_1} = 1/C_{p_2}$, and the limit (6.28) takes the value

$$\lim_{t\to\infty} \Psi(t) = \left(\int_{\mathbb{R}^d} |f_1(x)|\, dx \right)^{1/p_1} \left(\int_{\mathbb{R}^d} |f_2(x)|\, dx \right)^{1/p_2}. \tag{6.34}$$

Note that in this case inequality (6.33) reduces simply to the classical Hölder inequality.

Remark 16. As noticed by Brascamp and Lieb [28], Theorem 13 contains as special case the best possible improvement to Young's inequality. If $n = 3$ (6.33) reads

$$\int_{\mathbb{R}^{2d}} f(x)g(x-y)h(y)dx\,dy \le (C_p C_q C_s)^d \|f\|_{L^p}\|g\|_{L^q}\|h\|_{L^s}, \qquad (6.35)$$

where $1 \le p, q, s \le \infty$, $1/p + 1/q + 1/s = 2$, and equality holds when f, g, h are suitable Gaussian functions. Choosing

$$h(y) = (f * g(y))^{r-1}$$

leads to an equivalent form of (6.35)

$$\|f * g\|_{L^r} \le (C_p C_q C_{r'})^d \|f\|_{L^p}\|g\|_{L^q}, \qquad (6.36)$$

namely the standard form of Young's inequality [14, 28].

Also, repeated applications of (6.36) give

$$\|g_1 * g_2 * \cdots * g_n\|_r \le C_{r'}^d \prod_{j=1}^n C_{p_j}^d \|g_j\|_{p_j}, \qquad (6.37)$$

where $1 \le p_j \le \infty$ and $\sum_{j=1}^n 1/p_j = n - 1 + 1/r$.

6.2 Young inequality and its reverse

Theorem 13 shows the monotonicity in time of the L^∞-norm of the n-th convolution of type (6.27), as well as its convergence towards an explicitly computable limit value (in terms of the initial data). The key point in getting this result was the dilation property of the functional $\Psi(t)$.

To get a similar result for the L^r-norm of the n-th convolution $w(x,t)$, $r > 0$, and to obtain the (eventual) limit value, we need that the dilation property still holds for $\|w(t)\|_r$. By applying the same scaling $u_j(x) \to V_j(x) = a^d V(ax)$ to each function $u_j(x)$ in (6.2) we get

$$V_1^{\alpha_1} * V_2^{\alpha_2} * \cdots * V_n^{\alpha_n}(x) = a^{d\gamma} u_1^{\alpha_1} * u_2^{\alpha_2} * \cdots * u_n^{\alpha_n}(ax) = a^{d\gamma} w(ax),$$

where

$$\gamma = \sum_{j=1}^n \alpha_j - n + 1$$

Hence

$$\int_{\mathbb{R}^d} \left(V_1^{\alpha_1} * V_2^{\alpha_2} * \cdots * V_n^{\alpha_n}(x)\right)^r \, dx = \int_{\mathbb{R}^d} a^{dr\gamma} w^r(ax)\, dx,$$

and dilation invariance occurs if and only if $r\gamma = 1$, that is

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = n - 1 + \frac{1}{r}. \tag{6.38}$$

By analogy with condition (6.6), we will satisfy condition (6.38) in two separate cases. The first refers to fix, for $j = 1, \ldots, n$, positive real numbers p_j and r such that

$$p_j < 1, r < 1; \quad \sum_{j=1}^{n} \frac{1}{p_j} = n - 1 + \frac{1}{r}. \tag{6.39}$$

The second refers to fix, for $j = 1, \ldots, n$, positive real numbers p_j and r such that

$$1 < p_j \leq \infty, 1 < r \leq \infty; \quad \sum_{j=1}^{n} \frac{1}{p_j} = n - 1 + \frac{1}{r}. \tag{6.40}$$

In the following, we will analyze the time behaviour of $\|w(t)\|_r$ in the case (6.39). Then, the result for the case (6.40) will follow by the same line of proof.

Condition (6.39) implies that $p_j' < 0$ for all $j = 1, 2, \ldots, n$, and

$$\frac{1}{p_j} = \sum_{i \neq j} \frac{1}{p_i'} + \frac{1}{r}.$$

Making use of the proof of Lemma 10, let us set, for $j = 1, 2, \ldots, n$, the (positive) coefficients of diffusion

$$\kappa_j = \frac{1}{p_j |p_j'|}. \tag{6.41}$$

Then, by means of elementary computations we obtain

$$\sum_{l=1}^{n} \kappa_l = \sum_{l=1}^{n} \frac{1}{p_l |p_l'|} = \sum_{i \neq j} \frac{1}{p_i' |p_j'|} + \frac{1}{r |r'|}. \tag{6.42}$$

Since the real numbers p_j now satisfy condition (6.39), the quantity (6.18) considered in Lemma 10, with the same choice (6.21) of the coefficients $a_{i,j}$ takes the form

$$\mathfrak{L} = -\left(\sum_{j=1}^{n} \frac{Q_j}{p_j'} \right)^2 + \frac{1}{r|r'|} \sum_{i \neq j} a_{i,j} Q_i Q_j. \tag{6.43}$$

It is evident that in this case we cannot expect that $\mathfrak{L}$ has a definite sign. However, using expression (6.43) into (6.17) we obtain

$$\frac{\partial w(x,t)}{\partial t} = \int u_1^{1/p_1}(x - x_1) \ldots u_n^{1/p_n}(x_{n-1}) \mathfrak{L}(u_1, \cdots u_n) \, dx_1 \cdots dx_{n-1} =$$

$$-\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\left(\sum_{j=1}^n \frac{Q_j}{p_j'}\right)^2 dx_1\cdots dx_{n-1} + \frac{1}{r|r'|}\frac{\partial^2 w}{\partial x^2}. \qquad (6.44)$$

In fact, by formula (6.15)

$$\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\sum_{i\neq j} a_{i,j}Q_iQ_j\, dx_1\cdots dx_{n-1} =$$

$$\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\sum_{i\neq j} \frac{a_{i,j}}{p_ip_j}L_iL_j\, dx_1\cdots dx_{n-1} = \frac{\partial^2 w}{\partial x^2}.$$

Consequently, thanks to (6.44)

$$\frac{d}{dt}\int w^r(x,t)\,dx = r\int w^{r-1}(x,t)\frac{\partial w(x,t)}{\partial t}\,dx = \frac{1}{|r'|}\int w^{r-1}\frac{\partial^2 w}{\partial x^2}\,dx +$$

$$-r\int w^{r-1}\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\left(\sum_{j=1}^n \frac{Q_j}{p_j'}\right)^2 dx_1\cdots dx_{n-1}\,dx =$$

$$-r\int w^{r-1}\,dx\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\left(\sum_{j=1}^n \frac{Q_j}{p_j'}\right)^2 dx_1\cdots dx_{n-1} +$$

$$\frac{(1-r)^2}{r}\int w^{r-2}\left(\frac{\partial w}{\partial x}\right)^2 dx \qquad (6.45)$$

Surprisingly, the expression on (6.45) has a sign. This is consequence of the following Lemma, which generalizes a similar result that dates back to Blachman [20]. In case of convolution of two functions, analogous result has been obtained recently in [82]. For the sake of simplicity, we will present the proof in dimension $d = 1$.

Lemma 17. *Let $w(x)$ be the (smooth) n-th convolution defined by (6.27). Then, for any set of positive constants p_j and r, and positive constants λ_j, $j = 1, 2, \dots, n$ such that $\sum_{j=1}^n \lambda_j = 1$ it holds*

$$\int w^{r-2}\left(\frac{\partial w}{\partial x}\right)^2 \le \int w^{r-1}(x)\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\left(\sum_{j=1}^n \frac{\lambda_j}{p_j}L_j\right)^2. \qquad (6.46)$$

Moreover, equality in (6.46) holds if and only if any function u_j, $j = 1, 2, \dots, n$ is multiple of a Gaussian function of variance λ_j/p_j.

Proof. By property (6.14), if we take a set of positive constants λ_j, $j = 1, 2, \dots, n$ such that $\sum_{j=1}^n \lambda_j = 1$ we can express the first derivative of $w(x)$ as

$$w'(x) = \int u_1^{1/p_1}(x - x_1) \dots u_n^{1/p_n}(x_{n-1}) \sum_{j=1}^n \frac{\lambda_j}{p_j} L_j dx_1 \dots dx_{n-1},$$

where L_j is defined as in (6.16). Therefore, by denoting

$$d\mu_x(x_1, \dots, x_{n-1}) = \frac{u_1^{1/p_1}(x - x_1) \dots u_n^{1/p_n}(x_{n-1})}{w(x)} dx_1 \cdots dx_n, \tag{6.47}$$

we obtain

$$\frac{w'(x)}{w(x)} = \int \sum_{j=1}^n \frac{\lambda_j}{p_j} L_j d\mu_x(x_1, \dots, x_{n-1}).$$

Note that, for any $x \in \mathbb{R}$ the measure $d\mu$ defined in (6.47) is a unit measure on $\mathbb{R}^{n-1}$,

$$\int_{\mathbb{R}^{n-1}} d\mu_x(x_1, x_2, \cdots x_{n-1}) = 1.$$

Jensen's inequality then gives

$$\left(\frac{w'(x)}{w(x)}\right)^2 \le \int \left(\sum_{j=1}^n \frac{\lambda_j}{p_j} L_j\right)^2 d\mu_x(x_1, \dots, x_{n-1}). \tag{6.48}$$

Multiplying both sides of (6.48) by $w^r(x)$, and integrating over x proves the Lemma.

Note that, since equality in Jensen's inequality holds if and only if the argument is constant, equality in (6.48) holds if and only if

$$\sum_{j=1}^n \frac{\lambda_j}{p_j} L_j = const.$$

Hence, the reasoning of the last part of Lemma 10 can be repeated to show that there is equality in (6.46) if and only if any function u_j, $j = 1, 2, \dots, n$ is multiple of a Gaussian function of variance λ_j/p_j. □

Let us return to formula (6.45). Conditions (6.39) imply that

$$\sum_{j=1}^n \frac{1}{|p'_j|} = \frac{1}{|r'|}.$$

Hence

$$\frac{r}{1-r} \sum_{j=1}^n \frac{1}{|p'_j|} = 1.$$

Choosing then

$$\lambda_j = \frac{r}{1-r}\frac{1}{|p_j'|},$$

we obtain that (6.46) reads

$$\int w^{r-2}\left(\frac{\partial w}{\partial x}\right)^2 dx \le \frac{r^2}{(1-r)^2}\int w^{r-1}(x)\,dx\cdot$$

$$\cdot\int u_1^{1/p_1}(x-x_1)\dots u_n^{1/p_n}(x_{n-1})\left(\sum_{j=1}^n \frac{1}{p_j'}Q_j\right)^2 dx_1\cdots dx_n. \tag{6.49}$$

This shows that the quantity in (6.45) is negative. Hence, we proved that, if the positive constants p_j and r satisfy conditions (6.39), the functional

$$\Lambda(t) = \|w(t)\|_r = \left(\int (u_1^{1/p_1} * u_2^{1/p_2} * \cdots * u_n^{1/p_n})^r(x,t)\,dx\right)^{1/r} \tag{6.50}$$

is monotone decreasing. Since we know that, in view of conditions (6.39) on the constants p_j, the functional $\Lambda(t)$ is dilation invariant, we proved:

Theorem 18. *Let $\Lambda(t)$ be the functional (6.50), where the functions $u_j(x,t)$, $j = 1,2,\dots,n$, are solutions to the heat equation corresponding to the initial values $0 \le f_j(x) \in L^1(\mathbb{R}^d)$, $d \ge 1$. Then, if for each j the exponents p_j satisfy conditions (6.39) and the diffusion coefficients are given by $\kappa_j = (p_j|p_j'|)^{-1}$, $\Lambda(t)$ is decreasing in time from*

$$\Lambda(0) = \left(\int \left(f_1^{1/p_1} * f_2^{1/p_2} * \cdots * f_n^{1/p_n}(x)\right)^r dx\right)^{1/r}$$

to the limit value

$$\lim_{t\to\infty}\Lambda(t) = C_{r'}^d \prod_{j=1}^n C_{p_j}^d \left(\int_{\mathbb{R}^d} |f_j(x)|\,dx\right)^{1/p_j}. \tag{6.51}$$

The constants C_{p_j} in (6.28) are defined by

$$C_p^2 = \frac{p^{1/p}}{|p'|^{1/p'}}, \tag{6.52}$$

Moreover, $\Lambda(0) = \lim_{t\to\infty}\Lambda(t)$ if and only if $f_j(x)$, $j = 1,2,\dots,n$, is a multiple of a Gaussian density of variance $d\kappa_j$.

Proof. We know that the functional $\Lambda(t)$ is monotonically decreasing from $\Lambda(t = 0)$, unless the initial densities are Gaussian functions with the right variances. In addition, $\Lambda(t)$ is dilation invariant. As in Theorem 13, let us scale each function $u_j(x)$, $j = 1, 2, \dots, n$, according to (7.36). Therefore, by the central limit property, passing to the limit one obtains

$$\lim_{t\to\infty} \Lambda(t) = \prod_{j=1}^{n} \left(\int_{\mathbb{R}^d} |f_j(x)|\, dx \right)^{1/p_j} \left\| M_{\kappa_1}^{1/p_1} * M_{\kappa_2}^{1/p_2} * \cdots * M_{\kappa_n}^{1/p_n} \right\|_r . \qquad (6.53)$$

The value of the integral can be evaluated by using formula (6.31) of Theorem 13, with the additional difficulty to evaluate the norm of a Gaussian in L^r. Thanks to condition (6.39) we obtain

$$\left\| M_{\kappa_1}^{1/p_1} * M_{\kappa_2}^{1/p_2} * \cdots * M_{\kappa_n}^{1/p_n} \right\|_r = C_{r'}^d \prod_{j=1}^{n} C_{p_j}^d .$$

This concludes the proof of the theorem. □

The computations leading to Theorem 18 can be repeated step-by-step in the case in which the p_j's and r satisfy condition (6.40). In this case, however, the sign of $\mathfrak{L}$ changes, and we obtain

Theorem 19. *Let $\Lambda(t)$ be the functional (6.50), where the functions $u_j(x,t)$, $j = 1, 2, \dots, n$, are solutions to the heat equation corresponding to the initial values $0 \le f_j(x) \in L^1(\mathbb{R}^d)$, $d \ge 1$. Then, if for each j the exponents p_j satisfy conditions (6.40) and the diffusion coefficients are given by $\kappa_j = (p_j p_j')^{-1}$, $\Lambda(t)$ is increasing in time from*

$$\Lambda(0) = \left(\int \left(f_1^{1/p_1} * f_2^{1/p_2} * \cdots * f_n^{1/p_n}(x) \right)^r dx \right)^{1/r}$$

to the limit value

$$\lim_{t\to\infty} \Lambda(t) = C_{r'}^d \prod_{j=1}^{n} C_{p_j}^d \left(\int_{\mathbb{R}^d} |f_j(x)|\, dx \right)^{1/p_j} . \qquad (6.54)$$

The constants C_{p_j} in (6.28) are defined by (6.4). Moreover, $\Lambda(0) = \lim_{t\to\infty} \Lambda(t)$ if and only if $f_j(x)$, $j = 1, 2, \dots, n$, is a multiple of a Gaussian density of variance $d\kappa_j$.

