

Article

A Representation of the Relative Entropy with Respect to a Diffusion Process in Terms of Its Infinitesimal Generator

Olivier Faugeras * and James MacLaurin

INRIA Sophia Antipolis Mediterannee, 2004 Route Des Lucioles, Sophia Antipolis, France;

E-Mail: olivier.faugeras@inria.fr

* Author to whom correspondence should be addressed; E-Mail: james.maclaurin@inria.fr.

External Editor: Kevin H. Knuth

Received: 23 October 2014; in revised form: 17 December 2014 / Accepted: 18 December 2014 /

Published: 22 December 2014

Abstract: In this paper we derive an integral (with respect to time) representation of the relative entropy (or Kullback–Leibler Divergence) $R(\mu||P)$, where μ and P are measures on $C([0, T]; \mathbb{R}^d)$. The underlying measure P is a weak solution to a martingale problem with continuous coefficients. Our representation is in the form of an integral with respect to its infinitesimal generator. This representation is of use in statistical inference (particularly involving medical imaging). Since $R(\mu||P)$ governs the exponential rate of convergence of the empirical measure (according to Sanov’s theorem), this representation is also of use in the numerical and analytical investigation of finite-size effects in systems of interacting diffusions.

Keywords: relative entropy; Kullback–Leibler; diffusion; martingale formulation

1. Introduction

In this paper we derive an integral representation of the relative entropy $R(\mu||P)$, where μ is a measure on $C([0, T]; \mathbb{R}^d)$ and P governs the solution to a stochastic differential equation (SDE). The relative entropy is used to quantify the distance between two measures. It has considerable applications in statistics, imaging, information theory and communications. It has been used in the long-time analysis of Fokker–Planck equations [1,2], the analysis of dynamical systems [3] and the analysis of spectral density functions [4]. It has been used in financial mathematics to quantify the difference between martingale measures [5,6]. It has also been shown in [7] that the existence problem of the minimal relative

entropy martingale measure problem of birth and death processes can be reduced to the problem of solving the Hamilton–Jacobi–Bellman equation; furthermore the minimal entropy martingale measures (MEMMs) for geometric Levy processes are investigated in [8]. The finiteness of $R(\mu||P)$ has been shown to be equivalent to the invertibility of certain shifts on Wiener space, when P is the Wiener measure [9,10]. However, one of the most frequent uses of the relative entropy is in statistical inference (particularly in medical imaging) [11,12]. For example, in data fitting, it is a standard technique to select the parameters that minimise the relative entropy of two conditional probability distributions [13]. Modelling in medical imaging increasingly involves diffusion process with state space $C([0, T]; \mathbb{R}^d)$, for which the expression $R(\mu||P) = E^\mu[\log \frac{d\mu}{dP}]$ or the variational definition in Definition 1 may not always be tractable. Furthermore, it is not always clear that one may simply approximate the relative entropy by successively calculating it for the marginals over increasingly fine time-discretisations, since these expressions may asymptotically approach infinity (see (4) below).

Another very important application of the relative entropy is in the field of Large Deviations. Sanov’s theorem dictates that the empirical measure induced by independent samples governed by the same probability law P converge towards their limit exponentially fast; and the constant governing the rate of convergence is the relative entropy [14]. Large Deviations have been applied for example to spin glasses [15], neural networks [16–18] and mean-field models of interacting particles [19,20]. In the mean-field theory of neuroscience in particular, there has been a recent interest in the modelling of “finite size effects” [18,21], that is, the deviations from the limiting behaviour for a population of a particular size. Large Deviations provides a mathematically rigorous tool to do this. In this system, the limiting system is typically the law P of a stochastic process, and therefore the likelihood of the empirical measure of the system being “near” some measure μ is the relative entropy $R(\mu||P)$. However the numerical calculation of $R(\mu||P)$ is not straightforward: the results of this paper provide an alternative characterization of $R(\mu||P)$, which assists in this calculation.

For example, the rate function for the Large Deviation Principle of the interacting particle model of [20] is directly in terms of the relative entropy between two measures on the space of continuous functions (see in particular Theorem 5.2 of this paper). Similarly, the rate function in [18] (Theorem 10) may be expressed as a function of the relative entropy. In more detail, the rate function \check{J} in [18] (Theorem 10) is of the form $\check{J}(\mu) = \lim_{n \rightarrow \infty} \frac{1}{|V_n|} R(\mu^{V_n} || \Xi^{V_n})$. Here Ξ is the law of the process in [18] (Equation (31)), *i.e.*, the law of a \mathbb{Z}^d -indexed stochastic process, and μ^{V_n} and Ξ^{V_n} are the marginals over the finite hypercube V_n of side length $(2n + 1)$. The results of this paper give a means of evaluating $R(\mu^{V_n} || \Xi^{V_n})$ and therefore $\check{J}(\mu)$.

In this paper we derive a specific integral (with respect to time) representation of the relative entropy $R(\mu||P)$ when P is the law of a diffusion process. The representation is in terms of the infinitesimal generator of P . This P is the same as in [22] (Section 4). The representation makes use of regular conditional probabilities. We expect that in some circumstances, it ought to be more tractable than the standard definition in Definition 1, and thus it might be of practical use in the applications listed above.

2. Outline of Main Result

Let \mathcal{T} be the Banach Space $C([0, T]; \mathbb{R}^d)$ equipped with the norm

$$\|X\| = \sup_{s \in [0, T]} \{|X_s|\}, \tag{1}$$

where $|\cdot|$ is the standard Euclidean norm over \mathbb{R}^d . We let (\mathcal{F}_t) be the canonical filtration over $(\mathcal{T}, \mathcal{B}(\mathcal{T}))$. For some topological space \mathcal{X} , we let $\mathcal{B}(\mathcal{X})$ be the Borelian σ -algebra and $\mathcal{M}(\mathcal{X})$ the space of all probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Unless otherwise indicated, we endow $\mathcal{M}(\mathcal{X})$ with the topology of weak convergence. Let $\sigma = \{t_1, t_2, \dots, t_m\}$ be a finite set of elements such that $t_1 \geq 0$, $t_m \leq T$ and $t_j < t_{j+1}$. We term σ a partition, and denote the set of all such partitions by \mathbf{J} . The set of all partitions of the above form such that $t_1 = 0$ and $t_m = T$ is denoted \mathbf{J}_* . We define $|\sigma| = \sup_{1 \leq j \leq m-1} \{t_{j+1} - t_j\}$. For some $t \in [0, T]$ and $\sigma \in \mathbf{J}_*$, we define $\underline{\sigma}(t) = \sup\{s \in \sigma \mid s \leq t\}$. The following definition of relative entropy is standard.

Definition 1. Let (Ω, \mathcal{H}) be a measurable space, and μ, ν probability measures.

$$R_{\mathcal{H}}(\mu \parallel \nu) = \sup_{f \in \mathcal{E}} \{\mathbb{E}^\mu[f] - \log \mathbb{E}^\nu[\exp(f)]\} \in \mathbb{R} \cup \infty,$$

where \mathcal{E} is the set of all bounded functions. If the σ -algebra is clear from the context, we omit the \mathcal{H} and write $R(\mu \parallel \nu)$. If Ω is Polish and $\mathcal{H} = \mathcal{B}(\Omega)$, then we only need to take the supremum over the set of all continuous bounded functions.

Let $P \in \mathcal{M}(\mathcal{T})$ be the following law governing a Markov–Feller diffusion process on \mathcal{T} . Stipulate P to be a weak solution (with respect to the canonical filtration) of the local martingale problem with infinitesimal generator

$$\mathcal{L}_t(f) = \frac{1}{2} \sum_{1 \leq j, k \leq d} a^{jk}(t, x) \frac{\partial^2 f}{\partial x^j \partial x^k} + \sum_{1 \leq j \leq d} b^j(t, x) \frac{\partial f}{\partial x^j},$$

for $f(x)$ in $C^2(\mathbb{R}^d)$, *i.e.*, the space of twice continuously differentiable functions. The initial condition (governing P_0) is $\mu_I \in \mathcal{M}(\mathbb{R}^d)$. The coefficients $a^{jk}, b^j : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ are assumed to be continuous (over $[0, T] \times \mathbb{R}^d$), and the matrix $a(t, x)$ is strictly positive definite for all t and x . Here P is assumed to be the unique weak solution. We note that the above infinitesimal generator is the same as in [22] (p. 269) (note particularly its Remark 4.4). We note that P is the law of the solution $Y = (Y^j)$ to the following stochastic differential equation: for $j \in [1, d]$,

$$dY_t^j = b^j(t, Y)dt + \sum_{k=1}^d a^{jk}(t, Y)dW^k.$$

Here (W^k) are independent Wiener processes.

