

Article

A Dissipation of Relative Entropy by Diffusion Flows

Hiroaki Yoshida

Department of Information Sciences, Ochanomizu University, 2-1-1, Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan; yoshida.hiroaki@ocha.ac.jp; Tel.: +81-3-5978-5386

Academic Editor: Angelo Plastino

Received: 5 October 2016; Accepted: 22 December 2016; Published: 27 December 2016

Abstract: Given a probability measure, we consider the diffusion flows of probability measures associated with the partial differential equation (PDE) of Fokker–Planck. Our flows of the probability measures are defined as the solution of the Fokker–Planck equation for the same strictly convex potential, which means that the flows have the same equilibrium. Then, we shall investigate the time derivative for the relative entropy in the case where the object and the reference measures are moving according to the above diffusion flows, from which we can obtain a certain dissipation formula and also an integral representation of the relative entropy.

Keywords: relative entropy; relative Fisher information; Fokker–Planck equation; entropy dissipation; entropy gap

1. Introduction

We shall begin with the definitions and fix some notations. Several results in the literature that we will use later are also gathered in this section. Probability measures on \mathbb{R}^n in this paper are always assumed to be absolutely continuous with respect to the Lebesgue measure. Thus, we say that a probability measure μ has the continuous density f , which means $d\mu(x) = f(x) dx$, and the measure μ is sometimes identified with its density f . Throughout this paper, if we simply write an integral symbol \int without specifying any domain, which means $\int_{\mathbb{R}^n}$ is the integral over the whole space on \mathbb{R}^n .

Definition 1. For a probability measure μ on \mathbb{R}^n with the density f , the entropy of μ (or f) is defined by

$$H(\mu) (= H(f)) = - \int f \log f dx, \quad (1)$$

and, if the density f is smooth, then we define the Fisher information of μ (or f) as

$$I(\mu) (= I(f)) = \int \frac{\|\nabla f\|^2}{f} dx = \int \|\nabla \log f\|^2 f dx. \quad (2)$$

For an \mathbb{R}^n -valued random variable X , if X is distributed according to the probability measure μ , then we define the entropy $H(X)$ and the Fisher information $I(X)$ of X by $H(X) = H(\mu)$ and $I(X) = I(\mu)$, respectively.

In a one-dimensional case, the gradient $\nabla \log f$ in (2) is usually called the score function of X (or μ) and denoted by ρ_X (or ρ_μ). For a differentiable function ξ with bounded derivative, the score function behaves that

$$\int_{-\infty}^{\infty} \xi(x) \rho_\mu(x) d\mu(x) = - \int_{-\infty}^{\infty} \xi'(x) d\mu(x), \quad (3)$$

which is known as Stein's identity.

Let X be an \mathbb{R}^n -valued random variable distributed according to the probability measure μ , and Z be an n -dimensional standard (with mean vector $\mathbf{0}$ and identity covariance matrix I_n) Gaussian

random variable independent of X . Then, for $\tau > 0$, the independent sum $X + \sqrt{\tau}Z$ is called *the Gaussian perturbation* of X .

We denote by μ_τ the probability measure corresponding to the Gaussian perturbation $X + \sqrt{\tau}Z$, and f_τ stands for the density of μ_τ . It is fundamental that the density function f_τ satisfies *the heat equation*:

$$\frac{\partial}{\partial \tau} f_\tau = \frac{1}{2} \Delta f_\tau, \quad (4)$$

where $\Delta = \nabla \cdot \nabla$ is the Laplacian operator.

The remarkable relationship between the entropy and the Fisher information can be established by the Gaussian perturbation as follows (see, for instance, [1] or [2]), which is known as *the de Bruijn identity*.

Lemma 1. *Let X be an \mathbb{R}^n -valued random variable distributed according to the probability measure μ . Then, for the Gaussian perturbation, it holds that*

$$\frac{d}{d\tau} H(X + \sqrt{\tau}Z) = \frac{1}{2} I(X + \sqrt{\tau}Z) \quad \text{for } \tau > 0. \quad (5)$$

Namely, using the density f_τ of the Gaussian perturbed measure μ_τ , we can write

$$\frac{d}{d\tau} H(f_\tau) = \frac{1}{2} I(f_\tau) \quad \text{for } \tau > 0. \quad (6)$$

Definition 2. *Let μ and ν be probability measures on \mathbb{R}^n with $\mu \ll \nu$ (μ is absolutely continuous with respect to ν). We denote the probability density functions of μ and ν by f and g , respectively. Then, as the ways of indicating the difference between two measures, we shall introduce the following quantities: the relative entropy $H(\mu | \nu)$ of μ with respect to ν , $H(f | g)$ of f with respect to g , is defined by*

$$H(\mu | \nu) (= H(f | g)) = \int \left(\log \frac{f}{g} \right) f \, dx. \quad (7)$$

Although it does not appear to have received widespread attention, it is natural to define the relative Fisher information $I(\mu | \nu)$ of μ with respect to ν , $I(f | g)$ of f with respect to g as (see, for instance, [3])

$$I(\mu | \nu) (= I(f | g)) = \int \left\| \nabla \left(\log \frac{f}{g} \right) \right\|^2 f \, dx = 4 \int \left\| \nabla \sqrt{\frac{f}{g}} \right\|^2 g \, dx, \quad (8)$$

where the relative density f/g is assumed to be sufficiently smooth such that the above expressions make sense.

The relative entropy $H(f | g)$ and the relative Fisher information $I(f | g)$ take non-negative values and 0 if and only if $f(x) = g(x)$ for almost all $x \in \mathbb{R}^n$. Similar to Definition 1, for random variables X and Y with the distributions μ and ν , the relative entropy and the relative Fisher information of X with respect to Y are defined as $H(X | Y) = H(\mu | \nu)$ and $I(X | Y) = I(\mu | \nu)$, 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, Vérdu in [4] investigated the derivative in τ of $H(X + \sqrt{\tau}Z | Y + \sqrt{\tau}Z)$ for two Gaussian perturbations, and derived the following identity of the de Bruijn type via minimum mean-square error (MMSE) in estimation theory.

Lemma 2. *Let X and Y be \mathbb{R}^n -valued random variables distributed according to the probability measure μ and ν , respectively. Then, for the Gaussian perturbations, it holds that*

$$\frac{d}{d\tau} H(X + \sqrt{\tau}Z | Y + \sqrt{\tau}Z) = -\frac{1}{2} I(X + \sqrt{\tau}Z | Y + \sqrt{\tau}Z) \quad \text{for } \tau > 0, \quad (9)$$

that is,

$$\frac{d}{d\tau} H(\mu_\tau | \nu_\tau) = -\frac{1}{2} I(\mu_\tau | \nu_\tau) \quad \text{for } \tau > 0, \quad (10)$$

where μ_τ and ν_τ are the corresponding measures of the Gaussian perturbations.