Remark 20. The monotonicity property of the functional $\Lambda(t)$ defined by (6.50) have been noticed first by Bennett and Bez [16] by means of a different approach. Consequently, the results of both Theorems 18 and 19 also follow from their arguments. We note, however, that the dilation invariance property of $\Lambda(t)$, which is at the basis of the direct proof of the Theorems, has not explicitly taken into account before.

Remark 21. Theorems 18 and 19 show the monotonicity properties of the L^r-norm of the n-th convolution of powers of solutions to the heat equation. As discussed at the end of Theorem 13, apart from its intrinsic physical interest, this monotonicity can be rephrased in the form of inequalities for convolutions in sharp form. In particular, when $n = 2$, Theorem 18 contains the sharp form of Young inequality in the so-called reverse case

$$\|f * g\|_r \geq (C_p C_q C_{r'})^d \|f\|_p \|g\|_q, \tag{6.55}$$

where $0 < p, q, r < 1$ while $1/p + 1/q = 1 + 1/r$, and C_p is defined by (6.52).

Remark 22. A particular case of Theorem 19 implies Babenko's inequality [3] (cf. also Beckner [14]):

$$\|\mathfrak{F} f\|_{p'} \leq C_p^d \|f\|_p, \tag{6.56}$$

where C_p is defined as in (6.4), p' is an even integer $p' = 2, 4, 6, \dots$, and $\mathfrak{F} f$ denotes the Fourier transform of f. Here the Fourier transform is defined for integrable functions by

$$\mathfrak{F} f(\xi) = \int_{\mathbb{R}^d} \exp\{-2\pi i x \cdot \xi\} f(x)\, dx$$

Inequality (6.56) follows by choosing in Theorem 19 $r = 2$ and $1/p_j = 1/p = (2n-1)/2n$, $j = 1, 2, \dots, n$, which are such that condition (6.40) is satisfied. In this case, in fact, by setting $f_j = f^{2n/(2n-1)}$, for $j = 1, 2, \dots, n$, we obtain that f satisfies the inequality

$$\left(\int (\underbrace{f * f * \cdots * f}_{n})^2 \, dx \right)^{1/2} \leq C_p^{dn} \|f\|_p^n.$$

Since

$$\mathfrak{F} \left(\underbrace{f * f * \cdots * f}_{n} \right) = (\mathfrak{F} f)^n,$$

by Plancherel's identity we conclude that

$$\left(\int (\mathfrak{F} f)^{2n} \, d\xi \right)^{1/(2n)} \leq C_p^d \|f\|_p. \tag{6.57}$$

We remark that, as explicitly mentioned in [16], the monotonicity of the quantity in (6.57) also follows from the results in [18] (cf. also [16]). A further inside into Haussdorff–Young inequality, with an interesting discussion about the behavior in time of $\|\mathfrak{F} u^{1/p}(t)\|_{p'}$ can be found in [17].

6.3 Monotonicity and Prékopa–Leindler inequality

The analysis of the preceding section shows the monotonicity properties of the L^r norm of the n-th convolution of powers of the solutions to the heat equation. In particular Theorem 13 covers the L^∞ case, while Theorem 18 (respectively Theorem 19) cover the case $r < 1$ (respectively $r > 1$). Two limit cases remain to be examined, namely the cases $r \to 0$ and $r \to 1$. Here we will briefly discuss the first case, leaving the second to the next section.

Given a set of positive constants q_j, $j = 1, 2, \dots, n$, such that $\sum_{j=1}^n 1/q_j = 1$, and a constant $N \gg 1$, we choose in Theorem 18

$$p_j = \frac{q_j}{N}, \quad r = \frac{1}{N - (n-1)}. \tag{6.58}$$

Then, if $N \geq \max_j q_j + n$, conditions (6.39) are satisfied and the monotonicity of $\Lambda(t)^r$ is guaranteed. On the other hand, if (6.58) holds, one can write

$$w(x)^r = \left(u_1^{N/q_1} * u_2^{N/q_2} * \cdots * u_n^{N/q_n}(x,t)\right)^{1/(N-n+1)} =$$

$$\left(\int \left(u_1(x-x_1)^{1/q_1} \cdots u_n(x_{n-1})^{1/q_n}\right)^N dx_1 \cdots dx_{n-1}\right)^{1/(N-n+1)}. \tag{6.59}$$

We can render the dependence of $\Lambda(t)^r$ on N explicit by setting

$$\int w(x)^r \, dx = \Upsilon_N(t),$$

where

$$\Upsilon_N(t) = \left(\int \left(u_1(x-x_1)^{1/q_1} \cdots u_n(x_{n-1})^{1/q_n}\right)^N dx_1 \cdots dx_{n-1}\right)^{1/(N-n+1)}. \tag{6.60}$$

By Theorem 18, $\Upsilon_N(t)$ is monotonically decreasing in time for all values of $N \geq \max_j q_j + n$, provided the coefficients of diffusion are the correct ones. Moreover, note that, for any given N, the coefficients of diffusions κ_j depend on it, and

$$\kappa_j^N = \frac{N(N-q_j)}{q_j^2}.$$

On the other hand, Theorem 18 remains true if we multiply all coefficients of diffusion by the same constant. Therefore, without affecting the monotonicity of $\Upsilon_N(t)$ we can fix the coefficients of diffusion equal to

$$\tilde{\kappa}_j^N = \frac{N(N-q_j)}{N^2}\frac{1}{q_j^2} = \frac{1}{q_j^2} - \frac{1}{Nq_j}. \tag{6.61}$$

By letting $N \to \infty$, the coefficients of diffusion $\tilde{\kappa}_j^N$ tend to the values $\tilde{\kappa}_j^\infty = 1/q_j^2$. Consequently, recalling expression (6.59), for each $t > 0$, as $N \to \infty$ we have pointwise convergence of $w(x,t)^{r(N)}$, and

$$\lim_{N\to\infty} w(x,t)^{r(N)} = ess\ sup_{x_1,\dots,x_{n-1}} u_1(x-x_1,t)^{1/q_1} \cdots u_n(x_{n-1},t)^{1/q_n}\, dx, \tag{6.62}$$

where now $u_j(x,t)$, for $j = 1, 2\dots, n$ is solution to the heat equation (3.1) with diffusion coefficient $1/q_j^2$.

Suppose now that the initial data $f_j(x)$ in the heat equation are bounded and have bounded support. Then, they are bounded from above by multiples of some Gaussian functions, that is, for $j = 1, 2, \dots, n$

$$f_j(x) \le A_j M_{2\sigma_j}(x),$$

where A_j and σ_j are suitable positive constants. In this case, the solution to the heat equation with diffusion coefficient $\tilde{\kappa}_j^N$ satisfies the bound

$$u_j(x,t) = f_j * M_{2\tilde{\kappa}_j^N t}(x) \le A_j M_{2(\sigma_j + \tilde{\kappa}_j^N t)}. \tag{6.63}$$

Consequently,

$$w(x,t)^r \le \left(\prod_{j=1}^n A_j^{p_j} M_{2\kappa_1 t}^{p_1} * M_{2\kappa_2 t}^{p_2} * \cdots * M_{2\kappa_n t}^{p_n} \right)^r,$$

where the constants p_j and r are given as in (6.58), and

$$\kappa_j t = \sigma_j + \left(\frac{1}{q_j^2} - \frac{1}{N q_j} \right) t.$$

Then, we easily obtain that, for sufficiently large N,

$$w^r(x,t) \le C_1(t) M_{2C_2(t)}(x),$$

where the positive functions $C_1(t)$ and $C_2(t)$ do not depend on N. By the dominated convergence theorem it follows that

$$\lim_{N\to\infty} \Upsilon_N(t) = \Upsilon(t),$$

where

$$\Upsilon(t) = \int ess\ sup_{x_1,\dots,x_{n-1}} u_1(x-x_1)^{1/q_1} \cdots u_n(x_{n-1})^{1/q_n}\, dx. \tag{6.64}$$

Moreover, $\Upsilon(t)$ is monotonically decreasing in time if the coefficients of diffusion in the heat equations are given by $\kappa_j = 1/q_j^2$. Since the functional $\Upsilon(t)$ is invariant under dilation, we can pass to the limit to find the lower bound. By the same argument of the proof of Theorem 13, we conclude that the limit value is obtained by setting

$$u_j(x) = \int f_j(x)\, dx M_{1/q_j^2}.$$

Explicit computations then show that

$$\lim_{t\to\infty} \Upsilon(t) = \prod_{j=1}^{n} q_j^{d/q_j} \left(\int f_j(x)\, dx. \right)^{1/q_j} \tag{6.65}$$

The choice $f_j(x) = g_j(q_j x)$ then implies

$$\int_{\mathbb{R}^d} f_j(x)^{1/q_j}\, dx = q_j^{-d/q_j} \left(\int_{\mathbb{R}^d} g_j(x)^{1/q_j}\, dx \right)^{1/q_j}.$$

The case of general data follows by a density argument. We conclude with the following

Theorem 23. *Let $\Upsilon(t)$ denote the functional (6.64) where the functions $u_j(x,t)$, $j = 1, 2, \dots, n$, are solutions to the heat equation corresponding to the initial values $0 \le f_j(x) \in L^1(\mathbb{R}^d)$, $d \ge 1$. Then, if exponents q_j satisfy $\sum_{j=1}^n q_j^{-1} = 1$, and the diffusion coefficients are given by $\kappa_j = q_j^{-2}$, $\Upsilon(t)$ is decreasing in time from*

$$\Upsilon(0) = \int \text{ess sup}_{x_1,\dots,x_{n-1}} f_1(q_1(x - x_1))^{1/q_1} \cdots f_n(q_n(x_{n-1}))^{1/q_n}\, dx$$

to the limit value

$$\lim_{t\to\infty} \Upsilon(t) = \prod_{j=1}^{n} \left(\int_{\mathbb{R}^d} |f_j(x)|\, dx \right)^{1/q_j}. \tag{6.66}$$

Moreover, $\Upsilon(0) = \lim_{t\to\infty} \Upsilon(t)$ if and only if $f_j(x)$, $j = 1, 2, \dots, n$, is a multiple of a Gaussian density of variance $d\kappa_j$.

Remark 24. If $n = 2$ the monotonicity of the functional Υ proven in Theorem 23 implies the classical Prékopa–Leindler inequality. In this case, in fact one obtains the Prékopa–Leindler theorem [60, 72, 73] that reads

$$\|h\|_1 \ge \|f\|_1^{\lambda} \|g\|_1^{1-\lambda},$$

where

$$h(x|f,g) = ess\ sup_y f \left(\frac{x-y}{\lambda} \right)^{\lambda} g \left(\frac{y}{1-\lambda} \right)^{1-\lambda}.$$

The derivation of Prékopa–Leindler inequality from the Young's inequality has been obtained by Brascamp and Lieb [28]. Our result, however, enlightens a new meaning of this inequality, that is viewed as a consequence of the monotonicity of a Lyapunov functional of the convolution of two powers of the solution to the heat equation.

Remark 25. Theorem 23 is a corollary of the general result of Theorem 18. However, a direct proof of monotonicity could be possible by looking at the functional (6.64) directly.

6.4 Another proof of entropy power inequality

As noticed by Lieb [61], the entropy power inequality (1.4) can also be proven as a limit case of the Young inequality in the sharp form (6.36), by letting the parameters p, q and r tend to one. This result can be obtained as follows. Let $0 < a < 1$ denote a fixed constant. For a given (small) positive χ, let us consider Young's inequality (6.36) with

$$r = 1 + \chi, \qquad p = \frac{1+\chi}{1+a\chi}, \qquad q = \frac{1+\chi}{1+(1-a)\chi}, \tag{6.67}$$

which are such that

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}.$$

Note that, as $\chi \to 0$, $p, q, r \to 1$. Let f, g, h be smooth probability densities, and let us define $z(\chi) = \|h\|_{1+\chi}$. Then $z(0) = 1$, and thanks to the identity

$$z(\chi) = \exp\left\{ \frac{1}{1+\chi} \log \int_{\mathbb{R}^d} h^{1+\chi}\, dx \right\},$$

one evaluates straightforwardly

$$z'(\chi) = z(\chi) \left[-\frac{1}{(1+\chi)^2} \log \int h^{1+\chi} + \frac{1}{1+\chi} \frac{\int h^{1+\chi} \log h}{\int h^{1+\chi}} \right]. \tag{6.68}$$

Hence,

$$z'(0) = \int_{\mathbb{R}^d} h \log h\, dx = -H(h).$$

Owing to the smoothness of $f * g$, we can expand $\|f * g\|_{1+\chi}$ in Taylor's series of χ up to order one, to obtain

$$\|f * g\|_{1+\chi} = 1 - H(f * g)\chi + o_1(\chi), \tag{6.69}$$

where $o_1(\chi)$ is such that $o_1(\chi)/\chi \to 0$ as $\chi \to 0$. Analogous computations for the function

$$\omega(\chi) = \exp\left\{ d \log(C_p C_q C_{r'}) + \frac{1}{p} \log \int_{\mathbb{R}^d} f^p\, dx + \frac{1}{q} \log \int_{\mathbb{R}^d} g^q\, dx \right\}$$

where p and q are defined in (6.67), allow to conclude that

$$\omega'(0) = \frac{d}{2} \left(a \log a + (1-a) \log(1-a) \right) - (1-a) H(f) - a H(g). \tag{6.70}$$

Therefore, expanding again in Taylor's series of χ, we obtain

$$\omega(\chi) = 1 + \left(\frac{d}{2} \left(a \log a + (1-a) \log(1-a) \right) - (1-a) H(f) - a H(g) \right) \chi + o_2(\chi), \tag{6.71}$$

where again $o_2(\chi)/\chi \to 0$ as $\chi \to 0$. It is interesting to remark that the sharp constant $(C_p C_q C_{r'})^d$ furnishes an important contribution to formula (6.69). This contribution can be derived straightforwardly using the identity

$$\frac{d}{d\chi} \left(\frac{1}{p} \right) = - \frac{d}{d\chi} \left(\frac{1}{p'} \right).$$

This gives

$$\frac{d}{d\chi} \log C_p^2 = \frac{d}{d\chi} \left(\frac{1}{p} \log p - \frac{1}{p'} \log p' \right) = \left(-2 + \log \frac{p}{p'} \right) \frac{d}{d\chi} \left(\frac{1}{p} \right) =$$

$$\left(-2 + \log(p-1) \right) \frac{d}{d\chi} \left(\frac{1}{p} \right) = \frac{1-a}{(1+\chi)^2} \left(2 - \log \frac{(1-a)\chi}{1+a\chi} \right),$$

and

$$\frac{d}{d\chi} \log(C_p C_q C_{r'})^2 = \frac{1}{(1+\chi)^2} \left((1-a) \log \frac{1-a}{1+a\chi} + a \log \frac{a}{1+(1-a)\chi} \right).$$

In conclusion we have the following [61]:

Lemma 26. *Let the probability densities $f(x)$ and $g(x)$ $x \in \mathbb{R}^d$ possess bounded Shannon's entropy functional. Then, for any positive constant $0 < a < 1$ the following inequality holds*

$$H(f * g) \geq (1-a) H(f) + a H(g) - \frac{d}{2} \left(a \log a + (1-a) \log(1-a) \right). \tag{6.72}$$

Proof. The proof is a direct consequence of the sharp Young inequality (6.36). With our notations, Young inequality can be rephrased as $z(\chi) - \omega(\chi) \leq 0$. Using expansions (6.69) and (6.71), and letting $\chi \to 0$, inequality (6.72) follows for smooth densities. A standard density argument then concludes the proof. □

Shannon's entropy power inequality then follows by maximizing the right-hand side of inequality (6.72). A simple computation shows that the right-hand side, say $A(a, H(f), H(g))$ attains the maximum when

$$a = \bar{a} = \frac{\exp\left\{ 2 \left(H(g) - H(f) \right) / d \right\}}{1 + \exp\left\{ 2 \left(H(g) - H(f) \right) / d \right\}}, \tag{6.73}$$

and, for $a = \bar{a}$

$$A(\bar{a}, H(f), H(g)) = \frac{d}{2} \log \left\{ \exp\left(2H(f)/d\right) + \exp\left(2H(g)/d\right) \right\}. \tag{6.74}$$

With analogous computations, Shannon's entropy-power inequality can be easily extended to a convolution of n probability densities by means of Theorem 19.

