On the form of the large deviation rate function for the empirical measures of weakly interacting systems*

Markus Fischer†

Revised March 29, 2013

Abstract

A basic result of large deviations theory is Sanov's theorem, which states that the sequence of empirical measures of independent and identically distributed samples satisfies the large deviation principle with rate function given by relative entropy with respect to the common distribution. Large deviation principles for the empirical measures are also known to hold for broad classes of weakly interacting systems. When the interaction through the empirical measure corresponds to an absolutely continuous change of measure, the rate function can be expressed as relative entropy of a distribution with respect to the law of the McKean-Vlasov limit with measure-variable frozen at that distribution. We discuss situations, beyond that of tilted distributions, in which a large deviation principle holds with rate function in relative entropy form.

2000 AMS subject classifications: 60F10, 60K35, 94A17.

Key words and phrases: large deviations, empirical measure, particle system, mean field interaction, Laplace principle, relative entropy, Wiener measure.

1 Introduction

Weakly interacting systems are families of particle systems whose components, for each fixed number N of particles, are statistically indistinguishable and interact only through the empirical measure of the N-particle system.

*The author is grateful to Paolo Dai Pra for helpful comments, questions, and discussions. Thanks to referee, funding...

†Department of Mathematics, University of Padua, via Trieste 63, 35121 Padova, Italy.

The study of weakly interacting systems originates in statistical mechanics and kinetic theory; in this context, they are often referred to as mean field systems.

The joint law of the random variables describing the states of the N-particle system of a weakly interacting system is invariant under permutations of components, hence determined by the distribution of the associated empirical measure. For large classes of weakly interacting systems, the law of large numbers is known to hold, that is, the sequence of N-particle empirical measures converges to a deterministic probability measure as N tends to infinity. The limit measure can often be characterized in terms of a limit equation, which, by extrapolation from the important case of Markovian systems, is called McKean-Vlasov equation [cf. McKean, 1966]. As with the classical law of large numbers, different kinds of deviations of the prelimit quantities (the N-particle empirical measures) from the limit quantity (the McKean-Vlasov distribution) can be studied. Here we are interested in large deviations.

Large deviations for the empirical measures of weakly interacting systems, especially Markovian systems, have been the object of a number of works. The large deviation principle is usually obtained by transferring Sanov's theorem, which gives the large deviation principle for the empirical measures of independent and identically distributed samples, through an absolutely continuous change of measure. This approach works when the effect of the interaction through the empirical measure corresponds to a change of measure which is absolutely continuous with respect to some fixed reference distribution of product form. Sanov's theorem can then be transferred using Varadhan's lemma. In the case of Markovian dynamics, such a change-of-measure argument yields the large deviation principle on path space; see Léonard [1995a] for non-degenerate jump diffusions, Dai Pra and den Hollander [1996] for a model of Brownian particles in a potential field and random environment, and Del Moral and Guionnet [1998] for a class of discrete-time Markov processes. An extension of Varadhan's lemma tailored to the change of measure needed for empirical measures is given in Del Moral and Zajic [2003] and applied to a variety of non-degenerate weakly interacting systems. The large deviation rate function in all those cases can be written in relative entropy form, that is, expressed as relative entropy of a distribution with respect to the law of the McKean-Vlasov limit with measure-variable frozen at that distribution; cf. Remark 3.2 below.

In the case of Markovian dynamics, the large deviation principle on path space can be taken as the first step in deriving the large deviation principle for the empirical processes; cf. Léonard [1995a] or Feng [1994a,b]. In Dawson and Gärtner [1987], the large deviation principle for the empirical processes of weakly interacting Itô diffusions with non-degenerate and measure-independent diffusion matrix is established in Freidlin-Wentzell form starting from a process level representation of the rate function for non-interacting Itô diffusions. The large deviation principle for interacting diffusions is then derived by time discretization, local freezing of the measure variable and an absolutely continuous change of measure with respect to the resulting product distributions. A similar strategy is applied in Djehiche and Kaj [1995] to a class of pure jump processes.

A different approach is taken in the early work of Tanaka [1984], where the contraction principle is employed to derive the large deviation principle on path space for the special case of Itô diffusions with identity diffusion matrix. The contraction mapping in this case is actually a bijection. Using the invariance of relative entropy under bi-measurable bijections, the rate function is shown to be of relative entropy form. In Léonard [1995b], the large deviation upper bound, not the full principle, is derived by variational methods using Laplace functionals for certain pure jump Markov processes that do not allow for an absolutely continuous change of measure. In Budhiraja et al. [2012], the path space Laplace principle for weakly interacting Itô processes with measure-dependent and possibly degenerate diffusion matrix is established based on a variational representation of Laplace functionals, weak convergence methods and ideas from stochastic optimal control. The rate function is given in variational form.

The aim of this paper is to show that the large deviation principle holds with rate function in relative entropy form also for weakly interacting systems that do not allow for an absolutely continuous change of measure with respect to product distributions. The large deviation principle in that form is a natural generalization of Sanov's theorem. Two classes of systems will be discussed: noise-based systems to which the contraction principle is applicable, and systems described by weakly interacting Itô processes.

Remark 1.1. The random variables representing the states of the particles will be assumed to take values in a Polish space. The space of probability measures over a Polish space will be equipped, for simplicity, with the standard topology of weak convergence. Continuity of a functional with respect to the topology of weak convergence might be a rather restrictive condition. This restriction can be alleviated by considering the space of probability measures that satisfy an integrability condition (for instance, finite moments of a certain order), equipped with the topology of weak(-star) convergence with respect to the corresponding class of continuous functions [for instance, Section 2b) in Léonard, 1995a]. The results presented below can be adapted to this more general situation.

The rest of this paper is organized as follows. In Section 2, we collect basic definitions and results of the theory of large deviations in the context of Polish spaces that will be used in the sequel; standard references for our purposes are Dembo and Zeitouni [1998] and Dupuis and Ellis [1997]. In Section 3, we introduce a toy model of discrete-time weakly interacting systems to illustrate the use of Varadhan's lemma, which in turn yields, at least formally, a representation of the rate function in relative entropy form. In Section 4, a class of weakly interacting systems is presented to which the contraction principle is applicable but not necessarily the usual change-of-measure technique. The large deviation rate function is shown to be of the desired form thanks to a contraction property of relative entropy. In Section 5, we discuss the case of weakly interacting Itô diffusions with measure-dependent and possibly degenerate diffusion matrix studied in Budhiraja et al. [2012]. The variational form of the Laplace principle rate function established there is shown to be expressible in relative entropy form. As a by-product, one obtains a variational representation of relative entropy with respect to Wiener measure. The Appendix contains two results regarding relative entropy: the contraction property mentioned above, which extends a well-known invariance property (Appendix A), and a direct proof of the variational representation of relative entropy with respect to Wiener measure (Appendix B). In Appendix C, easily verifiable conditions entailing the hypotheses of the Laplace principle of Section 5 are given.