Our major result is the following. Let $\mu \in \mathcal{M}(\mathcal{T})$ govern a random variable $X \in \mathcal{T}$. For some $x \in \mathcal{T}$, we note $\mu_{|[0, s], x}$, the regular conditional probability (rcp) given $X_r = x_r$ for all $r \in [0, s]$. The marginal of $\mu_{|[0, s], x}$ at some time $t \geq s$ is noted $\mu_{t|[0, s], x}$.

Theorem 1. Let $(\sigma^{(m)})_{m \in \mathbb{Z}^+}$ be any series of partitions such that $\sigma^{(m)} \subseteq \sigma^{(m+1)}$ and $|\sigma^{(m)}| \rightarrow 0$ as $m \rightarrow \infty$. For $\mu \in \mathcal{M}(\mathcal{T})$,

$$\begin{aligned} R(\mu||P) &= R_{\mathcal{F}_0}(\mu||P) + \sup_{\sigma \in \mathbf{J}_*} \Gamma(\sigma) \\ &= R_{\mathcal{F}_0}(\mu||P) + \lim_{m \rightarrow \infty} \Gamma(\sigma^{(m)}) \end{aligned} \tag{2}$$

where

$$\begin{aligned} \Gamma(\sigma) &= E^{\mu(x)} \left[\int_0^T \sup_{f \in \mathcal{D}} \left\{ \frac{\partial}{\partial t} E^{\mu_t|[0, \underline{\sigma}(t)], x} [f] \right. \right. \\ &\quad \left. \left. - E^{\mu_t|[0, \underline{\sigma}(t)], x(y)} \left[\mathcal{L}_t f(y) + \frac{1}{2} \sum_{j,k=1}^d a^{jk}(t, y) \frac{\partial f}{\partial y^j} \frac{\partial f}{\partial y^k} \right] \right\} dt \right]. \end{aligned} \tag{3}$$

Here \mathcal{D} is the Schwartz space of compactly supported functions $\mathbb{R}^d \rightarrow \mathbb{R}$, possessing continuous derivatives of all orders. If $\frac{\partial}{\partial t} E^{\mu_t|[0, \underline{\sigma}(t)], x} [f]$ does not exist, then we consider it to be ∞ .

Our paper has the following format. In Section 3 we make some preliminary definitions, defining the process P against which the relative entropy is taken in this paper. In Section 4 we employ the projective limits approach of [22] to obtain the chief result of this paper: Theorem 1. This gives an explicit integral representation of the relative entropy. In Section 5 we apply the result in Theorem 1 to various corollaries, including the particular case when μ is the solution of a martingale problem. We finish by comparing our results to those of [19] and [20].

3. Preliminaries

We outline some necessary definitions. For $\sigma \in \mathbf{J}$ of the form $\sigma = \{t_1, t_2, \dots, t_m\}$, let $\sigma_{:j} = \{t_1, \dots, t_j\}$. We denote the number of elements in a partition σ by $m(\sigma)$. We let \mathbf{J}_s be the set of all partitions lying in $[0, s]$. For $0 < s < t \leq T$, we let $\mathbf{J}_{s,t}$ be the set of all partitions of the form $\sigma \cup t$, where $\sigma \in \mathbf{J}_s$.

Let $\pi_\sigma : \mathcal{T} \rightarrow \mathcal{T}_\sigma := \mathbb{R}^{d \times m(\sigma)}$ be the natural projection, i.e., such that $\pi_\sigma(x) = (x_{t_1}, \dots, x_{t_{m(\sigma)}})$. We similarly define the natural projection $\pi_{\alpha\gamma} : \mathcal{T}_\gamma \rightarrow \mathcal{T}_\alpha$ (for $\alpha \subseteq \gamma \in \mathbf{J}$), and we define $\pi_{[s,t]} : \mathcal{T} \rightarrow C([s, t]; \mathbb{R}^d)$ to be the natural restriction of $x \in \mathcal{T}$ to $[s, t]$. The expectation of some measurable function f with respect to a measure μ is written as $\mathbb{E}^{\mu(x)}[f(x)]$, or simply $\mathbb{E}^\mu[f]$ when the context is clear.

For $s < t$, we write $\mathcal{F}_{s,t} = \pi_{[s,t]}^{-1} \mathcal{B}(C([s, t]; \mathbb{R}^d))$ and $\mathcal{F}_\sigma = \pi_\sigma^{-1} \mathcal{B}(\mathcal{T}_\sigma)$. We define $\mathcal{F}_{s;t}$ to be the σ -algebra generated by \mathcal{F}_s and \mathcal{F}_γ (where $\gamma = [t]$). For $\mu \in \mathcal{M}(\mathcal{T})$, we denote its image laws by $\mu_\sigma := \mu \circ \pi_\sigma^{-1} \in \mathcal{M}(\mathcal{T}_\sigma)$ and $\mu_{[s,t]} := \mu \circ \pi_{[s,t]}^{-1} \in \mathcal{M}(C([s, t]; \mathbb{R}^d))$. Let $\mu \in \mathcal{M}(\mathcal{T})$ govern a random variable $X = (X_s) \in \mathcal{T}$. For $z \in \mathbb{R}^d$, we write the rcp given $X_s = z$ by $\mu|_{s,z}$. For $x \in C([0, s]; \mathbb{R}^d)$ or \mathcal{T} , the rcp given that $X_u = x_u$ for all $0 \leq u \leq s$ is written as $\mu|_{[0,s],x}$. The rcp given that $X_u = x_u$ for all $u \leq s$, and $X_t = z$, is written as $\mu|_{s,x;t,z}$. For $\sigma \in \mathbf{J}_s$ and $z \in (\mathbb{R}^d)^{m(\sigma)}$, the rcp given that $X_u = z_u$ for all $u \in \sigma$ is written as $\mu|_{\sigma,z}$. All of these measures are considered to be in $\mathcal{M}(C([s, T]; \mathbb{R}^d))$ (unless indicated otherwise in particular circumstances). The probability laws governing X_t (for $t \geq s$), for each

of these, are respectively $\mu_{t|s,z}$, $\mu_{t|[0,s],x}$ and $\mu_{t|\sigma,z}$. We clearly have $\mu_{s|s,z} = \delta_z$, for μ_s a.e. z , and similarly for the others.

REMARK. See [23] (Definition 5.3.16) for a definition of a rcp. Technically, if we let $\mu_{|s,z}^*$ be the rcp given $X_s = z$ according to this definition, then $\mu_{|s,z} = \mu_{s,z}^* \circ \pi_{[s,T]}^{-1}$ and $\mu_{t|s,z} = \mu_{s,z}^* \circ \pi_t^{-1}$. By [23] (Theorem 3.18), $\mu_{|s,z}$ is well-defined for μ_s a.e. z . Similar comments apply to the other rcp's defined above.

In the definition of the relative entropy, we abbreviate $R_{\mathcal{F}_\sigma}(\mu||P)$ by $R_\sigma(R||P)$. If $\sigma = \{t\}$, we write $R_t(\mu||P)$.

4. The Relative Entropy $R(\cdot||P)$ Using Projective Limits

In this section we derive an integral representation of the relative entropy $R(\mu||P)$, for arbitrary $\mu \in \mathcal{M}(\mathcal{T})$. We start with the standard result in Theorem 2, before adapting the projective limits approach of [22] to obtain the central result (Theorem 1).

We begin with a standard decomposition result for the relative entropy [24].

