Differentials of entropy and Fisher information along heat flow: a brief review of some conjectures

Michel Ledoux

Abstract

The note presents some known conjectures on successive derivatives of entropy and Fisher information along the heat flow (with a few partial results): the McKean Conjecture, the Completely Monotone Conjecture, the Log-Convexity Conjecture, the Entropy Power Conjecture, and the MMSE Conjecture.

Given a random vector X on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^n$, and G an independent standard normal vector, consider $X_t = X + \sqrt{t}\, G$, $t > 0$, with probability density f_t with respect to the Lebesgue measure on $\mathbb{R}^n$. That is, if μ denotes the distribution of X on the Borel sets of $\mathbb{R}^n$, f_t is the convolution

$$f_t(x) \,=\, p_t * \mu(x) \,=\, \int_{\mathbb{R}^n} p_t(x-y)d\mu(y), \quad x \in \mathbb{R}^n,$$

of μ with the Gaussian kernel

$$p_t(x) \,=\, \frac{1}{(2\pi t)^{n/2}}\, e^{-|x|^2/2t}, \quad t > 0,\ x \in \mathbb{R}^n.$$

Whenever defined, the entropy $\mathrm{H}(t)$ (with the negative sign convention) of X_t, rather f_t, is defined by

$$\mathrm{H}(t) \,=\, -\int_{\mathbb{R}^n} f_t \log f_t \, dx, \quad t > 0. \tag{1}$$

Since $\mathrm{H}(t) \geq -\log\big(\int_{\mathbb{R}^n} f_t^2 dx\big)$ by Jensen's inequality, $\mathrm{H}(t)$ takes its values in $(-\infty, +\infty]$[1].

Provided the entropy $\mathrm{H}(t)$, $t > 0$ is differentiable, the classical de Brujin formula expresses that

$$\frac{d}{dt}\,\mathrm{H}(t) \,=\, \frac{1}{2}\,\mathrm{I}(t) \tag{2}$$

[1]It may be verified for example that if μ has density $\frac{c}{y(\log y)^2}$ on $(2,\infty)$ with respect to the Lebesgue measure, the function $f_t \log f_t$ is not integrable.

where $\mathrm{I}(t)$ is the Fisher information

$$\mathrm{I}(t) \,=\, \int_{\mathbb{R}^n} \frac{|\nabla f_t|^2}{f_t}\, dx, \quad t > 0. \tag{3}$$

Note that with $v_t = \log f_t$, $t > 0$,

$$\mathrm{I}(t) \,=\, \int_{\mathbb{R}^n} f_t\, |\nabla v_t|^2\, dx.$$

It is simple to observe that $\mathrm{H}(t)$ and $\mathrm{I}(t)$ are invariant by the change of X into $X + a$ or $-X$.

The Fisher information $\mathrm{I}(t)$, $t > 0$, always exists and is infinitively differentiable on $(0, \infty)$. Indeed, by definition of f_t and the Lebesgue differentiability theorem, for all $x \in \mathbb{R}^n$,

$$\nabla f_t(x) \,=\, \int_{\mathbb{R}^n} \nabla p_t(x-y) d\mu(y) \,=\, -\frac{1}{t} \int_{\mathbb{R}^n} (x-y)\, p_t(x-y) d\mu(y).$$

By the Cauchy-Schwarz inequality,

$$\big|\nabla f_t(x)\big|^2 \,\leq\, \frac{1}{t^2}\, f_t(x) \int_{\mathbb{R}^n} |x-y|^2\, p_t(x-y) d\mu(y)$$

so that

$$\mathrm{I}(t) \,\leq\, \frac{1}{t^2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x-y|^2\, p_t(x-y) dx d\mu(y) \leq \frac{1}{t^2} \int_{\mathbb{R}^n} |x|^2\, p_t(x) dx \,<\, \infty.$$

Similar arguments show that I is infinitively differentiable on $(0, \infty)$ [7, 3].

Since $\mathrm{I}(t)$ is well-defined and smooth for any $t > 0$, the entropy may be extended by integration as $\mathrm{H}(t) - \mathrm{H}(t_0) = \frac{1}{2} \int_{t_0}^{t} \mathrm{I}(s) ds$, $0 < t_0 < t$, which coincides with the previous definition under suitable integrability assumptions. The conjectures below are nevertheless presented, for simplicity, for the Fisher information $\mathrm{I}(t)$, but in this way may be formulated equivalently on the entropy $\mathrm{H}(t)$.