2 Basic definitions and results

Let $\mathcal{S}$ be a Polish space (i.e., a separable topological space metrizable with a complete metric). Denote by $\mathcal{B}(\mathcal{S})$ the σ-algebra of Borel subsets of $\mathcal{S}$ and by $\mathcal{P}(\mathcal{S})$ the space of probability measures on $\mathcal{B}(\mathcal{S})$ equipped with the topology of weak convergence. For $\mu, \nu \in \mathcal{P}(\mathcal{S})$, let $R(\nu\|\mu)$ denote the *relative entropy* of ν with respect to μ, that is,

$$R(\nu\|\mu) \doteq \begin{cases} \int_{\mathcal{S}} \log\left(\frac{d\nu}{d\mu}(x)\right)\nu(dx) & \text{if } \nu \text{ absolutely continuous w.r.t. } \mu, \\ \infty & \text{else.} \end{cases}$$

Relative entropy is well defined as a $[0,\infty]$-valued function, it is lower semicontinuous as a function of both variables, and $R(\nu\|\mu) = 0$ if and only if $\nu = \mu$.

Let $(\xi^n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{S}$-valued random variables. A *rate function* on $\mathcal{S}$ is a lower semicontinuous function $\mathcal{S} \to [0,\infty]$. Let I be a rate function on $\mathcal{S}$. By lower semicontinuity, the sublevel sets of I, i.e., the sets $I^{-1}([0,c])$ for $c \in [0,\infty)$, are closed. A rate function is said to be *good* if its sublevel sets are compact.

Definition 2.1. The sequence $(\xi^n)_{n\in\mathbb{N}}$ satisfies the *large deviation principle* with rate function I if for all $B \in \mathcal{B}(\mathcal{S})$,

$$
\begin{aligned}
-\inf_{x\in B^\circ} I(x) &\le \liminf_{n\to\infty} \frac{1}{n}\log \mathbf{P}\left\{\xi^n \in B\right\} \\
&\le \limsup_{n\to\infty} \frac{1}{n}\log \mathbf{P}\left\{\xi^n \in B\right\} \le -\inf_{x\in \mathrm{cl}(B)} I(x),
\end{aligned}
$$

where $\mathrm{cl}(B)$ denotes the closure and B° the interior of B.

Definition 2.2. The sequence (ξ^n) satisfies the *Laplace principle* with rate function I iff for all $G \in \mathbf{C}_b(\mathcal{S})$,

$$
\lim_{n\to\infty} -\frac{1}{n}\log \mathbf{E}\left[\exp\left(-n\cdot G(\xi^n)\right)\right] = \inf_{x\in\mathcal{S}} \left\{I(x) + G(x)\right\},
$$

where $\boldsymbol{C}_b(\mathcal{S})$ denotes the space of all bounded continuous functions $\mathcal{S} \to \mathbb{R}$.

Clearly, the large deviation principle (or Laplace principle) is a distributional property. The rate function of a large deviation principle is unique; see, for instance, Lemma 4.1.4 in Dembo and Zeitouni [1998, p. 117]. The large deviation principle holds with a good rate function if and only if the Laplace principle holds with a good rate function, and the rate function is the same; see, for instance, Theorem 4.4.13 in Dembo and Zeitouni [1998, p. 146].

The fact that, for good rate functions, the large deviation principle implies the Laplace principle is a consequence of Varadhan's integral lemma; see Theorem 3.4 in Varadhan [1966]. Another consequence of Varadhan's lemma is the first of the following two basic transfer results, given here as Theorem 2.1; cf. Theorem II.7.2 in Ellis [1985, p. 52].

Theorem 2.1 (Change of measure, Varadhan)**.** *Let (ξ^n) be a sequence of $\mathcal{S}$-valued random variables such that (ξ^n) satisfies the large deviation principle with good rate function I. Let $(\tilde{\xi}^n)_{n\in\mathbb{N}}$ be a second sequence of $\mathcal{S}$-valued random variables. Suppose that, for every $n \in \mathbb{N}$, $\mathrm{Law}(\tilde{\xi}^n)$ is absolutely continuous with respect to $\mathrm{Law}(\xi^n)$ with density*

$$
\frac{d\,\mathrm{Law}(\tilde{\xi}^n)}{d\,\mathrm{Law}(\xi^n)}(x) = \exp\left(n\cdot F(x)\right), \quad x \in \mathcal{S},
$$

where $F\colon \mathcal{S} \to \mathbb{R}$ *is continuous and such that*

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[\mathbf{1}_{[L,\infty)}(F(\xi^n)) \cdot \exp\left(n \cdot F(\xi^n)\right)\right] = -\infty.$$

Then $(\tilde{\xi}^n)_{n\in\mathbb{N}}$ *satisfies the large deviation principle with good rate function* $I - F$*.*

The second basic transfer result is the contraction principle, given here as Theorem 2.2; see, for instance, Theorem 4.2.1 and Remark (c) in Dembo and Zeitouni [1998, pp. 126-127].

Theorem 2.2 (Contraction principle)**.** *Let* (ξ^n) *be a sequence of* $\mathcal{S}$*-valued random variables such that* (ξ^n) *satisfies the large deviation principle with good rate function* I*. Let* $\psi\colon \mathcal{S} \to \mathcal{Y}$ *be a measurable function,* $\mathcal{Y}$ *a Polish space. If* ψ *is continuous on* $I^{-1}([0,\infty))$*, then* $(\psi(\xi^n))$ *satisfies the large deviation principle with good rate function*

$$J(y) \doteq \inf_{x\in\psi^{-1}(y)} I(x), \quad y \in \mathcal{Y},$$

where $\inf \emptyset = \infty$ *by convention.*

Let $X_1, X_2, \dots$ be $\mathcal{S}$-valued independent and identically distributed random variables with common distribution $\mu \in \mathcal{P}(\mathcal{S})$ defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For $n \in \mathbb{N}$, let μ^n be the *empirical measure* of $X_1, \dots, X_n$, that is,

$$\mu^n(\omega) \doteq \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}, \quad \omega \in \Omega,$$

where δ_x denotes the Dirac measure concentrated in $x \in \mathcal{S}$. Sanov's theorem gives the large deviation principle for $(\mu^n)_{n\in\mathbb{N}}$ in terms of relative entropy. For a proof, see, for instance, Section 6.2 in Dembo and Zeitouni [1998, pp. 260-266] or Chapter 2 in Dupuis and Ellis [1997, 39-52]. Recall that $\mathcal{P}(\mathcal{S})$ is equipped with the topology of weak convergence of measures.