An alternative proof of this identity by direct calculation with integrations by part has been given in [5]. It should be noted here that the reference measure does move by the same heat equation in the formula of Lemma 1.

Other derivative formulas of the relative entropy have been investigated in [6–8], which are closely related to the theory of optimal transport and functional inequalities of informations. It is common in these fields that the reference measure is unchanged with the equilibrium measure. Here, we shall recall such a derivative formula and will list some useful related results.

Let V be a C^1 map on \mathbb{R}^n and consider the probability measure κ by

$$d\kappa(x) = \frac{1}{\mathcal{Z}} e^{-V(x)} dx, \quad (11)$$

where $\mathcal{Z} = \int e^{-V(x)} dx$, the normalization constant. Such a probability measure κ is called *the equilibrium (or Gibbs) measure for the potential function V* .

Given a probability measure μ_0 , we consider the diffusion flow of probability measures $(\mu_t)_{t \geq 0}$ associated with the gradient ∇V , that is, the density f_t of the measure μ_t ($t > 0$) is defined as the solution to the partial differential equation:

$$\frac{\partial}{\partial t} f_t = \nabla \cdot (\nabla f_t + f_t \nabla V), \quad (12)$$

which is called *the Fokker–Planck equation*. It is easily found that the long-time asymptotically stationary measure for Fokker–Planck Equation (12) is given by the above equilibrium (Gibbs) measure.

Setting the equilibrium measure as the reference, we can understand the relationship between the relative entropy and the relative Fisher information via the Fokker–Planck equation as follows (see, for instance, [8]):

Proposition 1. *Let $(\mu_t)_{t \geq 0}$ be a diffusion flow of the probability measure associated with the gradient ∇V , and let κ be the equilibrium measure for the potential function V . Then, it holds the differential formula:*

$$\frac{d}{dt} H(\mu_t | \kappa) = -I(\mu_t | \kappa). \quad (13)$$

Definition 3. *A C^2 function V on \mathbb{R}^n is called strictly K -convex if there exists a positive constant $K > 0$ such that $\text{Hess}(V) \geq K I_n$.*

In the case where the potential function V has the above convexity, we can obtain the inequality between the relative entropy and the relative Fisher information with respect to the equilibrium measure for the potential V , which is known as *the logarithmic Sobolev inequality* (see, for instance, [7,9,10]).

Theorem 1. *Let κ be the equilibrium measure for the potential function V . If the potential function V is strictly K -convex, then it follows that, for any probability measure $\mu (\ll \kappa)$,*

$$H(\mu | \kappa) \leq \frac{1}{2K} I(\mu | \kappa). \quad (14)$$

Combining Proposition 1 and Theorem 1, we can obtain the following convergence of the diffusion flow to the equilibrium:

Proposition 2. Let $(\mu_t)_{t \geq 0}$ be the diffusion flow of probability measures by the Fokker–Planck equation associated with the strictly K -convex potential V , and κ be the equilibrium for the potential V . Then, it follows that

$$\frac{d}{dt} H(\mu_t | \kappa) \leq -2K H(\mu_t | \kappa), \quad (15)$$

which implies that μ_t converges exponentially fast, as $t \rightarrow \infty$, to the equilibrium κ in the relative entropy. Namely,

$$H(\mu_t | \kappa) \leq e^{-2Kt} H(\mu_0 | \kappa) \text{ for } t > 0. \quad (16)$$

The diffusion flow by the quadratic potential $V(x) = \frac{\|x\|^2}{2}$ is called the Ornstein–Uhlenbeck flow, and the corresponding Fokker–Planck equation is reduced to

$$\frac{\partial}{\partial t} f_t = \nabla \cdot (\nabla f_t + f_t x) \text{ for } t > 0. \quad (17)$$

In this case, we can obtain the explicit solution f_t , and it follows that the equilibrium measure becomes the standard Gaussian.

Furthermore, it is known that the solution to Equation (17) can be represented in terms of random variables as follows: let X be a random variable on \mathbb{R}^n having the initial density f_0 , and Z be an n -dimensional standard Gaussian random variable independent of X . Then, the density function of the independent sum

$$X_t = \sqrt{e^{-2t}} X + \sqrt{1 - e^{-2t}} Z \text{ for } t > 0 \quad (18)$$

gives the solution f_t to partial differential Equation (17). Since the Ornstein–Uhlenbeck flow has Gaussian equilibrium, it has been widely used as the technical tool for the proof of Gross’s logarithmic Sobolev inequality [9] and Talagrand inequality [11].

Here, we shall mention one more useful result concerned with the convergence of the relative entropy, which is called the Csiszár–Kullback–Pinsker inequality (see, for instance, [12] or [13]).

Lemma 3. The convergence in the relative entropy is stronger than in L^1 -norm, that is, it holds for probability densities that

$$\frac{1}{2} \left(\int |f(x) - g(x)| dx \right)^2 \leq H(f | g). \quad (19)$$

The problem of finding the time derivative of the relative entropy between two densities under the same continuity equation has been investigated in [14,15]. In this paper, we will treat the Fokker–Planck equation with strictly convex potential as our continuity equation, which is because the first natural extension of the heat equation and the similar dissipation formula in Lemma 2 of Vérdu can be derived by the fundamental method, the integration by parts like in [5].

The time integration of our formula will give an integral representation of the relative entropy. Applying this representation to the Ornstein–Uhlenbeck flows, we can give an extension of the formula for entropy gap.

2. Dissipation of the Relative Entropy

We will calculate the time derivative of the relative entropy for the case where the objective and the reference measures are evolved by the Fokker–Planck equation with the same strictly convex potential. We shall begin with describing our situation of calculation precisely.

- Situation A:

Let μ_0 and ν_0 be Lebesgue absolutely continuous probability measures on \mathbb{R}^n with $\mu_0 \ll \nu_0$, and let μ_t and ν_t ($t \geq 0$) be the diffusion flows by the Fokker–Planck equation with the strictly K -convex potential function V starting from μ_0 and ν_0 , respectively. Here, the growth rate of the potential function V is assumed to be at most polynomial.

We assume that, for $t \geq 0$, the measures μ_t and ν_t have finite Fisher information $I(\mu_t) < \infty$ and $I(\nu_t) < \infty$, and are absolutely continuous with respect to the Lebesgue measure, the densities f_t and g_t of which are sufficiently smooth and rapidly decreasing at infinity. Furthermore, it is naturally required that $\mu_t \ll \nu_t$.

Here, we shall impose the following assumption on the relative densities, which does not cause any loss of generality but for simplicity of the proof.

- Assumption on the relative densities D:

Let $d\kappa(x) = e^{-V(x)} dx$ be the equilibrium measure of the potential function V , where the potential function V is normalized (shifted) so that $\mathcal{Z} = 1$. The Fokker–Planck equation will not be made of any effect by this normalization (shift) because it depends only on the gradient ∇V .