1 The McKean Conjecture

If X is a Gaussian vector with covariance $\mathrm{Cov}(X) = \sigma^2\, \mathrm{Id}$,

$$\mathrm{H}(t) \,=\, \frac{n}{2} \log\big[2\pi e(\sigma^2 + t)\big], \quad t > 0,$$

and

$$\mathrm{I}(t) \,=\, \frac{n}{\sigma^2 + t}\,, \quad t > 0.$$

The successive derivatives, $\ell \geq 1$, are given by

$$\frac{d^\ell}{dt^\ell}\, \mathrm{I}(t) \,=\, \frac{n(-1)^\ell\, \ell!}{(\sigma^2 + t)^{\ell+1}}\,, \quad t > 0.$$

It is classical that Gaussian vectors maximize the entropy for a given covariance. As a consequence, if $\mathrm{Cov}(X) = \sigma^2\, \mathrm{Id}$, then $\mathrm{H}(t) \leq \frac{n}{2} \log[2\pi e(\sigma^2 + t)]$ for every $t > 0$. In the same way, the standard Cramér-Rao lower bound expresses that Gaussian variables achieve the minimum of the Fisher information subject to the variance condition $\mathrm{Cov}(X) = \sigma^2\,\mathrm{Id}$. For a proof, assume, without any loss in generality, that X has mean zero. Then, setting $\sigma_t^2 = \sigma^2 + t$ for convenience,

$$0 \,\leq\, \int_{\mathbb{R}^n} f_t \Big|\nabla v_t + \frac{x}{\sigma_t^2}\Big|^2 dx \,=\, \mathrm{I}(t) + \frac{2}{\sigma_t^2}\int_{\mathbb{R}^n} x\cdot\nabla f_t\, dx + \frac{n}{\sigma_t^2}$$

where it has been used that $\int_{\mathbb{R}^n}|x|^2 f_t\,dx = n\sigma_t^2$. Next, $\int_{\mathbb{R}^n} x\cdot\nabla f_t\,dx = -n\int_{\mathbb{R}^n} f_t\,dx = -n$ by integration by parts, so that

$$0 \,\leq\, \mathrm{I}(t) - \frac{n}{\sigma_t^2}\,.$$

Hence $\mathrm{I}(t) \geq \frac{n}{\sigma^2+t}$ which is the value of the Fisher information of X_t whenever X is Gaussian with covariance $\sigma^2\,\mathrm{Id}$.

The time derivative of the Fisher information may be achieved from the Bakry-Émery calculus (cf. [1, 2]) as

$$\frac{d}{dt}\,\mathrm{I}(t) \,=\, -\int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx \tag{4}$$

where $\nabla^2 v_t$ is the matrix of the second derivatives of $v_t = \log f_t$. (Indeed, since $\partial_t v_t = \frac{1}{2f_t}\Delta f_t = \frac{1}{2}[\Delta v_t + |\nabla v_t|^2]$,

$$\begin{aligned}\frac{d}{dt}\,\mathrm{I}(t) \,&=\, \frac{d}{dt}\bigg(\int_{\mathbb{R}^n} f_t\,|\nabla v_t|^2\,dx\bigg)\\ &=\, \frac{1}{2}\int_{\mathbb{R}^n}\Delta f_t\,|\nabla v_t|^2 dx + \int_{\mathbb{R}^n} f_t\,\nabla v_t\cdot\nabla\big(\Delta v_t\big)dx + \int_{\mathbb{R}^n} f_t\,\nabla v_t\cdot\nabla\big(|\nabla v_t|^2\big)dx.\end{aligned}$$

Now $\nabla v_t\cdot\nabla(\Delta v_t) = -|\nabla^2 v_t|^2 + \frac{1}{2}\,\Delta\big(|\nabla v_t|^2\big)$ so that

$$\int_{\mathbb{R}^n} f_t\,\nabla v_t\cdot\nabla\big(\Delta v_t\big)dx \,=\, -\int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx + \frac{1}{2}\int_{\mathbb{R}^n}\Delta f_t\,|\nabla v_t|^2 dx$$

and therefore

$$\frac{d}{dt}\,\mathrm{I}(t) \,=\, -\int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx + \int_{\mathbb{R}^n}\Delta f_t\,|\nabla v_t|^2 dx + \int_{\mathbb{R}^n} f_t\,\nabla v_t\cdot\nabla\big(|\nabla v_t|^2\big)dx.$$

