Entropy **2012**, *14*, 1469-1477; doi:10.3390/e14081469

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

An Integral Representation of the Relative Entropy

Miku Hirata [1], Aya Nemoto [1] and Hiroaki Yoshida [2,*]

[1] Department of Mathematics, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan; E-Mails: hirata.miku@ocha.ac.jp (M.H.); g1240604@edu.cc.ocha.ac.jp (A.N.)

[2] Department of Information Sciences, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan

* Author to whom correspondence should be addressed; E-Mail: yoshida.hiroaki@ocha.ac.jp.

Received: 15 June 2012; in revised form: 28 July 2012 / Accepted: 2 August 2012 / Published: 8 August 2012

Abstract: Recently the identity of de Bruijn type between the relative entropy and the relative Fisher information with the reference moving has been unveiled by Verdú via MMSE in estimation theory. In this paper, we shall give another proof of this identity in more direct way that the derivative is calculated by applying integrations by part with the heat equation. We shall also derive an integral representation of the relative entropy, as one of the applications of which the logarithmic Sobolev inequality for centered Gaussian measures will be given.

Keywords: relative entropy; relative Fisher information; de Bruijn identity; logarithmic Sobolev inequality; Stam inequality

1. Introduction

Probability measures on $\mathbb{R}^n$ treated in this paper are absolutely continuous with respect to the standard Lebesgue measure and we shall identify them with their densities.

For a probability measure f, the entropy $H(f)$ and the Fisher information $J(f)$ can be introduced, which play important roles in information theory, probability, and statistics. For more details on these subjects see the famous book [1].

Hereafter, for an n-variables function $\phi(\boldsymbol{x}) = \phi(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$, the integral of ϕ over the whole $\mathbb{R}^n$ by the standard Lebesgue measure $d\boldsymbol{x} = dx_1\, dx_2 \cdots dx_n$ is abbreviated as

$$\int_{\mathbb{R}^n} \phi\, d\boldsymbol{x} = \iint \cdots \int_{\mathbb{R}^n} \phi(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

that is, we shall leave out $(x_1, x_2, \ldots, x_n)$ in the integrand in order to simplify the expressions.

Definition 1.1. Let f be a probability measure on $\mathbb{R}^n$. Then *the* (*differential*) *entropy of* f is defined by

$$H(f) = -\int_{\mathbb{R}^n} f\, \log f\, d\boldsymbol{x}.$$

For a random variable $\boldsymbol{X}$ on $\mathbb{R}^n$ with the density f, we write the entropy of $\boldsymbol{X}$ by $H(\boldsymbol{X}) = H(f)$.

The Fisher information for a differentiable density f is defined by

$$J(f) = \int_{\mathbb{R}^n} \frac{\left\|\nabla f\right\|^2}{f}\, d\boldsymbol{x} = \int_{\mathbb{R}^n} f \left\|\nabla(\log f)\right\|^2 d\boldsymbol{x}.$$

When the random variable $\boldsymbol{X}$ on $\mathbb{R}^n$ has the differentiable density f, we also write as $J(\boldsymbol{X}) = J(f)$.

The important result for a behavior of the Fisher information on convolution (sum of independent random variables) is the Stam inequality, which was first stated by Stam in [2] and subsequently proved by Blachman [3],

$$\frac{1}{J(f * g)} \geq \frac{1}{J(f)} + \frac{1}{J(g)} \tag{1}$$

where we have the equality if and only if f and g are Gaussian.

The importance of the Stam inequality can be found in its applications, for instance, the entropy power inequality [2]; the logarithmic Sobolev inequality [4]; Cercignani conjecture [5]; the Shannon conjecture on entropy and the central limit theorem [6,7].

For $t \geq 0$, we denote by $P_t f$ the convolution of f with the n-dimensional Gaussian density with mean vector $\mathbf{0}$ and covariance matrix $t\, I_n$, where I_n is the identity matrix. Namely, $\left(P_t\right)_{t \geq 0}$ is the heat semigroup acting on f and satisfies the partial differential equation

$$\frac{\partial}{\partial t} P_t f = \frac{1}{2}\, \Delta\big(P_t f\big) \tag{2}$$

which is called *the heat equation*. In this paper, we simply denote $P_t f$ by f_t and call it *the Gaussian perturbation* of f. Namely, letting $\boldsymbol{X}$ be the random variable on $\mathbb{R}^n$ with the density f and $\boldsymbol{Z}$ be an n-dimensional Gaussian random variable independent of $\boldsymbol{X}$ with mean vector $\mathbf{0}$ and covariance matrix I_n, the Gaussian perturbation f_t stands the density function $f(\boldsymbol{x}, t)$ of the independent sum $\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}$.