We may assume that the relative densities $\frac{f_t(x)}{e^{-V(x)}}$ and $\frac{g_t(x)}{e^{-V(x)}}$ are bounded away from zero and infinity for sufficiently large t . Namely, there exist uniform constants $0 < m_1 \leq M_1 < \infty$ and $0 < m_2 \leq M_2 < \infty$ for sufficiently large t such that

$$m_1 \leq \frac{f_t(x)}{e^{-V(x)}} \leq M_1 \text{ for } x \in \mathbb{R}^n \tag{20}$$

and

$$m_2 \leq \frac{g_t(x)}{e^{-V(x)}} \leq M_2 \text{ for } x \in \mathbb{R}^n. \tag{21}$$

Hence, the relative density $\frac{f_t}{g_t}$ is also bounded away from zero and infinity for sufficiently large t , that is, there exist uniform constants $0 < m_0 \leq M_0 < \infty$ such that

$$m_0 \leq \frac{f_t(x)}{g_t(x)} \leq M_0 \text{ for } x \in \mathbb{R}^n, \tag{22}$$

for sufficiently large t .

Remark 1. The above technical assumptions on the relative densities are due to the non-linear approximation argument given by Otto and Villani in [8] and the following fact: in our situation, it follows that the density f_t of the diffusion flow of probability measure by the Fokker–Planck equation converges to the equilibrium e^{-V} in L^1 -norm as $t \rightarrow \infty$ by combining Proposition 2 with Lemma 3—so does g_t .

Proposition 3. Let μ_t and ν_t ($t \geq 0$) be the flows of the probability measures on \mathbb{R}^n by the Fokker–Planck equation as in Situation A with the assumptions on the relative densities D. Then, it holds that, for $t > 0$,

$$\frac{d}{dt}H(\mu_t | \nu_t) = -I(\mu_t | \nu_t). \tag{23}$$

Proof. We expand the derivative of the relative entropy as

$$\frac{d}{dt}H(f_t | g_t) = \frac{d}{dt} \int (\log f_t) f_t dx - \frac{d}{dt} \int (\log g_t) f_t dx. \tag{24}$$

Since we know that f_t and g_t converge to the equilibrium e^{-V} in L^1 , and that the time derivatives $\partial_t f_t$ and $\partial_t g_t$ are converging to 0 as $t \rightarrow \infty$, and the densities f_t, g_t and $|\partial_t f_t|, |\partial_t g_t|$ are uniformly bounded for t . Furthermore, by our assumptions on the relative density, $\frac{f_t}{g_t}$ is bounded away from zero and infinity. Hence, we are allowed to exchange integration and t -differentiation, which is justified by a routine argument with the bounded convergence theorem (see, for instance, [2] and also [8]).

Then, the first term on the right-hand side of (24) is calculated with the Fokker–Planck equation as follows:

$$\begin{aligned}
 \frac{d}{dt} \int (\log f_t) f_t dx &= \int \frac{\partial_t f_t}{f_t} f_t dx + \int (\log f_t) (\partial_t f_t) dx \\
 &= \int \partial_t f_t dx + \int (\log f_t) (\Delta f_t + \nabla \cdot (f_t \nabla V)) dx \\
 &= \underbrace{\int \partial_t f_t dx}_{(I)} + \underbrace{\int (\log f_t) \Delta f_t dx}_{(II)} + \underbrace{\int (\log f_t) \nabla \cdot (f_t \nabla V) dx}_{(III)}.
 \end{aligned}
 \tag{25}$$

The integral (I) in (25) is clearly 0. By applying integration by parts, the integral (II) can be written by

$$(II) = - \int \nabla \log f_t \cdot \nabla f_t dx = - \int \left\| \frac{\nabla f_t}{f_t} \right\|^2 f_t dx.
 \tag{26}$$

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

$$(\log f_t) \nabla f_t = 2 (\sqrt{f_t} \log \sqrt{f_t}) \frac{\nabla f_t}{\sqrt{f_t}},
 \tag{27}$$

then, as μ_t has finite Fisher information $I(\mu_t) < \infty$, $\frac{\nabla f_t}{\sqrt{f_t}}$ has finite L^2 -norm in $L^2(\mathbb{R}^n, dx)$ and must be bounded at infinity. Furthermore, $\sqrt{f_t} \log \sqrt{f_t}$ will vanish at infinity by the limit formula, $\lim_{\xi \rightarrow +0} \xi \log \xi = 0$.

The integral (III) in (25) becomes

$$(III) = - \int \nabla f_t \cdot \nabla V dx,
 \tag{28}$$

by the following observations: since f_t is rapidly decreasing at infinity and the growth rate of the potential function V is at most polynomial by our assumption, we have $\lim_{|x| \rightarrow \infty} \sqrt{f_t} \nabla V = 0$. The limit

$\lim_{|x| \rightarrow \infty} \sqrt{f_t} \log \sqrt{f_t} = 0$ is the same as above. Thus,

$$(\log f_t) f_t \nabla V = 2 (\sqrt{f_t} \log \sqrt{f_t}) (\sqrt{f_t} \nabla V)
 \tag{29}$$

will vanish at infinity.

Substituting (26) and (28) into (25), we can obtain

$$\frac{d}{dt} \int (\log f_t) f_t dx = - \int \left\| \frac{\nabla f_t}{f_t} \right\|^2 f_t dx - \int \nabla f_t \cdot \nabla V dx.
 \tag{30}$$

Next, we shall see the second term on the right-hand side of (24), which can be reformulated by the Fokker–Planck equation as follows:

$$\begin{aligned}
 - \frac{d}{dt} \int (\log g_t) f_t dx &= - \int \frac{\partial_t g_t}{g_t} f_t dx - \int (\log g_t) (\partial_t f_t) dx \\
 &= - \int \frac{\Delta g_t + \nabla \cdot (g_t \nabla V)}{g_t} f_t dx - \int (\log g_t) (\Delta f_t + \nabla \cdot (f_t \nabla V)) dx \\
 &= \underbrace{- \int (\Delta g_t) \frac{f_t}{g_t} dx}_{(IV)} - \underbrace{\int \nabla \cdot (g_t \nabla V) \frac{f_t}{g_t} dx}_{(V)} \\
 &\quad - \underbrace{\int (\log g_t) \Delta f_t dx}_{(VI)} - \underbrace{\int (\log g_t) \nabla \cdot (f_t \nabla V) dx}_{(VII)}.
 \end{aligned}
 \tag{31}$$

The integral (IV) in (31) can be reformulated by applying integration by parts as follows:

$$\begin{aligned} \text{(IV)} &= \int \nabla g_t \cdot \nabla \left(\frac{f_t}{g_t} \right) dx = \int \nabla g_t \cdot \frac{g_t (\nabla f_t) - f_t (\nabla g_t)}{g_t^2} dx \\ &= \int \frac{\nabla g_t}{g_t} \cdot \frac{\nabla f_t}{f_t} f_t dx - \int \left\| \frac{\nabla g_t}{g_t} \right\|^2 f_t dx, \end{aligned} \quad (32)$$

where we can see that $(\nabla g_t) \frac{f_t}{g_t}$ vanishes at infinity by the following observation: factorize it as

$$(\nabla g_t) \frac{f_t}{g_t} = \frac{\nabla g_t}{\sqrt{g_t}} \sqrt{\frac{f_t}{g_t}} \sqrt{f_t}, \quad (33)$$

and then the boundedness of $\sqrt{\frac{f_t}{g_t}}$ is by our assumptions on the relative density and that of $\frac{\nabla g_t}{\sqrt{g_t}}$ comes from the finiteness of Fisher information $I(\nu_t) < \infty$.

