Heat Flow Derivatives and Minimum Mean-Square Error in Gaussian Noise

Michel Ledoux

Abstract—**We connect recent developments on Gaussian noise estimation and the Minimum Mean-Square Error to earlier results on entropy and Fisher information heat flow expansion. In particular, the derivatives of the Minimum mean-square error with respect to the noise parameter are related to the heat flow derivatives of the Fisher information and a special Lie algebra structure on iterated gradients. The results lead in particular to a partial answer to the Minimum mean-square error conjecture.**

Index Terms—**Minimum Mean-Square Error, Gaussian noise, entropy, heat flow, iterated gradient, Lie bracket.**

I. Introduction

RECENT works in information theory have been developed on the Mutual Mean Square Error (MMSE) in the estimation of a random variable from its observation perturbated by a Gaussian noise [5], [6]. The concept of MMSE has been very fruitful, in particular has allowed for a simple proof of Shannon's monotonicity of entropy [4], [10] (see also [9]), first established in [1].

Given a random vector X in $\mathbb{R}^n$, let $X_t = X + \sqrt{2t}N$, $t \geq 0$, where N is an independent standard normal on $\mathbb{R}^n$. The Minimum Mean-Square Error

$$\mathrm{MMSE}(t) = \mathbb{E}\Big(\big|X - \mathbb{E}\big(X|X_t\big)\big|^2\Big), \quad t \geq 0.$$

is an estimate of the input X of the model given the noisy output X_t (the notation $|\cdot|$ being the Euclidean norm in $\mathbb{R}^n$). The MMSE can be studied as a function of the signal-to-noise ratio and as a functional of the input distribution. In particular, the aforementioned articles (and the references therein) provide a study of the analytic properties of the MMSE and of its derivatives with respect to the parameter of the noise perturbation, in relation with entropy and Fisher information.

The MMSE actually appears as an alternate description of the standard Fisher information along the heat flow. Given a random vector X with smooth positive density f with respect to the Lebesgue measure, let

$$\mathrm{I}(X) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} dx$$

denote its Fisher information. For each $t > 0$, the law of X_t is a convolution with a Gaussian kernel and as such admits a smooth density f_t. With some abuse in notation, let then $\mathrm{I}(t) = \mathrm{I}(X_t)$ be the Fisher information of X_t. As emphasized in [5], the MMSE actually directly connects to the Fisher information $\mathrm{I}(t)$, $t > 0$, along the flow via the identity

$$4t^2\mathrm{I}(t) = 2nt - \mathrm{MMSE}(t), \quad t > 0,$$

recalled below in Section II.

The purpose of this note is to relate some of the conclusions of [5] and [6] to earlier results on entropy and Fisher information expansions under heat flow [7] by means of the so-called Γ-calculus for Markov diffusion operators (cf. [3]). In particular, the Lie algebra structure on the iterated Γ-gradients emphasized in [7] provides the suitable structure to understand the successive derivatives of the MMSE and an induction mechanism to compute arbitrary orders. This tool moreover allows for a partial answer to the MMSE conjecture of [6] stating that the MMSE/Fisher information characterizes (up to the sign) the distribution of the underlying random variable.

The first paragraph of the paper (Section II) relates, following [5], the standard notion of Fisher information to the MMSE. Section III describes via the Γ-calculus the successive derivatives of the Fisher information by the Lie brackets from [7], and the subsequent section interprets the main conclusion on the MMSE itself. Section V emphasizes conditional cumulants towards the representation of the derivatives of the MMSE. On the basis of this representation, the next section addresses the MMSE conjecture in a particular case. The final Section VII collects a few bounds on the Fisher information and the MMSE of independent interest already discussed in [6]. The Appendix (Section VII) presents proofs of some specific properties of the Lie brackets in this context.

II. MMSE and Fisher Information

We start by connecting, following [5], the MMSE to the Fisher information along the heat flow.

On a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let therefore X be a random vector and $X_t = X + \sqrt{2t}N$, $t \geq 0$, where N is an independent standard normal on $\mathbb{R}^n$ with the Gaussian distribution

$$d\gamma(x) = \mathrm{e}^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}}.$$

Denote by $p_t(x)$ the Gaussian kernel

$$p_t(x) = \frac{1}{(4\pi t)^{n/2}} \mathrm{e}^{-|x|^2/4t}, \quad t > 0, \ x \in \mathbb{R}^n,$$

Manuscript received January 2, 2015; revised January 3, 2016; accepted March 29, 2016. Date of publication April 22, 2016; date of current version May 18, 2016.

The author is with the Institut de Mathématiques de Toulouse and the Institut Universitaire de France, University of Toulouse, Toulouse 31062, France (e-mail: ledoux@math.univ-toulouse.fr).

Communicated by I. Nikiforov, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2016.2555809

so that in particular, for any measurable and bounded $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\big(\varphi(X_t)\big) &= \mathbb{E}\bigg(\int_{\mathbb{R}^n} \varphi\big(X + \sqrt{2t}x\big)d\gamma(x)\bigg)\\ &= \mathbb{E}\bigg(\int_{\mathbb{R}^n} \varphi(X+x)p_t(x)dx\bigg)\\ &= \mathbb{E}\bigg(\int_{\mathbb{R}^n} \varphi(x)p_t(x-X)dx\bigg).\end{aligned}$$

Therefore, for any $t > 0$, the law of X_t has a strictly positive C^∞ density f_t with respect to the Lebesgue measure which is given as the convolution of the law of X with p_t, that is

$$f_t(x) = \mathbb{E}\big(p_t(x-X)\big), \quad x \in \mathbb{R}^n.$$

The Fisher information $\mathrm{I}(X_t) = \mathrm{I}(t)$ of X_t is therefore represented along the heat flow as

$$\mathrm{I}(t) = \int_{\mathbb{R}^n} \frac{|\nabla f_t|^2}{f_t}dx, \quad t > 0. \tag{2.1}$$

As will be seen below via the connection with the MMSE, $\mathrm{I}(t)$ is actually well defined (finite) for every $t > 0$.

Define now, for any $t > 0$,

$$g_t(x) = \mathbb{E}\big(Xp_t(x-X)\big), \quad x \in \mathbb{R}^n.$$

Then, assuming first that X is integrable, the conditional expectation $\mathbb{E}(X|X_t)$ may be represented by

$$\mathbb{E}\big(X|X_t\big) = \frac{g_t}{f_t}(X_t)$$

(almost surely). Indeed, if $\varphi : \mathbb{R}^n \to \mathbb{R}$ is measurable and bounded,

$$\begin{aligned}\mathbb{E}\big(\varphi(X_t)\mathbb{E}\big(X|X_t\big)\big) &= \mathbb{E}\big(X\varphi(X_t)\big)\\ &= \mathbb{E}\bigg(\int_{\mathbb{R}^n} X\varphi(X+x)p_t(x)dx\bigg)\\ &= \int_{\mathbb{R}^n} \varphi(x)\mathbb{E}\big(Xp_t(x-X)\big)dx,\end{aligned}$$

while

$$\mathbb{E}\Big(\varphi(X_t)\frac{g_t}{f_t}(X_t)\Big) = \int_{\mathbb{R}^n} \varphi(x)g_t(x)dx$$

from which the claim follows.

As a consequence, if $h_t = x - \frac{g_t}{f_t}$, then

$$X_t - \mathbb{E}\big(X|X_t\big) = h_t(X_t).$$

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 $\sqrt{2t}\mathbb{E}(N|X_t)$ since

$$X_t = \mathbb{E}\big(X_t|X_t\big) = \mathbb{E}\big(X|X_t\big) + \sqrt{2t}\mathbb{E}\big(N|X_t\big).$$

In particular, $X - \mathbb{E}(X|X_t)$ makes also sense and has moments of all orders. Therefore, the Minimum Mean-Square Error

$$\mathrm{MMSE}(t) = \mathbb{E}\Big(\big|X - \mathbb{E}\big(X|X_t\big)\big|^2\Big), \quad t \geq 0. \tag{2.2}$$

is well-defined for every $t \geq 0$.

Now, observe that by the Lebesgue differentiation theorem,

$$2t\nabla f_t = \mathbb{E}\big((X-x)p_t(x-X)\big) = g_t - xf_t.$$

Hence

$$\begin{aligned}4t^2\mathrm{I}(t) &= \int_{\mathbb{R}^n} \frac{1}{f_t}\big|xf_t - g_t\big|^2dx\\ &= \int_{\mathbb{R}^n} f_t\Big|x - \frac{g_t}{f_t}\Big|^2dx\\ &= \mathbb{E}\Big(\big|X_t - \mathbb{E}\big(X|X_t\big)\big|^2\Big)\\ &= 2t\mathbb{E}\big(|N|^2\big) - \mathbb{E}\Big(\big|X - \mathbb{E}\big(X|X_t\big)\big|^2\Big)\end{aligned}$$

since $X_t - \mathbb{E}(X|X_t)$ is orthogonal to $X - \mathbb{E}(X|X_t)$.