Lemma 1. Let X be a Polish space with sub σ -algebras $\mathcal{G} \subseteq \mathcal{F} \subseteq \mathcal{B}(X)$. Let μ and ν be probability measures on (X, \mathcal{F}) , and their regular conditional probabilities over \mathcal{G} be (respectively) μ_ω and ν_ω . Then

$$R_{\mathcal{F}}(\mu||\nu) = R_{\mathcal{G}}(\mu||\nu) + \mathbb{E}^{\mu(\omega)} [R_{\mathcal{F}}(\mu_\omega||\nu_\omega)].$$

The following Theorem is a straightforward consequence of [25] (Theorem 6.6): we provide an alternative proof using the theory of Large Deviations in Section 6.

Theorem 2. If $\alpha, \sigma \in \mathbf{J}$ and $\alpha \subseteq \sigma$, then $R_\alpha(\mu||P) \leq R_\sigma(\mu||P)$. Furthermore,

$$R_{\mathcal{F}_{s,t}}(\mu||P) = \sup_{\sigma \in \mathbf{J} \cap [s,t]} R_\sigma(\mu||P), \tag{4}$$

$$R_{\mathcal{F}_{s,t}}(\mu||P) = \sup_{\sigma \in \mathbf{J}_{s,t}} R_\sigma(\mu||P). \tag{5}$$

It suffices for the supremums in (4) to take $\sigma \subset \mathcal{Q}_{s,t}$, where $\mathcal{Q}_{s,t}$ is any countable dense subset of $[s, t]$. Thus we may assume that there exists a sequence $\sigma^{(n)} \subset \mathcal{Q}_{s,t}$ of partitions such that $\sigma^{(n)} \subseteq \sigma^{(n+1)}$, $|\sigma^{(n)}| \rightarrow 0$ as $n \rightarrow \infty$ and

$$R_{\mathcal{F}_{s,t}}(\mu||P) = \lim_{n \rightarrow \infty} R_{\sigma^{(n)}}(\mu||P). \tag{6}$$

We now provide a technical lemma.

Lemma 2. Let $t > s$, $\alpha, \sigma \in \mathbf{J}_s$, $\sigma \subset \alpha$ and $s \in \sigma$. Then for μ_σ a.e. x , $R_t(\mu_{|\sigma,x}||P_{|s,x_s}) = R(\mu_{t|\sigma,x}||P_{t|s,x_s})$. Secondly,

$$E^{\mu_\sigma(x)} [R_t(\mu_{|\sigma,x}||P_{|s,x_s})] \leq E^{\mu_\alpha(z)} [R_t(\mu_{|\alpha,z}||P_{|s,z_s})].$$

Proof. The first statement is immediate from Definition 1 and the Markovian nature of P . For the second statement, it suffices to prove this in the case that $\alpha = \sigma \cup u$, for some $u < s$. We note that, using a property of regular conditional probabilities, for μ_σ a.e. x ,

$$\mu_{t|\sigma,x} = E^{\mu_{u|\sigma,x}(\omega)} [\mu_{t|\alpha,v(x,\omega)}], \tag{7}$$

where $v(x, \omega) \in \mathcal{T}_\alpha$, $v(x, \omega)_u = \omega$, $v(x, \omega)_r = x_r$ for all $r \in \sigma$.

We consider \mathfrak{A} to be the set of all finite disjoint partitions $\mathfrak{a} \subset \mathcal{B}(\mathbb{R}^d)$ of \mathbb{R}^d . The expression for the entropy in [26] (Lemma 1.4.3) yields

$$E^{\mu_\sigma(x)} [R(\mu_{t|\sigma,x} || P_{t|s,x_s})] = E^{\mu_\sigma(x)} \left[\sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} \mu_{t|\sigma,x}(A) \log \frac{\mu_{t|\sigma,x}(A)}{P_{t|s,x_s}(A)} \right].$$

Here the summand is considered to be zero if $\mu_{t|\sigma,x}(A) = 0$, and infinite if $\mu_{t|\sigma,x}(A) > 0$ and $P_{t|s,x_s}(A) = 0$. Making use of (7), we find that

$$\begin{aligned} & E^{\mu_\sigma(x)} [R(\mu_{t|\sigma,x} || P_{t|s,x_s})] \\ &= E^{\mu_\sigma(x)} \left[\sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} E^{\mu_{u|\sigma,x}(w)} [\mu_{t|\alpha,v(x,w)}(A)] \log \frac{\mu_{t|\sigma,x}(A)}{P_{t|s,x_s}(A)} \right] \\ &\leq E^{\mu_\sigma(x)} E^{\mu_{u|\sigma,x}(\omega)} \left[\sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} \mu_{t|\alpha,v(x,\omega)}(A) \log \frac{\mu_{t|\sigma,x}(A)}{P_{t|s,x_s}(A)} \right] \\ &= E^{\mu_\alpha(z)} \left[\sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A)}{P_{t|s,z_s}(A)} \right]. \end{aligned}$$

We note that, for μ_α a.e. z , if $\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A) = 0$ in this last expression, then $\mu_{t|\alpha,z}(A) = 0$ and we consider the summand to be zero. To complete the proof of the lemma, it is thus sufficient to prove that for μ_α a.e. z

$$\sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\alpha,z}(A)}{P_{t|s,z_s}(A)} \geq \sup_{\mathfrak{a} \in \mathfrak{A}} \sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A)}{P_{t|s,z_s}(A)}.$$

However, in turn, the above inequality will be true if we can prove that for each partition \mathfrak{a} such that $P_{t|s,z_s}(A) > 0$ and $\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A) > 0$ for all $A \in \mathfrak{a}$,

$$\sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\alpha,z}(A)}{P_{t|s,z_s}(A)} - \sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A)}{P_{t|s,z_s}(A)} \geq 0.$$

The left hand side is equal to $\sum_{A \in \mathfrak{a}} \mu_{t|\alpha,z}(A) \log \frac{\mu_{t|\alpha,z}(A)}{\mu_{t|\sigma,\pi_{\sigma\alpha}z}(A)}$. An application of Jensen’s inequality demonstrates that this is greater than or equal to zero. \square

REMARK. If, contrary to the definition, we briefly consider $\mu_{|[0,t],x}$ to be a probability measure on \mathcal{T} , such that $\mu(A) = 1$ where A is the set of all points y such that $y_s = x_s$ for all $s \leq t$, then it may be seen from the definition of R that

$$R_{\mathcal{F}_T}(\mu_{|[0,t],x} || P_{|[0,t],x}) = R_{\mathcal{F}_t,T}(\mu_{|[0,t],x} || P_{|[0,t],x}) = R_{\mathcal{F}_t,T}(\mu_{|[0,t],x} || P_{t,x_t}). \tag{8}$$

We have also made use of the Markov property of P . This is why our convention, to which we now return, is to consider $\mu_{|[0,t],x}$ to be a probability measure on $(C([t, T]; \mathbb{R}^d), \mathcal{F}_{t,T})$.

This leads us to the following expressions for $R(\mu || P)$.

Lemma 3. Each σ in the supremums below is of the form $\{t_1 < t_2 < \dots < t_{m(\sigma)-1} < t_{m(\sigma)}\}$ for some integer $m(\sigma)$.

$$R(\mu|P) = R_0(\mu|P) + \sum_{j=1}^{m(\sigma)-1} E^{\mu_{[0,t_j]}(x)} \left[R_{\mathcal{F}_{t_j,t_{j+1}}} \left(\mu_{|[0,t_j],x} || P_{|t_j,x_{t_j}} \right) \right], \tag{9}$$

$$R(\mu|P) = R_0(\mu|P) + \sup_{\sigma \in \mathbf{J}_*} \sum_{j=1}^{m(\sigma)-1} E^{\mu_{\sigma;j}(x)} \left[R_{t_{j+1}} \left(\mu_{t_{j+1}|\sigma;j,x} || P_{t_{j+1}|t_j,x_{t_j}} \right) \right], \tag{10}$$

$$E^{\mu_{[0,s]}(x)} \left[R_t \left(\mu_{t|[0,s],x} || P_{t|s,x_s} \right) \right] = \sup_{\sigma \in \mathbf{J}_s} E^{\mu_{\sigma}(y)} \left[R_t \left(\mu_{t|\sigma,y} || P_{t|s,y_s} \right) \right], \tag{11}$$

where in this last expression $0 \leq s < t \leq T$.