Applying integration by parts again, the integral (V) in (31) becomes

$$\begin{aligned} \text{(V)} &= \int g_t \nabla V \cdot \nabla \left(\frac{f_t}{g_t} \right) dx = \int g_t \nabla V \cdot \frac{g_t (\nabla f_t) - f_t (\nabla g_t)}{g_t^2} dx \\ &= \int \nabla V \cdot \nabla f_t dx - \int \nabla V \cdot \frac{\nabla g_t}{g_t} f_t dx. \end{aligned} \quad (34)$$

Because the growth rate of the function V is at most polynomial and f_t is rapidly decreasing at infinity, thus $g_t \nabla V \frac{f_t}{g_t} = f_t \nabla V$ will vanish at infinity.

The integral (VI) in (31) can be reformulated as

$$\text{(VI)} = \int (\nabla \log g_t) \cdot \nabla f_t = \int \frac{\nabla g_t}{g_t} \cdot \frac{\nabla f_t}{f_t} f_t dx, \quad (35)$$

where we can find that $(\log g_t) \nabla f_t$ vanishes at infinity by factorizing

$$(\log g_t) \nabla f_t = 2 (\sqrt{g_t} \log \sqrt{g_t}) \frac{\nabla f_t}{\sqrt{f_t}} \sqrt{\frac{f_t}{g_t}}, \quad (36)$$

with the assumption of $I(\mu_t) < \infty$ and the boundedness of $\frac{f_t}{g_t}$.

The last integral (VII) in (31) becomes

$$\text{(VII)} = \int (\nabla \log g_t) \cdot (f_t \nabla V) dx = \int \frac{\nabla g_t}{g_t} \cdot \nabla V f_t dx, \quad (37)$$

where $(\log g_t) f_t \nabla V$ will vanish with the following factorization by the same reasons as above:

$$(\log g_t) f_t \nabla V = 2 (\sqrt{g_t} \log \sqrt{g_t}) \sqrt{\frac{f_t}{g_t}} (\sqrt{f_t} \nabla V). \quad (38)$$

In reformulation of the integrals (VI) and (VII), we have, of course, used integration by parts.

Substituting the equations from (32) to (37) into (31), we can have

$$-\frac{d}{dt} \int (\log g_t) f_t dx = - \int \left\| \frac{\nabla g_t}{g_t} \right\|^2 f_t dx + 2 \int \frac{\nabla f_t}{f_t} \cdot \frac{\nabla g_t}{g_t} f_t dx + \int \nabla V \cdot \nabla f_t dx. \quad (39)$$

Finally, combining (30) and (39), we obtain that

$$\begin{aligned} \frac{d}{dt}H(f_t | g_t) &= - \int \left\| \frac{\nabla f_t}{f_t} \right\|^2 f_t dx + 2 \int \frac{\nabla f_t}{f_t} \cdot \frac{\nabla g_t}{g_t} f_t dx - \int \left\| \frac{\nabla g_t}{g_t} \right\|^2 f_t dx \\ &= - \int \|\nabla \log f_t - \nabla \log g_t\|^2 f_t dx = -I(f_t | g_t). \end{aligned} \tag{40}$$

□

Remark 2. The assumption of dropping surface terms on integrations by parts, that is, vanishing at infinity in the proof of Proposition 3, is rather common in various physics, which has been also repeatedly employed in a series of works by Plastino et al., for instance, in [16,17].

Next, we will see the convergence of the relative entropy for the pair of time evolutes by the same Fokker–Planck equations.

Proposition 4. Under Situation A with Assumption D, the relative entropy $H(f_t | g_t)$ converges exponentially fast to 0 as $t \rightarrow \infty$.

Proof. We first expand the relative entropy $H(f_t | g_t)$ as

$$\begin{aligned} H(f_t | g_t) &= \int (\log f_t(x) - \log g_t(x)) f_t(x) dx \\ &= \int \left((\log f_t(x) + V(x)) - (\log g_t(x) + V(x)) \right) f_t(x) dx, \end{aligned} \tag{41}$$

where $V(x)$ is the potential function of the Fokker–Planck equation. Then, we obtain

$$H(f_t | g_t) \leq \int (\log f_t(x) + V(x)) f_t(x) dx + \int |\log g_t(x) + V(x)| f_t(x) dx. \tag{42}$$

Since the first term on the right-hand side of (42) is the relative entropy $H(f_t | e^{-V})$, we concentrate our attention on the second term.

We put the set $P \subset \mathbb{R}^n$ as $P = \{x \in \mathbb{R}^n : \log g_t(x) + V(x) \geq 0\}$, and then we have

$$|\log g_t(x) + V(x)| = \begin{cases} \log \left(\frac{g_t(x)}{e^{-V(x)}} \right) \leq \frac{g_t(x)}{e^{-V(x)}} - 1 & \text{for } x \in P, \\ \log \left(\frac{e^{-V(x)}}{g_t(x)} \right) \leq \frac{e^{-V(x)}}{g_t(x)} - 1 & \text{for } x \in P^c = \mathbb{R}^n \setminus P. \end{cases} \tag{43}$$

Thus, for sufficiently large t , it can be evaluated as follows:

$$\begin{aligned} &\int |\log g_t(x) + V(x)| f_t(x) dx \\ &\leq \int_P \left(\frac{g_t(x)}{e^{-V(x)}} - 1 \right) f_t(x) dx + \int_{P^c} \left(\frac{e^{-V(x)}}{g_t(x)} - 1 \right) f_t(x) dx \\ &= \int_P |g_t(x) - e^{-V(x)}| \left(\frac{f_t(x)}{e^{-V(x)}} \right) dx + \int_{P^c} |g_t(x) - e^{-V(x)}| \left(\frac{f_t(x)}{g_t(x)} \right) dx \\ &\leq M_1 \int_P |g_t(x) - e^{-V(x)}| dx + M_0 \int_{P^c} |g_t(x) - e^{-V(x)}| dx, \end{aligned} \tag{44}$$

where in the last inequality is by virtue of the assumption on the relative densities. Consequently, we can have the estimation that

$$\int |\log g_t(x) + V(x)| f_t(x) dx \leq \overline{M} \int |g_t(x) - e^{-V(x)}| dx, \tag{45}$$

with the positive constant $\bar{M} = \max\{M_0, M_1\}$.