The remarkable relation between the entropy and the Fisher information can be established by a Gaussian perturbation (see, for instance, [1], [2] or [8]);

$$\frac{d}{dt} H(f_t) = \frac{1}{2} J(f_t) \quad \text{for} \ \ t > 0 \tag{3}$$

which is known as *the de Bruijn identity*.

Let f and g be probability measures on $\mathbb{R}^n$ such that $f \ll g$ (f is absolutely continuous with respect to g). Setting the probability measure g as a reference, the relative entropy and the relative Fisher information can be introduced as follows:

Definition 1.2. *The relative entropy of f with respect to g, $D(f \parallel g)$* is defined by

$$D(f \parallel g) = \int_{\mathbb{R}^n} f \left(\log \frac{f}{g} \right) d\boldsymbol{x} = \int_{\mathbb{R}^n} f \log f \, d\boldsymbol{x} - \int_{\mathbb{R}^n} f \log g \, d\boldsymbol{x},$$

which takes always a non-negative value.

We also define *the relative Fisher information of f with respect to g* by

$$J(f \parallel g) = \int_{\mathbb{R}^n} f \left\| \nabla \Big(\log \frac{f}{g} \Big) \right\|^2 d\boldsymbol{x} = \int_{\mathbb{R}^n} f \left\| \nabla(\log f) - \nabla(\log g) \right\|^2 d\boldsymbol{x},$$

which is also non-negative. When random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the densities f and g, respectively, the relative entropy and the relative Fisher information of $\boldsymbol{X}$ with respect to $\boldsymbol{Y}$ are defined by $D(\boldsymbol{X} \parallel \boldsymbol{Y}) = D(f \parallel g)$ and $J(\boldsymbol{X} \parallel \boldsymbol{Y}) = J(f \parallel g)$, respectively.

In view of the de Bruijn identity, one might expect that there is a similar connection between the relative entropy and the relative Fisher information. Indeed, the gradient formulas for the relative entropy functionals were obtained in [9–11], where the reference measures would not be changed in their cases.

Recently Verdú in [12], however, investigated the derivative in t of $D(f_t \parallel g_t)$ for two Gaussian perturbations f_t and g_t. Here we should note that the reference measure does move by the same time parameter in this case. The following identity of de Bruijn type

$$\frac{d}{dt} D(f_t \parallel g_t) = -\frac{1}{2} J(f_t \parallel g_t)$$

has been derived via MMSE in estimation theory (see also [13], for general perturbations).

The main aim in this paper is that we shall give an alternative proof of this identity by direct calculation with integrations by part, the method of which is similar to ones in [11,14]. Moreover, it will be easily found that the above identity yields an integral representation of the relative entropy. We shall also see the simple proof of the logarithmic Sobolev inequality for centered Gaussian in univariate ($n = 1$) case as an application of the integral representation.

2. An Integral Representation of the Relative Entropy

We shall make the Gaussian perturbations f_t and g_t, respectively, and consider the relative entropy $D(f_t \parallel g_t)$, where the absolute continuity $f_t \ll g_t$ remains true for $t > 0$.

Here, we regard $D(f_t \parallel g_t)$ as a function of t and calculate the derivative,

$$\frac{d}{dt} D(f_t \parallel g_t) = \frac{d}{dt} \int_{\mathbb{R}^n} f_t \, \log \frac{f_t}{g_t} \, d\boldsymbol{x} = \frac{d}{dt} \int_{\mathbb{R}^n} f_t \, \log f_t \, d\boldsymbol{x} - \frac{d}{dt} \int_{\mathbb{R}^n} f_t \, \log g_t \, d\boldsymbol{x} \tag{4}$$

by integrations by part with help of the heat equation.