Therefore the Fisher information I and the MMSE are related by the identity

$$4t^2\mathrm{I}(t) = 2nt - \mathrm{MMSE}(t), \quad t > 0. \tag{2.3}$$

In particular, the Fisher information $\mathrm{I}(t)$ is well-defined for every $t > 0$ whatsoever the underlying random vector X.

The Mutual Mean-Square Error MMSE of [5] and [6] is actually studied as a function of the noise parameter as

$$\begin{aligned}\mathrm{mmse}(s) &= \mathbb{E}\Big(\big|X - \mathbb{E}\big(X|\sqrt{s}X + N\big)\big|^2\Big)\\ &= \mathbb{E}\Big(\big|X - \mathbb{E}\big(X|X_t\big)\big|^2\Big) = \mathrm{MMSE}(t)\end{aligned} \tag{2.4}$$

for $t = t(s) = \frac{1}{2s}$, $s > 0$. It is indeed in this form that the successive derivatives of the Fisher information and of the MMSE will connect most suitably (as opposed to the relation (2.3)). The study of the function $\mathrm{mmse}(s)$, $s > 0$, will then require various, unfortunately somewhat heavy, changes of variables in the subsequent analysis.

It is shown in [6] that the mmse function is infinitely differentiable at every $s > 0$, and at $s = 0$ whenever X has moments of all orders (and thus also the MMSE and the Fisher information). We freely use these properties below.

III. Heat Flow Derivatives

We present here in the context of the heat semigroup on $\mathbb{R}^n$ the results of [7] on the successive derivatives of entropy and Fisher information.

The derivatives of entropy along the heat flow have been examined from various viewpoints in the literature, in particular in connection with logarithmic Sobolev inequalities (cf. e.g. [2], [3]). Given a random vector X and $X_t = X + \sqrt{2t}N$, $t > 0$, with probability density f_t with respect to the Lebesgue measure, consider, provided it exists, the entropy (with a positive sign convention) along the flow

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

The classical de Brujin formula expresses that

$$\frac{d}{dt}\mathrm{H}(t) = -\mathrm{I}(t) \tag{3.1}$$

(where we recall the Fisher information $\mathrm{I}(t)$ of (2.1)). Note that with $v = \log f_t$, $t > 0$,

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

Similarly, as is classical from the Bakry-Émery calculus [2], [3],

$$\frac{d^2}{dt^2}\mathrm{H}(t) = 2\int_{\mathbb{R}^n} f_t |\nabla^2 v|^2 dx \tag{3.2}$$

where $\nabla^2 v$ is the matrix of the second derivatives of v. (According to the end of Section II, it may be implicitly assumed here and below that entropy and Fisher information are infinitively differentiable on $(0,\infty)$.) Note that while the calculus of [2], [3], and [7] is developed along a semigroup from a given probability density, it makes sense in the same way along the densities f_t, $t > 0$, since the latter solve similarly the heat equation $\partial_t f_t = \Delta f_t$.

Nevertheless, it is known from [7] that the rule

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

cannot be iterated for $\ell \geq 3$. The work [7] 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. [3]). We recall here some of the main elements of this description.

Consider the bilinear symmetric carré du champ operator

$$\Gamma(u_1, u_2) = \nabla u_1 \cdot \nabla u_2$$

acting on smooth functions $u_1, u_2 : \mathbb{R}^n \to \mathbb{R}$. For simplicity, set $\Gamma(u) = \Gamma(u,u)$. Introduce the Lie bracket between the Laplace operator Δ (or more generally a linear operator L) and a j-multilinear ($j \geq 1$) symmetric form $B(u) = B(u, \ldots, u)$ on (smooth) functions by

$$2[\Delta, B](u) = \Delta B(u) - jB(\Delta u, u, \ldots, u).$$

Here $[\Delta, B]$ is actually a j-multilinear form which is defined by polarization from $[\Delta, B](u) = [\Delta, B](u, \ldots, u)$. For example if B is 2-linear,

$$\begin{aligned} 2[\Delta, B](u_1, u_2) &= \Delta B(u_1, u_2) - B(\Delta u_1, u_2) - B(u_1, \Delta u_2) \\ &= [\Delta, B](u_1 + u_2, u_1 + u_2) \\ &\quad - [\Delta, B](u_1, u_1) - [\Delta, B](u_2, u_2). \end{aligned}$$

It is classical and easy to see that if we define $\Gamma_\ell(u_1, u_2) = \nabla^\ell u_1 \cdot \nabla^\ell u_2$ (the dot product being understood as the scalar product between ℓ-tensors) with $\Gamma_1 = \Gamma$, for every $\ell \geq 1$,

$$[\Delta, \Gamma_\ell](u) = \Gamma_{\ell+1}(u) = \nabla^\ell u \cdot \nabla^\ell u = |\nabla^{\ell+1} u|^2. \tag{3.3}$$

However, in order to study the time derivatives of entropy, it is necessary to go one step further in the brackets and to consider

$$2[\Gamma, B](u) = 2\Gamma\big(B(u), u\big) - jB\big(\Gamma(u), u, \ldots, u\big).$$

Here, since Γ is 2-multilinear, $[\Gamma, B]$ is now $(j+1)$-multilinear and obtained by polizarization from $[\Gamma, B](u) = [\Gamma, B](u, \ldots, u)$. For example, if B is 2-linear,

$$\begin{aligned} 2[\Gamma, B](u_1, u_2, u_3) = \frac{1}{3}\sum_\sigma \Big[&\Gamma\big(B(u_{\sigma(1)}, u_{\sigma(2)}), u_{\sigma(3)}\big) \\ &- B\big(\Gamma(u_{\sigma(1)}, u_{\sigma(2)}), u_{\sigma(3)}\big)\Big] \end{aligned}$$

where the sum runs over all permutations σ of $\{1, 2, 3\}$. We refer to [7] for more details.

In this framework, the main conclusion of [7] is summarized in the following statement. According to the preceding, define for every $\ell \geq 2$,

$$\widetilde{\Gamma}_\ell = [\Delta + \Gamma, \widetilde{\Gamma}_{\ell-1}]$$

recursively from $\widetilde{\Gamma}_1(u) = \Gamma(u) = |\nabla u|^2$ (for smooth $u : \mathbb{R}^n \to \mathbb{R}$). Note that actually

$$\widetilde{\Gamma}_\ell = [\Delta, \widetilde{\Gamma}_{\ell-1}] + [\Gamma, \widetilde{\Gamma}_{\ell-1}]$$

which by induction yields a sum of j-multilinear forms with $2 \leq j \leq \ell$. Recall $v = \log f_t$.

Theorem 1: For every $\ell \geq 1$,

$$\frac{d^\ell}{dt^\ell}\mathrm{H}(t) = (-1)^\ell 2^{\ell-1}\int_{\mathbb{R}^n} f_t \widetilde{\Gamma}_\ell(v) dx.$$

For the further purposes, we will rather deal (equivalently by (3.1)) with the derivatives of the Fisher information. Theorem 1 therefore expresses that

$$\frac{d^\ell}{dt^\ell}\mathrm{I}(t) = (-2)^\ell \int_{\mathbb{R}^n} f_t \widetilde{\Gamma}_{\ell+1}(v) dx, \quad \ell \geq 0. \tag{3.4}$$

It is a main feature, at the root of the Bakry-Émery criterion for logarithmic Sobolev inequalities (cf. [2], [3], [7]), that $\widetilde{\Gamma}_2 = \Gamma_2$. Indeed,

$$\widetilde{\Gamma}_2 = [\Delta + \Gamma, \Gamma] = [\Delta, \Gamma] = \Gamma_2.$$

However, the whole point of Theorem 1 is that this identity does not extend to the higher orders. Specifically,

$$\begin{aligned} \widetilde{\Gamma}_2 &= \Gamma_2, \\ \widetilde{\Gamma}_3 &= \Gamma_3 + [\Gamma_1, \Gamma_2], \\ \widetilde{\Gamma}_4 &= \Gamma_4 + 2[\Gamma_1, \Gamma_3] + [\Gamma_1, [\Gamma_1, \Gamma_2]], \\ \widetilde{\Gamma}_5 &= \Gamma_5 + 3[\Gamma_1, \Gamma_4] + 2[\Gamma_2, \Gamma_3] + 3[\Gamma_1, [\Gamma_1, \Gamma_3]] \\ &\quad + [\Gamma_2, [\Gamma_1, \Gamma_2]] + [\Gamma_1, [\Gamma_1, [\Gamma_1, \Gamma_2]]], \\ &\ldots \end{aligned}$$

For example, in dimension one, $\widetilde{\Gamma}_1(u) = u'^2$, $\widetilde{\Gamma}_2(u) = u''^2$, but

$$\begin{aligned} \widetilde{\Gamma}_3(u) &= u'''^2 - 2u''^3, \\ \widetilde{\Gamma}_4(u) &= u^{(4)^2} - 12u''u'''^2 + 6u''^4, \\ \widetilde{\Gamma}_5(u) &= u^{(5)^2} - 30u'''^2u^{(4)} - 20u''u^{(4)^2} \\ &\quad + 120u''^2u'''^2 - 24u''^5. \end{aligned}$$

The successive expressions $\widetilde{\Gamma}_\ell(u)$ are more and more cumbersome, but precisely given by the Lie bracket structure.