While the result of Lieb [61] outlines an interesting connection between Young's inequality and the entropy power inequality, the proof of *EPI* via Young's inequality does not contain any connection with our idea about monotonicity properties of Lyapunov functionals for the solution to the heat equation. Indeed, a much simpler direct proof is available by making use of this idea. For the moment, let us fix the dimension equal to 1.

Let as usual $w(x,t)$ denote the n-th convolution

$$w(x,t) = u_1 * u_2 * \cdots * u_n(x,t), \tag{6.75}$$

where the functions $u_j(x,t)$, $j = 1, 2, \ldots, n$, are solutions to the heat equations, with coefficients of diffusion κ_j, corresponding to the initial probability densities $0 \le f_j(x)$ with bounded Shannon's entropy. It is important to note that, in view of the closure property of the Gaussian density (1.3) with respect to convolutions, $w(x,t)$ itself satisfies the heat equation (3.1) with coefficient of diffusion $\kappa = \sum_{j=1}^n \kappa_j$. For any set of positive values γ_j, $j = 1, 2, \ldots, n$, such that $\sum_{j=1}^n \gamma_j = 1$, we introduce the functional

$$\Phi(t) = H(w(t)) - \sum_{j=1}^n \gamma_j H(u_j(t)). \tag{6.76}$$

Let f_α be the scaled function defined as in (8.3). Since, for $\alpha > 0$

$$H(f_\alpha) = H(f) - \log \alpha,$$

the functional $\Phi(t)$ is dilation invariant. Given $t > 0$, let us evaluate the time derivative of $\Phi(t)$. We obtain

$$\frac{d}{dt} H(w(t)) = \kappa I(w(t)) - \sum_{j=1}^n \gamma_j \kappa_j I(u_j(t)), \tag{6.77}$$

where as usual $I(f)$ is the Fisher information of the density f, defined in (1.8). By setting in Lemma 17 $r = 1$ and $p_j = 1$, $j = 1, 2, \ldots, n$, which satisfy conditions (6.40), inequality (6.46) assumes the form

$$I(w) \le \int dx \int u_1(x - x_1) \ldots u_n(x_{n-1}) \left(\sum_{j=1}^n \lambda_j L_j \right)^2 = \sum_{j=1}^n \lambda_j^2 I(u_j). \tag{6.78}$$

Formula (6.78) follows simply owing to the definition of L_j, and applying Fubini's theorem. The proof of (6.78) in the case of the convolution of two functions goes back to Blachman [20], and it is contained into Lemma 5, with $r = 1$. Thanks to (6.78), by setting the constants $\gamma_j = k_j/k$, we have at once that these constants satisfy the condition $\sum_{j=1}^n \gamma_j = 1$, and that the sign of the derivative (6.77), consequent to this choice, is negative, unless the functions u_j are Gaussian. Since the functional $\Phi(t)$ is dilation invariant, we can pass to the limit $t \to \infty$ obtaining

$$\lim_{t\to\infty} \Phi(t) = H(M_\kappa) - \sum_{j=1}^n \frac{\kappa_j}{\kappa} H(M_{\kappa_j}). \tag{6.79}$$

Since

$$H(M_\sigma) = \frac{1}{2}\log 2\pi\sigma,$$

we obtain from (6.79)

$$\lim_{t\to\infty} \Phi(t) = -\frac{1}{2}\sum_{j=1}^n \frac{\kappa_j}{\kappa} \log\frac{\kappa_j}{\kappa}. \tag{6.80}$$

Clearly, the same result holds in dimension $d \geq 1$. Hence we proved the following:

Theorem 27. *Let $\gamma_j \geq 0$, $j = 1, 2, \dots, n$ be such that $\sum_{j=1}^n \gamma_j = 1$, and let $\Phi(t)$ be the functional (6.76), where the functions $u_j(x,t)$, $j = 1, 2, \dots, n$, are solutions to the heat equation corresponding to the initial probability densities $f_j(x) \in L^1(\mathbb{R}^d)$, $d \geq 1$. Then, if the diffusion coefficients $\kappa_j = C\gamma_j$, $j = 1, 2, \dots, n$ and $C > 0$, $\Phi(t)$ is decreasing in time from*

$$\Phi(0) = H(f_1 * f_2 * \cdots * f_n) - \sum_{j=1}^n \gamma_j H(f_j)$$

to the limit value

$$\lim_{t\to\infty} \Phi(t) = -\frac{d}{2}\sum_{j=1}^n \gamma_j \log\gamma_j. \tag{6.81}$$

Moreover, $\Phi(0) = \lim_{t\to\infty}\Phi(t)$ if and only if $f_j(x)$, $j = 1, 2, \dots, n$, is a Gaussian density of variance $d\kappa_j$.

Theorem 27 shows the monotonicity of a dilation invariant functional linked to the Shannon's entropy of a n-th convolution of probability density functions. A direct consequence of this monotonicity is the entropy power inequality. Indeed, the monotonicity of $\Phi(t)$ implies that, for any choice of the constants γ_j, with $\sum_{j=1}^n \gamma_j = 1$

$$H(f_1 * f_2 * \cdots * f_n) \geq \sum_{j=1}^n \gamma_j H(u_j(t)) - \frac{d}{2}\sum_{j=1}^n \gamma_j \log\gamma_j. \tag{6.82}$$

Inequality (6.82) generalizes to n functions the result of Lemma 26. Shannon's entropy power inequality then follows by maximizing the right-hand side of (6.82) over the sequence γ_j.

7 Further information inequalities

In this section we prove some recent generalizations of information inequalities, which follows by considering into the classical entropy power inequality density functions of particular type. The interest in this type of results is due to the fact that, while there are many proof of the entropy power inequality, to quantify the gap in this inequality is very difficult. In the rest of this section, we refer to the recent papers [85, 87]. The main interest here is to investigate about bounds for convolutions for the functional

$$J(X) = J(f) = \sum_{i,j=1}^{d} \int_{\{f>0\}} [\partial_{ij}(\log f)]^2 f\,dx = \sum_{i,j=1}^{d} \int_{\{f>0\}} \left[\frac{\partial_{ij} f}{f} - \frac{\partial_i f \partial_j f}{f^2}\right]^2 f\,dx. \tag{7.1}$$

We remark that, given a random vector X in $\mathbb{R}^d$, $d \geq 1$, the functional $J(X)$ is well-defined for a smooth, rapidly decaying probability density $f(x)$ such that $\log f$ has growth at most polynomial at infinity. As proven in Section 5.2 (cf. also Villani in [91]), $J(X)$ is related to Fisher information by the relationship

$$J(X + Z_{2t}) = -2\frac{d}{dt} I(X + Z_{2t}). \tag{7.2}$$

7.1 Log-concave functions and scores

We recall that a function f on $\mathbb{R}^d$ is log-concave if it is of the form

$$f(x) = \exp\{-\Phi(x)\}, \tag{7.3}$$

for some convex function $\Phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. A prime example is the Gaussian density, where $\Phi(x)$ is quadratic in x. Further, log-concave distributions include Gamma distributions with shape parameter at least one, $Beta(\alpha, \beta)$ distributions with $\alpha, \beta \geq 1$, Weibull distributions with shape parameter at least one, Gumbel, logistic and Laplace densities (see, for example, Marshall and Olkin [68]). Log-concave functions have a number of properties that are desirable for modelling. Marginal distributions, convolutions and product measures of log-concave distributions and densities are again log-concave (cf. for example, Dharmadhikari and Joag-Dev [48]).

A main consequence of log-concavity, which is at the basis of most computations in this paper, is the following. Consider the heat equation (3.1) in $\mathbb{R}^d$, $d \geq 1$. If $M_\sigma(x)$ denotes the Gaussian density defined as in (1.3), the solution at time t to the heat equation (3.1) coincides with $u = f * M_{2\kappa t}$.

Assume that the initial datum $f(x)$ is a non-negative, log-concave integrable function. Then, at each subsequent time $t > 0$, the solution $u(\cdot, t)$ to the heat equation, convolution of the log-concave functions f and the Gaussian density $M_{2\kappa t}$ defined in (1.3), is a non-negative integrable log-concave function. In other words, the heat equation propagates log-concavity. This simple remark, allows to proof things by using smooth functions with fast decay at infinity.

It is interesting to notice that the expressions of Shannon's entropy H, Fisher information I and the functional J defined in (7.1) take a very simple form if evaluated in correspondence to log-concave densities f, when written as in (7.3). In this case, if X is a random vector in $\mathbb{R}^d$ with density f, these functionals can be easily recognized as moments of $\Phi(X)$ or of its derivatives. It is immediate to reckon that Shannon's entropy H coincides with

$$H(f) = \int_{\mathbb{R}^d} \Phi(x) f(x)\, dx. \tag{7.4}$$

The Fisher information I reads

$$I(f) = \int_{\mathbb{R}^d} |\nabla \Phi(x)|^2 f(x)\, dx = \sum_{i=1}^{n} \int_{\mathbb{R}^d} |\partial_i \Phi(x)|^2 f(x)\, dx, \tag{7.5}$$

and, last, J takes the form

$$J(f) = \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} |\partial_{ij} \Phi(x)|^2 f(x)\, dx. \tag{7.6}$$

Thus, the functionals are well-defined in terms of the convex function Φ characterizing the log-concave function f.

For the log-concave Gaussian density (1.3)

$$\Phi(x) = \frac{|x|^2}{2\sigma} + \frac{d}{2} \log 2\pi\sigma,$$

which implies, for $i, j = 1, 2, \ldots, d$

$$\partial_i \Phi(x) = \frac{x_i}{\sigma}, \quad \partial_{ij} \Phi(x) = \frac{1}{\sigma} \delta_{ij},$$

where, as usual, δ_{ij} is the Kronecker delta.

According to the standard definition, given a random vector X in $\mathbb{R}^d$ distributed with with absolutely continuous probability density function $f(x)$

$$\rho(X) = \frac{\nabla f(X)}{f(X)}, \tag{7.7}$$

denotes the (almost everywhere defined) score function of the random variable [44] (cf. also [67] for further details). The score has zero mean, and its variance is just the Fisher information. For log-concave densities, which are expressed in the form (7.3)

$$\rho(X) = -\nabla\Phi(X) \tag{7.8}$$

In view of definition (7.1) and (7.7) one can think to introduce the concept of second-order score of a random vector X in $\mathbb{R}^d$, defined by the symmetric Hessian matrix $\mathcal{H}(X)$ of $-\log f(X)$, with elements

$$\Psi_{ij}(X) = \frac{\partial_i f \partial_j f(X)}{f^2(X)} - \frac{\partial_{ij} f(X)}{f(X)}. \tag{7.9}$$

Then, as the Fisher information coincides the second moment of the score function, the functional $J(X)$ in (7.1) is expressed by the moment of the trace of the product matrix $\mathcal{H}(X) \cdot \mathcal{H}(X)$. For a log-concave function, the element $\Psi_{i,j}$ of the Hessian matrix $\mathcal{H}(X)$ defining the second-order score function takes the simple expression

$$\Psi_{ij}(X) = \partial_{ij}\Phi(X). \tag{7.10}$$

Note that a Gaussian vector M_σ is uniquely defined by a *linear* score function $\rho(M_\sigma) = M_\sigma/\sigma$ and by a *constant* second-order score matrix $\mathcal{H}(M_\sigma) = I_d/\sigma$.

7.1.1 The one-dimensional case

For the moment, let us fix $d = 1$. In the rest of this section, we will only consider smooth log-concave probability densities $f(x)$ (cf. definition (7.3)) such that $\Phi(x) = -\log f(x)$ has growth at most polynomial at infinity. In order not to worry about derivatives of logarithms, which will often appear in the proof, we may also impose that

$$|\Phi^{(i)}(x)/\Phi(x)| \le C(1 + |x|^2), \quad i = 1, 2 \tag{7.11}$$

for some positive constant C. The general case will easily follow by a density argument [64]. If the convex function Φ satisfies (7.11), one can easily justify that, if

$$k(x) = f * g(x).$$

denoted the convolution product of the probability densities f and g, $k^{(i)}(x) = (f * g)^{(i)}(x) = (f^{(i)} * g)(x)$, $i = 1, 2$.

Let $i = 1$. For any given $\epsilon > 0$ we have

$$\frac{k(x+\epsilon) - k(x)}{\epsilon} = \int_{\mathbb{R}} \frac{f(x+\epsilon-y) - f(x-y)}{\epsilon} g(y)\, dy.$$

Note that, by the dominated convergence theorem, the right-hand side converges to $(f' * g)(x)$ as $\epsilon \to 0$ if f is a Lipschitz function, for which it suffices to show that $|f'|$ is bounded. Since $f(x) = e^{-\Phi(x)}$, $|f'(x)| = |\Phi'(x)|e^{-\Phi(x)}$. Then, if x is sufficiently large, $\Phi(x) > 0$ and $\Phi(x) \to \infty$ as $x \to +\infty$ by the convexity of Φ and the fact that f is a probability density. Thus we have $\Phi(x) \geq Dx$ for some positive number D and all sufficiently large x. Therefore we have

$$|f'(x)| \leq |\Phi'(x)|e^{-Dx} \leq C(1 + |x|^2)|\Phi(x)|e^{-Dx},$$

where the last inequality follows by the assumption on Φ. By the fact that Φ has growth at most polynomial at infinity, we have $|f'(x)| \to 0$ as $x \to +\infty$. By the same method we can prove that $|f'(x)| \to 0$ as $x \to -\infty$ as well, which means that $|f'|$ is bounded. A similar argument then applies to the second-order derivative of $k(x)$.