Proposition 2.1. *Let $f \ll g$ be probability measures on $\mathbb{R}^n$ with finite Fisher informations $J(f) < \infty$ and $J(g) < \infty$, and finite relative entropy $D(f \parallel g) < \infty$. Then we obtain*

$$\frac{d}{dt} D(f_t \parallel g_t) = -\frac{1}{2} J(f_t \parallel g_t) \quad \textit{for} \;\; t > 0.$$

Proof. First we should notice that the Fisher informations $J(f_t)$ and $J(g_t)$ are finite at any $t > 0$. Because, for instance, if an n-dimensional random variable $\boldsymbol{X}$ has the density f and $\boldsymbol{Z}$ is an n-dimensional Gaussian random variable independent of $\boldsymbol{X}$ with mean vector $\mathbf{0}$ and covariance matrix I_n, then by applying the Stam inequality (1) to independent random variables $\boldsymbol{X}$ and $\sqrt{t}\boldsymbol{Z}$, we have that

$$J(f_t) = J\big(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}\big) \leq \left(\frac{1}{J(\boldsymbol{X})} + \frac{1}{J(\sqrt{t}\boldsymbol{Z})}\right)^{-1} = \frac{J(\boldsymbol{X})}{1 + \frac{t}{n} J(\boldsymbol{X})} \leq J(f) < \infty \tag{5}$$

where $J(\boldsymbol{Z}) = n$ is by simple calculation. We shall also notice that the function $D(f_t \parallel g_t)$ is non-increasing in t, that is, for $t > 0$,

$$0 \leq D(f_t \parallel g_t) \leq D(f \parallel g) < \infty,$$

which can be found in [15] (p. 101). Therefore, $D(f_t \parallel g_t)$ is finite for $t > 0$. But by a nonlinear approximation argument in [11], we can impose a stronger assumption without loss of generality that

$$\text{"the relative density } \frac{f_t}{g_t} \text{ is bounded away from } 0 \text{ and } \infty \text{ on } \mathbb{R}^n\text{"} \tag{6}$$

Concerning the first term in the most right hand side of (4), it follows immediately that

$$\frac{d}{dt} \int_{\mathbb{R}^n} f_t \log f_t \, d\boldsymbol{x} = -\frac{1}{2} \int_{\mathbb{R}^n} \frac{\|\nabla f_t\|^2}{f_t} \, d\boldsymbol{x} \tag{7}$$

by the de Bruijn identity (3), hence, we shall concentrate our attention upon the second term.

Since the densities f_t and g_t satisfy the heat equation (2), the second term can be reformulated as follows:

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^n} f_t \log g_t \, d\boldsymbol{x} &= \int_{\mathbb{R}^n} f_t \left(\partial_t \log g_t\right) d\boldsymbol{x} + \int_{\mathbb{R}^n} \log g_t \left(\partial_t f_t\right) d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} f_t \frac{\partial_t g_t}{g_t} \, d\boldsymbol{x} + \int_{\mathbb{R}^n} \log g_t \left(\frac{1}{2} \Delta f_t\right) d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} \frac{f_t}{g_t} \left(\frac{1}{2} \Delta g_t\right) d\boldsymbol{x} + \int_{\mathbb{R}^n} \log g_t \left(\frac{1}{2} \Delta f_t\right) d\boldsymbol{x} \end{aligned} \tag{8}$$

In this reformulation, we have changed integration and differentiation at the first equality, which is justified by a routine argument with the bounded convergence theorem (see, for instance, [16]).

Applying integration by part to the first term in the last expression of (8), it becomes

$$\int_{\mathbb{R}^n} \frac{f_t}{g_t} \left(\frac{1}{2} \Delta g_t\right) d\boldsymbol{x} = -\frac{1}{2} \int_{\mathbb{R}^n} \nabla\left(\frac{f_t}{g_t}\right) \cdot \nabla g_t \, d\boldsymbol{x} \tag{9}$$

which can be asserted by the observation below. As g_t has finite Fisher information $J(g_t) < \infty$, $\frac{\nabla g_t}{\sqrt{g_t}}$ has finite 2-norm in $L^2(\mathbb{R}^n)$ and must be bounded at infinity. Furthermore, from our technical assumption (6), $\sqrt{\frac{f_t}{g_t}}$ is also bounded. Hence if we factorize as

$$\frac{f_t}{g_t}(\nabla g_t) = \sqrt{f_t}\sqrt{\frac{f_t}{g_t}}\frac{\nabla g_t}{\sqrt{g_t}},$$

then it can be found that $\frac{f_t}{g_t}(\nabla g_t)$ will vanish at infinity.