It will be useful to record the following rough description of the $\widetilde{\Gamma}_\ell(u)$, $\ell \geq 1$, in dimension one. Its proof is postponed to the Appendix, Section VII. Recall $\widetilde{\Gamma}_1(u) = \Gamma(u) = u'^2$.

Proposition 2: For every $\ell \geq 2$ and every smooth function $u : \mathbb{R} \to \mathbb{R}$,

$$\widetilde{\Gamma}_\ell(u) = \widetilde{R}_\ell\big(u^{(2)}, \dots, u^{(\ell)}\big) \tag{3.5}$$

where $u^{(j)}$, $j \geq 2$, denote the successive derivatives of u and where $\widetilde{R}_\ell = \widetilde{R}_\ell(Z_2, \dots, Z_\ell)$ is a polynomial in the variables $Z_2, \dots, Z_\ell$ of the form

$$\widetilde{R}_\ell(Z_2, \dots, Z_\ell) = Z_\ell^2 + R_{\ell-1}(Z_2, \dots, Z_{\ell-1})$$

where $R_{\ell-1}$ is some polynomial of $\ell - 2$ coordinates ($R_1 \equiv 0$). The polynomial $R_{\ell-1}$ is not explicit, although it may be noticed that it contains $(-1)^\ell(\ell-1)!Z_2^\ell$. Moreover $\widetilde{\Gamma}_\ell(u)$ is 2ℓ-homogeneous under the transformation $u(x) \mapsto u_\lambda(x) = u(\lambda x)$, $\lambda \in \mathbb{R}$, in the sense that

$$\widetilde{\Gamma}_\ell(u_\lambda)(x) = \lambda^{2\ell}\widetilde{\Gamma}_\ell(u)(\lambda x). \tag{3.6}$$

In particular, $\widetilde{R}_\ell(\lambda^2 Z_2, \dots, \lambda^\ell Z_\ell) = \lambda^{2\ell}\widetilde{R}_\ell(Z_2, \dots, Z_\ell)$.

For example, according to the above expressions for $\widetilde{\Gamma}_2, \widetilde{\Gamma}_3, \widetilde{\Gamma}_4, \widetilde{\Gamma}_5$,

$$\begin{aligned}
\widetilde{R}_2(Z_2) &= Z_2^2, \\
\widetilde{R}_3(Z_2, Z_3) &= Z_3^2 - 2Z_2^3, \\
\widetilde{R}_4(Z_2, Z_3, Z_4) &= Z_4^2 - 12Z_2Z_3^2 + 6Z_2^4, \\
\widetilde{R}_5(Z_2, Z_3, Z_4, Z_5) &= Z_5^2 - 30Z_3^2Z_4 - 20Z_2Z_4^2 \\
&\quad + 120Z_2^2Z_3^2 - 24Z_2^5.
\end{aligned}$$

IV. Derivatives of the MMSE

The preceding section provides a full description of the successive derivatives of the Fisher information $\mathrm{I}(t)$, $t > 0$. We develop here the connection with the derivatives of the mmse

$$\mathrm{mmse}(s) = \mathrm{MMSE}(t), \quad t = t(s) = \frac{1}{2s}, \quad s > 0,$$

of (2.4) via suitable changes of functions and variables.

Recall $v = v(t,x) = \log f_t(x)$, $t > 0$, $x \in \mathbb{R}^n$, which defines a C^∞ function of both the space and time variables. Recall also the Gaussian kernel $p_t(x) = \frac{1}{(4\pi t)^{n/2}}\mathrm{e}^{-|x|^2/4t}$ and set $h = h(t,x) = \frac{f_t(x)}{p_t(x)}$. Since $\partial_t p_t = \Delta p_t$ as well as $\partial_t f_t = \Delta f_t$, it is an easy exercise to check that h solves the pde

$$\partial_t h = \Delta h - \frac{1}{t}x \cdot \nabla h.$$

With $t = t(s) = \frac{1}{2s}$, set furthermore $k = k(s,x) = h\big(t, \sqrt{2t}x\big) = h\big(\frac{1}{2s}, \frac{x}{\sqrt{s}}\big)$. Then

$$\partial_s k = -\frac{1}{2s}\mathrm{L}k \tag{4.1}$$

where L is the Ornstein-Uhlenbeck generator acting on smooth function φ on $\mathbb{R}^n$ as

$$\mathrm{L}\varphi = \Delta\varphi - x \cdot \nabla\varphi.$$

The operator L has to be thought of as the Laplacian with a drift in such a way that the standard Gaussian measure γ is its invariant and symmetric measure.

Let then $q = \log k$ and introduce, for $s > 0$,

$$m(s) = \frac{1}{s}\int_{\mathbb{R}^n} k\Upsilon(q)d\gamma = \frac{1}{s}\int_{\mathbb{R}^n}\frac{\Upsilon(k)}{k}d\gamma \tag{4.2}$$

where $\Upsilon(q) = \Gamma(q) = |\nabla q|^2$ (the change to the letter Υ will be justified below). The function $m(s)$, $s > 0$, is closely linked to the mmse. To this aim, observe first that after a change of variables, with always $t = t(s) = \frac{1}{2s}$,

$$m(s) = 4t^2\int_{\mathbb{R}^n} f_t|\nabla w|^2dx.$$

where $w = \log h$. Indeed, recalling that $k = h\big(\frac{1}{2s}, \frac{x}{\sqrt{s}}\big)$, after the change of variables $\frac{x}{\sqrt{s}} = y$, and $t = \frac{1}{2s}$,

$$\begin{aligned}
&\frac{1}{s}\int_{\mathbb{R}^n} k\Upsilon(q)d\gamma \\
&= \frac{1}{s}\int_{\mathbb{R}^n}\frac{|\nabla k|^2}{k}d\gamma \\
&= \frac{1}{s^2}\int_{\mathbb{R}^n}\frac{1}{h\big(\frac{1}{2s}, \frac{x}{\sqrt{s}}\big)}|\nabla h|^2\big(\tfrac{1}{2s}, \tfrac{x}{\sqrt{s}}\big)\mathrm{e}^{-|x|^2/2}\frac{dx}{(2\pi)^{n/2}} \\
&= 4t^2\int_{\mathbb{R}^n}\frac{1}{h(t,y)}|\nabla h|^2(t,y)p_t(y)dy \\
&= \int_{\mathbb{R}^n} f_t\frac{|\nabla h|^2}{h^2}dy \\
&= 4t^2\int_{\mathbb{R}^n} f_t|\nabla w|^2dy.
\end{aligned}$$

Since $w = \log h = v - \log p_t$, going back to $v = \log f_t$,

$$\begin{aligned}
|\nabla w|^2 &= \Big|v + \frac{x}{2t}\Big|^2 = |\nabla v|^2 + \frac{1}{t}x \cdot \nabla v + \frac{|x|^2}{4t^2} \\
&= \frac{|\nabla f_t|^2}{f_t^2} + \frac{1}{tf_t}x \cdot \nabla f_t + \frac{|x|^2}{4t^2}.
\end{aligned}$$