Theorem 2.3 (Sanov)**.** *The sequence* $(\mu^n)_{n\in\mathbb{N}}$ *of* $\mathcal{P}(\mathcal{S})$*-valued random variables satisfies the large deviation principle with good rate function*

$$I(\theta) \doteq R\left(\theta \| \mu\right), \quad \theta \in \mathcal{P}(\mathcal{S}).$$

We are interested in analogous results for the empirical measures of weakly interacting systems. For $N \in \mathbb{N}$, let $X_1^N, \dots, X_N^N$ be $\mathcal{S}$-valued random variables defined on some probability space $(\Omega_N, \mathcal{F}_N, \mathbf{P}_N)$. Denote by μ^N the empirical measure of $X_1^N, \dots, X_N^N$.

Definition 2.3. The triangular array $(X_i^N)_{N\in\mathbb{N}, i\in\{1,\ldots,N\}}$ is called a *weakly interacting system* if the following hold:

(i) for each $N \in \mathbb{N}$, $X_1^N, \ldots, X_N^N$ is a finite exchangeable sequence;

(ii) the family $(\mu^N)_{N\in\mathbb{N}}$ of $\mathcal{P}(\mathcal{S})$-valued random variables is tight.

Recall that a finite sequence $Y_1, \ldots, Y_N$ of random variables with values in a common measurable space is called *exchangeable* if its joint distribution is invariant under permutations of the components, that is, $\mathrm{Law}(Y_1, \ldots, Y_N) = \mathrm{Law}(Y_{\sigma(1)}, \ldots, Y_{\sigma(N)})$ for every permutation σ of $\{1, \ldots, N\}$. A weakly interacting system (X_i^N) is said to satisfy the *law of large numbers* if there exists $\mu \in \mathcal{P}(\mathcal{S})$ such that (μ^N) converges to μ in distribution or, equivalently, $(\mathbf{P}_N \circ (\mu^N)^{-1})$ converges weakly to δ_μ. Weakly interacting systems are sometimes called *mean field systems*. In the situation of Theorem 2.3, setting $X_i^N \doteq X_i$, $N \in \mathbb{N}$, $i \in \{1, \ldots, N\}$, defines a weakly interacting system that satisfies the law of large numbers, the limit measure being the common sample distribution.

3 A toy model and the desired form of the rate function

For $N \in \mathbb{N}$, let $(Y_i^N(t))_{i\in\{1,\ldots,N\}, t\in\{0,1\}}$ be an independent family of standard normal real random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $b: \mathbb{R} \to \mathbb{R}$ be measurable; below we will assume b to be bounded and continuous. Define real random variables $X_1^N(t), \ldots, X_N^N(t)$, $t \in \{0, 1\}$, by

$$(3.1) \quad X_i^N(0) \doteq Y_i^N(0), \quad X_i^N(1) \doteq X_i^N(0) + \frac{1}{N}\sum_{j=1}^N b\left(X_j^N(0)\right) + Y_i^N(1).$$

We may interpret the variables $X_i^N(t)$ as the states of the components of an N-particle system at times $t \in \{0, 1\}$. This toy model can be obtained as the first two steps in a discrete time version of a system of weakly interacting Itô diffusions; cf. the discussion following Example 4.3 below. Let μ^N be the empirical measure of the N-particle system on "path space," that is,

$$\mu_\omega^N \doteq \frac{1}{N}\sum_{i=1}^N \delta_{X_i^N(\omega)} = \frac{1}{N}\sum_{i=1}^N \delta_{(X_i^N(0,\omega), X_i^N(1,\omega))}, \quad \omega \in \Omega.$$

Notice that the components of X^N are identically distributed and interact only through μ^N since

$$\frac{1}{N}\sum_{j=1}^N b\left(X_j^N(0)\right) = \int_{\mathbb{R}^2} b(x) d\mu^N(x, \tilde{x})$$

and the variables $Y_i^N(t)$ are independent and identically distributed. The sequence $X_1^N, \ldots, X_N^N$ of $\mathbb{R}^2$-valued random variables is exchangeable.

Let λ^N denote the empirical measure of $Y^N = (Y_1^N, \ldots, Y_N^N)$. By Sanov's theorem, $(\lambda^N)_{N\in\mathbb{N}}$ satisfies the large deviation principle with good rate function $R(.\|\gamma_0)$, where γ_0 is the bivariate standard normal distribution. Following the usual way of deriving the large deviation principle, we observe that, for every $N \in \mathbb{N}$, the law of μ^N is absolutely continuous with respect to the law of λ^N. To see this, set, for $\boldsymbol{y}, \tilde{\boldsymbol{y}} \in \mathbb{R}^N$, $\theta \in \mathcal{P}(\mathbb{R}^2)$,

$$\nu^N_{(\boldsymbol{y},\tilde{\boldsymbol{y}})} \doteq \frac{1}{N}\sum_{i=1}^N \delta_{(y_i,\tilde{y}_i)}, \quad m_b(\theta) \doteq \int b(x)d\theta(x,\tilde{x}),$$
$$\nu^N_{\boldsymbol{y}} \doteq \frac{1}{N}\sum_{i=1}^N \delta_{(y_i)}, \qquad \boldsymbol{m}_b(\boldsymbol{y}) \doteq \left(\int b(x)\nu^N_{\boldsymbol{y}}(dx), \ldots, \int b(x)\nu^N_{\boldsymbol{y}}(dx)\right)^{\mathsf{T}}.$$

Define functions $f\colon \mathcal{P}(\mathbb{R}^2)\times\mathbb{R}\to\mathbb{R}$ and $F\colon \mathcal{P}(\mathbb{R}^2)\to[-\infty,\infty)$ according to

$$f(\theta,(y,\tilde{y})) \doteq (y + m_b(\theta))\cdot\tilde{y} - \frac{1}{2}\left|y+m_b(\theta)\right|^2,$$
$$F(\theta) \doteq \begin{cases} \int_{\mathbb{R}^2} f(\theta,(y,\tilde{y}))d\theta(y,\tilde{y}) & \text{if } f(\theta,.\,) \text{ is } \theta\text{-integrable,} \\ -\infty & \text{otherwise.}\end{cases}$$

Then the law of X^N is absolutely continuous with respect to the law of Y^N with density given by

$$\begin{aligned}\frac{d\,\mathrm{Law}(X^N)}{d\,\mathrm{Law}(Y^N)}(\boldsymbol{y},\tilde{\boldsymbol{y}}) &= \exp\left(\langle \boldsymbol{y}+\boldsymbol{m}_b(\boldsymbol{y}),\tilde{\boldsymbol{y}}\rangle - \frac{1}{2}|\boldsymbol{y}+\boldsymbol{m}_b(\boldsymbol{y})|^2\right) \\ &= \exp\left(N\cdot F\left(\mu^N_{(\boldsymbol{y},\tilde{\boldsymbol{y}})}\right)\right).\end{aligned} \tag{3.2}$$

Since $\mu^N = \nu^N_{(X^N(0),X^N(1))}$ and $\lambda^N = \nu^N_{(Y^N(0),Y^N(1))}$, it follows from Equation (3.2) that

$$\frac{d\,\mathrm{Law}(\mu^N)}{d\,\mathrm{Law}(\lambda^N)}(\theta) = \exp\left(N\cdot F(\theta)\right), \quad \theta\in\mathcal{P}(\mathbb{R}^2). \tag{3.3}$$

The densities given by Equation (3.3) are of the form required by Theorem 2.1, the change of measure version of Varadhan's lemma. Assume from now on that b is bounded and continuous. Then F is upper semicontinuous and the tail condition in Theorem 2.1 is satisfied. However, F is discontinuous at any $\theta\in\mathcal{P}(\mathbb{R}^2)$ such that $F(\theta) > -\infty$. Indeed, let η be the univariate standard Cauchy distribution and set $\theta_n \doteq (1-\frac{1}{n})\theta + \frac{1}{n}\delta_0\otimes\eta$, $n\in\mathbb{N}$. Then $\theta_n\to\theta$ weakly, while $F(\theta_n) = -\infty$ for all n.