Applying integration by part to the second term in the last expression of (8), it becomes

$$\int_{\mathbb{R}^n} \log g_t \left(\frac{1}{2}\Delta f_t\right) d\boldsymbol{x} = -\frac{1}{2}\int_{\mathbb{R}^n} \frac{\nabla g_t}{g_t} \cdot \nabla f_t \, d\boldsymbol{x} \tag{10}$$

Here it should be noted that $\log g_t (\nabla f_t)$ will vanish at infinity by the following observation. Similarly, we factorize it as

$$\log g_t (\nabla f_t) = 2\big(\sqrt{g_t}\log\sqrt{g_t}\big)\sqrt{\frac{f_t}{g_t}}\left(\frac{\nabla f_t}{\sqrt{f_t}}\right).$$

Then the boundedness of $\frac{\nabla f_t}{\sqrt{f_t}}$ comes from that $J(f_t) < \infty$, and one of $\sqrt{\frac{f_t}{g_t}}$ is by the assumption (6) same as before. Furthermore, the limit formula $\lim_{\xi\to 0}\xi\log\xi = 0$ ensures that $\big(\sqrt{g_t}\log\sqrt{g_t}\big)$ will vanish at infinity.

Substitute the Equations (9) and (10) into (8), it follows that

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{R}^n} f_t \log g_t \, d\boldsymbol{x} &= -\frac{1}{2}\int_{\mathbb{R}^n} \nabla\Big(\frac{f_t}{g_t}\Big)\cdot\nabla g_t \, d\boldsymbol{x} - \frac{1}{2}\int_{\mathbb{R}^n}\frac{\nabla g_t}{g_t}\cdot\nabla f_t\, d\boldsymbol{x} \\
&= -\frac{1}{2}\int_{\mathbb{R}^n}\Big(\frac{\nabla f_t}{g_t} - f_t\frac{\nabla g_t}{g_t^2}\Big)\cdot\nabla g_t\, d\boldsymbol{x} - \frac{1}{2}\int_{\mathbb{R}^n} f_t\frac{\nabla g_t}{g_t}\cdot\frac{\nabla f_t}{f_t}\, d\boldsymbol{x} \\
&= -\int_{\mathbb{R}^n} f_t\frac{\nabla g_t}{g_t}\cdot\frac{\nabla f_t}{f_t}\, d\boldsymbol{x} + \frac{1}{2}\int_{\mathbb{R}^n} f_t\frac{\nabla g_t}{g_t}\cdot\frac{\nabla g_t}{g_t}\, d\boldsymbol{x}
\end{aligned} \tag{11}$$

Combining the Equations (7) and (11), we have that

$$\begin{aligned}
&\frac{d}{dt}\int_{\mathbb{R}^n} f_t \log f_t \, d\boldsymbol{x} - \frac{d}{dt}\int_{\mathbb{R}^n} f_t \log g_t \, d\boldsymbol{x} \\
&= -\frac{1}{2}\int_{\mathbb{R}^n} f_t\frac{\nabla f_t}{f_t}\cdot\frac{\nabla f_t}{f_t}\, d\boldsymbol{x} + \int_{\mathbb{R}^n} f_t\frac{\nabla g_t}{g_t}\cdot\frac{\nabla f_t}{f_t}\, d\boldsymbol{x} - \frac{1}{2}\int_{\mathbb{R}^n} f_t\frac{\nabla g_t}{g_t}\cdot\frac{\nabla g_t}{g_t}\, d\boldsymbol{x} \\
&= -\frac{1}{2}\int_{\mathbb{R}^n} f_t\left\|\frac{\nabla f_t}{f_t} - \frac{\nabla g_t}{g_t}\right\|^2 d\boldsymbol{x}
\end{aligned}$$

which means

$$\frac{d}{dt}D(f_t \,\|\, g_t) = -\frac{1}{2}\int_{\mathbb{R}^n} f_t\left\|\nabla(\log f_t) - \nabla(\log g_t)\right\|^2 d\boldsymbol{x} = -\frac{1}{2}J(f_t \,\|\, g_t).$$

□

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be n-dimensional random variables with the densities f and g, respectively, and $\boldsymbol{Z}$ be an n-dimensional Gaussian random variable independent of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with mean vector $\boldsymbol{0}$ and covariance matrix I_n.