The main argument here is due to Blachman [20], who proved in this way inequality (4.7). Since the result that follows constitute a generalization of this result, we give the complete proof. Since for any pair of positive constants a, b we have the identity

$$(a+b)k'(x) = a \int_{\mathbb{R}} f'(x-y)g(y)\, dy + b \int_{\mathbb{R}} f(x-y)g'(y)\, dy,$$

dividing by $k(x) > 0$ we obtain

$$(a+b)\frac{k'(x)}{k(x)} = \int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)} \right) d\mu_x(y).$$

We denoted

$$d\mu_x(y) = \frac{f(x-y)g(y)}{k(x)}\, dy.$$

Note that, for every $x \in \mathbb{R}$, $d\mu_x$ is a probability measure on $\mathbb{R}$. Consequently, by Jensen's inequality

$$(a+b)^2 \left(\frac{k'(x)}{k(x)} \right)^2 \leq \int_{\mathbb{R}} \left(a\frac{f'(x-y)}{f(x-y)} + b\frac{g'(y)}{g(y)} \right)^2 d\mu_x(y). \tag{7.12}$$

On the other hand, by analogous argument, for any pair of positive constants a, b we have the identity

$$(a+b)^2 k''(x) = a^2 \int_{\mathbb{R}} f''(x-y)g(y)\, dy +$$

$$b^2 \int_{\mathbb{R}} f(x-y)g''(y)\,dy + 2ab\int_{\mathbb{R}} f'(x-y)g'(y)\,dy.$$

Thus, dividing again by $k(x) > 0$ we obtain

$$\begin{aligned}(a+b)^2\frac{k''(x)}{k(x)} = \int_{\mathbb{R}} \Bigg(a^2\frac{f''(x-y)}{f(x-y)} + \\ b^2\frac{g''(y)}{g(y)} + 2ab\frac{f'(x-y)}{f(x-y)}\frac{g'(y)}{g(y)}\Bigg)\,d\mu_x(y).\end{aligned} \tag{7.13}$$

If we subtract identity (7.13) from inequality (7.12) we conclude with the inequality

$$\begin{aligned}(a+b)^2\left[\left(\frac{k'(x)}{k(x)}\right)^2 - \frac{k''(x)}{k(x)}\right] = -(a+b)^2(\log k(x))'' \le \\ -\int_{\mathbb{R}} \left\{a^2(\log f(x-y))'' + b^2(\log g(y))''\right\}\,d\mu_x(y).\end{aligned} \tag{7.14}$$

It is important to note that, since the functions f, g (and consequently k) are log-concave, both sides of inequality (7.14) are non-negative. Therefore, taking the square in (7.14), and using once more Jensen's inequality we end up with the inequality

$$\begin{aligned}(a+b)^4\left[(\log k(x))''\right]^2 \le \\ \int_{\mathbb{R}} \left\{a^2(\log f(x-y))'' + b^2(\log g(y))''\right\}^2\,d\mu_x(y).\end{aligned} \tag{7.15}$$

Multiplying both sides of (6.48) by $k(x)$, and integrating over $\mathbb{R}$ yields the inequality

$$(a+b)^4 J(k) \le a^4 J(f) + b^4 J(g) + 2a^2b^2 I(f)I(g).$$

Indeed, in one dimension, definition (7.1) of the functional $J(\cdot)$ reduces to

$$J(f) = \int_{\{f>0\}} \left[\log f(x))''\right]^2 f(x)\,dx. \tag{7.16}$$

Moreover

$$\begin{aligned}\int dx \int dy \log f(x-y))''(\log g(y))'' f(x-y)g(y) = \\ \int dx \int dy \left(\frac{(f'(x-y)}{f(x-y)}\right)^2\left(\frac{g'(y)}{g(y)}\right)^2 f(x-y)g(y) = \\ I(f)I(g),\end{aligned}$$

where $I(f)$ (respectively $I(g)$) denotes the Fisher information of f (respectively g)

$$I(f) = \int_{\{f>0\}} \frac{(f'(x))^2}{f(x)}\,dx.$$

The cases of equality in (7.12) and (7.15) are easily found resorting to the same argument used in the proof of Lemma 5. As in Lemma 5, the conclusion is that there is equality in (7.52) if and only if f and g are Gaussian densities, of variances ca and cb, respectively, for any given positive constant c. Note moreover that when f and g are Gaussian densities the coefficient of the probability measure $d\mu_x(y)$ in the integral appearing in (7.14) is constant, which guarantees equality in (6.48). We proved

Theorem 28. *Let $f(x)$ and $g(x)$ be log-concave probability density functions with values in $\mathbb{R}$, such that both $J(f)$ and $J(g)$, as given by (7.16) are bounded. Then, also $J(f*g)$ is bounded, and for any pair of positive constants a, b*

$$J(f*g) \le \frac{a^4}{(a+b)^4}J(f) + \frac{b^4}{(a+b)^4}J(g) + 2\frac{a^2b^2}{(a+b)^4}I(f)I(g). \tag{7.17}$$

Moreover, there is equality in (7.17) if and only if, up to translation and dilation f and g are Gaussian densities, $f(x) = M_a(x)$ and $g(x) = M_b(x)$.

Remark 29. The condition of log-concavity enters into the proof of Theorem 28 when we pass from inequality (7.14) to inequality (6.48). Without the condition of log-concavity, in fact, the left-hand side of (7.14) has no sign, and (6.48) does not hold true. Of course, this fact does not exclude the possibility that inequality (7.17) could hold also for other classes of probability densities, but if any, another method of proof has to be found, or a counterexample is needed.

Theorem 28 allows to prove a new inequality for the functional J, in the form of the Blachman–Stam inequality (1.9). To this aim, note that, for any pair of positive constants a, b

$$2\sqrt{J(f)J(g)} \le \frac{a}{b}J(f) + \frac{b}{a}J(g).$$

Moreover, as proven first by Dembo [46], and later on by Villani [91] with a proof based on McKean ideas [69]

$$J(f) \ge I(f)^2. \tag{7.18}$$

The proof of (7.18) is immediate and enlightening. Given the random variable X distributed with a sufficiently smooth density $f(x)$, consider the (almost everywhere defined) second-order score variable (cf. definition (7.9))

$$\Psi(X) = \left(\frac{f'(X)}{f(X)}\right)^2 - \frac{f''(X)}{f(X)}. \tag{7.19}$$

Then, denoting with $\langle Y\rangle$ the mathematical expectation of the random variable Y, under condition (7.11) on Φ it holds

$$I(f) = I(X) = \langle \Psi(X)\rangle, \qquad J(f) = J(X) = \langle \Psi(X)^2\rangle.$$

Then, (7.18) coincides with the standard inequality $\langle \Psi(X)^2 \rangle \geq \langle \Psi(X) \rangle^2$. Note moreover that equality in (7.18) holds if and only if $\Psi(X)$ is constant, or, what is the same, if

$$\frac{d^2}{dx^2} \log f(x) = c.$$

As observed in the proof of Theorem 28 this implies that X is a Gaussian variable.

Grace to inequality (7.18)

$$2I(f)I(g) \leq 2\sqrt{J(f)J(g)} \leq \frac{a}{b}J(f) + \frac{b}{a}J(g). \tag{7.20}$$

Using (7.20) to bound from above the last term in inequality (7.17) we obtain

$$J(f * g) \leq \frac{a^3}{(a+b)^3}J(f) + \frac{b^3}{(a+b)^3}J(g). \tag{7.21}$$

Optimizing over $z = a/(a+b)$, with $z \in [0,1]$, one finds that the minimum of the right-hand side is obtained when

$$z = \bar{z} = \frac{\sqrt{J(g)}}{\sqrt{J(f)} + \sqrt{J(g)}}, \tag{7.22}$$

Thus we proved

Corollary 30. *Let X and Y be independent random variables with log-concave probability density functions with values in $\mathbb{R}$, such that both $J(X)$ and $J(Y)$, as given by (7.16) are bounded. Then*

$$\frac{1}{\sqrt{J(X+Y)}} \geq \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}}. \tag{7.23}$$

Moreover, there is equality if and only if, up to translation and dilation X and Y are Gaussian variables.

Remark 31. Inequality (7.17) implies in general a stronger inequality in Corollary 30. In fact, to obtain inequality (7.21) we discarded the (non-positive) term

$$-R_{a,b}(f,g) = \frac{a^2b^2}{(a+b)^4}\left(2I(f)I(g) - \frac{a}{b}J(f) - \frac{b}{a}J(g)\right). \tag{7.24}$$

By evaluating the value of $R(f,g,a,b)$ in $z = \bar{z}$, one shows that inequality (7.23) is improved by the following

$$\frac{1}{\sqrt{J(X+Y)}} \geq \left(\frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}}\right)\mathcal{R}(X,Y), \tag{7.25}$$

where

$$1 \leq \mathcal{R}(X,Y) = \left(1 - 2\frac{\sqrt{J(X)}\sqrt{J(Y)} - I(X)I(Y)}{(\sqrt{J(X)} + \sqrt{J(Y)})^2}\right)^{-1/2}.$$

As before, $\mathcal{R}(X,Y) = 1$ if and only if X and Y are Gaussian random variables.

Note that the non-negative remainder $R(f,g,a,b)$ can be bounded from below in terms of other expressions. In particular, one of these bounds is particularly significative. Adding and subtracting to the right-hand side of (7.24) the positive quantity $aI^2(f)/b + bI^2(g)/a$ one obtains the bound

$$R_{a,b}(f,g) \geq \frac{a^2b^2}{(a+b)^4}\left[\frac{a}{b}(J(f) - I^2(f)) + \frac{b}{a}(J(g) - I^2(g))\right]. \tag{7.26}$$

This implies that (7.21) can be improved by the following

$$\begin{aligned} J(f*g) \leq \frac{a^3}{(a+b)^3}J(f) + \frac{b^3}{(a+b)^3}J(g) + \\ -\frac{a^2b^2}{(a+b)^4}\left[\frac{a}{b}(J(f) - I^2(f)) + \frac{b}{a}(J(g) - I^2(g))\right]. \end{aligned} \tag{7.27}$$

7.1.2 The general case

With few variants, the proof in the multi-dimensional case follows along the same lines of the one-dimensional one. Let $f(x)$ and $g(x)$, with $x \in \mathbb{R}^d$ be multidimensional log-concave functions, and let $k(x) = f*g(x)$ be their log-concave convolution. In addition, let us suppose that both $f(x)$ and $g(x)$ are sufficiently smooth and decay at infinity in such a way to justify computations. To simplify notations, given a function $f(x)$, with $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $n > 1$, we denote its partial derivatives as

$$f_i(x) = \partial_i f(x), \qquad f_{ij}(x) = \partial_{ij} f(x).$$

For any given vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, and positive constants a, b we have the identity

$$(a+b)\sum_{i=1}^d \alpha_i \frac{k_i(x)}{k(x)} =$$

$$\int_{\mathbb{R}^d} \sum_{i=1}^d \alpha_i \left(a\frac{f_i(x-y)}{f(x-y)} + b\frac{g_i(y)}{g(y)}\right) d\mu_x(y),$$

where now, for every $x \in \mathbb{R}^d$

$$d\mu_x(y) = \frac{f(x-y)g(y)}{k(x)}\, dy,$$

is a probability measure on $\mathbb{R}^d$. Therefore, by Jensen's inequality

$$
\begin{aligned}
&(a+b)^2\left(\sum_{i=1}^d \alpha_i \frac{k_i(x)}{k(x)}\right)^2 = (a+b)^2 \sum_{i,j=1}^d \alpha_i\alpha_j \frac{k_i}{k}\frac{k_j}{k} \leq \\
&\int \sum_{i,j=1}^d \alpha_i\alpha_j \left(a^2 \frac{f_i(x-y)}{f(x-y)}\frac{f_j(x-y)}{f(x-y)} + b^2 \frac{g_i(y)}{g(y)}\frac{g_j(y)}{g(y)} + \right. \\
&\qquad \left. +2ab\, \frac{f_i(x-y)}{f(x-y)}\frac{g_j(y)}{g(y)}\right) d\mu_x(y).
\end{aligned}
\tag{7.28}
$$

Likewise, thanks to the identity

$$
\begin{aligned}
(a+b)^2\frac{k_{ij}(x)}{k(x)} = \int \left(a^2 \frac{f_{ij}(x-y)}{f(x-y)} + b^2\frac{g_{ij}(y)}{g(y)} + \right. \\
\left. +2ab\frac{f_i(x-y)}{f(x-y)}\frac{g_j(y)}{g(y)}\right) d\mu_x(y),
\end{aligned}
$$

we have

$$
\begin{aligned}
&(a+b)^2 \sum_{i,j=1}^d \alpha_i\alpha_j \frac{k_{ij}(x)}{k(x)} = \\
&\int \sum_{i,j=1}^d \alpha_i\alpha_j \left(a^2 \frac{f_{ij}(x-y)}{f(x-y)} + b^2\frac{g_{ij}(y)}{g(y)} + \right. \\
&\qquad \left. + 2ab\frac{f_i(x-y)}{f(x-y)}\frac{g_j(y)}{g(y)}\right) d\mu_x(y).
\end{aligned}
\tag{7.29}
$$

Finally, subtracting (7.29) from (7.28), for any given vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, and positive constants a, b we obtain the inequality

$$
\begin{aligned}
&(a+b)^2 \sum_{i,j=1}^d \alpha_i\alpha_j \left(\frac{k_i(x)}{k(x)}\frac{k_j(x)}{k(x)} - \frac{k_{ij}(x)}{k(x)}\right) \leq \\
&\sum_{i,j=1}^d \alpha_i\alpha_j \int_{\mathbb{R}^d} \left[a^2\left(\frac{f_i(x-y)}{f(x-y)}\frac{f_j(x-y)}{f(x-y)} - \frac{f_{ij}(x-y)}{f(x-y)}\right) + \right. \\
&\qquad \left. b^2\left(\frac{g_i(y)}{g(y)}\frac{g_j(y)}{g(y)} - \frac{g_{ij}(y)}{g(y)}\right)\right] d\mu_x(y).
\end{aligned}
\tag{7.30}
$$

Inequality (7.30) says that, for any d-dimensional row vector α one has

$$
\alpha \mathcal{M} \alpha^\top \leq \alpha \mathcal{N} \alpha^\top,
$$

where $\mathcal{M}$ and $\mathcal{N}$ are the matrices with elements

$$m_{i,j} = \frac{k_i(x)}{k(x)}\frac{k_j(x)}{k(x)} - \frac{k_{ij}(x)}{k(x)},$$

and, respectively

$$n_{ij} = \int_{\mathbb{R}^d} \left[a^2 \left(\frac{f_i}{f}\frac{f_j}{f} - \frac{f_{ij}}{f} \right)(x-y) + \right.$$
$$\left. + \, b^2 \left(\frac{g_i}{g}\frac{g_j}{g} - \frac{g_{ij}}{g} \right)(y) \right] d\mu_x(y).$$

Consequently, both the matrix $\mathcal{N}$ and the matrix $\mathcal{N} - \mathcal{M}$ are symmetric positive semi-definite matrices. Finally, the product matrix $(\mathcal{N} - \mathcal{M})(\mathcal{N} + \mathcal{M})$ is positive semi-definite, which implies that its trace is non-negative. Now, in view of classical properties of the trace of a matrix we obtain

$$\mathtt{tr}(\mathcal{N} - \mathcal{M})(\mathcal{N} + \mathcal{M}) = \mathtt{tr}\mathcal{N}\mathcal{N} + \mathtt{tr}\mathcal{N}\mathcal{M} - \mathtt{tr}\mathcal{M}\mathcal{N} - \mathtt{tr}\mathcal{M}\mathcal{M} =$$

$$\mathtt{tr}\mathcal{N}\mathcal{N} - \mathtt{tr}\mathcal{M}\mathcal{M} = \sum_{i,j=1}^d n_{ij}^2 - \sum_{i,j=1}^d m_{ij}^2 \geq 0.$$