Although Theorem 2.1 cannot be applied directly, an approximation argument based on Varadhan's lemma could be used to show (cf. Remark 3.1 below) that the sequence of empirical measures $(\mu^N)_{N\in\mathbb{N}}$ satisfies the large deviation principle with good rate function

$$I(\theta) \doteq R(\theta\|\gamma_0) - F(\theta), \quad \theta \in \mathcal{P}(\mathbb{R}^2). \tag{3.4}$$

The function I in (3.4) can be rewritten in terms of relative entropy as follows. Define a mapping $\psi\colon \mathcal{P}(R^2) \times \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\psi(\theta, (y, \tilde{y})) \doteq (y, y + m_b(\theta) + \tilde{y}). \tag{3.5}$$

For $\theta \in \mathcal{P}(\mathbb{R}^2)$, let $\Psi_{\gamma_0}(\theta)$ be the image measure of γ_0 under $\psi(\theta, .)$. Then $\Psi_{\gamma_0}(\theta)$ is equivalent to γ_0 with density given by

$$\frac{d\Psi_{\gamma_0}(\theta)}{d\gamma_0}(y, \tilde{y}) = \exp\left(f(\theta, (y, \tilde{y}))\right).$$

If θ is not absolutely continuous with respect to $\Psi_{\gamma_0}(\theta)$, then $R(\theta\|\Psi_{\gamma_0}(\theta)) = \infty = R(\theta\|\gamma_0)$. If θ is absolutely continuous with respect to $\Psi_{\gamma_0}(\theta)$, then

$$\begin{aligned} R\big(\theta\|\Psi_{\gamma_0}(\theta)\big) &= \int \log\left(\frac{d\theta}{d\Psi_{\gamma_0}(\theta)}\right) d\theta \\ &= \int \log\left(\frac{d\theta}{\gamma_0}\right) d\theta - \int \log\left(\frac{d\Psi_{\gamma_0}(\gamma_0)}{d\gamma_0}\right) d\theta \\ &= R\big(\theta\|\gamma_0\big) - F(\theta). \end{aligned}$$

Consequently, for all $\theta \in \mathcal{P}(\mathbb{R}^2)$,

$$I(\theta) = R\big(\theta\|\Psi_{\gamma_0}(\theta)\big). \tag{3.6}$$

Notice that $\Psi_{\gamma_0}(\theta)$ is the law of a one-particle system with measure variable frozen at θ; $\Psi_{\gamma_0}(\theta)$ can also be interpreted as the solution of the McKean-Vlasov equation for the toy model with measure variable frozen at θ.

Remark 3.1. A version of Varadhan's lemma (or Theorem 2.1) that allows to rigorously derive the large deviation principle for (μ^N) with rate function in relative entropy form is provided by Lemma 1.1 in Del Moral and Zajic [2003]. Observe that the density of $\mathrm{Law}(X^N)$ may be computed with respect to product measures different from $\mathrm{Law}(Y^N) = \otimes^N \gamma_0$. A natural alternative is the product $\otimes^N \Psi_{\gamma_0}(\mu_*)$, where μ_* is the (unique) solution of the fixed point equation $\mu = \Psi_{\gamma_0}(\mu)$; μ_* can be seen as the McKean-Vlasov distribution of the toy model. We do not give the details here. The results of Section 4, based on different arguments, will imply that $(\mu^N)_{N\in\mathbb{N}}$ satisfies the large deviation principle with good rate function I as given by Equation (3.6); see Example 4.1 below.

Remark 3.2. Equation (3.6) gives the desired form of the rate function in terms of relative entropy. More generally, suppose that $\Psi\colon \mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{S})$ is continuous, where $\mathcal{S}$ is a Polish space. Then the function

$$J(\theta) \doteq R\big(\theta \| \Psi(\theta)\big), \quad \theta \in \mathcal{P}(\mathcal{S}),$$

is lower semicontinuous with values in $[0, \infty]$, hence a rate function, and it is in relative entropy form. The lower semicontinuity of J follows from the lower semicontinuity of relative entropy jointly in both its arguments and the continuity of Ψ. If, in addition, $\mathrm{range}(\Psi) \doteq \{\Psi(\theta) : \theta \in \mathcal{P}(\mathcal{S})\}$ is compact in $\mathcal{P}(\mathcal{S})$, then the sublevel sets of J are compact and J is a good rate function. Indeed, compactness of $\mathrm{range}(\Psi)$ implies tightness, and the compactness of the sublevel sets of J, which are closed by lower semicontinuity, follows as in the proof of Lemma 1.4.3(c) in Dupuis and Ellis [1997, pp. 29-31].

4 Noise-based systems

Let $\mathcal{X}$, $\mathcal{Y}$ be Polish spaces. For $N \in \mathbb{N}$, let $X_1^N, \ldots, X_N^N$ be $\mathcal{X}$-valued random variables defined on some probability space $(\Omega_N, \mathcal{F}_N, \mathbf{P}_N)$. Denote by μ^N the empirical measure of $X_1^N, \ldots, X_N^N$. We suppose that there are a probability measure $\gamma_0 \in \mathcal{P}(\mathcal{Y})$ and a Borel measurable mapping $\psi\colon \mathcal{P}(\mathcal{X}) \times \mathcal{Y} \to \mathcal{X}$ such that the following representation for the triangular array $(X_i^N)_{i \in \{1,\ldots,N\}, N \in \mathbb{N}}$ holds: For each $N \in \mathbb{N}$, there is a sequence $Y_1^N, \ldots, Y_N^N$ of independent and identically distributed $\mathcal{Y}$-valued random variables on $(\Omega_N, \mathcal{F}_N, \mathbf{P}_N)$ with common distribution γ_0 such that for all $i \in \{1, \ldots, N\}$,

$$X_i^N(\omega) = \psi\big(\mu_\omega^N, Y_i^N(\omega)\big), \quad \mathbf{P}_N\text{-almost all } \omega \in \Omega_N. \tag{4.1}$$