Therefore

$$\begin{aligned}
m(s) &= 4t^2\int_{\mathbb{R}^n} f_t|\nabla w|^2dx \\
&= 4t^2\mathrm{I}(t) + 4t\int_{\mathbb{R}^n} x \cdot \nabla f_t dx + \int_{\mathbb{R}^n}|x|^2f_tdx \\
&= 4t^2\mathrm{I}(t) - 4nt + \int_{\mathbb{R}^n}|x|^2f_tdx
\end{aligned}$$

where we used that

$$\begin{aligned}
\int_{\mathbb{R}^n} x \cdot \nabla f_tdx &= -\int_{\mathbb{R}^n} x \cdot \mathbb{E}\Big(\frac{x-X}{2t}p_t(x-X)\Big)dx \\
&= -\frac{1}{2t}\mathbb{E}\bigg(\int_{\mathbb{R}^n} x \cdot (x-X)p_t(x-X)dx\bigg) \\
&= -\frac{1}{2t}\mathbb{E}\bigg(\int_{\mathbb{R}^n}(x+X) \cdot xp_t(x)dx\bigg) \\
&= -\frac{1}{2t}\int_{\mathbb{R}^n}|x|^2p_t(x)dx = -n.
\end{aligned}$$

Using in addition that

$$\int_{\mathbb{R}^n}|x|^2f_tdx = \mathbb{E}\big(|X_t|^2\big) = \mathbb{E}\big(|X|^2\big) + 2nt,$$

it follows that

$$m(s) = 4t^2\mathrm{I}(t) - 2nt + \mathbb{E}\big(|X|^2\big) = -\mathrm{MMSE}(t) + \mathbb{E}\big(|X|^2\big).$$

Hence

$$m(s) = -\mathrm{mmse}(s) + \mathbb{E}\big(|X|^2\big). \tag{4.3}$$

This identity requires a second moment on X. However, as will be clear from the main Theorem 7 below, this condition is no more necessary at the first (and next) derivatives (for which $\mathbb{E}(|X|^2)$ cancels out).

The analysis of the derivatives of the mmse is thus brought back to the derivatives of the function m of (4.2). The function m is of the form of a Fisher information but, with respect to the setting of Section II, the underlying generator and semigroup are the Ornstein-Uhlenbeck ones rather than the standard heat (Brownian) generator and semigroup, with the Gaussian measure γ as invariant and symmetric measure. The principle of proof emphasized in [7] is nevertheless exactly the same, with simply some variations in the expression of the iterated gradients.

Denote by Υ_ℓ, $\ell \geq 1$, the iterated gradients associated to Ornstein-Uhlenbeck operator $\mathrm{L} = \Delta - x \cdot \nabla$ defined by $\Upsilon_\ell = [\mathrm{L}, \Upsilon_{\ell-1}]$, $\ell \geq 2$, with $\Upsilon_1(u) = \Gamma(u) = |\nabla u|^2$. As developed in [7], these iterated gradients are closely related to the standard gradients Γ_ℓ, $\ell \geq 1$, of (3.3) in the following way. Let Q_ℓ the polynomial in the variable Z

$$Q_\ell(Z) = \prod_{i=0}^{\ell-1}(Z-i) = \sum_{i=1}^{\ell} a_i^\ell Z^i,$$

and set by extension

$$Q_\ell(\Upsilon) = \sum_{i=1}^{\ell} a_i^\ell \Upsilon_i.$$

Then, for every $\ell \geq 1$,

$$Q_\ell(\Upsilon) = \Gamma_\ell.$$

For example, $Q_1(\Upsilon) = \Upsilon_1 = \Gamma_1$, $Q_2(\Upsilon) = \Upsilon_2 - \Upsilon_1 = \Gamma_2$, $Q_3(\Upsilon) = \Upsilon_3 - 3\Upsilon_2 + 2\Upsilon_1 = \Gamma_3$.

As in the standard Laplacian case, introduce the brackets $\widetilde{\Upsilon}_\ell = [\mathrm{L}+\Gamma, \widetilde{\Upsilon}_{\ell-1}]$, $\ell \geq 2$, starting from $\widetilde{\Upsilon}_1 = \Upsilon_1 = \Gamma_1 = \Gamma$, and set then

$$Q_\ell(\widetilde{\Upsilon}) = \sum_{i=1}^{\ell} a_i^\ell \widetilde{\Upsilon}_i.$$

Note that $\widetilde{\Upsilon}_2 = \Upsilon_2$ $(= \Gamma_2 + \Gamma)$.

The differentiation process expressed by Theorem 1 then takes the same form. At the first and second steps, the derivatives of

$$m(s) = \frac{1}{s}\int_{\mathbb{R}^n} k\Upsilon(q)d\gamma, \quad s > 0,$$

are given by

$$\begin{aligned}
\frac{d}{ds}m(s) &= \frac{1}{s^2}\int_{\mathbb{R}^n} k\big[\Upsilon_2(q) - \Upsilon_1(q)\big]d\gamma \\
&= \frac{1}{s^2}\int_{\mathbb{R}^n} kQ_2(\widetilde{\Upsilon})(q)d\gamma, \\
\frac{d^2}{ds^2}m(s) &= \frac{1}{s^3}\int_{\mathbb{R}^n} k\big[\widetilde{\Upsilon}_3(q) - 3\Upsilon_2(q) + 2\Upsilon_1(q)\big]d\gamma \\
&= \frac{1}{s^3}\int_{\mathbb{R}^n} kQ_3(\widetilde{\Upsilon})(q)d\gamma.
\end{aligned}$$

By induction, we end up with the following conclusion [7].

Theorem 3: In the preceding notation, for every $\ell \geq 0$,

$$\frac{d^\ell}{ds^\ell}m(s) = \frac{1}{s^{\ell+1}}\int_{\mathbb{R}^n} kQ_{\ell+1}(\widetilde{\Upsilon})(q)d\gamma.$$

For the proof, we may refer more precisely to the key relation (3.7) from [7] which expresses that

$$\begin{aligned}
\frac{d}{ds}\int_{\mathbb{R}^n} kQ_\ell(\widetilde{\Upsilon})(q)d\gamma &= \frac{\ell}{s}\int_{\mathbb{R}^n} kQ_\ell(\widetilde{\Upsilon})(q)d\gamma \\
&\quad + \frac{1}{s}\int_{\mathbb{R}^n} kQ_{\ell+1}(\widetilde{\Upsilon})(q)d\gamma,
\end{aligned}$$

from which the inductive step follows immediately.

Theorem 3 provides a full description of the derivatives of the function m, and therefore of the mmse, in terms of the iterated gradients Υ_ℓ and their brackets. While the following is not strictly necessary towards the main conclusion, it will be convenient to record it here as the calculus with the standard gradients Γ_ℓ is somewhat more direct than the one with the Υ_ℓ's. Actually, in the same way as $Q_\ell(\Upsilon) = \Gamma_\ell$, we also have

Proposition 4: For every $\ell \geq 1$,

$$Q_\ell(\widetilde{\Upsilon}) = \widetilde{\Gamma}_\ell.$$

We postpone the proof of this proposition to the appendix.

As a consequence of Proposition 4, and Theorem 3,

$$\frac{d^\ell}{ds^\ell}m(s) = \frac{1}{s^{\ell+1}}\int_{\mathbb{R}^n} k\widetilde{\Gamma}_{\ell+1}(q)d\gamma.$$

According to the change of variables (3.6), we therefore obtain a first description of the derivatives of the mmse. Recall $t = t(s) = \frac{1}{2s}$ and $w = \log\frac{f_{t(s)}}{p_{t(s)}}$.

Corollary 5: For every $\ell \geq 1$,

$$\frac{d^\ell}{ds^\ell}\mathrm{mmse}(s) = -\frac{1}{s^{2(\ell+1)}}\int_{\mathbb{R}^n} f_{t(s)}\widetilde{\Gamma}_{\ell+1}(w)dx.$$

V. Conditional Cumulants

In order to make Corollary 5 effective, we express in this section the quantities $\widetilde{\Gamma}_\ell(w)$, $\ell \geq 1$, in terms of conditional cumulants. For simplicity, we deal with dimension one and analyze first the spatial derivatives of $v = v(t,x) = \log f_t(x)$, $t > 0$, denoted $v^{(\ell)}$, $\ell \geq 1$.