Finally, applying property (7.28) to (7.30) implies

$$\begin{aligned}
&(a+b)^4 \sum_{i,j=1}^d \left(\frac{k_i(x)}{k(x)}\frac{k_j(x)}{k(x)} - \frac{k_{ij}(x)}{k(x)} \right)^2 \leq \\
&\sum_{i,j=1}^d \left\{ \int_{\mathbb{R}^d} \left[a^2 \left(\frac{f_i}{f}\frac{f_j}{f} - \frac{f_{ij}}{f} \right)(x-y) + \right.\right. \\
&\qquad \left.\left. b^2 \left(\frac{g_i}{g}\frac{g_j}{g} - \frac{g_{ij}}{g} \right)(y) \right] d\mu_x(y) \right\}^2 \leq \\
&\sum_{i,j=1}^d \int_{\mathbb{R}^d} \left[a^2 \left(\frac{f_i}{f}\frac{f_j}{f} - \frac{f_{ij}}{f} \right)(x-y) + \right. \\
&\qquad \left. b^2 \left(\frac{g_i}{g}\frac{g_j}{g} - \frac{g_{ij}}{g} \right)(y) \right]^2 d\mu_x(y),
\end{aligned} \tag{7.31}$$

where the last inequality in (7.31) follows by Jensen's inequality. Expanding the squares, and then proceeding as in the proof of Theorem 28 we easily arrive to inequality (7.52). Then, the cases of equality are found by the same argument of the one-dimensional proof. Indeed, inequality (7.30) reduces to an equality if and only if the two (i,j)-vectors $\int \frac{\partial_i f \partial_j f}{f}$ and $\int \frac{\partial_i g \partial_j g}{g}$ are linearly dependent, which is equivalent to the fact that f and g have proportional covariance matrices. Thus, there is equality in (7.28) if and only if f and g are Gaussian densities, of covariance matrices aC, and bC respectively. Note that this choice guarantees the equality sign also in (7.31). Further, when f and g are Gaussian densities the coefficient of the probability measure $d\mu_x(y)$ in the integral appearing in (7.31) is constant, which guarantees equality in the consequent Jensen's inequality leading to (7.52). We conclude with the following

Theorem 32. *Let $f(x)$ and $g(x)$ be log-concave probability density functions with values in $\mathbb{R}^d$, with $d > 1$, such that both $J(f)$ and $J(g)$, as given by (7.1) are bounded. Then, $J(f * g)$ is bounded, and for any pair of positive constants a, b*

$$J(f * g) \leq \frac{a^4}{(a+b)^4} J(f) + \frac{b^4}{(a+b)^4} J(g) + 2\frac{a^2b^2}{(a+b)^4} H(f,g), \tag{7.32}$$

where

$$H(f,g) = \sum_{i,j=1}^{d} \int_{\{f>0\}} \frac{\partial_i f \partial_j f}{f} \, dx \int_{\{g>0\}} \frac{\partial_i g \partial_j g}{g} \, dx. \tag{7.33}$$

Moreover, there is equality in (7.32) if and only if f and g are Gaussian densities with proportional covariance matrices aC and bC respectively.

As for the one-dimensional case, given the random vector X distributed with density $f(x)$, $x \in \mathbb{R}^d$, consider the generic element of the second-order score function $\mathcal{H}(X)$, given by (7.9)

$$\Psi_{ij}(X) = \frac{f_i(X)}{f(X)} \frac{f_j(X)}{f(X)} - \frac{f_{ij}(X)}{f(X)}.$$

Then, for each pair of i, j we have the identities

$$\langle \Psi_{ij}(X) \rangle = \int_{\{f>0\}} \frac{\partial_i f \partial_j f}{f} \, dx,$$

and

$$\langle \Psi_{i,j}(X)^2 \rangle = \int_{\{f>0\}} \left[\frac{\partial_{ij} f}{f} - \frac{\partial_i f \partial_j f}{f^2} \right]^2 f \, dx.$$

Then, the standard inequality $\langle \Psi_{ij}(X)^2 \rangle \geq \langle \Psi_{ij}(X) \rangle^2$ gives

$$\begin{aligned} J(X) = \sum_{i,j=1}^{d} \langle \Psi_{ij}(X)^2 \rangle \geq \\ \sum_{i,j=1}^{d} \langle \Psi_{ij}(X) \rangle^2 = \sum_{i,j=1}^{d} \left[\int_{\{f>0\}} \frac{\partial_i f \partial_j f}{f} \, dx \right]^2. \end{aligned} \tag{7.34}$$

Using the Cauchy-Schwarz inequality, (7.34) gives

$$H(f,g) = \sum_{i,j=1}^{d} \int_{\{f>0\}} \frac{\partial_i f \partial_j f}{f}\, dx \int_{\{g>0\}} \frac{\partial_i g \partial_j g}{g}\, dx \le$$

$$\left[\sum_{i,j=1}^{d} \left(\int_{\{f>0\}} \frac{\partial_i f \partial_j f}{f}\, dx\right)^2\right]^{1/2} \cdot \tag{7.35}$$

$$\cdot \left[\sum_{i,j=1}^{d} \left[\int_{\{g>0\}} \frac{\partial_i g \partial_j g}{g}\, dx\right)^2\right]^{1/2} \le \sqrt{J(f)}\sqrt{J(g)}.$$

Hence, we can proceed as in the proof of Corollary 30 to obtain

Corollary 33. *Let X and Y be independent multi-dimensional random variables with log-concave probability density functions with values in $\mathbb{R}^d$, such that both $J(X)$ and $J(Y)$, as given by (7.1) are bounded. Then*

$$\frac{1}{\sqrt{J(X+Y)}} \ge \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}}.$$

Moreover, there is equality if and only if X and Y are Gaussian densities with proportional covariance matrices.

7.1.3 A strengthened entropy power inequality

In this section, we will study the evolution in time of the functional $\Lambda(t)$ defined in (6.76), that is

$$\Lambda(t) = \Lambda(f(t), g(t)) = H(f(t) * g(t)) - \kappa H(f(t)) - (1-\kappa)H(g(t)).$$

Here, κ is a positive constant, with $0 < \kappa < 1$, while $f(x,t)$ (respectively $g(x,t)$) are the solutions to the heat equation (3.1) with diffusion constant κ (respectively $1-\kappa$), corresponding to the initial data f and g, log-concave probability densities in $\mathbb{R}^d$. It is a simple exercise to verify that $\Lambda(t)$ is dilation invariant. As in the other cases, this property allows to identify the limit, as $t \to \infty$ of the functional $\Lambda(t)$. If

$$U(x,t) = \left(\sqrt{1+2t}\right)^d u(x\sqrt{1+2t}, t). \tag{7.36}$$

$U(x,t)$ tends in any Sobolev space towards the limit Gaussian function

$$\lim_{t\to\infty} U(x,t) = M_\kappa(x) \int_{\mathbb{R}^d} f(x)\, dx = M_\kappa(x). \tag{7.37}$$

and

$$\lim_{t\to\infty} \Lambda(t) = H(M_1) - \kappa H(M_\kappa) - (1-\kappa)H(M_{1-\kappa}) = \\ -\frac{d}{2}\left[\kappa \log \kappa + (1-\kappa)\log(1-\kappa)\right]. \tag{7.38}$$

If we differentiate $\Lambda(t)$ with respect to time, by the Bruijn's identity we obtain

$$\Lambda'(t) = I(f(t) * g(t)) - \kappa^2 I(f(t)) - (1-\kappa)^2 I(g(t)). \tag{7.39}$$

In view of Blachman-Stam inequality (1.9) [20, 78]

$$I(f * g) \leq \frac{a^2}{(a+b)^2} I(f) + \frac{b^2}{(a+b)^2} I(g), \quad a, b > 0$$

with equality if and only if f and g are Gaussian densities, $\Lambda'(t) \leq 0$. Hence $\Lambda(t)$ is monotonically decreasing in time from $\Lambda(0)$ to $\Lambda(\infty)$, given by (7.38).

Differentiating again with respect to time, from (7.39) we obtain

$$\Lambda''(t) = -\frac{1}{2}\left(J(f(t) * g(t)) - \kappa^3 J(f(t)) - (1-\kappa)^3 J(g(t))\right). \tag{7.40}$$

Therefore, by inequality (7.21), if f and g are log-concave, $\Lambda''(t) \geq 0$, and the convexity property of $\Lambda(t)$ follows.

On the other hand, proceeding as in the proof of Corollary 30, we obtain from inequality (7.32) the bound

$$J(f(t) * g(t)) \leq \kappa^3 J(f(t)) + (1-\kappa)^3 J(g(t)) - 2\kappa^2(1-\kappa)^2 P(f(t), g(t)), \tag{7.41}$$

where

$$P(f, g) = \sqrt{J(f)}\sqrt{J(g)} - H(f, g) \geq 0 \tag{7.42}$$

in view of inequality (7.35). In addition, equality to zero holds if and only if both f and g are Gaussian densities.

Integrating (7.41) from t to ∞, we obtain for the Fisher information of two log-concave densities the strengthened inequality

$$I(f(t) * g(t)) \leq \kappa^2 I(f(t)) - (1-\kappa)^2 I(g(t)) - \kappa^2(1-\kappa)^2 \int_t^\infty P(f(s), g(s))\, ds. \tag{7.43}$$

In fact, by the central limit property,

$$\lim_{t\to\infty} \left[I(f(t) * g(t)) - \kappa^2 I(f(t)) - (1-\kappa)^2 I(g(t))\right] =$$

$$I(M_1) - \kappa^2 I(M_\kappa) - (1-\kappa)^2 I(M_{1-\kappa}) = 0$$

Last, integrating (7.43) from 0 to ∞ we obtain for Shannon's entropy of the two log-concave densities the strengthened inequality

$$H(f*g)-\kappa H(f)-(1-\kappa)H(g)-\frac{d}{2}\left[\kappa\log\kappa+(1-\kappa)\log(1-\kappa)\right]\geq\mathcal{P}_\kappa(f,g), \tag{7.44}$$

where

$$\mathcal{P}_\kappa(f,g)=\kappa^2(1-\kappa)^2\int_0^\infty d\tau\int_\tau^\infty P(f(s),g(s))\,ds\geq 0. \tag{7.45}$$

Choosing now $\kappa=\bar\kappa$ as given by (6.73) we end up with inequality

$$N(X+Y)\geq[N(X)+N(Y)]\,R(X,Y), \tag{7.46}$$

where the quantity $R(X,Y)\geq 1$ can be interpreted as a measure of the *non-Gaussianity* of the two random vectors X,Y. Indeed, $R(X,Y)=1$ if and only if both X and Y are Gaussian random vectors. Clearly

$$R(X,Y)=\exp\left\{\frac{2}{d}\mathcal{P}_{\bar\kappa}(f,g)\right\}>1. \tag{7.47}$$

Consequently, $R(X,Y)=1$ if and only if both X and Y are Gaussian random vectors.

Remark 34. In general, the expression of the term $R(X,Y)$ is very complicated, due to the fact that it is given in terms of integrals of nonlinear functionals evaluated along solutions to the heat equations which depart from the densities of X and Y. It would be certainly interesting to be able to express the term $R(X,Y)$ (or to bound it from below) in terms of some distance of X and Y from the space of Gaussian vectors. This problem is clearly easier in one dimension, where one can use the remainder as given by inequality (7.27), namely as the sum of the two contributions of the type $J(f)-I^2(f)$. In this case, one would know if, for some distance $d(f,g)$ between two probability densities f and g and some positive constant C

$$J(f)-I^2(f)\geq C\inf_{M\in\mathcal{M}}d(f,M),$$

where $\mathcal{M}$ denotes the space of Gaussian densities.

7.2 More about Fisher information

In this section, we refer to the heat equation with coefficien diffusion $\kappa=1/2$. In this case, if X is a random vector in $\mathbb{R}^d$, $d\geq 1$ with probability density f, the solution to equation (3.1) coincides with the density function of $X+Z_t$, where Z_t is a Gaussian random vector independent of X. This choice simplifies the computations that follow. In the previous sections we showed Shannon's entropy power inequality (EPI), that gives a lower bound on Shannon's entropy power of the sum of independent random vector X, Y in $\mathbb{R}^d$ with densities

$$N(X+Y) \geq N(X) + N(Y),$$

with equality if and only X and Y are Gaussian random vectors with proportional covariance matrices.

Likewise, we showed Blachman–Stam inequality, that gives a lower bound on the reciprocal of Fisher information of the sum of independent random vectors with (smooth) densities

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)},$$

still with equality if and only X and Y are Gaussian random vectors with proportional covariance matrices.

In analogy with the definition of entropy power, let us introduce the (normalized) reciprocal of Fisher information

$$\tilde{I}(X) = \frac{d}{I(X)}. \tag{7.48}$$

By construction, since $I(Z_\sigma) = d/\sigma$, $\tilde{I}(\cdot)$ is linear at Gaussian random vectors, with $\tilde{I}(Z_\sigma) = \sigma$. Moreover, in terms of $\tilde{I}$, Blachman–Stam inequality reads

$$\tilde{I}(X+Y) \geq \tilde{I}(X) + \tilde{I}(Y). \tag{7.49}$$

Therefore, both the entropy power and the reciprocal of Fisher information $\tilde{I}$, as given by (7.48), share common properties when evaluated on Gaussian random vectors and on sums of independent random vectors.

By pushing further this analogy, in agreement with Costa's result on entropy power, in [87] it has been proven that the quantity $\tilde{I}(X+Z_t)$ has the concavity property

$$\frac{d^2}{dt^2}\tilde{I}(X+Z_t) \leq 0. \tag{7.50}$$

Unlike the concavity property of the entropy power, the proof of (7.50) is restricted to log-concave random vectors. Similarly to (1.6), equality to zero in (7.50) holds if and only if X is a Gaussian random vector, $X = Z_1$.

The estimates obtained in the proof of (7.50) can be fruitfully employed to study the third derivative of $N(X+Z_t)$. The surprising result is that, at least for log-concave probability densities, the third derivative has a sign, and

$$\frac{d^3}{dt^3}N(X+Z_t) \geq 0. \tag{7.51}$$

Once again, equality to zero in (7.51) holds if and only if X is a Gaussian random variable, $X = N(0, \sigma I_d)$. Considering that

$$\frac{d}{dt}N(X + Z_t) \geq 0,$$

the new inequality (7.51) seems to indicate that the subsequent derivatives of $N(X+Z_t)$ alternate in sign, even if a proof of this seems prohibitive.

The concavity property of the reciprocal of Fisher information is a consequence of the results proven in Section 7.1.2 related to the functional $J(X)$ defined in (7.1). We rewrite here in a sdifferent form the new inequality (7.32) proven in Theorem 32. Let X and Y are independent random vectors in $\mathbb{R}^d$, such that their probability densities f and g are *log-concave*, and $J(X)$, $J(Y)$ are well defined. For any constant α, with $0 \leq \alpha \leq 1$, it holds

$$J(X + Y) \leq \alpha^4 J(X) + (1-\alpha)^4 J(Y) + 2\alpha^2(1-\alpha)^2 H(X,Y), \tag{7.52}$$

where $H(X,Y)$ is defined in (7.42). Note that, in one-dimension $H(f,g) = I(f)I(g)$. Inequality (7.52) is sharp. Indeed, there is equality if and only if X and Y are d-dimensional Gaussian vectors with covariance matrices proportional to αI_d and $(1-\alpha)I_d$ respectively.

Even if inequality (7.52) is restricted to the set of log-concave densities, this set includes many of the most commonly-encountered parametric families of probability density functions [68].

Inequality (7.52) implies a Blachman-Stam type inequality for $\sqrt{J(\cdot)}$ [85]

$$\frac{1}{\sqrt{J(X+Y)}} \geq \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}}, \tag{7.53}$$

where, also in this case, equality holds if and only if both X and Y are Gaussian random vectors.