Since the relative entropy is scale invariant, it follows that

$$D(\boldsymbol{X} + \sqrt{t}\,\boldsymbol{Z} \parallel \boldsymbol{Y} + \sqrt{t}\,\boldsymbol{Z}) = D\big(\tfrac{1}{\sqrt{t}}\boldsymbol{X} + \boldsymbol{Z} \parallel \tfrac{1}{\sqrt{t}}\boldsymbol{Y} + \boldsymbol{Z}\big).$$

We know that both of $\frac{1}{\sqrt{t}}X + Z$ and $\frac{1}{\sqrt{t}}Y + Z$, as $t \to \infty$ converge to Z in distribution. Thus, we have

$$\lim_{t\to\infty} D(f_t \parallel g_t) = 0,$$

and the following integral representation for the relative entropy can be obtained:

Theorem 2.2. *Let $f \ll g$ be probability measures with finite Fisher informations and finite relative entropy $D(f \parallel g)$. Then we have the integral representation,*

$$D(f \parallel g) = \frac{1}{2}\int_0^\infty J(f_t \parallel g_t)\,dt.$$

3. An Application to the Logarithmic Sobolev Inequality

In this section, we shall give a proof of the logarithmic Sobolev inequality for a centered Gaussian measure in case of $n = 1$. Although several proofs of the logarithmic Sobolev inequality have already been given in many literatures (see, for instance, [10,17]), we shall give it here again as an application of the integral representation in Theorem 2.2.

Theorem 3.1. *Let g be the centered Gaussian measure of variance σ^2. Then for any probability measure f on $\mathbb{R}$ of finite moment of order 2 with finite Fisher information $J(f) < \infty$, the following inequality holds:*

$$D(f \parallel g) \le \frac{\sigma^2}{2} J(f \parallel g).$$

Proof. It is clear that the perturbed measure g_t is the centered Gaussian of variance $\sigma^2 + t$ and the score of which is given by

$$\big(\partial_x \log g_t\big) = -\frac{x}{\sigma^2 + t}.$$

Then using the Stein relation (see, for instance, [15]), the relative Fisher information $J(f_t \parallel g_t)$ can be expanded as follows:

$$\begin{aligned} J(f_t \parallel g_t) &= \int_{\mathbb{R}} \big\{\big(\partial_x \log f_t\big) - \big(\partial_x \log g_t\big)\big\}^2 f_t\,dx \\ &= J(f_t) + 2\int_{\mathbb{R}} \Big\{\partial_x\Big(-\frac{x}{\sigma^2+t}\Big)\Big\} f_t\,dx + \int_{\mathbb{R}} \Big(-\frac{x}{\sigma^2+t}\Big)^2 f_t\,dx \\ &= J(f_t) - \frac{2}{\sigma^2+t}\int_{\mathbb{R}} f_t\,dx + \frac{1}{(\sigma^2+t)^2}\int_{\mathbb{R}} x^2 f_t\,dx \qquad (12) \end{aligned}$$

As it was seen in (5), by Stam inequality, we have that

$$J(f_t) \le \Big(\frac{1}{J(f)} + t\Big)^{-1} = \frac{1}{(1/\alpha) + t} \qquad (13)$$

where we put $\alpha = J(f) < \infty$.

Since f has finite moment of order 2, if we put the second moment of f as $\beta = m_2(f) < \infty$, then it is easy to see that the second moment of f_t is given by

$$m_2(f_t) = \int x^2 f_t \, dx = \beta + t \tag{14}$$

Substitute (13) and (14) into (12) and we obtain that

$$J(f_t \parallel g_t) \leq \frac{1}{(1/\alpha) + t} - \frac{2}{\sigma^2 + t} + \frac{\beta + t}{(\sigma^2 + t)^2} = \frac{1}{(1/\alpha) + t} - \frac{1}{\sigma^2 + t} + \frac{\beta - \sigma^2}{(\sigma^2 + t)^2}.$$

Integrating for $t \geq 0$, we have

$$\begin{aligned} \frac{1}{2}\int_0^\infty J(f_t \parallel g_t)\, dt &\leq \frac{1}{2}\int_0^\infty \left(\frac{1}{(1/\alpha) + t} - \frac{1}{\sigma^2 + t} + \frac{\beta - \sigma^2}{(\sigma^2 + t)^2} \right) dt \\ &= \frac{1}{2}\left[\log\left(\frac{(1/\alpha) + t}{\sigma^2 + t} \right) - \frac{\beta - \sigma^2}{\sigma^2 + t} \right]_0^\infty \\ &= \frac{1}{2}\left(\log(\sigma^2 \alpha) + \frac{\beta}{\sigma^2} - 1 \right). \end{aligned}$$