Denote by $\kappa_t^\ell = \kappa_t^\ell(x)$, $\ell \geq 1$, the conditional cumulants given $X_t = x$ defined by the expansions in $\lambda \in \mathbb{R}$,

$$\log\frac{\mathbb{E}(\mathrm{e}^{\lambda X}p_t(x-X))}{f_t(x)} = \sum_{\ell=1}^{\infty}\frac{\lambda^\ell}{\ell!}\kappa_t^\ell(x). \tag{5.1}$$

The first (conditional) cumulant is the mean

$$\kappa_t^1(x) = \mathbb{E}\big(X|X_t = x\big)$$

while the second cumulant is the variance

$$\kappa_t^2(x) = \mathbb{E}\big(X^2|X_t = x\big) - \mathbb{E}\big(X|X_t = x\big)^2.$$

By the kernel representation of the heat semigroup, for every $t > 0$ and $\lambda \in \mathbb{R}$,

$$\begin{aligned}
f_t(x + 2t\lambda) &= \mathbb{E}\big(p_t(x + 2t\lambda - X)\big) \\
&= \frac{1}{\sqrt{4\pi t}}\mathbb{E}\big(\mathrm{e}^{-(x+2t\lambda-X)^2/4t}\big) \\
&= \mathrm{e}^{-\lambda^2 t - \lambda x}\mathbb{E}\big(\mathrm{e}^{\lambda X}p_t(x-X)\big).
\end{aligned}$$

Hence

$$\begin{aligned} v(x+2t\lambda) &= \log f_t(x+2t\lambda) \\ &= -\lambda^2 t - \lambda x + \log \mathbb{E}\big(\mathrm{e}^{\lambda X} p_t(x-X)\big) \end{aligned}$$

that is, by (5.1),

$$v(x+2t\lambda) - v(x) = -\lambda^2 t - \lambda x + \sum_{\ell=1}^{\infty} \frac{\lambda^\ell}{\ell!} \kappa_t^\ell(x).$$

Since by a (formal) Taylor expansion, $v(x + 2t\lambda) = \sum_{\ell=0}^{\infty} \frac{(2t\lambda)^\ell}{\ell!} v^{(\ell)}(x)$, it follows by comparison that $2tv' = \kappa_t^1 - x$, $4t^2 v'' = \kappa_t^2 - 2t$ and, for every $\ell \geq 3$, $(2t)^\ell v^{(\ell)} = \kappa_t^\ell$. Recalling that

$$w = v - \log p_t = v - \frac{x^2}{4t} - \frac{1}{2}\log(4\pi t),$$

we therefore reach the following conclusion.

Proposition 6: For every $\ell \geq 1$,

$$(2t)^\ell w^{(\ell)} = \kappa_t^\ell. \tag{5.2}$$

We can now go back to Corollary 5. Recall the polynomials $\widetilde{R}_\ell$, $\ell \geq 1$, from Proposition 2 so that

$$\begin{aligned} \widetilde{\Gamma}_{\ell+1}(w) &= \widetilde{R}_{\ell+1}\big(w^{(2)}, \ldots, w^{(\ell+1)}\big) \\ &= (2t)^{-2(\ell+1)} \widetilde{R}_{\ell+1}\big(\kappa_t^2, \ldots, \kappa_t^{\ell+1}\big). \end{aligned}$$

Hence, with $t = t(s) = \frac{1}{2s}$,

$$\frac{d^\ell}{ds^\ell}\mathrm{mmse}(s) = -\int_{\mathbb{R}^n} f_t \widetilde{R}_{\ell+1}(\kappa_t^2, \ldots, \kappa_t^{\ell+1}) dx. \tag{5.3}$$

Let K_s^ℓ be the conditional cumulants of X given $\sqrt{s}X + N$, equivalently $X_{t(s)}$ to shorten the notation, that is with generating series in $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}\big(\mathrm{e}^{\lambda X} | X_{t(s)}\big) = \sum_{\ell=1}^{\infty} \frac{\lambda^\ell}{\ell!} K_s^\ell.$$

Note that all these conditional cumulants, besides the first one, have moments of all orders (on the same scheme as for (2.2)). Indeed, as is classical, $\frac{K_s^\ell}{\ell!}$, $\ell \geq 2$, is the coefficient in front of λ^ℓ of the formal series expansion

$$-\sum_{k\geq 1} \frac{1}{k}\bigg(-\sum_{m\geq 2} \frac{\lambda^m}{m!} M_s^m\bigg)^k$$

where

$$M_s^m = \mathbb{E}\Big(\big[X - \mathbb{E}\big(X | X_{t(s)}\big)\big]^m \big| X_{t(s)}\Big), \quad m \geq 2, \tag{5.4}$$

are the conditional centered moments. For example,

$$K_s^2 = M_s^2, \quad K_s^3 = M_s^3, \quad K_s^4 = M_s^4 - 3\big(M_s^2\big)^2.$$

In addition, $\mathbb{E}(K_s^2) = \mathbb{E}(M_s^2) = \mathrm{MMSE}(t(s))$.

Recalling that X_t has law $f_t dx$, the identities (5.3) lead to the following main conclusion describing the successive derivatives of the mmse in terms of the conditional cumulants K_s^ℓ, $\ell \geq 1$.

Theorem 7: In the preceding notation, for every $\ell \geq 1$, at $s > 0$,

$$\frac{d^\ell}{ds^\ell}\mathrm{mmse}(s) = -\mathbb{E}\big(\widetilde{R}_{\ell+1}(K_s^2, \ldots, K_s^{\ell+1})\big). \tag{5.5}$$

As an illustration, the first derivatives are given by

$$\begin{aligned} \frac{d}{ds}\mathrm{mmse}(s) &= -\mathbb{E}\big((K_s^2)^2\big), \\ \frac{d^2}{ds^2}\mathrm{mmse}(s) &= -\mathbb{E}\big((K_s^3)^2 - 2(K_s^2)^3\big), \\ \frac{d^3}{ds^3}\mathrm{mmse}(s) &= -\mathbb{E}\big((K_s^4)^2 - 12K_s^2(K_s^3)^2 + 6(K_s^2)^4\big), \\ \frac{d^4}{ds^4}\mathrm{mmse}(s) &= -\mathbb{E}\big((K_s^5)^2 - 30(K_s^3)^2 K_s^4 - 20K_s^2(K_s^4)^2 \\ &\qquad + 120(K_s^2)^2(K_s^3)^2 - 24(K_s^2)^5\big). \end{aligned}$$

These identities are in accordance with the formulas of [6] and [8] expressed in terms of the conditional central moments (5.4). Indeed,

$$\begin{aligned} \frac{d}{ds}\mathrm{mmse}(s) &= -\mathbb{E}\big((M_s^2)^2\big), \\ \frac{d^2}{ds^2}\mathrm{mmse}(s) &= -\mathbb{E}\big((M_s^3)^2 - 2(M_s^2)^3\big), \\ \frac{d^3}{ds^3}\mathrm{mmse}(s) &= -\mathbb{E}\big((M_s^4)^2 - 6M_s^4(M_s^2)^2 \\ &\qquad - 12M_s^2(M_s^3)^2 + 15(M_s^2)^4\big) \\ \frac{d^4}{ds^4}\mathrm{mmse}(s) &= -\mathbb{E}\big((M_s^5)^2 - 20M_s^2 M_s^3 M_s^5 - 30(M_s^3)^2 M_s^4 \\ &\qquad - 20M_s^2(M_s^4)^2 + 120(M_s^2)^3 M_s^4 \\ &\qquad + 310(M_s^2)^2(M_s^3)^2 - 204(M_s^2)^5\big). \end{aligned}$$

To illustrate these identities, observe for example that if X is centered normal with variance σ^2, then

$$\mathrm{mmse}(s) = \frac{\sigma^2}{1+\sigma^2 s}, \quad s > 0,$$

while $K_s^1 = \frac{\sigma^2 s}{1+\sigma^2 s} X_{t(s)}$, $K_s^2 = \frac{\sigma^2}{1+\sigma^2 s}$, and $K_s^\ell = 0$ if $\ell \geq 2$, so that

$$\begin{aligned} \frac{d^\ell}{ds^\ell}\mathrm{mmse}(s) &= (-1)^\ell \ell! \Big(\frac{\sigma^2}{1+\sigma^2 s}\Big)^{\ell+1} \\ &= -(-1)^{\ell+1} \ell! \mathbb{E}\big((K_s^2)^{\ell+1}\big). \end{aligned}$$

VI. On the MMSE Conjecture

The work [6] raises the conjecture that the knowledge of the Minimum Mean-Square Error $\mathrm{MMSE}(t)$, $t \geq 0$ (equivalently the entropy $\mathrm{H}(t)$ or the Fisher information $\mathrm{I}(t)$, $t > 0$, along the flow), determines the underlying distribution of a centered real-valued random variable X up to the sign, namely either the law of X or of $-X$. Note indeed that

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

is invariant by translation of X by a constant, so does not characterize the mean, and by the change of X into $-X$.