Inequality (7.53) shows that, at least if applied to log-concave probability densities, the functional $1/\sqrt{J(\cdot)}$ behaves with respect to convolutions like Shannon's entropy power and the reciprocal of Fisher information. The fact that these inequalities share a common nature is further clarified by noticing that, when evaluated in correspondence to the Gaussian vector Z_σ,

$$N(Z_\sigma) = \tilde{I}(Z_\sigma) = \sqrt{n/J(Z_\sigma)} = \sigma.$$

In addition to the present results, other inequalities related to Fisher information in one-dimension have been recently obtained in [38]. In particular, the sign of the subsequent derivatives of Shannon's entropy $H(X + Z_t)$ up to order four have been computed explicitly. Since these derivatives alternate in sign, it is conjectured in

[38] that this property has to hold for all subsequent derivatives. This is an old conjecture that goes back at least to McKean [69], who investigated derivatives of Shannon's entropy up to the order three. Despite the title, however, in [38] the sign of the subsequent derivatives of the entropy power $N(X+Z_t)$ is not investigated.

7.2.1 A concavity property of Fisher information

Let us assume that the random vector X has a log-concave density $f(x)$, $x \in \mathbb{R}^d$. Then, for any $t > 0$, the random vector $X + Z_t$ has a log-concave smooth density function. Let us evaluate the derivatives of $\tilde{I}(X+Z_t)$, with respect to t, $t > 0$. Thanks to (7.2) we obtain

$$\frac{d}{dt}\tilde{I}(X+Z_t) = d\,\frac{J(X+Z_t)}{I^2(X+Z_t)}, \tag{7.54}$$

and

$$\frac{d^2}{dt^2}\tilde{I}(X+Z_t) = d\left(2\,\frac{J^2(X+Z_t)}{I^3(X+Z_t)} - \frac{K(X+Z_t)}{I^2(X+Z_t)}\right). \tag{7.55}$$

In (7.55) we defined

$$K(X+Z_t) = -\frac{d}{dt}I(X+Z_t). \tag{7.56}$$

Hence, to prove concavity we need to show that, for log-concave densities

$$K(X+Z_t) \ge 2\frac{J^2(X+Z_t)}{I(X+Z_t)}. \tag{7.57}$$

Note that

$$I(Z_\sigma) = \frac{d}{\sigma}, \quad J(Z_\sigma) = \frac{d}{\sigma^2}, \quad K(Z_\sigma) = 2\frac{d}{\sigma^3}. \tag{7.58}$$

Consequently, inequality (7.57) is verified with the equality sign in correspondence to a Gaussian random vector.

Using the second identity in (7.58) into (7.53) it is immediate to recover a lower bound for $K(\cdot)$. This idea goes back to Dembo [46], and has been presented in Section 5.3. Let $\sigma, t > 0$. By choosing $X = W + Z_\sigma$ and $Y = Z_t$, inequality (7.53) becomes

$$\frac{1}{\sqrt{J(W+Z_\sigma+Z_t)}} \ge \frac{1}{\sqrt{J(W+Z_\sigma)}} + \frac{t}{\sqrt{d}}.$$

Then, for all $t > 0$

$$\frac{1}{t}\left(\frac{1}{\sqrt{J(W+Z_\sigma+Z_t)}} - \frac{1}{\sqrt{J(W+Z_\sigma)}}\right) \ge \frac{1}{\sqrt{d}},$$

and this implies, passing to the limit $t \to 0^+$

$$\frac{1}{2}\frac{K(W+Z_\sigma)}{J^{3/2}(W+Z_\sigma)} \ge \frac{1}{\sqrt{d}},$$

for any $\sigma > 0$. Hence, a direct application of inequality (7.53) shows that $K(X+Z_t)$ is bounded from below, and

$$K(X+Z_t) \ge 2\frac{J^{3/2}(X+Z_t)}{\sqrt{d}}. \tag{7.59}$$

Unfortunately, inequality (7.59) is weaker than (7.57), since it is known that, for all random vectors X and Z_t independent from each other [46, 91]

$$J(X+Z_t) \ge \frac{I^2(X+Z_t)}{d}, \tag{7.60}$$

and (7.60) implies

$$\frac{J^2(X+Z_t)}{I(X+Z_t)} \ge \frac{J^{3/2}(X+Z_t)}{\sqrt{d}}.$$

To achieve the right result, we will work directly on inequality (7.52). Let us fix $Y = Z_t$. Then, since for if $i \neq j$

$$\int_{\mathbb{R}^d} \frac{\partial_i M_t(x)\partial_j M_t(x)}{M_t(x)}\,dx = \int_{\mathbb{R}^d} \frac{x_i x_j}{t^2} M_t(x)\,dx = 0,$$

one obtains

$$H(X, Z_t) = \sum_{i=1}^d \int_{\{f>0\}} \frac{f_i^2}{f}\,dx \int_{\mathbb{R}^d} \frac{x_i^2}{t^2} M_t\,dx = I(X)\frac{1}{d}I(Z_t) = \frac{1}{t}I(X). \tag{7.61}$$

Hence, using (7.58) and (7.61), inequality (7.52) takes the form

$$J(X+Z_t) \le \alpha^4 J(X) + (1-\alpha)^4\frac{d}{t^2} + 2\alpha^2(1-\alpha)^2\frac{1}{t}I(X). \tag{7.62}$$

We observe that the function

$$\Lambda(\alpha) = \alpha^4 J(X) + (1-\alpha)^4\frac{d}{t^2} + 2\alpha^2(1-\alpha)^2\frac{1}{t}I(X)$$

is convex in a, $0 \le \alpha \le 1$. This fact follows by evaluating the sign of $\Lambda''(\alpha)$, where

$$\frac{1}{12}\Lambda''(\alpha) = \alpha^2 J(X) + (\alpha-1)^2\frac{d}{t^2} + \frac{1}{3}[(1-\alpha)^2 + 4\alpha(\alpha-1) + \alpha^2]\frac{1}{t}I(X).$$

Clearly both $\Lambda''(0)$ and $\Lambda''(1)$ are strictly bigger than zero. Hence, $\Lambda(\alpha)$ is convex if, for $r = \alpha/(1-\alpha)$

$$r^2 J(X) + \frac{d}{t^2} + \frac{1}{3}(r^2 - 4r + 1)\frac{1}{t} I(X) \geq 0.$$

Now,

$$\left(J(X) + \frac{1}{3t} I(X)\right) r^2 - \frac{4}{3}\frac{1}{t} I(X) r + \left(\frac{d}{t^2} + \frac{1}{3t} I(X)\right) \geq$$
$$J(X) r^2 - 2\frac{I(X)}{t} r + \frac{d}{t^2} \geq 0.$$

The last inequality follows from (7.60).

The previous computations show that, for any given value of $t > 0$, there exists a unique point $\bar{\alpha} = \bar{\alpha}(t)$ in which the function $\Lambda(\alpha)$ attains the minimum value. In correspondence to this optimal value, inequality (7.62) takes the equivalent optimal form

$$J(X + Z_t) \leq \bar{\alpha}(t)^4 J(X) + (1 - \bar{\alpha}(t))^4 \frac{d}{t^2} + 2\bar{\alpha}(t)^2 (1 - \bar{\alpha}(t))^2 \frac{1}{t} I(X). \tag{7.63}$$

The evaluation of $\bar{\alpha}(t)$ requires to solve a third order equation. However, since we are interested in the value of the right-hand side of (7.63) for small values of the variable t, it is enough to evaluate in an exact way the value of $\bar{\alpha}(t)$ up to the order one in t. By substituting

$$\bar{\alpha}(t) = c_0 + c_1 t + o(t)$$

in the third order equation $\Lambda'(\alpha) = 0$, and equating the coefficients of t at the orders 0 and 1 we obtain

$$c_0 = 1, \quad c_1 = -\frac{J(x)}{I(X)}. \tag{7.64}$$

Consequently, for $t \ll 1$

$$\Lambda(\bar{\alpha}(t)) = J(X) - 2\frac{J^2(X)}{I(X)} t + o(t). \tag{7.65}$$

Finally, by using expression (7.65) into inequality (7.63) we obtain

$$\frac{1}{\sqrt{J(X + Z_\sigma + Z_t)}} \geq \frac{1}{\sqrt{J(X + Z_\sigma) - 2\frac{J^2(X+Z_\sigma)}{I(X+Z_\sigma)} t + o(t)}} = \frac{1}{\sqrt{J(X + Z_\sigma)}} + \frac{\sqrt{J(X + Z_\sigma)}}{I(X + Z_\sigma)} t + o(t), \tag{7.66}$$

which implies, for all $\sigma > 0$, the inequality

$$\lim_{t\to 0^+} \frac{1}{t}\left(\frac{1}{\sqrt{J(X+Z_\sigma+Z_t)}} - \frac{1}{\sqrt{J(X+Z_\sigma)}}\right) \ge \frac{\sqrt{J(X+Z_\sigma)}}{I(X+Z_\sigma)}. \tag{7.67}$$

At this point, inequality (7.57) follows from (7.67) simply by evaluating the derivative of

$$J(X+Z_t) = \left(\frac{1}{\sqrt{J(X+Z_t)}}\right)^{-2}.$$

This gives

$$\begin{aligned} K(X+Z_t) = -\frac{d}{dt}J(X+Z_t) = -\frac{d}{dt}\left(\frac{1}{\sqrt{J(X+Z_t)}}\right)^{-2} = \\ 2\,J(X+Z_t)^{3/2}\frac{d}{dt}\frac{1}{\sqrt{J(X+Z_t)}} \ge 2\,\frac{J^2(X+Z_t)}{I(X+Z_t)}. \end{aligned} \tag{7.68}$$

Hence we proved

Theorem 35. *Let X be a random vector in $\mathbb{R}^d$, $d \ge 1$, such that its probability density $f(x)$ is log-concave. Then the reciprocal of the Fisher information of $X+Z_t$, where X and Z_t are independent each other, is concave in t, i.e.*

$$\frac{d^2}{dt^2}\frac{1}{I(X+Z_t)} \le 0.$$

7.2.2 An improvement of Costa's entropy power inequality

The computations of Section 7.2.1 can be fruitfully used to improve the concavity property of the entropy power $N(X+Z_t)$. To this aim, let us compute the derivatives in t of $N(X+Z_t)$, up to the third order. The first derivative can be easily evaluated by resorting to de Bruijn identity

$$\frac{d}{dt}H(X+Z_t) = \frac{1}{2}I(X+Z_t).$$

Then, identities (7.2) and (7.56) can be applied to compute the subsequent ones. By setting $X+Z_t = W_t$ one obtains

$$\frac{d}{dt}N(W_t) = \frac{1}{d}N(W_t)I(W_t), \tag{7.69}$$

and, respectively

$$\frac{d^2}{dt^2}N(W_t) = \frac{1}{d}N(W_t)\left(\frac{I(W_t)^2}{n} - J(W_t)\right), \tag{7.70}$$

and

$$\frac{d^3}{dt^3}N(W_t) = \frac{1}{d}N(W_t)\left(K(W_t) + \frac{I(W_t)^3}{d^2} - 3\frac{I(W_t)J(W_t)}{d}\right). \tag{7.71}$$

Note that, by virtue of identities (7.60) and (7.58), the right-hand sides of both (7.70) and (7.71) vanish if W_t is a Gaussian random vector. Using inequality (7.57) we get

$$K(W_t) + \frac{I(W_t)^3}{d^2} - 3\frac{I(W_t)J(W_t)}{d} \geq 2\frac{J^2(W_t)}{I(W_t)} + \frac{I(W_t)^3}{d^2} - 3\frac{I(W_t)J(W_t)}{d}.$$

Thus, by setting $p = dJ(W_t)/I^2(W_t)$, the sign of the expression on the right-hand side of (7.71) will coincide with the sign of the expression

$$2p^2 - 3p + 1. \tag{7.72}$$

Since $p \geq 1$ in view of the inequality (7.60) [46, 91], $2p^2 - 3p + 1 \geq 0$, and the result follows. Last, the cases of equality coincide with the cases in which there is equality both in (7.57) and (7.60), namely if and only if W_t is a Gaussian random vector.
We proved

Theorem 36. *Let X be a random vector in $\mathbb{R}^d$, $d \geq 1$, such that its probability density $f(x)$ is log-concave. Then the entropy power of $X + Z_t$, where X and Z_t are independent each other, has the derivatives which alternate in sign up to the order three. In particular $N(X + Z_t)$ is concave in t, and*

$$\frac{d^3}{dt^3}N(X + Z_t) \geq 0.$$

8 Nonlinear diffusion equations

8.1 Rényi entropies

Given a probability density $f(x)$, $x \in \mathbb{R}^d$, and a positive constant p the Rényi entropy of order p of f is defined by [47]:

$$\mathcal{R}_p(f) = \frac{1}{1-p}\log\left(\int_{\mathbb{R}^d} f^p(x)\,dx\right). \tag{8.1}$$

This concept of entropy has been introduced by Rényi in [74] for a discrete probability measure to generalize the classical logarithmic entropy, by maintaining at the same time most of its properties. Indeed, the Rényi entropy of order 1, defined as the limit as $p \to 1$ of $\mathcal{R}_p(f)$ is

$$\lim_{p\to 1} \mathcal{R}_p(f) = \mathcal{R}(u) = -\int_{\mathbb{R}^d} f(x) \log f(x)\, dx. \tag{8.2}$$

Therefore, the standard (Shannon) entropy of a probability density [77] is included in the set of Rényi entropies, and it is identified with the Rényi entropy of index $p = 1$.

Among other properties, the Rényi entropy (8.1) behaves as the Shannon entropy (8.2) with respect to the scaling for dilation of the probability density. As usual, for any given density $f(x)$ and positive constant a, we define the *dilation* of f by a, as the mass-preserving scaling

$$f(x) \to f_a(x) = a^d f\,(ax)\,. \tag{8.3}$$

Then, for any $p \ge 0$ it holds

$$\mathcal{R}_p(f_a) = \mathcal{R}_p(f) - d \log a. \tag{8.4}$$

This characteristic differentiates the Rényi entropy from other generalizations of the Shannon entropy, which have been introduced later on on the literature. For example, the Tsallis entropy of order p [88]:

$$\mathcal{T}_p(f) = \frac{1}{1-p} \int_{\mathbb{R}^d} \left(f^p(x) - f(x)\right)\, dx, \tag{8.5}$$

which is extensively used by physicists in statistical mechanics [89], does not satisfy property (8.4). Property (8.4) is one of the main ingredients to work with the Rényi entropy functional and to derive from it inequalities in sharp form. Thus, in our opinion, the definition (8.1) introduced by Rényi represents a very coherent generalization of the Shannon entropy.

The Shannon entropy is naturally coupled to the heat equation (3.1), as soon as the initial datum given is assumed to be a probability density. As recently noticed in [84], the deep link between the Shannon entropy and the heat equation started to be used as a powerful instrument to obtain mathematical inequalities in sharp form in the years between the late fifties to mid sixties. To our knowledge, the first application of this idea can be found in two papers by Linnik [63] and Stam [78] (cf. also Blachman [20]), published in the same year and concerned with two apparently disconnected arguments. Stam [78] was motivated by the finding of a rigorous proof of Shannon's entropy power inequality [77], while Linnik [63] used the information measures of Shannon and Fisher in a proof of the central limit theorem of probability theory. Also, in the same years, the heat equation has been used in the context of kinetic theory of rarefied gases by McKean [69] to investigate that large-time behaviour of Kac caricature of a Maxwell gas. There, various monotonicity properties of the derivatives of the Shannon entropy along the solution to the heat equation have been derived.