The above representation entails by symmetry that, for N fixed, the sequence $X_1^N, \ldots, X_N^N$ is exchangeable. Representation (4.1) also implies that μ^N satisfies the equation

$$\mu^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\psi(\mu^N, Y_i^N)} \quad \mathbf{P}_N\text{-almost surely.} \tag{4.2}$$

In order to describe the limit behavior of the sequence of empirical measures $(\mu^N)_{N \in \mathbb{N}}$, define a mapping $\Psi\colon \mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ by

$$(\gamma, \mu) \mapsto \Psi_\gamma(\mu) \doteq \gamma \circ \psi^{-1}(\mu, .). \tag{4.3}$$

Thus $\Psi_\gamma(\mu)$ is the image measure of γ under the mapping $\mathcal{Y} \ni y \mapsto \psi(\mu, y)$. Equivalently, $\Psi_\gamma(\mu) = \mathrm{Law}(\psi(\mu, Y))$ with Y any $\mathcal{Y}$-valued random variable with distribution γ. Limit points of $(\mu^N)_{N\in\mathbb{N}}$ will be described in terms of solutions to the fixed point equation

$$\mu = \Psi_{\gamma}(\mu). \tag{4.4}$$

Assume that there is a Borel measurable set $\mathcal{D} \subset \mathcal{P}(\mathcal{Y})$ such that the following properties hold:

(A1) Equation (4.4) has a unique fixed point $\mu_*(\gamma)$ for every $\gamma \in \mathcal{D}$, and the mapping $\mathcal{D} \ni \gamma \mapsto \mu_*(\gamma) \in \mathcal{P}(\mathcal{X})$ is Borel measurable.

(A2) For all $N \in \mathbb{N}$,

$$\otimes^N \gamma_0 \left\{ (y_1, \ldots, y_N) \in \mathcal{Y}^N : \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i} \in \mathcal{D} \right\} = 1.$$

(A3) If $\gamma \in \mathcal{P}(\mathcal{Y})$ is such that $R(\gamma\|\gamma_0) < \infty$, then $\gamma \in \mathcal{D}$ and $\mu_{*|\mathcal{D}}$ is continuous at γ .

Assumption (A2) implies that Equation (4.4) possesses a unique solution for almost all (with respect to products of γ_0) probability measures of empirical measure form. Such probability measures are therefore in the domain of definition of the mapping $\mu_{*|\mathcal{D}}$. According to Assumption (A3), also all probability measures γ with finite γ_0-relative entropy are in the domain of definition of $\mu_{*|\mathcal{D}}$, which is continuous at any such γ in the topology of weak convergence.

Theorem 4.1. *Grant (A1) – (A3). Then the sequence $(\mu^N)_{N\in\mathbb{N}}$ satisfies the large deviation principle with good rate function $I\colon \mathcal{P}(\mathcal{X}) \to [0,\infty]$ given by*

$$I(\eta) = \inf_{\gamma\in\mathcal{D}:\mu_*(\gamma)=\eta} R\big(\gamma\|\gamma_0\big),$$

where $\inf\emptyset = \infty$ *by convention.*

Proof. The assertion follows from Sanov's theorem and the contraction principle. To see this, let λ^N denote the empirical measure of $Y_1^N, \ldots, Y_N^N$. Then for $\mathbf{P}_N$-almost all $\omega \in \Omega_N$,

$$\mu_\omega^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\psi(\mu_\omega^N, Y_i^N(\omega))} = \lambda_\omega^N \circ \psi^{-1}\big(\mu_\omega^N, .\,\big) = \Psi_{\lambda_\omega^N}\big(\mu_\omega^N\big). \tag{4.5}$$

Thus $\mu^N = \Psi_{\lambda^N}(\mu^N)$ with probability one. For $\mathbf{P}_N$-almost all $\omega \in \Omega_N$, $\lambda_\omega^N \in \mathcal{D}$ by Assumption (A2) and, by uniqueness according to (A1), $\mu_*(\lambda_\omega^N) = \mu_\omega^N$.

By Theorem 2.3 (Sanov), $(\lambda^N)_{N\in\mathbb{N}}$ satisfies the large deviation principle with good rate function $R(.\|\gamma_0)$. By Assumption (A3), $\mu_*(.)$ is defined and continuous on $\{\gamma \in \mathcal{P}(\mathcal{Y}) : R(\gamma\|\gamma_0) < \infty\}$. Theorem 2.2 (contraction principle) therefore applies, and it follows that $(\mu_*(\lambda^N))_{N\in\mathbb{N}}$, hence $(\mu^N)_{N\in\mathbb{N}}$, satisfies the large deviation principle with good rate function

$$\mathcal{P}(\mathcal{X}) \ni \eta \mapsto \inf_{\gamma\in\mathcal{D}:\mu_*(\gamma)=\eta} R\bigl(\gamma\|\gamma_0\bigr).$$

□

The rate function of Theorem 4.1 can be expressed in relative entropy form as in Remark 3.2. The key observation is the contraction property of relative entropy established in Lemma A.1 in the appendix.

Corollary 4.2. *Let I be the rate function of Theorem 4.1. Then for all $\eta \in \mathcal{P}(\mathcal{X})$,*

$$I(\eta) = R\bigl(\eta\|\Psi_{\gamma_0}(\eta)\bigr).$$

Proof. Let $\eta \in \mathcal{P}(\mathcal{X})$. The mapping $\mathcal{P}(\mathcal{Y}) \ni \gamma \mapsto \Psi_\gamma(\eta) \in \mathcal{P}(\mathcal{X})$ is Borel measurable. Since $\{\gamma \in \mathcal{P}(\mathcal{X}) : R(\gamma\|\gamma_0) < \infty\} \subset \mathcal{D}$ and $\inf \emptyset = \infty$,

$$\inf_{\gamma\in\mathcal{D}:\mu_*(\gamma)=\eta} R\bigl(\gamma\|\gamma_0\bigr) = \inf_{\gamma\in\mathcal{D}:\Psi_\gamma(\eta)=\eta} R\bigl(\gamma\|\gamma_0\bigr) = \inf_{\gamma\in\mathcal{P}(\mathcal{Y}):\Psi_\gamma(\eta)=\eta} R\bigl(\gamma\|\gamma_0\bigr).$$

By Lemma A.1, it follows that

$$\inf_{\gamma\in\mathcal{P}(\mathcal{Y}):\Psi_\gamma(\eta)=\eta} R\bigl(\gamma\|\gamma_0\bigr) = R\bigl(\eta\|\Psi_{\gamma_0}(\eta)\bigr).$$