As we have mentioned in Lemma 3 that the relative entropy controls the L^1 -norm, we have

$$\int |g_t(x) - e^{-V(x)}| dx \leq \sqrt{2H(g_t | e^{-V})}. \quad (46)$$

Thus, we obtain that, for sufficiently large t ,

$$H(f_t | g_t) \leq H(f_t | e^{-V}) + \bar{M} \sqrt{2H(g_t | e^{-V})}. \quad (47)$$

Taking the limit $t \rightarrow \infty$, it follows that $H(f_t | g_t) \rightarrow 0$ exponentially fast because $H(f_t | e^{-V})$ and $H(g_t | e^{-V})$ converge to 0 in exponentially fast with rate $2K$. \square

By the dissipation formula in Proposition 3, together with the above convergence, we can obtain the following integral representation of relative entropy.

Theorem 2. *Let f_t and g_t ($t \geq 0$) be the flows of probability densities on \mathbb{R}^n by the Fokker–Planck equation under situation A with the assumptions on relative densities D . Then, we have the integral representation of the relative entropy*

$$H(f_0 | g_0) = \int_0^\infty I(f_t | g_t) dt. \quad (48)$$

If we choose particularly the equilibrium e^{-V} as the initial measure of the reference g_0 , then it is stationary such that $g_t = e^{-V}$ ($t \geq 0$). Hence, as the direct consequence of the above theorem, we have the following integral formula:

Corollary 1.

$$H(f_0 | e^{-V}) = \int_0^\infty I(f_t | e^{-V}) dt. \quad (49)$$

3. An Application to the Entropy Gap

In this section, we shall apply the formula of the time integration in Theorem 2 to the Ornstein–Uhlenbeck flows, which gives an extension of the formula of the entropy gap. For simplicity, we will consider the one-dimensional case in this section.

Among random variables with unit variance, the Gaussian has the largest entropy. Let X be a standardized (mean 0 and variance 1) random variable, and let Z be a standard Gaussian random variable. Then, the quantity $H(Z) - H(X)$ is called *the entropy gap* or *the non-Gaussianity*, which coincides, of course, with the relative entropy $H(X | Z)$. It is known (see, for instance, [18]) that this entropy gap can be written as the integration of the Fisher information. Namely,

$$H(X | Z) = \int_0^\infty (I(X_t) - 1) dt, \quad (50)$$

where X_t is the time evolute at t of the random variable X by the Ornstein–Uhlenbeck semigroup in (17). It is easy to find that our formula (48) of Theorem 2 covers (50) as one of the special cases.

In the formula (48) of Theorem 2, even for the case of the quadratic potential $V(x) = \frac{x^2}{2}$, we can choose the initial reference measure ν_0 more freely other than the standard Gaussian as we will illustrate below. Let X be a centered random variable of variance σ^2 (not a unit in general) and G be a centered Gaussian of the same variance σ^2 . Then, applying the integral formula with the potential function $V(x) = \frac{x^2}{2}$, the relative entropy $H(X | G)$, which is equal to the entropy gap $H(G) - H(X)$, can be written by the integration

$$H(X | G) = \int_0^\infty I(X_t | G_t) dt, \quad (51)$$

where X_t and G_t are the time evolutes by the Ornstein–Uhlenbeck semigroup of X and G , respectively. By formula (18), X_t and G_t can be given as

$$\begin{cases} X_t &= \sqrt{e^{-2t}}X + \sqrt{1 - e^{-2t}}Z, \\ G_t &= \sqrt{e^{-2t}}G + \sqrt{1 - e^{-2t}}Z, \end{cases} \tag{52}$$

where Z is a standard Gaussian random variable independent of X and G .

Since the time evolute G_t becomes a Gaussian random variable of variance $e^{-2t}\sigma^2 + (1 - e^{-2t})$, the score function of which is given by

$$\rho_{G_t}(x) = -\frac{x}{\text{Var}(G_t)} = -\frac{x}{e^{-2t}\sigma^2 + (1 - e^{-2t})}, \tag{53}$$

and it is easy to find that the time evolute X_t has the same variance as of G_t , $\text{Var}(X_t) = \text{Var}(G_t)$.

We denote by μ_t the probability distribution of X_t , and let ρ_{X_t} and ρ_{G_t} be the score functions of the random variables X_t and G_t , respectively. Then, by direct calculation, we obtain that

$$\int_{-\infty}^{\infty} x \rho_{X_t}(x) d\mu_t(x) = -1, \tag{54}$$

which corresponds to the special case of the Stein’s identity in (3).

With the above observations, we can reformulate the relative Fisher information $I(X_t | G_t)$ as follows:

$$\begin{aligned} I(X_t | G_t) &= \int_{-\infty}^{\infty} (\rho_{X_t}(x) - \rho_{G_t}(x))^2 d\mu_t(x) \\ &= \int_{-\infty}^{\infty} (\rho_{X_t}(x)^2 - 2\rho_{X_t}(x)\rho_{G_t}(x) + \rho_{G_t}(x)^2) d\mu_t(x) \\ &= \int_{-\infty}^{\infty} \rho_{X_t}(x)^2 d\mu_t(x) + \frac{2}{\text{Var}(G_t)} \int_{-\infty}^{\infty} x \rho_{X_t}(x) d\mu_t(x) + \frac{1}{\text{Var}(G_t)^2} \int_{-\infty}^{\infty} x^2 d\mu_t(x) \\ &= I(X_t) - \frac{2}{\text{Var}(G_t)} + \frac{\text{Var}(X_t)}{\text{Var}(G_t)^2} \\ &= I(X_t) - \frac{1}{\text{Var}(X_t)}, \end{aligned} \tag{55}$$

where we have used formula (54) and the fact that $\text{Var}(G_t) = \text{Var}(X_t)$ in the last equality. Now, the following formula can be obtained as a direct consequence of Theorem 2.