The idea expressed by the preceding investigation is that the MMSE could act as a kind of moment generating function, its successive derivatives (at 0) determining the moments or the cumulants of the underlying distribution. To make use of Theorem 7 following this principle, observe first that the expected conditional cumulants $\mathbb{E}(K_s^\ell)$, $\ell \geq 1$, converge as

$s \to 0$ to the cumulants $K^\ell(X)$ of X, provided all moments of X are finite. It might be of interest to briefly justify this intuitive statement. To this task, let us first show that $\mathbb{E}(X|X_t) \to \mathbb{E}(X)$ almost surely as $t \to \infty$ whenever X is integrable. Indeed, in the notation of Section II,

$$\mathbb{E}\big(X|X_t\big) = \frac{g_t}{f_t}(X_t) = \frac{\widetilde{\mathbb{E}}(\widetilde{X} p_t(X_t - \widetilde{X}))}{\widetilde{\mathbb{E}}(p_t(X_t - \widetilde{X}))} = \frac{\widetilde{\mathbb{E}}(\widetilde{X} \mathrm{e}^{-(X_t - \widetilde{X})^2/4t})}{\widetilde{\mathbb{E}}(\mathrm{e}^{-(X_t - \widetilde{X})^2/4t})}$$

where $\widetilde{X}$ is an independent copy of X. Now, if $\mathbb{E}(|X|) = \mathbb{E}(|\widetilde{X}|) < \infty$, by dominated convergence, almost surely,

$$\lim_{t\to\infty} \widetilde{\mathbb{E}}\big(\widetilde{X}\mathrm{e}^{-(X_t-\widetilde{X})^2/4t}\big) = \mathrm{e}^{-N^2/2}\mathbb{E}(X)$$

while

$$\lim_{t\to\infty} \widetilde{\mathbb{E}}\big(\mathrm{e}^{-(\widetilde{X}^2 - 2\widetilde{X}X_t)/4t}\big) = \mathrm{e}^{-N^2/2}$$

from which the claim follows.

Assuming all moments of X are finite, for each m, the family $[X - \mathbb{E}(X|X_t)]^m$, $t \geq 0$, is uniformly integrable. Therefore, as $s \to 0$ (recall $s = \frac{1}{t} = \frac{1}{2t(s)}$), the expected conditional moments $\mathbb{E}(M_s^m)$ converge to the centered moments $\mathbb{E}([X - \mathbb{E}(X)]^m)$ of X. (In particular, $\mathrm{MMSE}(t) \to K^2(X)$ as $t \to \infty$.) Now, the same type of arguments shows that, for every $k \geq 1$, and every $m_1, \ldots, m_k \geq 1$,

$$\mathbb{E}(M_s^{m_1} \cdots M_s^{m_k}) \to \mathbb{E}\big(\big[X - \mathbb{E}(X)\big]^{m_1}\big) \cdots \mathbb{E}\big(\big[X - \mathbb{E}(X)\big]^{m_k}\big)$$

and similarly with the conditional cumulants K_s^ℓ,

$$\mathbb{E}(K_s^{m_1} \cdots K_s^{m_k}) \to K^{m_1}(X) \cdots K^{m_k}(X).$$

On the basis of Theorem 7 and the latter, a partial result towards the MMSE conjecture may therefore be emphasized. Recall from Proposition 2 that $\widetilde{R}_2(Z_2) = Z_2^2$ and, for every $\ell \geq 2$,

$$\widetilde{R}_{\ell+1}(Z_2, \ldots, Z_{\ell+1}) = Z_{\ell+1}^2 + R_\ell(Z_2, \ldots, Z_\ell).$$

Hence, in the limit as $s \to 0$ in (5.5), the quantities

$$\big(K^{\ell+1}(X)\big)^2 + R_\ell\big(K^2(X), \ldots, K^\ell(X)\big), \quad \ell \geq 1$$

are determined by the mmse/MMSE. While these expressions are recursive, the knowledge of $K^2(X), \ldots, K^\ell(X)$ only determines $K^{\ell+1}(X)$ up to the sign. Therefore, only the case of non-negative cumulants may be handled.

Corollary 8: Let X be a centered random variable determined by its moments such that all its cumulants $K^\ell(X)$, $\ell \geq 0$, are non-negative. Then the MMSE characterizes the distribution of X.

Since $K^\ell(-X) = (-1)^\ell K_\ell(X)$, if X is a centered random variable determined by its moments such as both $K^\ell(X)$ and $K^\ell(-X)$, $\ell \geq 0$, are non-negative, then necessarily $K^\ell(X) = 0$ for odd ℓ's and the law of X is symmetric.

It may be pointed out that, within this corollary, the knowledge of all derivatives of MMSE/mmse at zero is enough to determine the law of X. Since it is not known whether mmse(s) is real-analytic at zero, which is proved in [5] only under extra assumptions, this might be weaker than knowing the function mmse.

A simple example of a random variable for which the hypotheses of the corollary are satisfied is the case of a symmetric infinitively divisible random variable X (determined by its moments). That is, the law of X has Fourier transform of the form e^ψ where

$$\psi(t) = -\frac{\sigma^2 t^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{itx} - 1) d\mu(x)$$

where $\sigma^2 \geq 0$ and μ is a symmetric measure on $\mathbb{R}$ such that $\mu(\{0\}) = 0$ and $\int_{\mathbb{R}} x^{2k} d\mu(x) < \infty$ for every $k \geq 1$. Then $K^2(X) = \sigma^2 + \int_{\mathbb{R}} x^2 d\mu(x)$, $K^{2k+1}(X) = 0$, $k \geq 1$, and

$$K^{2k}(X) = \int_{\mathbb{R}} x^{2k} d\mu(x), \quad k \geq 2.$$

One may also consider the case of a Lévy measure μ supported on $\mathbb{R}_+$. Concrete examples may be achieved as compound Poisson distributions. Let Y be a symmetric or positive random variables and P an independent Poisson variable with parameter $\theta > 0$. Considering $Y_0, Y_1, \ldots$, independent copies of Y, set

$$X = \sum_{k=0}^{P} Y_k.$$

Then $K^\ell(X) = \theta\mathbb{E}(Y^\ell) \geq 0$, $\ell \geq 1$.

It might be worthwhile mentioning in addition that the knowledge of the cumulants up to the sign is not enough to characterize the distribution in general. For example, the hyperbolic distribution $\frac{dx}{2\pi\cosh(\pi x/2)}$ has Laplace transform $\frac{1}{\cos(\lambda)}$ while the Laplace transform of the symmetric Bernoulli measure on $\{-1, +1\}$ is $\cosh(\lambda)$. The odd cumulants of these two distributions are zero, while if $K^{2\ell}$, $\ell \geq 1$, denote the even cumulants of the Bernoulli law, those of the hyperbolic distribution are $(-1)^{\ell+1}K^{2\ell}$.

The multivariate version of the MMSE conjecture is also worth consideration, and might involve multidimensional versions of the Fisher information and cumulants. Some formulas in this regard are displayed in the next section.

VII. Some Monotonicity Properties of the MMSE

In this last section, we collect a few simple monotonicity properties of the Fisher information and the MMSE, some of them already emphasized in [6].

We first mention from the identity

$$4t^2\mathrm{I}(t) = 2nt - \mathrm{MMSE}(t), \quad t > 0,$$

of (2.3) that, for every $t > 0$,

$$\mathrm{I}(t) \leq \frac{n}{2t} \tag{7.1}$$

which connects in this Euclidean setting to the classical Li-Yau parabolic inequality on heat kernels in Riemannian manifolds.[1]

The derivative formulas of the preceding sections may be extended to higher dimension, at the expense however of somewhat cumbersome computations and notation.

[1]The Li-Yau inequality states that if f_t is a positive solution of the heat equation on a n-dimensional Riemannian manifold with non-negative Ricci curvature, then $\Delta \log f_t \geq -\frac{n}{2t}$ which integrated yields (7.1) (see [3]).

For simplicity, let us stick here at the level of the Fisher information and its first derivative.