□

Example 4.1. Consider the toy model of Section 3. Suppose that $b \in \boldsymbol{C}_b(\mathbb{R})$. Then $\theta \mapsto m_b(\theta) \doteq \int b(x)d\theta(x,\tilde{x})$ is bounded and continuous as a mapping $\mathcal{P}(\mathbb{R}^2) \to \mathbb{R}$. Observe that $m_b(\theta)$ depends only on the first marginal of θ. Set $\mathcal{X} \doteq \mathbb{R}^2$, $\mathcal{Y} \doteq \mathbb{R}^2$, let γ_0 be the bivariate standard normal distribution, and define $\psi\colon \mathcal{P}(\mathbb{R}^2)\times\mathbb{R}^2 \to \mathbb{R}^2$ according to (3.5). Recalling Equation (3.1), one sees that the toy model satisfies representation (4.1). Based on ψ, define Ψ according to (4.3). Given any $\gamma \in \mathcal{P}(\mathbb{R}^2)$, the mapping $\mu \mapsto \Psi_\gamma(\mu)$ possesses a unique fixed point $\mu_*(\gamma)$. To see this, suppose that $\theta \in \mathcal{P}(\mathbb{R}^2)$ is a fixed point, that is, $\theta = \Psi_\gamma(\theta) = \gamma \circ \psi(\theta, .)^{-1}$. Let $X = (X(0), X(1))$, $Y = (Y(0), Y(1))$ be two $\mathbb{R}^2$-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with distribution θ and γ, respectively. By the fixed point property, $\mathrm{Law}(X) = \mathrm{Law}(\psi(\theta, Y))$. By definition of ψ, $\mathrm{Law}(X(0)) = \mathrm{Law}(Y(0))$. Since $m_b(\theta)$ depends on $\theta = \mathrm{Law}(X)$ only through its first marginal, which is equal to $\text{Law}(X(0)) = \text{Law}(Y(0))$, we have $m_b(\theta) = m_b(\gamma)$. It follows that, for all $B_0, B_1 \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{P}\left(X(1) \in B_1 | X(0) \in B_0\right) = \mathbf{P}\left(Y(0) + m_b(\gamma) + Y(1) \in B_1 | Y(0) \in B_0\right).$$

This determines the conditional distribution of $X(1)$ given $X(0)$ and, since $\text{Law}(X(0)) = \text{Law}(Y(0))$, also the joint law of $X(0)$ and $X(1)$. In fact, $\text{Law}(X) = \Psi_\gamma(\gamma)$. Consequently, $\mu_*(\gamma) \doteq \Psi_\gamma(\gamma)$ is the unique solution of Equation (4.4). By the extended mapping theorem for weak convergence [Theorem 5.5 in Billingsley, 1968, p. 34] and since $m_b(.) \in \boldsymbol{C}_b(\mathcal{P}(\mathbb{R}^2))$, the mapping $(\gamma, \mu) \mapsto \Psi_\gamma(\gamma)$ is continuous as a function $\mathcal{P}(\mathbb{R}^2) \times \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$. It follows that the mapping $\gamma \mapsto \mu_*(\gamma) = \Psi_\gamma(\gamma)$ is continuous. Assumptions (A1)$-$(A3) are therefore satisfied with the choice $\mathcal{D} \doteq \mathcal{P}(\mathbb{R}^2)$. By Corollary 4.2, the sequence of empirical measures (μ^N) for the toy model satisfies the large deviation principle with good rate function I given by (3.6). Observe that the distribution γ_0 need not be the bivariate standard normal distribution for the large deviation principle to hold; it can be any probability measure on $\mathcal{B}(\mathbb{R}^2)$.

Example 4.2. Consider the following variation on the toy model of Section 3 and Example 4.1. For $N \in \mathbb{N}$, let $(Y_i^N(t))_{i\in\{1,\dots,N\},t\in\{0,1\}}$ be independent standard normal real random variables as above. Denote by γ_0 the bivariate standard normal distribution and let $B \in \mathcal{B}(\mathbb{R})$ be a γ_0-continuity set, that is, $\gamma_0(\partial(B \times \mathbb{R})) = 0$, where $\partial(B \times \mathbb{R})$ is the boundary of $B \times \mathbb{R}$. Define real random variables $X_1^N(t), \dots, X_N^N(t)$, $t \in \{0, 1\}$, by

$$X_i^N(0) \doteq Y_i^N(0), \quad X_i^N(1) \doteq X_i^N(0) + \frac{1}{N}\left(\sum_{j=1}^{N} \mathbf{1}_B(X_j^N(0))\right) \cdot Y_i^N(1).$$

For this new toy model, define $\psi \colon \mathcal{P}(\mathbb{R}^2) \times \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\psi(\mu, (y, \tilde{y})) \doteq (y, y + \mu(B \times \mathbb{R}) \cdot \tilde{y}).$$

With this choice of ψ, representation (4.1) holds and ψ is measurable as composition of measurable maps since $\mu \mapsto \mu(B \times \mathbb{R})$ is measurable with respect to the Borel σ-algebra induced by the topology of weak convergence. Based on ψ, define Ψ according to (4.3). As in Example 4.1, one checks that the fixed point equation (4.4) possesses a unique solution $\mu_*(\gamma) \doteq \Psi_\gamma(\gamma)$ for every $\gamma \in \mathcal{P}(\mathbb{R}^2)$. However, if $\partial(B \times \mathbb{R}) \neq \emptyset$, then $\mu_*(.)$ is not continuous on $\mathcal{P}(\mathbb{R}^2)$. On the other hand, if $\gamma \in \mathcal{P}(\mathbb{R}^2)$ is such that $R(\gamma \| \gamma_0) < \infty$, then γ is absolutely continuous with respect to γ_0, so that $B \times \mathbb{R}$ is also a γ-continuity set. By the extended mapping theorem, it follows that $\mu_*(.)$ is continuous at any such γ. Assumptions (A1) – (A3) are therefore satisfied, again with the choice $\mathcal{D} \doteq \mathcal{P}(\mathbb{R}^2)$, and Corollary 4.2 yields the large deviation principle. In this example, if $\gamma_0(B \times \mathbb{R}) < 1$, then the distribution of μ^N, the empirical measure of $X_1^N, \ldots, X_N^N$, is not absolutely continuous with respect to λ^N, the empirical measure of $Y_1^N, \ldots, Y_N^N$. Indeed, in this case, the event $\{\nu^N_{(\boldsymbol{y},\boldsymbol{y})} : \boldsymbol{y} \in \mathbb{R}^N\} \subset \mathcal{P}(\mathbb{R}^2)$, where $\nu^N_{(\boldsymbol{y},\boldsymbol{y})}$ is defined as in Section 3, has strictly positive probability with respect to $\mathbf{P} \circ (\mu^N)^{-1}$, while it has probability zero with respect to $\mathbf{P} \circ (\lambda^N)^{-1}$.