Proof. Consider the sub σ -algebra $\mathcal{F}_{0,t_{m(\sigma)-1}}$. We then find, through an application of Lemma 1 and (8), that

$$R(\mu|P) = R_{\mathcal{F}_{0,t_{m(\sigma)-1}}}(\mu|P) + E^{\mu_{[0,t_{m(\sigma)-1}]}(x)} \left[R_{\mathcal{F}_{t_{m(\sigma)-1},t_{m(\sigma)}}} \left(\mu_{|[0,t_{m(\sigma)-1}],x} || P_{|t_{m(\sigma)-1},x_{t_{m(\sigma)-1}}} \right) \right].$$

We may continue inductively to obtain the first identity.

We use Theorem 2 to prove the second identity. It suffices to take the supremum over \mathbf{J}_* , because $R_{\sigma}(\mu|P) \geq R_{\gamma}(\mu|P)$ if $\gamma \subset \sigma$. It thus suffices to prove that

$$R_{\sigma}(\mu|P) = R_0(\mu|P) + \sum_{j=1}^{m(\sigma)-1} E^{\mu_{\sigma;j}(x)} \left[R_{t_{j+1}} \left(\mu_{t_{j+1}|\sigma;j,x} || P_{t_{j+1}|t_j,x_{t_j}} \right) \right]. \tag{12}$$

However, this also follows from repeated application of Lemma 1. To prove the third identity, we firstly note that

$$\begin{aligned} R_{\mathcal{F}_{s;t}}(\mu|P) &= R_0(\mu|P) + \sup_{\sigma \in \mathbf{J}_{s;t}} \sum_{j=1}^{m(\sigma)-1} E^{\mu_{\sigma;j}(x)} \left[R_{t_{j+1}} \left(\mu_{t_{j+1}|\sigma;j,x} || P_{t_{j+1}|t_j,x_{t_j}} \right) \right]. \\ &= \sup_{\sigma \in \mathbf{J}_s} \left\{ R_{\sigma}(\mu|P) + E^{\mu_{\sigma}(x)} \left[R_t \left(\mu_{t|\sigma,x} || P_{t|s,x_s} \right) \right] \right\}. \end{aligned}$$

The proof of this is entirely analogous to that of the second identity, except that it makes use of (5) instead of (4). However, after another application of Lemma 1, we also have that

$$R_{\mathcal{F}_{s;t}}(\mu|P) = R_{\mathcal{F}_{0,s}}(\mu|P) + E^{\mu_{[0,s]}(x)} \left[R_t \left(\mu_{t|[0,s],x} || P_{t|s,x_s} \right) \right].$$

On equating these two different expressions for $R_{\mathcal{F}_{s;t}}(\mu|P)$, we obtain

$$\begin{aligned} E^{\mu_{[0,s]}(x)} \left[R_t \left(\mu_{t|[0,s],x} || P_{t|s,x_s} \right) \right] &= \sup_{\sigma \in \mathbf{J}_s} \left\{ (R_{\sigma}(\mu|P) - R_{\mathcal{F}_{0,s}}(\mu|P)) \right. \\ &\quad \left. + E^{\mu_{\sigma}(x)} \left[R_t \left(\mu_{t|\sigma,x} || P_{t|s,x_s} \right) \right] \right\}. \end{aligned}$$

Let $(\sigma^{(k)}) \subset \mathbf{J}_s$, $\sigma^{(k-1)} \subseteq \sigma^{(k)}$ be such that $\lim_{k \rightarrow \infty} R_{\sigma^{(k)}}(\mu||P) = R_{\mathcal{F}_{0,s}}(\mu||P)$. Such a sequence exists by (4). Similarly, let $(\gamma^{(k)}) \subseteq \mathbf{J}_s$ be a sequence such that $E^{\mu_{\gamma^{(k)}}(x)} [R_t(\mu_{t|\gamma^{(k)},x}||P_{t|s,x_s})]$ is strictly non-decreasing and, as $k \rightarrow \infty$, asymptotically approaches $\sup_{\sigma \in \mathbf{J}_s} E^{\mu_{\sigma}(x)} [R_t(\mu_{t|\sigma,x}||P_{t|s,x_s})]$. Lemma 2 dictates that

$$E^{\mu_{\sigma^{(k)} \cup \gamma^{(k)}}(x)} [R_t(\mu_{t|\sigma^{(k)} \cup \gamma^{(k)},x}||P_{t|s,x_s})]$$

asymptotically approaches the same limit as well. Clearly $\lim_{k \rightarrow \infty} R_{\sigma^{(k)} \cup \gamma^{(k)}}(\mu||P) = R_{\mathcal{F}_{0,s}}(\mu||P)$ because of the identity at the start of Theorem 2. This yields the third identity. \square

4.1. Proof of Theorem 1

In this section we work towards the proof of Theorem 1, making use of some results in [22]. However, we first require some more definitions.

If $K \subset \mathbb{R}^d$ is compact, let \mathcal{D}_K be the set of all $f \in \mathcal{D}$ whose support is contained in K . The corresponding space of real distributions is \mathcal{D}' , and we denote the action of $\theta \in \mathcal{D}'$ by $\langle \theta, f \rangle$. If $\theta \in \mathcal{M}(\mathbb{R}^d)$, then clearly $\langle \theta, f \rangle = \mathbb{E}^\theta[f]$. We let $C_0^{2,1}(\mathbb{R}^d)$ denote the set of all continuous functions, possessing continuous spatial derivatives of first and second order, a continuous time derivative of first order, and of compact support. For $f \in \mathcal{D}$ and $t \in [0, T]$, we define the random variable $\nabla_t f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $(\nabla_t f(y))^i = \sum_{j=1}^d a^{ij}(t, y) \frac{\partial f}{\partial y^j}$ (for $x \in \mathcal{T}$, we may also understand $\nabla_t f(x) := \nabla_t f(x_t)$). Let a_{ij} be the components of the matrix inverse of a^{ij} . For random variables $X, Y : \mathcal{T} \rightarrow \mathbb{R}^d$, we define the inner product $(X, Y)_{t,x} = \sum_{i,j=1}^d X^i(x) Y^j(x) a_{ij}(t, x_t)$, with associated norm $|X|_{t,x}^2 = (X(x), X(x))_{t,x}$. We note that $|\nabla_t f|_{t,x}^2 = \sum_{i,j=1}^d a^{ij}(t, x_t) \frac{\partial f}{\partial z^i}(x_t) \frac{\partial f}{\partial z^j}(x_t)$.

Let \mathfrak{M} be the space of all continuous maps $[0, T] \rightarrow \mathcal{M}(\mathbb{R}^d)$, equipped with the topology of uniform convergence. For $s \in [0, T]$, $\vartheta \in \mathfrak{M}$ and $\nu \in \mathcal{M}(\mathbb{R}^d)$ we define $n(s, \vartheta, \nu) \geq 0$ and such that

$$n(s, \vartheta, \nu)^2 = \sup_{f \in \mathcal{D}} \left\{ \langle \vartheta, f \rangle - \frac{1}{2} E^{\nu(y)} [|\nabla_t f|_{t,y}^2] \right\}. \tag{13}$$

This definition is taken from [22] (Equation (4.7))—we note that n is convex in ϑ . For $\gamma \in \mathcal{M}(\mathcal{T})$, we may naturally write $n(s, \gamma, \nu) := n(s, \omega, \nu)$, where ω is the projection of γ onto \mathfrak{M} , i.e., $\omega(s) = \gamma_s$. It is shown in [22] that this projection is continuous. The following two definitions, lemma and two propositions are all taken (with some small modifications) from [22].

Definition 2. Let I be an interval of the real line. A measure $\mu \in \mathcal{M}(\mathcal{T})$ is called absolutely continuous if for each compact set $K \subset \mathbb{R}^d$ there exists a neighbourhood U of 0 in K and an absolutely continuous function $H_K : I \rightarrow \mathbb{R}$ such that

$$|\mathbb{E}^{\mu_u}[f] - \mathbb{E}^{\mu_v}[f]| \leq |H_K(u) - H_K(v)|,$$

for all $u, v \in I$ and $f \in U_K$.