Proposition 5. *Let X be a centered random variable of finite variance σ^2 , and let G be a centered Gaussian variable of the same variance σ^2 . For $t \geq 0$, we denote by X_t the evolute of $X_0 = X$ by the Ornstein–Uhlenbeck semigroup, that is,*

$$X_t = \sqrt{e^{-2t}}X + \sqrt{1 - e^{-2t}}Z, \tag{56}$$

where Z stands for the standard Gaussian random variable independent of X . Then, it follows that

$$H(X | G) = \int_0^{\infty} \left(I(X_t) - \frac{1}{\text{Var}(X_t)} \right) dt = \int_0^{\infty} \left(I(X_t) - \frac{1}{(\sigma^2 - 1)e^{-2t} + 1} \right) dt. \tag{57}$$

Remark 3. *The Ornstein–Uhlenbeck model can be regarded as the time-dependent dilation of Gaussian perturbation. Thus, we can rewrite formula (57) in terms of Gaussian perturbation. Since the Fisher information behaves for dilation of a random variable as $I(\alpha Y) = \frac{1}{\alpha^2} I(Y)$, we obtain*

$$I(X_t) = I\left(\sqrt{e^{-2t}}X + \sqrt{1 - e^{-2t}}Z\right) = e^{2t}I\left(X + \sqrt{\frac{1 - e^{-2t}}{e^{-2t}}}Z\right). \tag{58}$$

Changing the variables by $\tau = \frac{1 - e^{-2t}}{e^{-2t}}$, the integral on the right-hand side of (56) becomes

$$\int_0^\infty \left\{ (1 + \tau) I(X + \sqrt{\tau}Z) - \frac{1 + \tau}{\sigma^2 + \tau} \right\} \frac{d\tau}{2(1 + \tau)} = \frac{1}{2} \int_0^\infty \left\{ I(X + \sqrt{\tau}Z) - \frac{1}{\sigma^2 + \tau} \right\} d\tau. \tag{59}$$

Hence, we obtain

$$H(G) - H(X) = \frac{1}{2} \int_0^\infty \left\{ I(X + \sqrt{\tau}Z) - \frac{1}{\sigma^2 + \tau} \right\} d\tau, \tag{60}$$

that is,

$$H(X) = \frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \int_0^\infty \left(I(X + \sqrt{\tau}Z) - \frac{1}{\sigma^2 + \tau} \right) d\tau, \tag{61}$$

which is the known integral representation for the entropy of a random variable X by the Fisher information of Gaussian perturbation derived by Barron in Section 2 in [2].

4. Some Numerical Examples

In this section, we will give numerical examples for the case of the Ornstein–Uhlenbeck flows, which are given by the potential $V = \frac{x^2}{2}$, and, hence, have the standard Gaussian equilibrium. As we mentioned in (18), the densities of the flows at time t can be calculated analytically by convolution of the scaled initial measures.

Example 1. In the first numerical example, we take the uniform distribution on the interval $(-1, 1)$ as the initial objective measure. Namely, the density $f_0(x) = u(x)$ is given by

$$u(x) = \begin{cases} \frac{1}{2} & (-1 \leq x \leq 1), \\ 0 & \text{otherwise,} \end{cases}$$

which has the mean 0 and the variance $\frac{1}{3}$. We set the centered Gaussian of variance $\frac{1}{3}$ as the initial reference measure. Namely, $g_0(x) = \varphi(x, \frac{1}{3})$, where the function φ is defined by

$$\varphi(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Here, we can calculate the densities f_t and g_t of the Ornstein–Uhlenbeck flows at time t analytically as follows: we rescale the time parameter t as $\tau = \sqrt{e^{-2t}}$ and put

$$u_\tau(x) = \frac{1}{\tau} u\left(\frac{x}{\tau}\right), \quad \varphi_\tau(x) = \frac{1}{\sqrt{1-\tau^2}} \varphi\left(\frac{x}{\sqrt{1-\tau^2}}, 1\right).$$

Then, f_t and g_t are given by

$$\begin{aligned} f_t(x) &= (u_\tau * \varphi_\tau)(x) = \int_{-\infty}^\infty u_\tau(y) \varphi_\tau(x - y) dy \\ &= \frac{1}{2\tau\sqrt{1-\tau^2}} \int_{-\tau}^\tau \varphi\left(\frac{x-y}{\sqrt{1-\tau^2}}, 1\right) dy, \end{aligned}$$

and

$$g_t(x) = \varphi\left(x, 1 - \frac{2}{3}\tau^2\right),$$

respectively. Now, we shall illustrate the convergence of $H(f_t|g_t)$ and the upper bound of the right-hand side in (47) numerically by graphs. We claim that the constant \bar{M} can be assumed to be 1 because, in our assumptions (20), (21), and (22), the relative densities converge uniformly to 1. The convergence of $H(f_t|g_t)$ is illustrated in Figure 1.

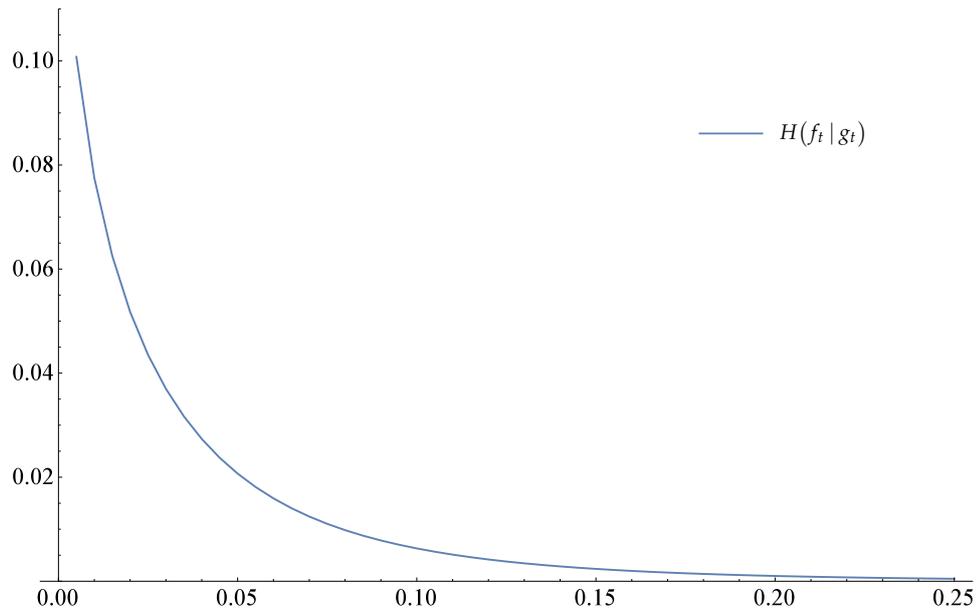


Figure 1. The value of $H(f_t | g_t)$.

In Figure 2, the dashed curve indicates the convergence of $H(f_t | g_t)$ in Figure 1.