Let thus X be a random vector in $\mathbb{R}^n$ and $X_t = X + \sqrt{2t}N$, $t \geq 0$, as above. With

$$Z_t = \mathrm{Cov}(X|X_t) = \mathbb{E}\Big(\big[X - \mathbb{E}(X|X_t)\big] \otimes \big[X - \mathbb{E}(X|X_t)\big]|X_t\Big) = \mathbb{E}\big(X \otimes X|X_t\big) - \mathbb{E}\big(X|X_t\big) \otimes \mathbb{E}\big(X|X_t\big),$$

it holds from Sections II and III that

$$4t^2 \mathrm{I}(t) = 2nt - \mathrm{MMSE}(t) = -\mathbb{E}\big(\mathrm{Tr}(Z_t - 2t\mathrm{Id})\big) \quad (7.2)$$

and

$$8t^4 \frac{d}{dt}\mathrm{I}(t) = -\mathbb{E}\big(|Z_t - 2t\mathrm{Id}|^2\big) \quad (7.3)$$

(where $|\cdot|$ is the Hilbert-Schmidt norm on matrices). Indeed, from (3.2),

$$\frac{d}{dt}\mathrm{I}(t) = -2\int_{\mathbb{R}^n} f_t |\nabla^2 v|^2 dx = -2\int_{\mathbb{R}^n} f_t \left| \frac{\nabla^2 f_t}{f_t} - \frac{\nabla f_t \otimes \nabla f_t}{f_t^2}\right|^2 dx.$$

Now, in the same form as

$$2t\frac{\nabla f_t}{f_t} = \frac{g_t}{f_t} - x$$

where $g_t(x) = \mathbb{E}(Xp_t(x - X))$ and $\mathbb{E}(X|X_t) = \frac{g_t}{f_t}(X_t)$, it holds true that

$$4t^2\frac{\nabla^2 f_t}{f_t} = x \otimes x - 2t\mathrm{Id} - 2x \otimes \frac{g_t}{f_t} + \frac{h_t}{f_t}$$

where $h_t(x) = \mathbb{E}(X \otimes X p_t(x - X))$ and $\mathbb{E}(X \otimes X|X_t) = \frac{h_t}{f_t}(X_t)$. Hence

$$4t^2\left(\frac{\nabla^2 f_t}{f_t} - \frac{\nabla f_t \otimes \nabla f_t}{f_t^2}\right) = \frac{h_t}{f_t} - \frac{g_t}{f_t} \otimes \frac{g_t}{f_t} - 2t\mathrm{Id}$$

from which the claim (7.3) follows by definition of Z_t.

Combining (7.2) and (7.3) yields some interesting consequences. First,

$$2t^2\frac{d}{dt}\mathrm{MMSE}(t) = 4nt^2 + 4t\mathbb{E}\big(\mathrm{Tr}(Z_t - 2t\mathrm{Id})\big) + \mathbb{E}\big(|Z_t - 2t\mathrm{Id}|^2\big) = \mathbb{E}\big(|Z_t|^2\big) \geq 0,$$

so that $\mathrm{MMSE}(t)$, $t \geq 0$, is increasing.

This property can be made somewhat more precise. Indeed,

$$\mathbb{E}\big(\mathrm{Tr}(Z_t - 2t\mathrm{Id})\big)^2 \leq n\mathbb{E}\big(|Z_t - 2t\mathrm{Id}|^2\big)$$

from which it follows that

$$2t^2\frac{d}{dt}\mathrm{MMSE}(t) \geq \frac{1}{n}\mathrm{MMSE}(t)^2.$$

Hence

$$\frac{d}{dt}\Big(\frac{1}{\mathrm{MMSE}(t)}\Big) \leq -\frac{1}{2nt^2}.$$

Therefore, for all $0 < t \leq t'$,

$$\frac{1}{\mathrm{MMSE}(t')} - \frac{1}{\mathrm{MMSE}(t)} \leq \frac{1}{2n}\Big(\frac{1}{t'} - \frac{1}{t}\Big).$$

In particular, as $t' \to \infty$, $\mathrm{MMSE}(t') \to M^2(X) = \mathbb{E}(|X - \mathbb{E}(X)|^2)$, so that for every $t \geq 0$,

$$\mathrm{MMSE}(t) \leq \frac{2ntM^2(X)}{2nt + M^2(X)}. \quad (7.4)$$

Note that the right-hand side of (7.4) is precisely the MMSE of a random vector distributed according to the standard normal law γ, which thus achieves the maximum of the MMSE over centered random variables with finite second moment.

Another property was emphasized in [6] as the single crossing property. Namely, if $\mathrm{MMSE}(t_0)^2 \geq -2nt_0^2\alpha(t_0)$ at $t_0 > 0$ for some (smooth) function $\alpha(t)$, $t > 0$, then

$$\frac{d}{dt}\mathrm{MMSE}(t)\Big|_{t=t_0} \geq -\alpha(t_0).$$

Indeed, setting $F(t) = \beta(t) + \mathrm{MMSE}(t)$, $t > 0$, where β is an anti-derivative of α, by (7.2),

$$\begin{aligned} F'(t_0) &= \alpha(t_0) + 2n + \frac{2}{t_0}\mathbb{E}\big(\mathrm{Tr}(Z_{t_0} - 2t_0\mathrm{Id})\big) \\ &\quad + \frac{1}{2t_0^2}\mathbb{E}\big(|Z_{t_0} - 2t_0\mathrm{Id}|^2\big) \\ &\geq -\frac{1}{2nt_0^2}\mathbb{E}\big(\mathrm{Tr}(Z_{t_0} - 2t_0\mathrm{Id})\big)^2 + \frac{1}{2t_0^2}\mathbb{E}\big(|Z_{t_0} - 2t_0\mathrm{Id}|^2\big) \\ &\geq 0. \end{aligned}$$

As discussed in [6], a sensible choice for the function α is the MMSE of a standard normal given by the right-hand side of (7.4).

Appendix
Proofs of Propositions 2 and 4

Proof of Proposition 2: Although the proposition is presented for functions on the real line, we use the general (multi-dimensional) notation. The proof of (3.5) of Proposition 2 relies on the following statement. The more precise description of the multilinear forms V_ℓ therein is not strictly necessary for the purpose. Property (3.6) also follows from this statement.

Proposition 9: For every $\ell \geq 3$, $\widetilde{\Gamma}_\ell(u) = \Gamma_\ell(u) + V_\ell(u)$ where

$$V_\ell(u) = \sum_{j=3}^{\ell} \sum_{\substack{\alpha_1,\ldots,\alpha_j \in \{2,\ldots,\ell-1\} \\ \alpha_1+\cdots+\alpha_j = 2\ell}} c_{\alpha_1,\ldots,\alpha_j} u^{(\alpha_1)} \cdots u^{(\alpha_j)} \quad (8.1)$$

for real coefficients $c_{\alpha_1,\ldots,\alpha_j}$, $\alpha_1, \ldots, \alpha_j \in \{2, \ldots, \ell - 1\}$, $j = 3, \ldots, \ell$. In addition, $c_{\alpha_1,\ldots,\alpha_\ell} = (-1)^\ell(\ell - 1)!$.

It might be of interest to understand the combinatorial structure of the coefficients c.

Proof: The proof is performed by induction. Assuming that $\widetilde{\Gamma}_\ell(u) = \Gamma_\ell(u) + V_\ell(u)$, we examine $[\Delta, \widetilde{\Gamma}_\ell](u) = [\Delta, \Gamma_\ell](u) + [\Delta, V_\ell](u)$ and $[\Gamma, \widetilde{\Gamma}_\ell](u) = [\Gamma, \Gamma_\ell](u) + [\Gamma, V_\ell](u)$ (as $\widetilde{\Gamma}_{\ell+1} = [\Delta + \Gamma, \widetilde{\Gamma}_\ell]$).

First $[\Delta, \Gamma_\ell](u) = \Gamma_{\ell+1}(u)$, providing the term $\Gamma_{\ell+1}(u)$ in $\widetilde{\Gamma}_{\ell+1}(u)$. Next

$$[\Gamma, \Gamma_\ell](u) = \Gamma\big(\Gamma_\ell(u), u\big) - \Gamma_\ell\big(\Gamma(u), u\big) = u'\big(u^{(\ell)2}\big)' - \big(u'^2\big)^{(\ell)} u^{(\ell)}$$

which is of the form $V_{\ell+1}(u)$.