Lemma 4. [22] (Lemma 4.2) If μ is absolutely continuous over an interval I , then its derivative exists (in the distributional sense) for Lebesgue a.e. $t \in I$. That is, for Lebesgue a.e. $t \in I$, there exists $\dot{\mu}_t \in \mathcal{D}'$ such that for all $f \in \mathcal{D}$

$$\lim_{h \rightarrow 0} \frac{1}{h} (\langle \mu_{t+h}, f \rangle - \langle \mu_t, f \rangle) = \langle \dot{\mu}_t, f \rangle.$$

Definition 3. For $\nu \in \mathcal{M}(C([s, t]; \mathbb{R}^d))$, and $0 \leq s < t \leq T$, let $L^2_{s,t}(\nu)$ be the Hilbert space of all measurable maps $h : [s, t] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ with inner product

$$[h_1, h_2] = \int_s^t \mathbb{E}^{\nu_u(x)} [(h_1(u, x), h_2(u, x))_{u,x}] du.$$

We denote by $L^2_{s,t,\nabla}(\nu)$ the closure in $L^2_{s,t}(\nu)$ of the linear subset generated by maps of the form $(x, u) \rightarrow \nabla_u f$, where $f \in C^{2,1}_0([s, t], \mathbb{R}^d)$. We note that functions in $L^2_{s,t,\nabla}(\nu)$ only need to be defined $du \otimes \nu_u(dx)$ almost everywhere.

Recall that \mathbf{n} is defined in (13), and note that $\langle * \mathcal{L}_t \mu_t, f \rangle := \langle \mu_t, \mathcal{L}_t f \rangle$.

Proposition 1. Assume that $\mu \in \mathcal{M}(C([r, s]; \mathbb{R}^d))$, such that $\mu_r = \delta_y$ for some $y \in \mathbb{R}^d$ and $0 \leq r < s \leq T$. We have that [22] (Equation 4.9 and Lemma 4.8)

$$\int_r^s \mathbf{n}(t, \dot{\mu}_t - * \mathcal{L}_t \mu_t, \mu_t)^2 dt = \sup_{f \in C^{2,1}_0(\mathbb{R}^d)} \left\{ \mathbb{E}^{\mu_s(x)} [f(s, x)] - f(r, y) - \int_r^s \mathbb{E}^{\mu_t(x)} \left[\left(\frac{\partial}{\partial t} + \mathcal{L}_t \right) f(t, x) + \frac{1}{2} |\nabla_t f(t, x)|^2_{t,x} \right] dt \right\}. \quad (14)$$

It clearly suffices to take the supremum over a countable dense subset. Assume now that $\int_r^s \mathbf{n}(t, \dot{\mu}_t - * \mathcal{L}_t \mu_t, \mu_t)^2 dt < \infty$. Then for Lebesgue a.e. t , $\dot{\mu}_t = * \mathcal{K}_t \mu_t$, where [22] (Lemma 4.8(3))

$$\mathcal{K}_t f(\cdot) = \mathcal{L}_t f(\cdot) + \sum_{1 \leq j \leq d} (h^\mu(t, \cdot))^j \frac{\partial f}{\partial x^j}(\cdot), \quad (15)$$

for some $h^\mu \in L^2_{r,s,\nabla}(\mu)$ that satisfies [22] (Lemma 4.8(4))

$$\int_r^s \mathbf{n}(t, \dot{\mu}_t - * \mathcal{L}_t \mu_t, \mu_t)^2 dt = \frac{1}{2} \int_r^s \mathbb{E}^{\mu_t(x)} [|h^\mu(t, x)|^2_{t,x}] dt < \infty. \quad (16)$$

REMARK. We reach (17) from the proof of Lemma 9 in [22] (Eq 4.10). One should note also that in the equation (4.10) of [22] the relative entropy R as $L^{(1)}_\nu$. To reach (18), we also use the equivalence between (4.7) and (4.8) in [22].

Proposition 2. Assume that $\mu \in \mathcal{M}(\mathcal{T})$, such that $\mu_r = \delta_y$ for some $y \in \mathbb{R}^d$ and $0 \leq r < s \leq T$. If $R_{\mathcal{F}_{r,s}}(\mu || P_{[r,y]}) < \infty$, then μ is absolutely continuous on $[r, s]$, and [22] (Lemma 4.9)

$$R_{\mathcal{F}_{r,s}}(\mu || P_{[r,y]}) \geq \int_r^s \mathbf{n}(t, \dot{\mu}_t - * \mathcal{L}_t \mu_t, \mu_t)^2 dt. \quad (17)$$

Here the derivative $\dot{\mu}_t$ is defined in Lemma 4. For all $f \in \mathcal{D}$, [22] (Eq. (4.35))

$$\mathbb{E}^{\mu_s}[f] - \log \mathbb{E}^{P_{s|r,y}}[\exp(f)] \leq \int_r^s \mathbf{n}(t, \dot{\mu}_t - {}^* \mathcal{L}_t \mu_t, \mu_t)^2 dt. \tag{18}$$

We are now ready to prove Theorem 1 (the central result).

Proof. Fix a partition $\sigma = \{t_1, \dots, t_m\}$. We may conclude from (9) and (17) that

$$R(\mu||P) \geq R_0(\mu||P) + \sum_{j=1}^{m-1} E^{\mu_{[0,t_j]}(x)} \int_{t_j}^{t_{j+1}} \mathbf{n}(t, \dot{\mu}_{t|[0,t_j],x} - {}^* \mathcal{L}_t \mu_{t|[0,t_j],x}, \mu_{t|[0,t_j],x})^2 dt. \tag{19}$$

The integrand on the right hand side is measurable with respect to $E^{\mu_{[0,t_j]}(x)}$ due to the equivalent expression (14). We may infer from (18) that

$$\begin{aligned} E^{\mu_{[0,t_j]}(x)} \int_{t_j}^{t_{j+1}} \mathbf{n}(t, \dot{\mu}_{t|[0,t_j],x} - {}^* \mathcal{L}_t \mu_{t|[0,t_j],x}, \mu_{t|[0,t_j],x})^2 dt \\ \geq \mathbb{E}^{\mu_{[0,t_j]}(x)} \left[\sup_{f \in \mathcal{D}} \left\{ \mathbb{E}^{\mu_{t_{j+1}}|[0,t_j],x}[f] - \log \mathbb{E}^{P_{t_{j+1}}|t_j,x t_j}[\exp(f)] \right\} \right] \\ = \mathbb{E}^{\mu_{[0,t_j]}(x)} \left[\sup_{f \in C_b(\mathbb{R}^d)} \left\{ \mathbb{E}^{\mu_{t_{j+1}}|[0,t_j],x}[f] - \log \mathbb{E}^{P_{t_{j+1}}|t_j,x t_j}[\exp(f)] \right\} \right]. \tag{20} \end{aligned}$$

This last step follows by noting that if $\nu \in \mathcal{M}(\mathbb{R}^d)$, and $f \in C_b(\mathbb{R}^d)$, and the expectation of f with respect to ν is finite, then there exists a series $(K_n) \subset \mathbb{R}^d$ of compact sets such that

$$\int_{\mathbb{R}^d} f(x) d\nu(x) = \lim_{n \rightarrow \infty} \int_{K_n} f(x) d\nu(x).$$

In turn, for each n there exist $(f_n^{(m)}) \in \mathcal{D}_{K_n}$ such that we may write

$$\int_{K_n} f(x) d\nu(x) = \lim_{m \rightarrow \infty} \int_{K_n} f_n^{(m)}(x) d\nu(x).$$

This allows us to conclude that the two supremums are the same. The last expression in (20) is merely

$$E^{\mu_{[0,t_j]}(x)} \left[R_{t_{j+1}} \left(\mu_{t_{j+1}}|[0,t_j],x || P_{t_{j+1}}|t_j,x t_j \right) \right].$$

By (11), this is greater than or equal to

$$E^{\mu_{\sigma;j}(y)} \left[R_{t_{j+1}} \left(\mu_{t_{j+1}}|\sigma;j,y || P_{t_{j+1}}|t_j,y t_j \right) \right].$$

We thus obtain the theorem using (10). \square

5. Some Corollaries

We state some corollaries of Theorem 1. In the course of this section we make progressively stronger assumptions on the nature of μ , culminating in the elegant expression for $R(\mu||P)$ when μ is a solution of a martingale problem. We finish by comparing our work with that of [19,20].