Example 4.3 (Discrete time systems). Let $T \in \mathbb{N}$. Let $\mathcal{X}_0$, $\mathcal{Y}_0$ be Polish spaces, and let $\mathcal{X}$, $\mathcal{Y}$ be the Polish product spaces $\mathcal{X} \doteq (\mathcal{X}_0)^{T+1}$ and $\mathcal{Y} \doteq (\mathcal{Y}_0)^{T+1}$, respectively. Let

$$\varphi_0 \colon \mathcal{Y}_0 \to \mathcal{X}_0, \qquad \varphi \colon \{1, \ldots, T\} \times \mathcal{X}_0 \times \mathcal{P}(\mathcal{X}_0) \times \mathcal{Y}_0 \to \mathcal{X}_0$$

be measurable maps. Let $\gamma_0 \in \mathcal{P}(\mathcal{Y})$ and, for $N \in \mathbb{N}$, let $Y_1^N, \ldots, Y_N^N$ be independent and identically distributed $\mathcal{Y}$-valued random variables defined on some probability space $(\Omega_N, \mathcal{F}_N, \mathbf{P}_N)$ with common distribution γ_0. Write $Y_i^N = (Y_i^N(t))_{t \in \{0,\ldots,T\}}$ and define $\mathcal{X}$-valued random variables $X_1^N, \ldots, X_N^N$ with $X_i^N = (X_i^N(t))_{t \in \{0,\ldots,T\}}$ recursively by

$$\begin{aligned} X_i^N(0) &\doteq \varphi_0\big(Y_i^N(0)\big), \\ X_i^N(t+1) &\doteq \varphi\big(t+1, X_i^N(t), \mu^N(t), Y_i^N(t+1)\big), \quad t \in \{0, \ldots, T-1\}, \end{aligned} \tag{4.6}$$

where $\mu^N(t) \doteq \frac{1}{N} \sum_{i=1}^N \delta_{X_i^N(t)}$ is the empirical measure of $X_1^N, \ldots, X_N^N$ at marginal (or time) t. In analogy with (4.6), define $\psi : \mathcal{P}(\mathcal{X}) \times \mathcal{Y} \to \mathcal{X}$ according to $(\mu, y) = (\mu, (y_0, \ldots, y_T)) \mapsto \psi(\mu, y) \doteq x$ with $x = (x_0, \ldots, x_T)$ given by

$$\begin{aligned} x_0 &\doteq \varphi_0(y_0), \\ x_{t+1} &\doteq \varphi\big(t+1, x_t, \mu(t), y_{t+1}\big), \ t \in \{0, \ldots, T-1\}, \end{aligned} \tag{4.7}$$

where $\mu(t)$ is the marginal of $\mu \in \mathcal{P}(\mathcal{X})$ at time t. Then ψ is measurable as a composition of measurable maps, and representation (4.1) holds. Based on ψ, define Ψ according to (4.3). Using the recursive structure of (4.6) and the components of ψ according to (4.7), one checks that the fixed point equation (4.4) has a unique solution $\mu_*(\gamma)$ given any $\gamma \in \mathcal{D} \doteq \mathcal{P}(\mathcal{Y})$. To be more precise, define functions $\bar{\varphi}_t \colon \mathcal{P}(\mathcal{X}_0)^t \times \mathcal{Y} \to \mathcal{X}_0$, $t \in \{0, \ldots, T\}$, recursively by

$$\begin{aligned} \bar{\varphi}_0(y) &\doteq \varphi_0(y_0), \\ \bar{\varphi}_t\big((\alpha_0, \ldots, \alpha_{t-1}), y\big) &\doteq \varphi\big(t, \bar{\varphi}_{t-1}((\alpha_0, \ldots, \alpha_{t-2}), y), \alpha_{t-1}, y_t\big). \end{aligned} \tag{4.8}$$

Notice that $\bar{\varphi}_t$ depends on $y = (y_0, \ldots, y_T) \in \mathcal{Y}$ only through $(y_0, \ldots, y_t)$. Given $\gamma \in \mathcal{P}(\mathcal{Y})$, recursively define probability measures $\alpha_t(\gamma) \in \mathcal{P}(\mathcal{X}_0)$, $t \in \{0, \ldots, T\}$, according to

$$
\begin{aligned}
&\alpha_0(\gamma) \doteq \gamma \circ \bar{\varphi}_0^{-1}, \\
&\alpha_t(\gamma) \doteq \gamma \circ \bar{\varphi}_t\big((\alpha_0(\gamma), \ldots, \alpha_{t-1}(\gamma)), .\big)^{-1}, \quad t \in \{1, \ldots, T\}.
\end{aligned} \tag{4.9}
$$

The mapping $\mathcal{P}(\mathcal{Y}) \ni \gamma \mapsto \alpha_t(\gamma) \in \mathcal{P}(\mathcal{X}_0)$ is measurable for every $t \in \{0, \ldots, T\}$. Define $\Phi \colon \mathcal{P}(\mathcal{X}_0)^T \times \mathcal{Y} \to \mathcal{X}$ by

$$
\Phi\big((\alpha_0, \ldots, \alpha_{T-1}), y\big) \doteq \big(\bar{\varphi}_0(y), \bar{\varphi}_1(\alpha_0, y), \ldots, \bar{\varphi}_T((\alpha_0, \ldots, \alpha_{T-1}), y)\big). \tag{4.10}
$$

Then the mapping $\gamma \mapsto \gamma \circ \Phi((\alpha_0(\gamma), \ldots, \alpha_{T-1}(\gamma)), .)^{-1}$ is measurable and provides the unique fixed point of Equation (4.4) with noise distribution $\gamma \in \mathcal{P}(\mathcal{Y})$. In fact,

$$
\mu_*(\gamma) = \gamma \circ \Phi\big((\alpha_0(\gamma), \ldots, \alpha_{T-1}(\gamma)), .\big)^{-1}. \tag{4.11}
$$

Writing $\mu_*(t, \gamma)$ for the t-marginal of $\mu_*(\gamma)$, we also notice that

$$
\mu_*(t, \gamma) = \alpha_t(\gamma) = \gamma \circ \bar{\varphi}_t\big((\alpha_0(\gamma), \ldots, \alpha_{t-1}(\gamma)), .\big)^{-1}.
$$

If φ_0, φ are continuous maps, then it follows by (4.11) and the extended mapping theorem that $\mu_*(.)$ is continuous on $\mathcal{D} = \mathcal{P}(\mathcal{Y})$, and Corollary 4.2 yields the large deviation principle for the sequence of "path space" empirical measures $(\mu^N)_{N\in\mathbb{N}}$.