Likewise, the Rényi entropy od order p is strongly coupled to the nonlinear diffusion equation of order p posed in the whole $\mathbb{R}^d$

$$\frac{\partial v(x,t)}{\partial t} = \kappa \Delta v^p(x,t), \tag{8.6}$$

still with the initial datum which is assumed to be a probability density. This maybe not so well-known link has been outlined in some recent papers [31, 76], where various results valid for the Shannon entropy have been shown to hold also for the Rényi entropies, and applied to the study of the large-time behavior of the solution to equation (8.6). We aim in this note to highlight this connection.

8.2 Self-similar solutions and Rényi entropies

To start with, let us recall briefly some essential features of the nonlinear diffusion equation (8.6). Existence an uniqueness of the solution of (8.6) to the initial value problem posed in the whole space is well-known [21, 90], and we address the interested reader to these references for details. The forthcoming analysis will be restricted to initial data which are probability densities with finite variance, and it will include both the case $p > 1$, usually known as porous medium equation, and the case $p < 1$, the fast diffusion equation. In dimension $d \geq 1$, the range of exponents which ensure the boundedness of the second moment of the solution is $p > \bar{p}$ with $\bar{p} = d/(d+2)$, which contains a part of the so-called fast diffusion range $p < 1$. The particular subinterval of p is motivated by the existence of a precise solution, found by Zel'dovich, Kompaneets and Barenblatt in the fifties (briefly called here Barenblatt solution) [7, 8, 93], which serves as a model for the asymptotic behavior of a wide class of solutions with finite second moment. In the case $p > 1$ (see [21] for $p < 1$) the Barenblatt (also called self-similar or generalized Gaussian solution) departing from $x = 0$ takes the self-similar form

$$M_p(x,t) := \frac{1}{t^{n/\mu}} \tilde{M}_p \left(\frac{x}{t^{1/\mu}} \right), \tag{8.7}$$

where

$$\mu = 2 + d(p-1)$$

and $\tilde{M}_p(x)$ is the time-independent function

$$\tilde{M}_p(x) = \left(C - \lambda \left| x \right|^2 \right)_+^{\frac{1}{p-1}}. \tag{8.8}$$

In (8.8) $(s)_+ = \max\{s, 0\}$, $\lambda = \frac{1}{2\mu}\frac{p-1}{p}$, and the constant C can be chosen to fix the mass of the source-type Barenblatt solution equal to one.

The solution to equation (8.6) satisfies mass and momentum conservations, so that

$$\int_{\mathbb{R}^d} v(x,t)\, dx = 1\,; \qquad \int_{\mathbb{R}^d} x\, v(x,t)\, dx = 0\,; \quad t \geq 0. \tag{8.9}$$

Hence, without loss of generality, one can always assume that $v_0(x)$ is a probability density of first moment equal to zero. Let us define by $E(v(t))$ the second moment of the solution:

$$E(v(t)) = \int_{\mathbb{R}^d} |x|^2 v(x,t) dx. \tag{8.10}$$

Then, $E(v(t))$ increases in time from $E_0 = E(v_0)$, and its evolution is given by the nonlinear law

$$\frac{dE(v(t))}{dt} = 2d \int_{\mathbb{R}^d} v^p(x,t)\, dx \geq 0, \tag{8.11}$$

which is not explicitly integrable unless $p = 1$. The second moment of the solution to equation (8.6) has an important role in connection with the knowledge of the large time behavior of the solution. Also, in presence of a finite second moment we can immediately establish a deep connection between equation (8.6) and the Rényi entropy of the same order p.

Indeed, let us consider the evolution in time of the Rényi entropy of order p along the solution of the nonlinear diffusion equation (8.6). Integration by parts immediately yields

$$\frac{d}{dt}\mathcal{R}_p(v(\cdot,t)) = \mathcal{I}_p(v(\cdot,t)), \quad t > 0, \tag{8.12}$$

where, for a given probability density $f(x)$

$$\mathcal{I}_p(f) := \frac{1}{\int_{\mathbb{R}^d} f^p\, dx} \int_{\{f>0\}} \frac{|\nabla f^p(x)|^2}{f(x)}\, dx. \tag{8.13}$$

When $p \to 1$, identity (8.12) reduces to DeBruijn's identity, which connects Shannon's entropy functional with the Fisher information

$$\mathcal{I}(f) = \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)}\, dx. \tag{8.14}$$

via the heat equation [20, 31, 78]. Since $\mathcal{I}_p(f) > 0$, identity (8.12) shows that the Rényi entropy of the solution to equation (8.6) is increasing in time.

Since the energy scales under the dilation (8.4) of f according to

$$E(f_a) = \int_{\mathbb{R}^d} |x|^2 f_a(x)\, dx = \frac{1}{a^2} E(f),$$

if the probability density has bounded second moment, a dilation invariant functional is obtained by coupling Rényi entropy of f with the logarithm of the second moment of f

$$\Lambda_p(f) = \mathcal{R}_p(f) - \frac{d}{2} \log E(f). \tag{8.15}$$

Let $v(x,t)$ be a solution to equation (8.6). If we now compute the time derivative of $\Lambda_p(v(t))$, we obtain

$$\frac{d}{dt}\Lambda_p(v(t)) = \mathcal{I}_p(v(t)) - d^2 \frac{\int_{\mathbb{R}^d} v^p(t)\, dx}{E(v(t))}, \tag{8.16}$$

which is a direct consequence of both identities (8.11) and (8.12).

The right-hand side of (8.16) is nonnegative. This can be easily shown by an argument which is often used in this type of proofs, and goes back at least to McKean [69]. One obtains

$$0 \le \int_{\{v>0\}} \left(\frac{\nabla v^p(x)}{v(x)} + d\, x \frac{\int v^p(x)}{E(x)} \right)^2 \frac{v(x)}{\int v^p}\, dx =$$

$$\mathcal{I}_p(x) + d^2 \frac{\int v^p}{E(x)^2} \int_{\mathbb{R}^d} |x|^2 v(x)\, dx + 2d \frac{\int v^p}{E(x)} \int_{\{v>0\}} x \cdot \nabla v(x)\, dx =$$

$$\mathcal{I}_p(x) + d^2 \frac{\int v^p}{E(x)} - 2d^2 \frac{\int v^p}{E(x)} = \mathcal{I}_p(x) - d^2 \frac{\int v^p}{E(x)}. \tag{8.17}$$

Note that equality to zero in (8.17) holds if and only if, when $v(x) > 0$

$$\frac{\nabla v^p(x)}{v(x)} + dx \frac{\int v^p}{E(x)} = 0.$$

This condition can be rewritten as

$$\nabla \left(v^{p-1} + \frac{p-1}{2p} |x|^2 \frac{d \int v^p}{E(x)} \right) = 0 \tag{8.18}$$

which identifies the probability density $v(x)$ as a Barenblatt density in $\mathbb{R}^d$ (cf. equation (8.8)). Also, (8.17) shows that, among all densities with the same second moment, Fisher information of order p takes its minimum value in correspondence to a Barenblatt density.

We proved that the functional (8.15) is monotonically increasing in time along the solution to the nonlinear diffusion. The dilation invariance can now be used to identify the limit value. The computation of the limit value uses in a substantial way the scaling invariance property. Indeed, it is well-known that the solution to equation (8.6) converges towards the self-similar Barenblatt solution (8.8) in $L_1(\mathbb{R}^d)$

at an explicitly computable rate [21, 22, 32, 45]. By definition, the second moment of the self-similar solution increases in time, and it is infinite as time goes to infinity. However, by dilation invariance, the value of the functional (8.15) in correspondence to a Barenblatt function does not depend on its second moment. In other words, we can scale at each time, without changing the value of the functional, in such a way to fix a certain value of the second moment of the Barenblatt when time goes to infinity [83, 84].

The argument we presented is twofold. From one side, it represents a notable tool to study the large-time behavior of solutions to nonlinear diffusion equations. From the other side, it allows to find inequalities by means of solutions to these nonlinear diffusions. Indeed, we proved that, for any probability density function f with bounded second moment

$$\mathcal{R}_p(f) - \frac{d}{2}\log E(f) \le \mathcal{R}_p(B_{p,\sigma}) - \frac{d}{2}\log E(B_{p,\sigma}), \tag{8.19}$$

where, for $\sigma > 0$, we denoted by $B_{p,\sigma}(x)$ the Barenblatt density defined in (8.7), of second moment equal to σ. Clearly (8.19) implies that, under a variance constraint, the Rényi entropy power of order p is maximized by a Barenblatt type density.

Inequality (8.19) can be rephrased in a slightly different way. Let $f(x)$ be a probability density function in $\mathbb{R}^d$, and let $\mathcal{N}_p(f)$ denote the entropy power of f associated to the Rényi entropy of order p:

$$\mathcal{N}_p(f) = \exp\left\{\left(\frac{2}{d} + p - 1\right)\mathcal{R}_p(f)\right\}. \tag{8.20}$$

Then, if $p > d/(d+2)$,

$$\frac{\mathcal{N}_p(f)}{E(f)^{1+d(p-1)/2}} \le \frac{\mathcal{N}_p(B_{p,\sigma})}{E(B_{p,\sigma})^{1+d(p-1)/2}}. \tag{8.21}$$

We note that the definition (8.20) of p-Rényi entropy power, proposed recently in [76], coincides with the classical definition of Shannon entropy power [77], valid when $p = 1$. This definition requires $p > (d-2)/d$, in which case $2/d + p - 1 > 0$. The range of the parameter p for which we can introduce our notion of Rényi entropy power, coincides with the range for which there is mass conservation for the solution of (8.6) [21]. This range includes the cases in which the Barenblatt has bounded second moment, since $(d-2)/d < d/(d+2)$. We observe that inequality (8.21) has been derived by a completely different method in [41, 65, 66].

8.3 The concavity of Rényi entropy power

The physical idea behind the concavity of the Shannon entropy power is clear. If we evaluate the entropy power in correspondence to a Gaussian density like (1.3), we obtain

$$\mathcal{N}(M_\sigma) = 2\pi\sigma e.$$

Hence, the entropy power of the self similar solution to the heat equation, namely a Gaussian density of variance $2t$, is a linear function of time, and its second derivative (with respect to time) is equal to zero. This property is restricted to Gaussian densities. Any other solution to the heat equation, different from the self-similar one, is such that its entropy power is concave.

Having in mind to extend the concavity property to the Rényi entropy power, and making use of the result of Section 8.2, in which we established a connection of the Rényi entropy with the solution of the nonlinear diffusion equation, the starting point for the proof of such a property would be a definition of Rényi entropy power (of order p) which is consistent with the fact that, when evaluated in correspondence to the Barenblatt self-similar solution (8.7) to the nonlinear diffusion of order p, the value of the Rényi entropy power is linear with respect to t. It is a simple exercise to verify that, owing to definition (8.20), this is true, since

$$\mathcal{N}_p(M_p(t)) = \mathcal{N}_p(\tilde{M}_p) \cdot t. \tag{8.22}$$

In [76], starting from definition (8.20), we proved that the Rényi entropy power of order p has the concavity property when evaluated along the solution to the nonlinear diffusion (8.6). The precise result is the following:

Theorem 37. *(**[76]**) Let $p > (d-2)/d$ and let $u(\cdot, t)$ be probability densities in $\mathbb{R}^d$ solving (8.6) for $t > 0$. Then the p-th Rényi entropy power defined by (8.20) satisfies*

$$\frac{d^2}{dt^2}\mathcal{N}_p(v(\cdot, t)) \leq 0 \tag{8.23}$$

Like in the Shannon's case, inequality (8.23) lieds to sharp isoperimetric inequalities. The (isoperimetric) inequality for the p-th Rényi entropy is contained into the following

Theorem 38. *(**[76]**) If $p > d/(d+2)$ every smooth, strictly positive and rapidly decaying probability density f satisfies*

$$\mathcal{N}_p(f)\,\mathcal{I}_p(f) \geq \mathcal{N}_p(\tilde{M}_p)\,\mathcal{I}_p(\tilde{M}_p) = \gamma_{d,p}. \tag{8.24}$$

We remark that $\mathcal{I}_p(f)$ is the generalized Fisher information defined in (8.13). Once again, it is immediate to show that the product in (8.24) is invariant under dilation, which allows to reckon explicitly the value of the constant by using the same argument of Section 8.2. If $p > 1$ the value of the constant $\gamma_{d,p}$ is

$$\gamma_{d,p} = d\pi \frac{2p}{p-1} \left(\frac{\Gamma\left(\frac{p+1}{p}\right)}{\Gamma\left(\frac{d}{2} + \frac{p+1}{p}\right)} \right)^{2/d} \left(\frac{(d+2)p - d}{2p} \right)^{\frac{2+d(p-1)}{d(p-1)}}. \tag{8.25}$$

In the remaining set of the parameter p, that is if $d/(d+2) < p < 1$,

$$\gamma_{d,p} = d\pi \frac{2p}{1-p} \left(\frac{\Gamma\left(\frac{1}{1-p} - \frac{d}{2}\right)}{\Gamma\left(\frac{1}{1-p}\right)} \right)^{2/d} \left(\frac{(d+2)p - d}{2p} \right)^{\frac{2+d(p-1)}{d(p-1)}}. \tag{8.26}$$

Inequality (8.24) can be rewritten in a form more suitable to functional analysis. Let $f(x)$ be a probability density in $\mathbb{R}^d$. Then, if $p > d/(d+2)$

$$\int_{\mathbb{R}^d} \frac{|\nabla f^p(x)|^2}{f(x)}\, dx \geq \gamma_{d,p} \left(\int_{\mathbb{R}^d} f^p(x)\, dx \right)^{\frac{2+2d(p-1)}{d(p-1)}}. \tag{8.27}$$

If $d > 2$, the case $p = (d-1)/d$ is distinguished from the others, since it leads to

$$\frac{2+2d(p-1)}{d(p-1)} = 0, \quad \nu = \frac{1}{d},$$

and

$$\mathcal{N}_{1-1/d}(f) = \int_{\mathbb{R}^d} f^{1-1/d}(x)\, dx.$$

In this case the concavity of $\mathcal{N}_{1-1/d}$ along (8.6) has been already known and has a nice geometric interpretation in terms of transport distances, see [71].

Note that the restriction $d > 2$ implies $(d-1)/d > d/(d+2)$. Hence, for $p = (d-1)/d$ we obtain that the probability density f satisfies the inequality

$$\int_{\mathbb{R}^d} \frac{|\nabla f^{(d-1)/d}(x)|^2}{f(x)}\, dx \geq \gamma_{d,(d-1)/d}. \tag{8.28}$$

The substitution $f = g^{2^*}$, where $2^* = 2d/(d-2)$, yields

$$\int_{\mathbb{R}^d} \frac{|\nabla f^{(d-1)/d}(x)|^2}{f(x)}\, dx = \left(\frac{2d-2}{d-2} \right)^2 \int_{\mathbb{R}^d} |\nabla g(x)|^2\, dx.$$

Therefore, for any given function $g \geq 0$ such that $g(x)^{2^*}$ is a probability density in $\mathbb{R}^d$, with $d > 2$, we obtain the inequality

$$\int_{\mathbb{R}^d} |\nabla g(x)|^2\, dx \geq \left(\frac{d-2}{2d-2} \right)^2 \gamma_{d,(d-1)/d}. \tag{8.29}$$

Since

$$\gamma_{d,(d-1)/d} = d\pi \frac{2^2(d-1)^2}{d-2} \left(\frac{\Gamma(d/2)}{\Gamma(d)} \right)^{2/d},$$

a simple scaling argument finally shows that, if $g(x)^{2^*}$ has a mass different from 1, g satisfies the *Sobolev* inequality [2], [79]

$$\int_{\mathbb{R}^d} |\nabla g(x)|^2 \, dx \geq \mathcal{S}_d \left(\int_{\mathbb{R}^d} g(x)^{2^*} \, dx \right)^{2/2^*}, \tag{8.30}$$

where

$$\mathcal{S}_d = d(d-2)\pi \left(\frac{\Gamma(d/2)}{\Gamma(d)} \right)^{2/d}$$

is the sharp Sobolev constant. Hence, Sobolev inequality with the sharp constant is a consequence of the concavity of Rényi entropy power of parameter $p = (d-1)/d$, when $d > 2$.