The claim follows since by integration by parts,

$$\int_{\mathbb{R}^n}\Delta f_t\,|\nabla v_t|^2 dx \,=\, -\int_{\mathbb{R}^n}\nabla f_t\cdot\nabla\big(|\nabla v_t|^2\big)dx \,=\, -\int_{\mathbb{R}^n} f_t\,\nabla v_t\cdot\nabla\big(|\nabla v_t|^2\big)dx.)$$

Then developing and integrating by parts as before, for any $\lambda\in\mathbb{R}$,

$$\begin{aligned}0 \,&\leq\, \int_{\mathbb{R}^n} f_t\big|\nabla^2 v_t + \lambda\mathrm{Id}\big|^2 dx\\ &=\, \int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx + 2\lambda\int_{\mathbb{R}^n} f_t\,\Delta v_t\,dx + \lambda^2 n\\ &=\, \int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx - 2\lambda\,\mathrm{I}(t) + \lambda^2 n.\end{aligned}$$

For the optimal value $\lambda = \frac{1}{n}\,\mathrm{I}(t)$,

$$-\frac{d}{dt}\,\mathrm{I}(t) \,=\, \int_{\mathbb{R}^n} f_t\,|\nabla^2 v_t|^2 dx \,\geq\, \frac{1}{n}\,\mathrm{I}(t)^2. \tag{5}$$

Since $-\frac{d}{dt}\,\mathrm{I}(t) = \frac{n}{\sigma_t^4}$ for a normal vector with covariance $\sigma^2\,\mathrm{Id}$, it follows from the first level $\mathrm{I}(t) \geq \frac{n}{\sigma_t^2}$ that Gaussian variables achieve the minimum of $-\frac{d}{dt}\,\mathrm{I}(t)$ subject to the variance condition $\mathrm{Cov}(X) = \sigma^2\,\mathrm{Id}$.

According to these first steps, the following, largely open conjecture, has been inspired by H.-P. McKean [11].

McKean's Conjecture. *Subject to* $\mathrm{Cov}(X) = \sigma^2\,\mathrm{Id}$*, Gaussian variables achieve the minimum of* $(-1)^\ell \frac{d^\ell}{dt^\ell}\,\mathrm{I}(t)$ *for any* $\ell \geq 0$*. In other words,*

$$(-1)^\ell \frac{d^\ell}{dt^\ell}\,\mathrm{I}(t) \,\geq\, \frac{n\,\ell!}{(\sigma^2+t)^{\ell+1}}\,, \quad t > 0.$$

The difficulty is that, following the preceding path, the rule

$$\frac{d^\ell}{dt^\ell}\,\mathrm{I}(t) \,=\, (-1)^\ell \int_{\mathbb{R}^n} f_t\,|\nabla^\ell v_t|^2 dx$$

cannot be iterated for $\ell \geq 2$ (cf. [8]). For example

$$\frac{d^2}{dt^2}\,\mathrm{I}(t) \,=\, \int_{\mathbb{R}^n} f_t\big[|\nabla^3 v_t|^2 - 2T_3(v_t)\big]dx$$

where $T_3(v_t) = \sum_{i,j,k=1}^n \partial_{ij}v_t\,\partial_{ik}v_t\,\partial_{jk}v_t$. The work [8] (see also [7, 9]) actually develops (in the context of an abstract Markov diffusion operator) the suitable algebraic framework to express the successive derivatives of entropy and Fisher information along the heat flow via the Γ-calculus on the iterated gradients (cf. [2]). The resulting formulas however do not clearly reveal any indication towards the conjecture, or even the Completely Monotone Conjecture of the next section.

It actually appears that some subtil linear algebra might underly these conjectures, as witnessed by the recent contributions [3, 13]. In the latter [13], using linear matrix inequalities, the authors show that McKean's Conjecture holds true in dimension one for $\ell = 2$, 3 and 4, provided the density f of the law of X is log-concave.

2 The Completely Monotone Conjecture

A perhaps milder conjecture than the McKean Conjecture would be to ask whether the sign of the derivatives of the Fisher information is alternating.

Completely Monotone Conjecture. *For any* $\ell \geq 0$,

$$(-1)^\ell \frac{d^\ell}{dt^\ell} \mathrm{I}(t) \, \geq \, 0.$$

It has been a remarkable and somewhat overlooked achievement by F. Cheng and Y. Geng [3], relying on clever integrations by parts and quadratic factorizations, that the Completely Monotone Conjecture holds true for $\ell = 2$ and 3 in dimension one.[2, 3]

3 The Log-Convexity Conjecture

As mentioned in [3], if the Completely Monotone Conjecture holds true, then by Schoenberg's theorem, H and I are Laplace transforms of finite measures. It is a result of A. M. Fink [5] that if a (positive) function φ is completely monotone, then φ is log-convex, that is $\log \varphi$ is convex. The following conjecture would thus be a consequence of the Completely Monotone Conjecture. It is satisfied in the Gaussian case for which $\mathrm{I}(t) = \frac{n}{\sigma^2 + t}$.