Corollary 1. *Suppose that $\mu \in \mathcal{M}(\mathcal{T})$ and $R(\mu||P) < \infty$. Then for all s and μ a.e. x , $\mu|_{[0,s],x}$ is absolutely continuous over $[s, T]$. For each $s \in [0, T]$ and μ a.e. $x \in \mathcal{T}$, for Lebesgue a.e. $t \geq s$*

$$\dot{\mu}_{t|[0,s],x} = {}^* \mathcal{K}_{t|s,x}^\mu \mu_{t|[0,s],x} \tag{21}$$

where for some $h_{s,x}^\mu \in L^2_{s,T,\nabla}(\mu|_{[0,s],x})$

$$\mathcal{K}_{t|s,x}^\mu f(y) = \mathcal{L}_t f(y) + \sum_{j=1}^d h_{s,x}^{\mu,j}(t, y) \frac{\partial f}{\partial y^j}(y). \tag{22}$$

Furthermore,

$$R(\mu||P) = R_0(\mu||P) + \frac{1}{2} \sup_{\sigma \in \mathcal{J}_*} \int_0^T E^{\mu(w)} E^{\mu_{t|[0,\underline{\sigma}(t)],w}(z)} \left[\left| h_{\underline{\sigma}(t),w}^\mu(t, z) \right|_{t,z}^2 \right] dt. \tag{23}$$

For any dense countable subset $\mathcal{Q}_{0,T}$ of $[0, T]$, there exists a series of partitions $\sigma^{(n)} \subset \sigma^{(n+1)} \in \mathcal{Q}_{0,T}$, such that as $n \rightarrow \infty$, $|\sigma^{(n)}| \rightarrow 0$, and

$$R(\mu||P) = R_0(\mu||P) + \frac{1}{2} \lim_{n \rightarrow \infty} \int_0^T E^{\mu(w)} E^{\mu_{t|[0,\underline{\sigma}^{(n)}(t)],w}(z)} \left[\left| h_{\underline{\sigma}^{(n)}(t),w}^\mu(t, z) \right|_{t,z}^2 \right] dt. \tag{24}$$

REMARK. *It is not immediately clear that we may simplify (23) further (barring further assumptions). The reason for this is that we only know that*

$E^{\mu_{t|[0,\underline{\sigma}(t)],w}(z)} \left[\left| h_{\underline{\sigma}(t),w}^\mu(t, z) \right|_{t,z}^2 \right]$ *is measurable (as a function of w), but it has not been proven that $h_{\underline{\sigma}(t),w}^\mu(t, z)$ is measurable (as a function of w).*

Proof. Let $\sigma = \{0 = t_1, \dots, t_m = T\}$ be an arbitrary partition. For all $j < m$, we find from Lemma 3 that $R_{\mathcal{F}_{t_j, t_{j+1}}}(\mu|_{[0,t_j],x}||P|_{t_j, x_{t_j}}) < \infty$ for $\mu|_{[0,t_j]}$ a.e. $x \in C([0, t_j]; \mathbb{R}^d)$. We thus find that, for all such x , $\mu|_{[0,t_j],x}$ is absolutely continuous on $[t_j, t_{j+1}]$ from Proposition 2. We are then able to obtain (21) and (22) from Propositions 1 and 2. From (2), (16) and (21) we find that

$$R(\mu||P) = R_0(\mu||P) + \frac{1}{2} \sup_{\sigma \in \mathcal{J}_*} E^{\mu(x)} \int_0^T E^{\mu_{t|[0,\underline{\sigma}(t)],x}(z)} \left[\left| h_{\underline{\sigma}(t),x}^\mu(t, z) \right|_{t,z}^2 \right] dt. \tag{25}$$

The above integral must be finite (since we are assuming $R(\mu||P)$ is finite). Furthermore $E^{\mu_{t|[0,\underline{\sigma}(t)],x}(z)} \left[\left| h_{\underline{\sigma}(t),x}^\mu(t, z) \right|_{t,z}^2 \right]$ is (t, x) measurable as a consequence of the equivalent form (14). This allows us to apply Fubini's theorem to obtain (23). The last statement on the sequence of maximising partitions follows from Theorem 2. \square

Corollary 2. Suppose that $R(\mu||P) < \infty$. Suppose that for all $s \in \mathcal{Q}_{0,T}$ (any countable, dense subset of $[0, T]$), for μ a.e. x and Lebesgue a.e. t , $h_{s,x}^\mu(t, x_t) = E^{\mu|_{[0,s],x;t,x_t}(w)} h^\mu(t, w)$ for some progressively measurable random variable $h^\mu : [0, T] \times \mathcal{T} \rightarrow \mathbb{R}^d$. Then

$$R(\mu||P) = R_0(\mu||P) + \frac{1}{2} \int_0^T E^{\mu(w)} \left[|h^\mu(t, w)|_{t,w_t}^2 \right] dt.$$

Proof. Let $\mathcal{G}^{s,x;t,y}$ be the sub σ -algebra consisting of all $B \in \mathcal{B}(\mathcal{T})$ such that for all $w \in B$, $w_r = x_r$ for all $r \leq s$ and $w_t = y$. Thus $h_{s,x}^\mu(t, y) = E^{\mu|_{[0,s],x;t,y}(w)} h^\mu(t, w) = E^\mu [h^\mu(t, \cdot) | \mathcal{G}^{s,x;t,y}]$. By [27] (Corollary 2.4), since $\cap_{s < t} \mathcal{G}^{s,x;t,x_t} = \mathcal{G}^{t,x;t,x_t}$ (restricting to $s \in \mathcal{Q}_{0,T}$), for μ a.e. x ,

$$\lim_{s \rightarrow t^-} E^{\mu|_{[0,s],x;t,x_t}(w)} h^\mu(t, w) = h^\mu(t, x), \tag{26}$$

where $s \in \mathcal{Q}_{0,T}$. By the properties of the regular conditional probability, we find from (24) that

$$R(\mu||P) = R_0(\mu||P) + \frac{1}{2} \lim_{n \rightarrow \infty} \int_0^T E^{\mu(w)} \left[\left| E^{\mu|_{[0,\sigma^{(n)}(t)],w;t,w_t}(v)} [h^\mu(t, v)] \right|_{t,w_t}^2 \right] dt. \tag{27}$$

By assumption, the above limit is finite. Thus by Fatou’s lemma, and using the properties of the regular conditional probability,

$$R(\mu||P) \geq R_0(\mu||P) + \frac{1}{2} \int_0^T E^{\mu(w)} \left[\underline{\lim}_{n \rightarrow \infty} \left| E^{\mu|_{[0,\sigma^{(n)}(t)],w;t,w_t}(v)} [h^\mu(t, v)] \right|_{t,w_t}^2 \right] dt.$$

Through use of (26),

$$R(\mu||P) \geq R_0(\mu||P) + \frac{1}{2} \int_0^T E^{\mu(w)} \left[|h^\mu(t, w)|_{t,w_t}^2 \right] dt.$$

Conversely, through an application of Jensen’s inequality to (27)

$$R(\mu||P) \leq R_0(\mu||P) + \frac{1}{2} \lim_{n \rightarrow \infty} \int_0^T E^{\mu(w)} \left[E^{\mu|_{[0,\sigma^{(n)}(t)],w;t,w_t}(v)} \left[|h^\mu(t, v)|_{t,w_t}^2 \right] \right] dt.$$

A property of the regular conditional probability yields

$$R(\mu||P) \leq R_0(\mu||P) + \frac{1}{2} \int_0^T E^{\mu(w)} \left[|h^\mu(t, w)|_{t,w_t}^2 \right] dt.$$

□

REMARK. The condition in the above corollary is satisfied when μ is a solution to a martingale problem—see Lemma 5.