Since $\log y$ is dominated as $\log y \leq y - 1$ for $y > 0$, it follows that

$$\frac{1}{2}\int_0^\infty J(f_t \parallel g_t)\, dt \leq \frac{1}{2}\left(\sigma^2 \alpha - 2 + \frac{\beta}{\sigma^2} \right) \tag{15}$$

On the other hand, the relative Fisher information $J(f \parallel g)$ can be given as

$$\begin{aligned} J(f \parallel g) &= \int_{\mathbb{R}} \left(\partial_x \log f - \left(-\frac{x}{\sigma^2} \right) \right)^2 f\, dx \\ &= \int_{\mathbb{R}} \left(\partial_x \log f \right)^2 f\, dx - \frac{2}{\sigma^2}\int_{\mathbb{R}} f\, dx + \frac{1}{(\sigma^2)^2}\int_{\mathbb{R}} x^2 f\, dx \\ &= J(f) - \frac{2}{\sigma^2} + \frac{m_2(f)}{(\sigma^2)^2} = \alpha - \frac{2}{\sigma^2} + \frac{\beta}{(\sigma^2)^2} \end{aligned} \tag{16}$$

Combining (15) and (16), we have

$$\frac{1}{2}\int_0^\infty J(f_t \parallel g_t)\, dt \leq \frac{\sigma^2}{2} J(f \parallel g),$$

which means our desired inequality by Theorem 2.2.

□

Remark 3.2. Similar way to the proof of Theorem 3.1 can be found in the paper by Stam [2], where it is not for relative case. Namely, based on convolution inequalities and the de Bruijn identity, the isoperimetric inequality on entropy for a standardized random variable X on $\mathbb{R}$,

$$(2\pi e)e^{-2H(X)} \leq J(X) \tag{17}$$

was shown. This inequality is essentially the same as the logarithmic Sobolev inequality for the standard Gaussian measure, where the left hand side in (17) is the reciprocal of the entropy power.

Acknowledgments

The authors are grateful to the anonymous reviewers for their correcting inaccuracies, useful suggestions, and valuable comments. Especially, the extension to the $\mathbb{R}^n$-version of Proposition 2.1 is based on the reviewer's comments.

References

1. Cover, T.; Thomas, J. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
2. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.* **1959**, *2*, 101–112.
3. Blachman, N.M. The convolution inequality for entropy powers. *IEEE Trans. Inform. Theor.* **1965**, *2*, 267–271.
4. Carlen, E. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.* **1991**, *101*, 194–211.
5. Villani, C. Cercignani's conjecture is sometimes true and always almost true. *Commun. Math. Phys.* **2003**, *234*, 455–490.
6. Madiman, M.; Barron, A. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theor.* **2007**, *53*, 2317–2329.
7. Johnson, O.; Barron, A. Fisher information inequalities and the central limit theorem. *Probab. Theor. Relat. Field.* **2004**, *129*, 391–409.
8. Dembo, A.; Thomas, J.; Cover, T. Information theoretic inequalities. *IEEE Trans. Inform. Theor.* **1991**, *37*, 1501–1518.
9. Arnold, A.; Markowich, P.; Toscani, G.; Unterreiter A. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations. *Comm Part. Differ. Equat.* **2001**, *26*, 43–100.
10. Bakry, D.; Émery, M. Diffusions hypercontractives. In *Séminar de Probabilités XIX, 1983/84, Lecture Notes in Math. 1123*; Springer-Verlag: Berlin, Germany, 1985; pp. 177–206.
11. Otto, F.; Villani, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **2000**, *173*, 361–400.
12. Verdú, S. Mismatched estimation and relative entropy. *IEEE Trans. Inform. Theor.* **2010**, *56*, 3712–3719.
13. Guo, D.; Shamai, S.; Verdú, S. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theor.* **2005**, *51*, 1261–1283.
14. Villani, C. A short proof of the concavity of entropy power. *IEEE Trans. Inform. Theor.* **2000**, *46*, 1695–1696.
15. Johnson, O. *Information Theory and the Central Limit Theorem*; Imperial College Press: London, UK, 2004.
16. Barron, A. Entropy and the central limit theorem. *Ann. Probab.* **1986**, *14*, 336–342.

17. Gross, L. Logarithmic Sobolev inequalities. *Amer. J. Math.* **1975**, *97*, 1061–1083.