Log-Convexity Conjecture. *The function* $\log \mathrm{I}(t)$, $t > 0$, *is convex.*

In other words,

$$\mathrm{I}(t) \, \frac{d^2}{dt^2} \, \mathrm{I}(t) \, \geq \, \bigg(\frac{d}{dt} \, \mathrm{I}(t) \bigg)^2. \tag{6}$$

The Log-Convexity Conjecture is proved in dimension one in [10].[4]

4 The Entropy Power Conjecture

Related to entropy and Fisher information, recent developments have concerned the entropy power

$$\mathrm{N}(t) \, = \, e^{\frac{2}{n} \mathrm{H}(t)}, \quad t > 0, \tag{7}$$

in particular for log-concave distributions.

Clearly by the de Brujin formula, $\frac{d}{dt} \, \mathrm{N}(t) = \frac{1}{n} \, \mathrm{N}(t) \, \mathrm{I}(t) \geq 0$. It has been shown by M. Costa [4] that $\frac{d^2}{dt^2} \, \mathrm{N}(t) \leq 0$, which actually amounts to (5) since

$$\frac{d^2}{dt^2} \, \mathrm{N}(t) \, = \, \frac{1}{n^2} \, \mathrm{N}(t) \Big[n \, \frac{d}{dt} \, \mathrm{I}(t) + \mathrm{I}(t)^2 \Big].$$

[2] L. Guo, C. M. Yuan, X. S. Gao, *Entropy* 24 (2022), 1155, proved the conjecture for $\ell = 2$ in dimension $n = 2, 3, 4$, and for $\ell = 3$ in dimension 2, by semidefinite programming.

[3] In the work "A higher-order Otto calculus approach to the Gaussian completely monotone conjecture" (2024), G. Wang established the conjecture for orders up to $\ell = 5$, assuming that the initial distribution is log-concave, in any dimension.

[4] It has been established by J. Liu and X. Gao, *Entropy* 25 (2023), 558, that the square root of the Fisher information is convex in dimension two, using again semidefinite programming.

It has been proved in [12] that $\frac{d^3}{dt^3}\,\mathrm{N}(t) \geq 0$ whenever the distribution of X is log-concave. In other words,

$$n\,\frac{d^2}{dt^2}\,\mathrm{I}(t) \;\geq\; -3\,\mathrm{I}(t)\,\frac{d}{dt}\,\mathrm{I}(t) - \frac{1}{n}\,\mathrm{I}(t)^3.$$

By Costa's inequality (5), it follows that

$$n\,\frac{d^2}{dt^2}\,\mathrm{I}(t) \;\geq\; -2\,\mathrm{I}(t)\,\frac{d}{dt}\,\mathrm{I}(t),$$

which shows that McKean's Conjecture for $\ell = 2$

$$\frac{d^2}{dt^2}\,\mathrm{I}(t) \;\geq\; \frac{2n}{(\sigma^2 + t)^3}$$

whenever the covariance is fixed to $\sigma^2\,\mathrm{Id}$ is satisfied in this case.

It is also shown in [12] that, under the log-concavity assumption, $\frac{1}{\mathrm{I}(t)}$ is concave. That is,

$$\mathrm{I}(t)\,\frac{d^2}{dt^2}\,\mathrm{I}(t) \;\geq\; 2\bigg(\frac{d}{dt}\,\mathrm{I}(t)\bigg)^2$$

which has to be compared to (6), and therefore implies, in this case, the Log-Convexity Conjecture.

The following alternating sign conjecture might be proposed.

Entropy Power Conjecture. *For any $\ell \geq 0$,*

$$(-1)^\ell\,\frac{d^\ell}{dt^\ell}\,\mathrm{N}(t) \;\leq\; 0.$$

It might be that the Entropy Power Conjecture is stronger than the McKean Conjecture[5].

5 The MMSE Conjecture

In this section, X is a real-valued random variable. As before, let $X_t = X + \sqrt{t}\,G$, $t > 0$, where G an independent standard normal variable, with probability density f_t with respect to the Lebesgue measure on $\mathbb{R}$.

MMSE Conjecture. *In dimension one, the knowledge of* $\mathrm{I}(t)$*, $t > 0$, characterizes the law of X up to the change of X into $X + a$ or $-X$.*[6]

[5]This has been confirmed by G. Wang in "The entropy power conjecture implies the McKean conjecture" (2024).