We may further simplify the expression in Theorem 1 when μ is a solution to the following martingale problem. Let $\{e^{jk}, e^j\}$ be progressively measurable functions $[0, T] \times \mathcal{T} \rightarrow \mathbb{R}$. We suppose that

$c^{jk} = c^{kj}$. For all $1 \leq j, k \leq d$, $c^{jk}(t, x)$ and $e^j(t, x)$ are assumed to be bounded for $x \in L$ (where L is compact) and all $t \in [0, T]$. For $f \in C_0^2(\mathbb{R}^d)$ and $x \in \mathcal{T}$, let

$$\mathcal{M}_u(f)(x) = \sum_{1 \leq j, k \leq d} c^{jk}(u, x) \frac{\partial^2 f}{\partial y^j \partial y^k}(x_u) + \sum_{1 \leq j \leq d} e^j(u, x) \frac{\partial f}{\partial y^j}(x_u).$$

We assume that for all such f , the following is a continuous martingale (relative to the canonical filtration) under μ

$$f(X_t) - f(X_0) - \int_0^t \mathcal{M}_u f(X) du. \tag{28}$$

The law governing X_0 is stipulated to be $\nu \in \mathcal{M}(\mathbb{R}^d)$.

From now on we switch from our earlier convention and we consider $\mu_{|[0,s],x}$ to be a measure on \mathcal{T} such that, for μ a.e. $x \in \mathcal{T}$, $\mu_{|[0,s],x}(A_{s,x}) = 1$, where $A_{s,x}$ is the set of all $X \in \mathcal{T}$ satisfying $X_t = x_t$ for all $0 \leq t \leq s$. This is a property of a regular conditional probability (see Theorem 3.18 in [23]). Similarly, $\mu_{|s,x;t,y}$ is considered to be a measure on \mathcal{T} such that for μ a.e. $x \in \mathcal{T}$, $\mu_{|s,x;t,y}(B_{s,x;t,y}) = 1$, where $B_{s,x;t,y}$ is the set of all $X \in A_{s,x}$ such that $X_t = y$. We may apply Fubini's Theorem (since f is compactly supported and bounded) to (28) to find that

$$\langle \mu_{t|[0,s],x}, f \rangle - f(x_s) = \int_s^t \mathbb{E}^{\mu_{|[0,s],x}} [\mathcal{M}_u f] du.$$

This ensures that $\mu_{|[0,s]}$ is absolutely continuous over $[s, T]$, and that

$$\langle \dot{\mu}_{t|[0,s],x}, f \rangle = \mathbb{E}^{\mu_{|[0,s],x}} [\mathcal{M}_t f]. \tag{29}$$

Lemma 5. *If $R(\mu||P) < \infty$ then for Lebesgue a.e. $t \in [0, T]$ and μ a.e. $x \in \mathcal{T}$,*

$$a(t, x_t) = c(t, x). \tag{30}$$

If $R(\mu||P) < \infty$ then

$$R(\mu||P) = R(\nu||\mu_T) + \frac{1}{2} \mathbb{E}^{\mu(x)} \left[\int_0^T |b(s, x_s) - e(s, x)|_{s,x_s}^2 ds \right]. \tag{31}$$

Proof. It follows from $R(\mu||P) < \infty$, (21) and (22) that for all s and μ a.e. x , for Lebesgue a.e. $t \geq s$

$$E^{\mu_{|s,x;t,x_t}} [c(t, \cdot)] = a(t, x_t). \tag{32}$$

Let us take a countable dense subset $\mathcal{Q}_{0,T}$ of $[0, T]$. There thus exists a null set $\mathcal{N} \subseteq [0, T]$ such that for every $s \in \mathcal{Q}_{0,T}$, μ a.e. x and every $t \notin \mathcal{N}$ the above equation holds. We may therefore conclude (30) using [27] (Corollary 2.4) and taking $s \rightarrow t^-$. From (29), we observe that for all $s \in [0, T]$ and μ a.e. x , for Lebesgue a.e. t

$$h_{s,x}^\mu(t, x_t) = E^{\mu_{|[0,s],x;t,x_t}} [e(t, \cdot)].$$

Equation (31) thus follows from Corollary 2. \square

5.1. Comparison of our Results to Those of Fischer et al. [19,20]

We have already noted in the introduction that one may infer a variational representation of the relative entropy from [19,20] by assuming that the coefficients of the underlying stochastic process are independent of the empirical measure in these papers. The assumptions in [20] on the underlying process P are both more general and more restrictive than ours. His assumptions are more general insofar as the coefficients of the SDE may depend on the past history of the process and the diffusion coefficient is allowed to be degenerate. However, our assumptions are more general insofar as we only require P to be the unique (in the sense of probability law) weak solution of the SDE, whereas [20] requires P to be the unique strong solution of the SDE. Of course when both sets of assumptions are satisfied, one may infer that the expressions for the relative entropy are identical.

6. Proof of Theorem 2

The following is an alternative proof to that of [25] (Theorem 6.6) employing the theory of Large Deviations. The fact that, if $\alpha \subseteq \sigma$, then $R_\alpha(\mu||P) \leq R_\sigma(\mu||P)$, follows from Lemma 1. We prove the first expression (4) in the case $s = 0, t = T$ (the proof of the second identity (5) is analogous).

Definition 4. A series of probability laws Γ^N on some topological space Ω equipped with its Borelian σ -algebra is said to satisfy a strong Large Deviation Principle with rate function $I : \Omega \rightarrow \mathbb{R}$ if for all open sets O ,

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log \Gamma^N(O) \geq - \inf_{x \in O} I(x)$$

and for all closed sets F

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \Gamma^N(F) \leq - \inf_{x \in F} I(x).$$

If furthermore the set $\{x : I(x) \leq \alpha\}$ is compact for all $\alpha \geq 0$, we say that I is a good rate function.

We define the following empirical measures.

Definition 5. For $x \in \mathcal{T}^N, y \in \mathcal{T}_\sigma^N$, let

$$\hat{\mu}^N(x) = \frac{1}{N} \sum_{1 \leq j \leq N} \delta_{x_j} \in \mathcal{M}(\mathcal{T}), \quad \hat{\mu}_\sigma^N(y) = \frac{1}{N} \sum_{1 \leq j \leq N} \delta_{y_j} \in \mathcal{M}(\mathcal{T}_\sigma).$$

Clearly $\hat{\mu}_\sigma^N(x_\sigma) = \pi_\sigma(\hat{\mu}^N(x))$. The image law $P^{\otimes N} \circ (\hat{\mu}^N)^{-1}$ is denoted by $\Pi_{s,t}^N \in \mathcal{M}(\mathcal{M}(\mathcal{T}))$. Similarly, for $\sigma \in \mathbf{J}$, the image law of $P_\sigma^{\otimes N} \circ (\hat{\mu}_\sigma^N)^{-1}$ on $\mathcal{M}(\mathcal{T}_\sigma)$ is denoted by $\Pi_\sigma^N \in \mathcal{M}(\mathcal{M}(\mathcal{T}_\sigma))$. Since \mathcal{T} and \mathcal{T}_σ are Polish spaces, we have by Sanov’s theorem (see Theorem 6.2.10 in [14]) that Π^N satisfies a strong Large Deviation Principle with good rate function $R(\cdot||P)$. Similarly, Π_σ^N satisfies a strong Large Deviation Principle on $\mathcal{M}(\mathcal{T}_\sigma)$ with good rate function $R_{\mathcal{F}_\sigma}(\cdot||P)$.

We now define the projective limit $\underline{\mathcal{M}}(\mathcal{T})$. If $\alpha, \gamma \in \mathbf{J}, \alpha \subset \gamma$, then we may define the projection $\pi_{\alpha\gamma}^{\mathcal{M}} : \mathcal{M}(\mathcal{T}_\gamma) \rightarrow \mathcal{M}(\mathcal{T}_\alpha)$ as $\pi_{\alpha\gamma}^{\mathcal{M}}(\xi) := \xi \circ \pi_{\alpha\gamma}^{-1}$. An element of $\underline{\mathcal{M}}(\mathcal{T})$ is then a member $\otimes_\sigma \zeta(\sigma)$ of the Cartesian product $\otimes_{\sigma \in \mathbf{J}} \mathcal{M}(\mathcal{T}_\sigma)$ satisfying the consistency condition $\pi_{\alpha\gamma}^{\mathcal{M}}(\zeta(\gamma)) = \zeta(\alpha)$ for all $\alpha \subset \gamma$.