We then study $[\Delta, V_\ell](u)$ and $[\Gamma, V_\ell](u)$ for some V_ℓ of the form (8.1). Let

$$B(u) = B(u, \ldots, u) = u^{(\alpha_1)} \cdots u^{(\alpha_j)}$$

be j-multilinear symmetric in smooth functions u where $\alpha_1, \ldots, \alpha_j$ are integers greater than 2 (satisfying $\alpha_1 + \cdots + \alpha_j = 2\ell$). By definition,

$$\begin{aligned}
2[\Delta, B](u) &= \Delta\big(B(u)\big) - jB(\Delta u, u, \ldots, u) \\
&= \big(u^{(\alpha_1)} \cdots u^{(\alpha_j)}\big)'' - \sum_{i=1}^{j} u^{(\alpha_1)} \cdots u^{(\alpha_i+2)} \cdots u^{(\alpha_j)} \\
&= \sum_{i,i'=1, i\neq i'}^{j} u^{(\alpha_1)} \cdots u^{(\alpha_i+1)} \cdots u^{(\alpha_{i'}+1)} \cdots u^{(\alpha_j)}.
\end{aligned}$$

Hence $[\Delta, V_\ell](u)$ has the form $V_{\ell+1}(u)$. In the same way,

$$\begin{aligned}
2[\Gamma, B](u) &= 2\Gamma\big(u, B(u)\big) - jB\big(\Gamma(u), u, u, \ldots, u\big) \\
&= 2u' \sum_{i=1}^{j} u^{(\alpha_1)} \cdots u^{(\alpha_i+1)} \cdots u^{(\alpha_j)} \\
&\quad - \sum_{i=1}^{j} u^{(\alpha_1)} \cdots (u'^2)^{(\alpha_i)} \cdots u^{(\alpha_j)}.
\end{aligned}$$

Now $(u'^2)^{(\alpha_i)} = 2u'u^{(\alpha_i+1)} + P(u)$ where $P(u)$ only involves derivatives of u up to the order α_i, with total degree $\alpha_i + 2$. Therefore $[\Gamma, V_\ell](u)$ is also of the form $V_{\ell+1}(u)$. Moreover, if $j = \ell$ so that necessarily $\alpha_1 = \cdots = \alpha_\ell = 2$, then $2[\Gamma, B](u) = -2\ell {u^{(2)}}^{\ell+1}$, the only term of this form in $V_{\ell+1}(u)$, justifying the last assertion of the statement. The proposition is established. □

Proof of Proposition 4: We use induction on ℓ. By construction of the polynomials

$$Q_\ell(Z) = \prod_{i=0}^{\ell-1}(Z - i) = \sum_{i=1}^{\ell} a_i^\ell Z^i, \quad \ell \geq 1,$$

it holds that $Q_{\ell+1}(Z) = \sum_{i=1}^{\ell} a_i^\ell Z^{i+1} - \ell Q_\ell(Z)$. Hence,

$$Q_{\ell+1}(\widetilde{\Upsilon}) = [\mathrm{L} + \Upsilon, Q_\ell(\widetilde{\Upsilon})] - \ell Q_\ell(\widetilde{\Upsilon}).$$

Then, by the induction hypothesis,

$$\begin{aligned}
Q_{\ell+1}(\widetilde{\Upsilon}) &= [\mathrm{L} + \Gamma, \widetilde{\Gamma}_\ell] - \ell\widetilde{\Gamma}_\ell \\
&= [\mathrm{L}, \widetilde{\Gamma}_\ell] + [\Gamma, \widetilde{\Gamma}_\ell] - \ell\widetilde{\Gamma}_\ell.
\end{aligned}$$

Since $\widetilde{\Gamma}_{\ell+1} = [\Delta + \Gamma, \widetilde{\Gamma}_\ell]$, it suffices to show that

$$[\mathrm{L}, \widetilde{\Gamma}_\ell] = [\Delta, \widetilde{\Gamma}_\ell] + \ell\widetilde{\Gamma}_\ell. \tag{8.2}$$

We establish (8.2) also by induction. On the one hand,

$$[\mathrm{L}, \widetilde{\Gamma}_{\ell+1}] = [\mathrm{L}, [\Delta, \widetilde{\Gamma}_\ell]] + [\mathrm{L}, [\Gamma, \widetilde{\Gamma}_\ell]].$$

On the other hand, by the induction hypothesis,

$$\begin{aligned}
[\Delta, \widetilde{\Gamma}_{\ell+1}] &= [\Delta, [\Delta, \widetilde{\Gamma}_\ell]] + [\Delta, [\Gamma, \widetilde{\Gamma}_\ell]] \\
&= [\Delta, [\mathrm{L}, \widetilde{\Gamma}_\ell]] - \ell[\Delta, \widetilde{\Gamma}_\ell] + [\Delta, [\Gamma, \widetilde{\Gamma}_\ell]].
\end{aligned}$$

The identity (8.2) at the order $\ell + 1$ reads

$$[\mathrm{L}, \widetilde{\Gamma}_{\ell+1}] = [\Delta, \widetilde{\Gamma}_{\ell+1}] + (\ell+1)[\Delta + \Gamma, \widetilde{\Gamma}_\ell]$$

so that in order it holds true, it amounts to show that

$$\begin{aligned}
&[\mathrm{L}, [\Delta, \widetilde{\Gamma}_\ell]] + [\mathrm{L}, [\Gamma, \widetilde{\Gamma}_\ell]] \\
&\quad = [\Delta, [\mathrm{L}, \widetilde{\Gamma}_\ell]] + [\Delta, \widetilde{\Gamma}_\ell] + [\Delta, [\Gamma, \widetilde{\Gamma}_\ell]] + (\ell+1)[\Gamma, \widetilde{\Gamma}_\ell].
\end{aligned}$$

Since $[\Delta, \mathrm{L}] = -\Delta$, by the Lie bracket property (cf. [7]),

$$[\Delta, [\mathrm{L}, \widetilde{\Gamma}_\ell]] + [\Delta, \widetilde{\Gamma}_\ell] - [\mathrm{L}, [\Delta, \widetilde{\Gamma}_\ell]] = 0$$

so that it is left to show that

$$[\mathrm{L}, [\Gamma, \widetilde{\Gamma}_\ell]] = [\Delta, [\Gamma, \widetilde{\Gamma}_\ell]] + (\ell+1)[\Gamma, \widetilde{\Gamma}_\ell]. \tag{8.3}$$

Again by the Lie bracket property, and the fact that $[\mathrm{L}, \Gamma] = \Upsilon_2 = \Gamma_2 + \Gamma = [\Delta, \Gamma] + \Gamma$,

$$\begin{aligned}
[\mathrm{L}, [\Gamma, \widetilde{\Gamma}_\ell]] &= -[\widetilde{\Gamma}_\ell, [\mathrm{L}, \Gamma]] + [\Gamma, [\mathrm{L}, \widetilde{\Gamma}_\ell]] \\
&= -[\widetilde{\Gamma}_\ell, [\Delta, \Gamma]] + [\Gamma, \widetilde{\Gamma}_\ell] + [\Gamma, [\mathrm{L}, \widetilde{\Gamma}_\ell]]
\end{aligned}$$

while

$$[\Delta, [\Gamma, \widetilde{\Gamma}_\ell]] = -[\widetilde{\Gamma}_\ell, [\Delta, \Gamma]] + [\Gamma, [\Delta, \widetilde{\Gamma}_\ell]].$$

Since by the induction hypothesis $[\mathrm{L}, \widetilde{\Gamma}_\ell] = [\Delta, \widetilde{\Gamma}_\ell] + \ell\widetilde{\Gamma}_\ell$, it immediately follows that (8.3) holds true, thereby completing the proof of Proposition 4.

Acknowledgment

The author is grateful to O. Johnson, and I. Nourdin and G. Peccati, for pointing out the reference [6], and to D. Bakry, M. Capitaine and G. Letac for useful discussions. He thanks the referees for a careful reading of the earlier manuscript and suggestions of improvements in the exposition.

References

[1] S. Artstein, K. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *J. Amer. Math. Soc.*, vol. 17, no. 4, pp. 975–982, 2004.

[2] D. Bakry and M. Émery, "Diffusions hypercontractives," in *Séminaire de Probabilités* (Lecture Notes in Mathematics), vol. 1123. Berlin, Germany: Springer, 1985, pp. 177–206.

[3] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators* (Grundlehren der Mathematischen Wissenschaften), vol. 348. Berlin, Germany: Springer, 2014.

[4] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2317–2329, Jul. 2007.

[5] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[6] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[7] M. Ledoux, "L'algèbre de Lie des gradients itérés d'un générateur Markovien—Développements de moyennes et entropies," *Ann. Sci. l'École Normale Supérieure*, vol. 28, no. 4, pp. 435–460, 1995.

[8] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[9] D. Shlyakhtenko, "A free analogue of Shannon's problem on monotonicity of entropy," *Adv. Math.*, vol. 208, no. 2, pp. 824–833, 2007.

[10] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4295–4297, Sep. 2006.

Michel Ledoux, biography not available at the time of publication.