[6]When I mentioned this conjecture to a close colleague, I got the following answer: to characterize a probability distribution, there is a convenient tool, the Fourier transform.

Some partial results on the conjecture are described in [6, 7, 9][7], relying in particular on the description of the successive derivatives of $\mathrm{I}(t)$. The multi-dimensional case may also be addressed.

The MMSE conjecture came up in information theory in the works [6, 7] by D. Guo, Y. Wu, S. Shamai and S. Verdú on the estimation of a random variable from its observation perturbated by a Gaussian noise. Consider indeed

$$\mathrm{MMSE}(t) \,=\, \mathbb{E}\Big(\big[X - \mathbb{E}\big(X \,|\, X_t\big)\big]^2\Big), \quad t > 0, \tag{8}$$

known as the Minimum Mean-Square Error (MMSE), which actually represents an alternate formulation of the Fisher information. Note that although $\mathbb{E}(X \,|\, X_t)$ may not be well-defined if X is not integrable, it makes sense to consider the integrable random variable $X_t - \mathbb{E}(X \,|\, X_t)$ which is identified to $\sqrt{t}\,\mathbb{E}(G \,|\, X_t)$ since

$$X_t \,=\, \mathbb{E}\big(X_t \,|\, X_t\big) \,=\, \mathbb{E}\big(X \,|\, X_t\big) + \sqrt{t}\,\mathbb{E}\big(G \,|\, X_t\big).$$

In particular, $X - \mathbb{E}(X \,|\, X_t)$ makes also sense and has moments of all orders. The MMSE connects to the Fisher information $\mathrm{I}(t)$, $t > 0$, along the heat flow via the identity

$$t^2\,\mathrm{I}(t) \,=\, t - \mathrm{MMSE}(t), \quad t > 0 \tag{9}$$

(cf. [6, 9]). (Note that the invariances by translation and symmetry are immediate on this representation.) It might be observed furthermore from the latter (9) that $\mathrm{I}(t) \leq \frac{1}{t}$, and $\mathrm{I}(t) \leq \frac{n}{t}$ in $\mathbb{R}^n$ together with the analogous identity.

References

[1] D. Bakry, M. Émery. Diffusions hypercontractives. *Séminaire de Probabilités XIX, Lecture Notes in Math.* 1123 (1985), 177–206. Springer.

[2] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften 348, Springer, Berlin, 2014.

[3] F. Cheng, Y. Geng. Higher order derivatives in Costa's entropy power inequality. *IEEE Trans. Inform. Theory* 61 (2015), 5892–5905.

[4] M. Costa. A new entropy power inequality. *IEEE Trans. Inform. Theory* 31 (1985), 751–760.

[5] A. M. Fink. Kolmogorov-Landau inequalities for monotone functions. *J. Math. Anal. Appl.* 90 (1982), 251–258.

[7] See also the further development "Derivatives of entropy and the MMSE conjecture" (2023) by P. Mansanarez, G. Poly, Y. Swan.

[6] D. Guo, S. Shamai, S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory* 51 (2005), 1261–1282.

[7] D. Guo, Y. Wu, S. Shamai, S. Verdú. Estimation in Gaussian noise: properties of the minimum mean-square error. *IEEE Trans. Inform. Theory* 57 (2011), 2371–2385.

[8] M. Ledoux. L'algèbre de Lie des gradients itérés d'un générateur markovien – Développements de moyennes et entropies. *Ann. scient. Éc. Norm. Sup.* 28 (1995), 435–460.

[9] M. Ledoux. Heat flow derivatives and Minimum Mean-Square Error in Gaussian noise. *IEEE Trans. Inform. Theory* 62 (2016), 3401–3409.

[10] M. Ledoux, C. Nair, Y. N. Wang. Log-convexity of Fisher information along heat flow (2020).

[11] H. P. McKean. Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Ration. Mech. Anal.* 21 (1966), 343–367.

[12] G. Toscani. A concavity property for the reciprocal of Fisher information and its consequences on Costa's EPI. *Phys. A Stat. Mech. Appl.* 432 (2015), 35–42.

[13] X. Zhang, V. Anantharam, Y. Geng. Gaussian optimality for derivatives of differential entropy using linear matrix inequalities. *Entropy* 20 (2018), 182.

Institut de Mathématiques de Toulouse

Université de Toulouse – Paul-Sabatier, F-31062 Toulouse, France

ledoux@math.univ-toulouse.fr