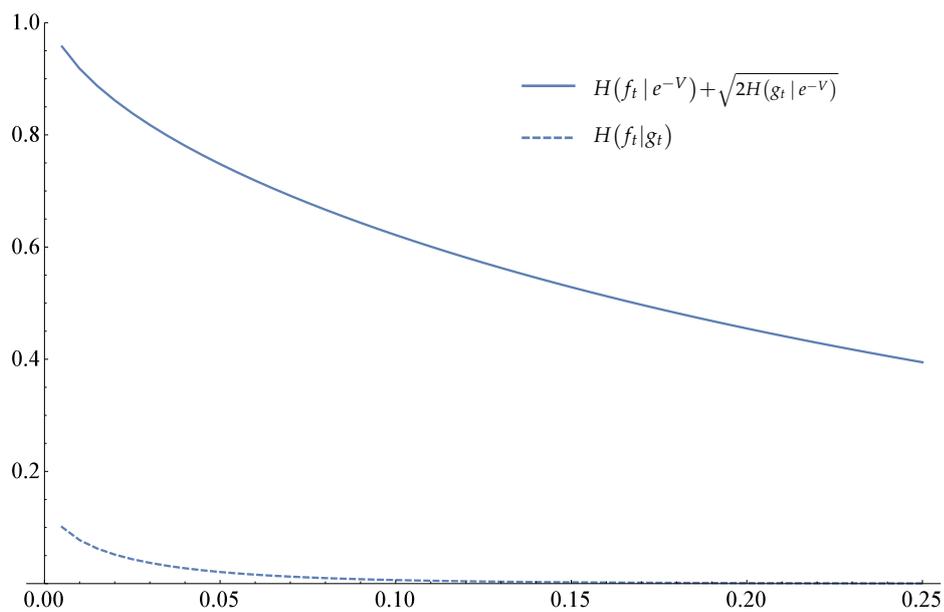


Figure 2. The value of $H(f_t | e^{-V}) + \sqrt{2H(g_t | e^{-V})}$.

Example 2. In the second example, we put the initial reference measure as $g_0(x) = \varphi(x, 3)$, that is, we take the centered Gaussian of variance 3 as the initial reference measure, but $f_0(x) = u(x)$ is unchanged. In Figure 3, the convergence of $H(f_t | g_t)$ is illustrated.

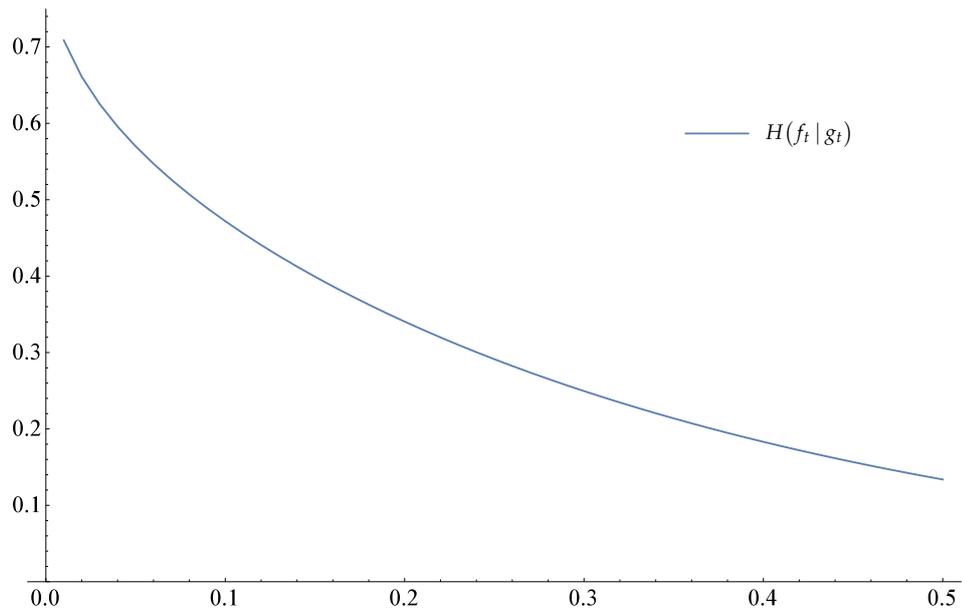


Figure 3. The value of $H(f_t | g_t)$.

In Figure 4, the dashed curve indicates the convergence of $H(f_t | g_t)$ in Figure 3.

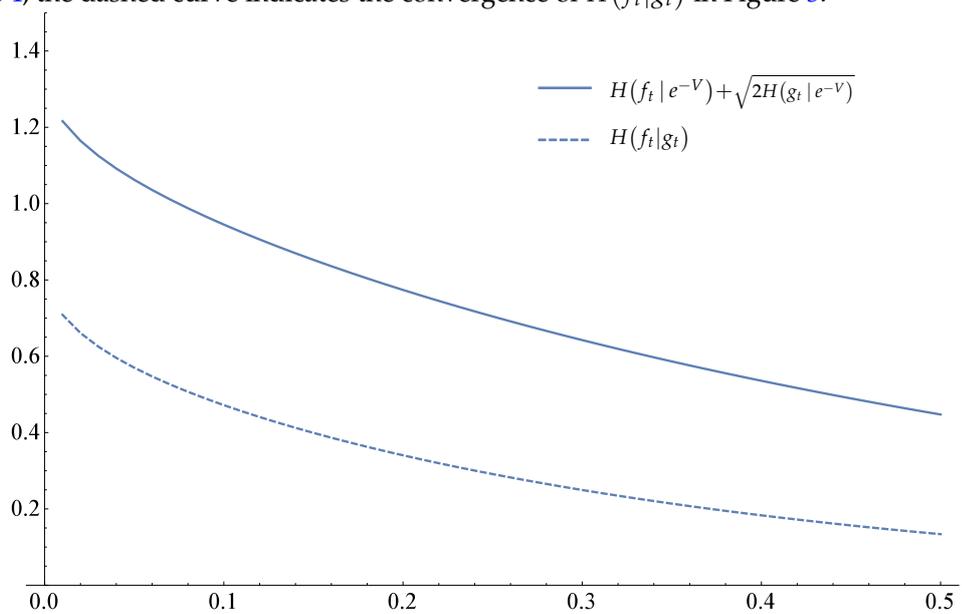


Figure 4. The value of $H(f_t | e^{-V}) + \sqrt{2H(g_t | e^{-V})}$.

Example 3. In the third numerical example, the initial objective and the initial reference measures are given as the uniform distributions on the intervals $(-1, 1)$ and $(-\frac{1}{2}, \frac{1}{2})$, respectively. Namely, we set the densities as $f_0(x) = u(x)$ and $g_0 = 2u(2x)$. We illustrate the convergence of $H(f_t | g_t)$ in Figure 5.