In all the other cases, the concavity of Rényi entropy power leads to Gagliardo-Nirenberg type inequalities with sharp constants, like the ones recently studied by Del Pino and Dolbeault [45], and Cordero-Erausquin, Nazaret, and Villani, [39] with different methods.

References

[1] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, *Comm. Partial Differential Equations*, **26** (2001) 43–100.

[2] T. Aubin, Problémes isopérimétriques et espaces de Sobolev. *J. Differential Geometry* **11**, (1976) 573–598.

[3] K.I. Babenko, An inequality in the theory of Fourier analysis. *Izv. Akad. Nauk SSSR, Ser. Mat.* **25** (1961), 531–542; English transl. *Amer. Math. Soc. Transl.* (2) **44** (1965), 115–128.

[4] D. Bakry and M. Emery, *Diffusions hypercontractives*, *Sém. Proba. XIX*, Lect. Notes in Math. n. 1123, Springer Verlag, Berlin, pp. 177–206, 1985.

[5] K. Ball, F. Barthe, and A. Naor, Entropy jumps in the presence of a spectral gap. *Duke Math. J.* **119** (2003) 41–63.

[6] K. Ball, and V.H. Nguyen, Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.*, **213** (1)(2012) 81–96.

[7] G. I. Barenblatt, On some unsteady motions of a liquid and gas in a porous medium. *Akad. Nauk SSSR. Prikl. Mat. Meh.*, **16** (1952) 67–78.

[8] G.I. Barenblatt, *Scaling. Self-Similarity and Intermediate Asymptotics*. Cambridge Univ. Press, Cambridge 1996.

[9] A.R. Barron, Entropy and the central limit theorem. *Ann. Probab.*, **14** (1986) 336–342.

[10] F. Barthe, Optimal Young's inequality and its converse: a simple proof. *Geom. Funct. Anal.* **8** (1998) 234–242.

[11] F. Barthe and D. Cordero-Erausquin, Inverse Brascamp-Lieb inequalities along the heat equation. In: *Geometric Aspects of Functional Analysis*. Lecture Notes in Math., vol. 1850, 65–71. Springer, Berlin 2004.

[12] F. Barthe and N. Huet, On Gaussian Brunn-Minkowski inequalities. *Stud. Math.* **191** (2009) 283–304.

[13] J-P. Bartier, A. Blanchet, J. Dolbeault and M. Escobedo, Improved intermediate asymptotics for the heat equation. *Appl. Math. Letters*, **24** (2011) 76–81.

[14] W. Bechner, Inequalities in Fourier Analysis on $\mathbb{R}^d$. *Proc. Nat. Acad. Sci. USA* **72** (1975) 638–641.

[15] J. Bennett, Heat-flow monotonicity related to some inequalities in euclidean analysis. *Contemporary Mathematics* **505** (2010) 85–96.

[16] J. Bennett and N. Bez,Closure Properties of Solutions to Heat Inequalities. *J. Geom. Anal.* **19** (2009) 584–600.

[17] J. Bennett, N. Bez and A. Carbery, Heat-flow monotonicity related to the Hausdorff–Young inequality, *Bull. London Math. Soc.* **41** (6) (2009) 971–979.

[18] J. Bennett, A. Carbery, M. Christ and T. Tao, The Brascamp–Lieb Inequalities: Finiteness, Structure and Extremals. *Geometric And Functional Analysis* **17** (2007) 1343–1415.

[19] P.L. Bhatnagar, E.P. Gross and M. Krook, A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Phys. Rev.* **94** (1954) 511–525.

[20] N.M. Blachman, The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory* **2** (1965) 267–271.

[21] A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo and J. L Vázquez, Asymptotics of the Fast Diffusion Equation via Entropy Estimates. *Arch. Ration. Mech. Anal.*, **191** (2009) 347–385.

[22] M. Bonforte, J. Dolbeault, G. Grillo, and J.-L. Vázquez, Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. *Proc. Natl. Acad. Sci. USA*, **107** (2010) 16459–16464.

[23] S.G. Bobkov, and M. Madiman, Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, **349** (2011) 201–204.

[24] S.G. Bobkov, and M. Madiman, Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures, *J. Funct. Anal.* **262** (2012) 3309Ũ-3339.

[25] S.G. Bobkov, and M. Madiman, On the problem of reversibility of the entropy power inequality, in P. Eichelsbacher et al., editors, *Limit Theorems in Probability, Statistics, and Number Theory (in honor of Friedrich Götze)*, vol. 42 of Springer Proceedings in Mathematics and Statistics. Springer-Verlag, Berlin 2013.

[26] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Akademie der Wissenschaften*, **66** 275–370, in *Lectures on Gas Theory.* Berkeley: University of California Press (1964). Translated by S.G. Brush. Reprint of the 1896–1898 Edition. Reprinted by Dover Publ.

[27] C. Borell, The Ehrhard inequality. *C.R. Math. Acad. Sci. Paris* **337** (2003) 663–666.

[28] H.J. Brascamp and E.H. Lieb, Best constants in Young's inequality, its converse and its generalization to more than three functions. *Adx. Math.* **20** (1976) 151–173.

[29] E.A. Carlen, E.H. Lieb, and M. Loss, A sharp analog of Young's inequality on S^N and related entropy inequalities. *J. Geom. Anal.* **14** (2004) 487–520.

[30] E.A. Carlen, and A. Soffer, Propogation of localization, optimal entropy production, and convergence rates for the central limit theorem, Preprint, arXiv:1106.2256, (2011).

[31] J.A. Carrillo, G. Toscani, Exponential convergence toward equilibrium for homogeneous Fokker-Planck-type equations. *Math. Methods Appl. Sci.*, **21** (1998) 1269–1286.

[32] J.A. Carrillo, G. Toscani, Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, **49** (2000) 113–141.

[33] J.A. Carrillo, G. Toscani, Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Riv. Mat. Univ. Parma* (7) **6**(2007) 75–198.

[34] K.M. Case, and P.F. Zweifel *Linear Transport Theory.* Addison-Wesley, Reading 1967.

[35] S. Chandrasekhar, Stochastic problems in physics and astronomy. *Rev. Modern Phys.* **15**, (1943) 1–110.

[36] S. Chandrasekhar *Radiative Transfer.* Reprinted by Clarendon Press, Oxford 1960, Dover Publications, New York 1950.

[37] S. Chapman, and T.G. Cowling, *The Mathematical Theory of Non- Uniform Gases*, Cambridge University Press, Cambridge 1958.

[38] F. Cheng, and Y. Geng, Higher order derivatives in Costa's entropy power inequality. *IEEE Trans. Inform. Theory* **61** (2015) 5892–5905.

[39] D. Cordero-Erausquin, B. Nazaret, and C. Villani, A mass–transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities. *Advances in Mathematics*, **182** (2004) 307–332.

[40] M.H.M. Costa, *A new entropy power inequality.* IEEE Trans. Inform. Theory, **IT-31**, (6) (1985) 751–760.

[41] J. Costa, A. Hero and C. Vignat, On solutions to multivariate maximum alpha-entropy problems. *Lecture Notes in Computer Science*, **2683**, no. EMMCVPR 2003, Lisbon, 7-9 July 2003, 211Ű-228, 2003.

[42] T.M. Cover and J.A. Thomas, *Elements of Information Theory.* 2nd ed. John Wiley & Sons, New York, 2006.

[43] T.M. Cover, and Z. Zhang, On the maximum entropy of the sum of tweo dependent random variables. *IEEE Trans. Inform. Theory*, **40**, (1994) 1244–1246.

[44] D.R. Cox, and D.V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London 1974

[45] M. Del Pino, and J. Dolbeault, Best constants for Gagliardo-Nirenberg inequalities and application to nonlinear diffusions. *J. Math. Pures Appl.*, **81** (2002) 847–875.

[46] A. Dembo, A simple proof of the concavity of the entropy power with respect to the variance of additive normal noise. *IEEE Trans. Inform. Theory* **35**, (1989) 887–888.

[47] A. Dembo, T.M. Cover, and J.A. Thomas, Information theoretic inequalities. IEEE Trans. Inform. Theory **37** (6) (1991) 1501–1518.

[48] S. Dharmadhikari, and K. Joag-Dev, *Unimodality, Convexity, and Applications*, Academic Press, Boston (1988).

[49] J. Dolbeault, and G. Toscani, Nonlinear diffusions: extremal properties of Barenblatt profiles, best matching and delays, *Nonlinear Anal. Series A* **138** (2016) 31–43.

[50] E. Gabetta, P.A. Markowich, and A. Unterreiter, A note on the entropy production of the radiative transfer equation. *Appl. Math. Lett.*, **12** (1999) 111–116.

[51] E. Gabetta, G. Toscani, and B. Wennberg, Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation. *J. Statist. Phys.* **81** (1995) 901–934.

[52] R.J. Gardner, The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc.* **39** (2002) 355–405.

[53] L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** (1975) 1061–1083.

[54] T. Goudon, S. Junca, and G. Toscani, Fourier-based distances and Berry–Esseen like inequalities for smooth densities. *Monatsh. Math.* **135** (2002) 115–136.

[55] D. Guo, S. Shamai, and S. Verdú, Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, **51**, (4) (2005) 1261–1282.

[56] D. Guo, S. Shamai, and S. Verdú, A Simple Proof of the Entropy-Power Inequality. *IEEE Trans. Inform. Theory*, **52**, (5) (2006) 2165–2166.

[57] O. Johnson, and A.R. Barron, Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, **129** (3), (2004) 391–409.

[58] S. Kullback, *Information Theory and Statistics.* John Wiley & Sons, New York 1959.

[59] R. G. Laha and V. K. Rohatgi, *Probability theory.* John Wiley & Sons, New York, 1979.

[60] L. Leindler, On a certain converse of Hölder's inequality. II. *Acta. Sci. Math. Szeged* **33** (1972), 217–223.

[61] E.H. Lieb, Proof of an entropy conjecture of Wehrl. *Commun. Math. Phys.*, **62** (1978) 35–41.

[62] E.H. Lieb, Gaussian kernels have only gaussian maximizers. *Invent. Math.*, **102** (1990) 179–208.

[63] Yu.V. Linnik, An information-theoretic proof of the central limit theorem with the Lindeberg condition. *Theory Probab. Appl.* **4** (1959) 288-299.

[64] P.L. Lions, G. Toscani, A strenghened central limit theorem for smooth densities. *J. Funct. Anal.* **129** (1995) 148–167.

[65] E. Lutwak, D. Yang, and G. Zhang, Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Trans. Inform. Theory*, **51** (2005) 473–478.

[66] E. Lutwak, D. Yang and G. Zhang, Moment-entropy inequalities for a random vector. *IEEE Trans. Inform. Theory*, **53** (2007) 1603–1607.

[67] M. Madiman, and A. Barron, Generalized entropy power inequalities and monotonicity properties of information, *IEEE Trans. Inform. Theory*, **53** (2007) 2317–2329.

[68] A.W. Marshall, and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* Academic Press, Orlando 1979.

[69] H.P. McKean Jr., Speed of approach to equilibrium for Kacï£¡s caricature of a Maxwellian gas. *Arch. Rat. Mech. Anal.* **21** (1966) 343–367.

[70] J. Nash, Continuity of solutions of parabolic and elliptic equations, *Amer. J. Math.* **80**, (1958) 931–954.

[71] F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Commun. Part. Diff. Eq.*, **26** (2001) 101–174.

[72] A. Prékopa, Logarithmic concave measures with application to stochastic programming. *Acta. Sci. Math. Szeged* **32** (1971) 301–315.

[73] A. Prékopa, On logarithmic measures and functions. *Acta. Sci. Math. Szeged* **34** (1973) 335–343.

[74] A. Rényi, On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1*, 547–561. University of California Press, Berkeley 1961.

[75] O. Rioul, A simple proof of the entropy-power inequality via properties of mutual information, in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, June 24-29-th, 2007.

[76] G. Savaré, and G. Toscani, The concavity of Renyi entropy power, *IEEE Trans. Inform. Theory*, **60** (5) (2014) 2687–2693.

[77] C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27** Jul.(1948), 379–423; Oct. (1948), 623–656.

[78] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.* **2** (1959) 101–112.

[79] G. Talenti, Best constant in Sobolev inequality. *Ann. Mat. Pura Appl.*, **110** (1976) 353–372.

[80] G. Toscani, Sur l'iégalité logarithmique de Sobolev. *C. R. Acad. Sci. Paris Sér. I Math.*, **324**, Série I (1997) 689–694.

[81] G. Toscani, Entropy production and the rate of convergence to equilibrium for the Fokker-Planck equation, *Quarterly of Appl. Math.*, Vol. LVII (1999) 521–541.

[82] G. Toscani, An information-theoretic proof of Nash's inequality. *Rend. Lincei Mat. Appl.*, **24** (2013) 83–93.

[83] G. Toscani, Lyapunov functionals for the heat equation and sharp inequalities. *Atti Acc. Peloritana Pericolanti, Classe Sc. Fis. Mat. e Nat.*, **91** (2013) 1–10.

[84] G. Toscani, Heat equation and convolution inequalities, *Milan J. Math.* **82** (2) (2014) 183–212.

[85] G. Toscani, A strengthened entropy power inequality for log-concave densities, *IEEE Trans. Inform. Theory*, **61** (12) (2015) 6550–6559.

[86] G. Toscani, Rényi entropies and nonlinear diffusion equations, *Acta. Appl. Math.*, **132** (2014) 595–604.

[87] G. Toscani, A concavity property for the reciprocal of Fisher information and its consequences on Costa's EPI, *Physica A*, **432** (2015) 35–42.

[88] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, **52** (1988) 479–487.

[89] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics - Approaching a Complex World*. Springer Verlag, New York 2009.

[90] J.L. Vazquez, *The Porous Medium Equation: Mathematical Theory*, Oxford University Press, Oxford 2007.

[91] C. Villani, A Short Proof of the Concavity of Entropy Power. *IEEE Trans. Inform. Theory*, **46** (2000) 1695–1696.

[92] R. Zamir and M. Feder, A generalization of the entropy power inequality with applications. *IEEE Trans. Inf. Theory*, **39** (5) (1993) 1723–1728.

[93] Ya.B. Zel'dovich, and A.S. Kompaneetz, Towards a theory of heat conduction with thermal conductivity depending on the temperature. *Collection of Papers Dedicated to 70th Birthday of Academician* A. F. Ioffe, *Izd. Akad. Nauk SSSR*, Moscow (1950) 61Ű-71.