The topology on $\underline{\mathcal{M}}(\mathcal{T})$ is the minimal topology necessary for the natural projection $\underline{\mathcal{M}}(\mathcal{T}) \rightarrow \mathcal{M}(\mathcal{T}_\alpha)$ to be continuous for all $\alpha \in \mathbf{J}$. That is, it is generated by open sets of the form

$$A_{\gamma,O} = \{\otimes_\sigma \zeta(\sigma) \in \underline{\mathcal{M}}(\mathcal{T}) : \zeta(\gamma) \in O\}, \quad (33)$$

for some $\gamma \in \mathbf{J}$ and open O (with respect to the weak topology of $\mathcal{M}(\mathcal{T}_\gamma)$).

We may continuously embed $\mathcal{M}(\mathcal{T})$ into the projective limit $\underline{\mathcal{M}}(\mathcal{T})$ of its marginals, letting ι denote this embedding. That is, for any $\sigma \in \mathbf{J}$, $(\iota(\mu))(\sigma) = \mu_\sigma$. We note that ι is continuous because $\iota^{-1}(A_{\gamma,O})$ is open in $\mathcal{M}(\mathcal{T})$, for all $A_{\gamma,O}$ of the form in (33). We equip $\underline{\mathcal{M}}(\mathcal{T})$ with the Borelian σ -algebra generated by this topology. The embedding ι is measurable with respect to this σ -algebra because the topology of $\mathcal{M}(\mathcal{T})$ has a countable base. The embedding induces the image laws $(\Pi^N \circ \iota^{-1})$ on $\mathcal{M}(\underline{\mathcal{M}}(\mathcal{T}))$. For $\sigma \in \mathbf{J}$, it may be seen that $\Pi_\sigma^N = \Pi^N \circ \iota^{-1} \circ (\pi_\sigma^{\mathcal{M}})^{-1} \in \mathcal{M}(\mathcal{M}(\mathcal{T}_\sigma))$, where $\pi_\sigma^{\mathcal{M}}(\otimes_\alpha \mu(\alpha)) = \mu(\sigma)$.

It follows from [22] (Thm 3.3) that $\Pi^N \circ \iota^{-1}$ satisfies a Large Deviation Principle with rate function $\sup_{\sigma \in \mathbf{J}} R_\sigma(\mu||P)$. However, we note that ι is 1-1, because any two measures $\mu, \nu \in \mathcal{M}(\mathcal{T})$ such that $\mu_\sigma = \nu_\sigma$ for all $\sigma \in \mathbf{J}$ must be equal. Furthermore, ι is continuous. Because of Sanov's theorem, (Π^N) is exponentially tight (see Defn 1.2.17, Exercise 1.2.19 in [14] for a definition of exponential tightness and proof of this statement). These facts mean that we may apply the inverse contraction principle [14] (Thm 4.2.4) to infer that Π^N satisfies a Large Deviation Principle with the rate function $\sup_{\sigma \in \mathbf{J}} R_\sigma(\mu||P)$. Since rate functions are unique [14] (Lemma 4.1.4), we obtain the first identity in conjunction with Sanov's theorem. The second identity (5) follows similarly. We may repeat the argument above, while restricting to $\sigma \subset \mathcal{Q}_{s,t}$. We obtain the same conclusion because the σ -algebra generated by $(\mathcal{F}_\sigma)_{\sigma \subset \mathcal{Q}_{s,t}}$ is the same as $\mathcal{F}_{s,t}$. The last identity follows from the fact that, if $\alpha \subseteq \sigma$, then $R_\alpha(\mu||P) \leq R_\sigma(\mu||P)$.

Acknowledgments

This work was supported by INRIA FRM, ERC-NERVI number 227747, European Union Project # FP7-269921 (BrainScales), and Mathemacis # FP7-ICT-2011.9.7

Author Contributions

Both authors contributed to all the article. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Plastino, A.; Miller, H.; Plastino, A. Minimum Kullback entropy approach to the Fokker-Planck equation. *Phys. Rev. E* **1997**, *56*, 3927–3934.
2. Desvillettes, L.; Villani, C. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems. Part 1: The Linear Fokker-Planck Equation. *Comm. Pure Appl. Math.* **2001**, *54*, 1–42.

3. Yu, S.; Mehta, P. The Kullback-Leibler Rate Metric for Comparing Dynamical Systems. In Proceedings of Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, China, 16–18 December 2009; pp. 8363–8368.
4. Georgiou, T.T.; Lindquist, A. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Inf. Theory* **2003**, *49*, 2910–2917.
5. Frittelli, M. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Math. Finance* **2000**, *10*, 39–52.
6. Grandits, P.; Rheinlander, T. On the minimal entropy martingale measure. *Ann. Probab.* **2002**, *30*, 1003–1038.
7. Miyahara, Y. Minimal Relative Entropy Martingale Measure of Birth and Death Process. In *Discussion Papers in Economics*; Nagoya City University: Nagoya, Japan, 2000.
8. Miyahara, Y. On the Minimal Entropy Martingale Measures for Geometric Lévy Processes. In *Discussion Papers in Economics*; Nagoya City University: Nagoya, Japan, 2000.
9. Ustunel, A.S. Entropy, invertibility and variational calculus of the adapted shifts on Wiener Space. *J. Funct. Anal.* **2009**, *257*, 3655–3689.
10. Lassalle, R. Invertibility of adapted perturbations of the identity on abstract Wiener space. *J. Funct. Anal.* **2012**, *262*, 2734–2776.
11. Akaike, H. Likelihood of a model and information criteria. *J. Econometrics* **1981**, *16*, 3–14.
12. Do, M.; Vetterli, M. Wavelet-based Texture Retrieval Using Generalized Gaussian Density and Kullback–Leibler Distance. *IEEE Trans. Image Process.* **2002**, *11*, 146–158.
13. Bozdogan, H. Akaike’s Information Criterion and Recent Developments in Information Complexity. *J. Math. Psychol.* **2000**, *44*, 62–91.
14. Dembo, A.; Zeitouni, O. *Large Deviations Techniques*, 2nd ed.; Springer: Berlin, Germany, 1997.
15. Ben-Arous, G.; Guionnet, A. Large deviations for Langevin spin glass dynamics. *Probab. Theor. Relat. Field.* **1995**, *102*, 455–509.
16. Moynot, O.; Samuelides, M. Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probab. Theor. Relat. Field.* **2002**, *123*, 41–75.
17. Faugeras, O.; MacLaurin, J. A large deviation principle for networks of rate neurons with correlated synaptic weights. **2013**, arXiv:1302.1029.
18. Faugeras, O.; MacLaurin, J. Large Deviations of an Ergodic Synchronous Neural Network with Learning. **2014**, arXiv:1404.0732v3[math.PR].
19. Budhiraja, A.; Dupuis, P.; Fischer, M. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.* **2012**, *40*, 74–102.
20. Fischer, M. On the form of the large deviation rate function for the empirical measures of weakly interacting systems. *Bernoulli* **2014**, *20*, 1765–1801.
21. Baladron, J.; Fasoli, D.; Faugeras, O.; Touboul, J. Mean field description of and propagation of chaos in recurrent multipopulation networks of Hodgkin-Huxley and Fitzhugh-Nagumo neurons. **2011**, arXiv:1110.4294.
22. Dawson, D.A.; Gärtner, J. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **1987**, *20*, 247–308.

23. Karatzas, I.; Shreve, S.E. *Brownian Motion and Stochastic Calculus*, 2nd ed.; Graduate Texts in Mathematics, Volume 113; Springer-Verlag: New York, NY, USA, 1991.
24. Donsker, M.; Varadhan, S. Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, IV. *Comm. Pure Appl. Math.* **1983**, *XXXVI*, 183–212.
25. Xanh, N.X.; Zessin, H. Ergodic Theorems for Spatial Processes. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **1979**, *48*, 133–158.
26. Dupuis, P.; Ellis, R.S. *A Weak Convergence Approach to the Theory of Large Deviations*; John Wiley & Sons: London, UK, 1997.
27. Revuz, D.; Yor, M. *Continuous Martingales and Brownian Motion*, 2nd ed.; Springer-Verlag: Berlin, Germany, 1991.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).