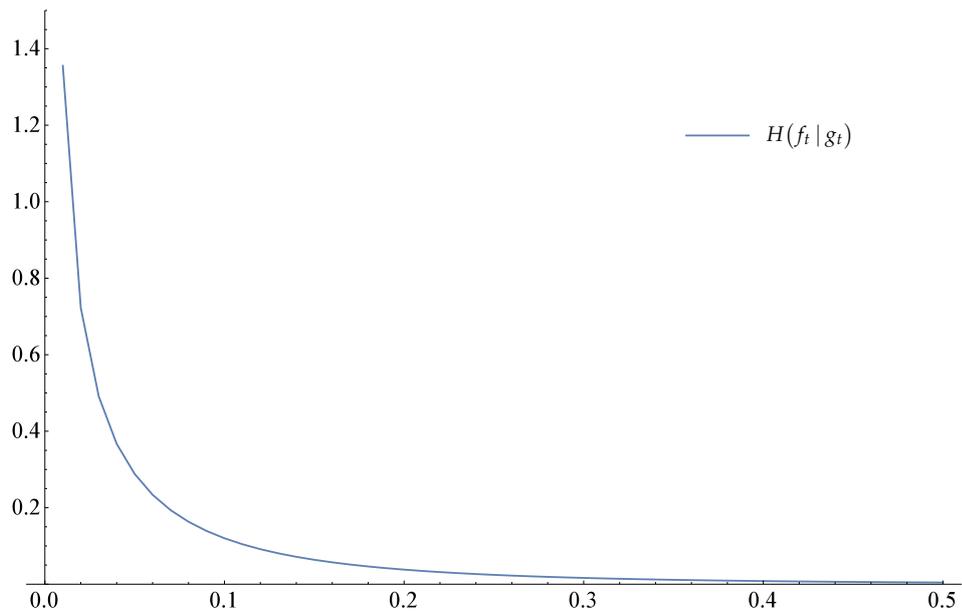


Figure 5. The value of $H(f_t | g_t)$.

In Figure 6, the dashed curve indicates the convergence of $H(f_t | g_t)$ in Figure 5.

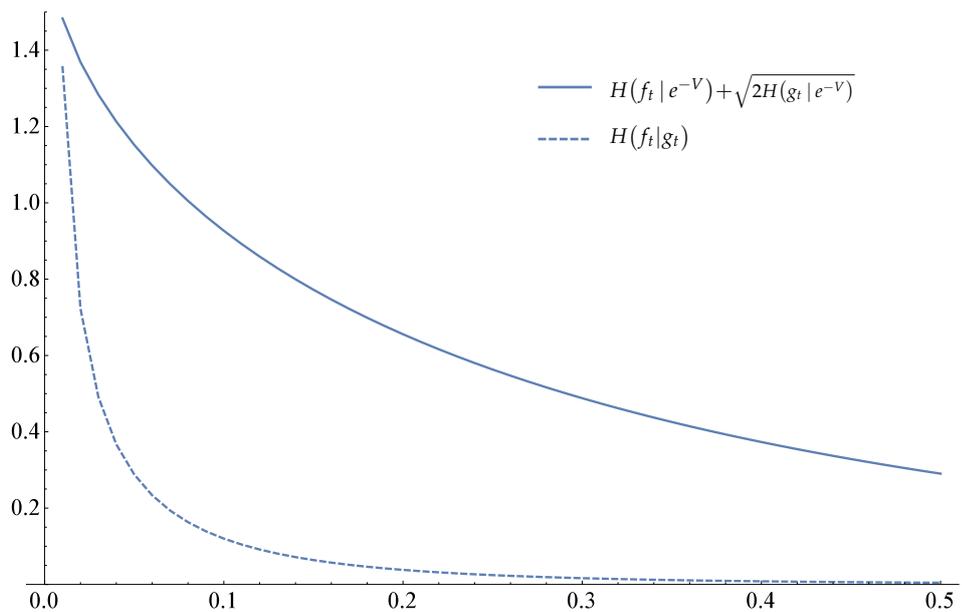


Figure 6. The value of $H(f_t | e^{-V}) + \sqrt{2H(g_t | e^{-V})}$.

5. Conclusions

The partial differential equation of Fokker–Planck describes the flow of the probability measures for diffusion process. The diffusion by the Fokker–Planck equation with the strictly convex potential V has the long-time asymptotic stationary measure e^{-V} . In the case of the relative entropy endowed with the stationary measure e^{-V} as reference, the dissipation formula of the relative entropy of the diffusion flow by the Fokker–Planck equation with the potential V is known in literature.

In this paper, we have derived the similar dissipation formula under the more flexible situation. Namely, we have considered the situation that the reference measure is also evolved by the Fokker–Planck equation with the same potential function V for the objective measure. Then, we

have obtained another integral representation of the relative entropy, which gives an extension of the formula of the entropy gap.

Acknowledgments: The author is grateful for comments and suggestions by anonymous reviewers. In particular, the numerical simulations are based on their valuable comments.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.* **1959**, *2*, 101–112.
2. Barron, A. Entropy and the central limit theorem. *Ann. Probab.* **1986**, *14*, 336–342.
3. Villani, C. *Topics in Optimal Transportation*; American Mathematical: Providence, RI, USA, 2003.
4. Verdú, S. Mismatched estimation and relative entropy. *IEEE Trans. Inf. Theory* **2010**, *56*, 3712–3719.
5. Hirata, M.; Nemoto, A.; Yoshida, H. An integral representation of the relative entropy. *Entropy* **2012**, *14*, 1469–1477.
6. Arnold, A.; Markowich, P.; Toscani, G.; Unterreiter, A. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations. *Commun. Partial Differ. Equ.* **2001**, *26*, 43–100.
7. Bakry, D.; Émery, M. Diffusions hypercontractives. In *Séminar de Probabilités XIX*; Springer: Berlin/Heidelberg, Germany, 1985; pp. 177–206.
8. Otto, F.; Villani, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **2000**, *173*, 361–400.
9. Gross, L. Logarithmic Sobolev inequalities. *Am. J. Math.* **1975**, *97*, 1061–1083.
10. Carlen, E. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.* **1991**, *101*, 194–211.
11. Talagrand, M. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **1996**, *6*, 587–600.
12. Csiszar, I. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.* **1967**, *2*, 299–318.
13. Csiszar, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
14. Yanomo, T. De Bruijn-type identity for systems with flux. *Eur. J. Phys. B* **2013**, *86*, 363.
15. Yanomo, T. Phase space gradient of dissipated work and information: A role of relative Fisher information. *J. Math. Phys.* **2013**, *54*, 113301.
16. Daffertshofer, A.; Plastino, A.R.; Plastino, A. Classical No-Cloning Theorem. *Phys. Rev. Lett.* **2002**, *88*, 210601.
17. Plastino, A.R.; Daffertshofer, A. Liouville Dynamics and the Conservation of Classical Information. *Phys. Rev. Lett.* **2004**, *93*, 138701.
18. Carlen, E.A.; Soffer, A. Entropy production by block variable summation and central limit theorems. *Commun. Math. Phys.* **1991**, *140*, 339–371.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).