Example 4.3 comprises a large class of discrete time weakly interacting systems. The sequence of $(\mathcal{X}_0)^N$-valued random variables $X^N(0), \ldots, X^N(T)$ given by (4.6) enjoys the Markov property if the $(\mathcal{Y}_0)^N$-valued random variables $Y^N(0), \ldots, Y^N(T)$ are independent (since $Y^N_1, \ldots, Y^N_N$ are assumed to be independent and identically distributed with common distribution γ_0, this amounts to requiring that γ_0 be of product form, that is, $\gamma_0 = \otimes_{t=0}^{T} \nu_t$ for some $\nu_0, \ldots, \nu_T \in \mathcal{P}(\mathcal{Y}_0)$). In particular, discrete time versions of weakly interacting Itô processes as considered in Section 5 are covered by Example 4.3. More precisely, assuming coefficients of diffusion type and using a standard Euler-Maruyana scheme for the system of stochastic differential equations (5.1) and the corresponding limit equation (5.2), one would choose $T \in \mathbb{N}$ and $h > 0$ so that $h \cdot T$ corresponds to the continuous time horizon, set $\mathcal{X}_0 \doteq \mathbb{R}^d$, $\mathcal{Y}_0 \doteq \mathbb{R}^{d_1}$, define $\varphi \colon \{1, \ldots, T\} \times \mathbb{R}^d \times \mathcal{P}(\mathcal{X}_0) \times \mathcal{Y}_0 \to \mathcal{X}_0$ according to

$$
\varphi(t, x, \nu, y) \doteq x + \tilde{b}\big((t-1)h, x, \nu\big)h + \sqrt{h} \cdot \tilde{\sigma}\big((t-1)h, x, \nu\big)y,
$$

and set $\gamma_0 \doteq \otimes^{T+1}\nu$ for some $\nu \in \mathcal{P}(\mathbb{R}^{d_1})$ with mean zero and identity covariance matrix (in particular, $\nu \doteq N(0, \mathrm{Id}_{d_1})$ the d_1-variate standard normal distribution). In Section 5 we assume for simplicity that all component processes have the same deterministic initial condition; this corresponds to setting $\varphi_0 \equiv x_0$ for some $x_0 \in \mathbb{R}^d$. If the drift coefficient b and the dispersion coefficient σ are continuous, then so is φ, and Corollary 4.2 applies.

Example 4.3 also applies to finite state discrete time weakly interacting Markov chains, which arise as discrete time versions of the mean field systems found, for instance, in the analysis of large communication networks, especially WLANs [cf. Duffy, 2010, for an overview]. In this situation, the functions φ_0, φ are in general discontinuous in $y \in \mathcal{Y}_0$; yet the hypotheses of Corollary 4.2 are still satisfied. To be more precise, let $\mathcal{S} \doteq \{s_1, \ldots, s_M\}$ be a finite set, and let $\iota\colon \mathcal{S} \to \{1, \ldots, M\}$ be the natural bijection between elements of $\mathcal{S}$ and their indices (thus $\iota(s_i) = i$ for every $i \in \{1, \ldots, M\}$). The space of probability measures $\mathcal{P}(\mathcal{S})$ can be identified with $\{p \in [0,1]^M : \sum_{k=1}^M p_k = 1\}$ endowed with the standard metric. For $t \in \mathbb{N}_0$, let $a_{ij}(t,.)\colon \mathcal{P}(\mathcal{S}) \to [0,1]$, $i,j \in \{1, \ldots, M\}$, be measurable maps such that, for every $p \in \mathcal{P}(\mathcal{S})$, $A(t,p) \doteq (a_{ij}(t,p))_{i,j\in\{1,\ldots,M\}}$ is a transition probability matrix on $\mathcal{S} \equiv \{1, \ldots, M\}$. Let $q \in \mathcal{P}(\mathcal{S})$. Using the notation of Example 4.3, fix $T \in \mathbb{N}$, set $\mathcal{X}_0 \doteq \mathcal{S}$, $\mathcal{Y}_0 \doteq [0,1]$, $\mathcal{X} \doteq \mathcal{X}_0^{T+1}$, and $\mathcal{Y} \doteq \mathcal{Y}_0^{T+1}$; define $\varphi\colon \{1, \ldots, T\} \times \mathcal{X}_0 \times \mathcal{P}(\mathcal{X}_0) \times \mathcal{Y}_0 \to \mathcal{X}_0$ by

$$\varphi(t,x,p,y) \doteq \sum_{j=1}^{M} s_j \cdot \mathbf{1}_{\left(\sum_{k=1}^{j-1} a_{\iota(x)k}(t-1,p), \sum_{k=1}^{j} a_{\iota(x)k}(t-1,p)\right]}(y),$$

and let $\varphi_0\colon \mathcal{Y}_0 \to \mathcal{X}_0$ be given by

$$\varphi_0(y) \doteq \sum_{j=1}^{M} s_j \cdot \mathbf{1}_{\left(\sum_{k=1}^{j-1} q_k, \sum_{k=1}^{j} q_k\right]}(y).$$

Set $\gamma_0 \doteq \otimes^{T+1}\lambda_{[0,1]}$ with $\lambda_{[0,1]}$ Lebesgue measure on $\mathcal{B}([0,1]) = \mathcal{B}(\mathcal{Y}_0)$. For $N \in \mathbb{N}$, let $Y_1^N, \ldots, Y_N^N$ be independent and identically distributed $\mathcal{Y}$-valued random variables with common distribution γ_0, and define $\mathcal{X}$-valued random variables $X_1^N, \ldots, X_N^N$ with $X_i^N = (X_i^N(t))_{t\in\{0,\ldots,T\}}$ recursively by (4.6). Observe that $X_1^N(0), \ldots, X_N^N(0)$ are independent and identically distributed with common distribution q. Moreover, for all $t \in \{0, \ldots, T-1\}$, all $\boldsymbol{z} \in \mathcal{S}^N$,

(4.12)

$$\begin{aligned}\mathbf{P}_N\left(X^N(t+1) = \boldsymbol{z} \middle| X^N(0), \ldots, X^N(t)\right) &= \prod_{i=1}^{N} a_{\iota(X_i^N(t))\iota(z_i)}(t, \mu^N(t)) \\ &= \exp\left(N \cdot \int_{\mathcal{S}} \log\left(a_{\iota(x)\iota(z_i)}(t, \mu^N(t))\right) \mu^N(t, dx)\right),\end{aligned}$$

where $\log(0) = -\infty$, $e^{-\infty} = 0$. It follows that $(X^N(t))_{t\in\{0,\ldots,T\}}$ is a Markov chain with state space $\mathcal{S}^N$. Equation (4.12) also implies that $(\mu^N(t))_{t\in\{0,\ldots,T\}}$ is a Markov chain with transition probabilities given by

$$\mathbf{P}_N\left(\mu^N(t+1) = \frac{1}{N}\sum_{i=1}^{N}\delta_{z_i}\middle| X^N(0),\ldots,X^N(t)\right)$$
$$= \sum_{\tilde{\boldsymbol{z}}\in\mathfrak{p}(\boldsymbol{z})} \exp\left(N\cdot\int_{\mathcal{S}}\log\left(a_{\iota(x)\iota(\tilde{z}_i)}(t,\mu^N(t))\right)\mu^N(t,dx)\right),$$

where $\mathfrak{p}(\boldsymbol{z})$ indicates the set of elements of $\mathcal{S}^N$ that arise by permuting the components of $\boldsymbol{z}\in\mathcal{S}^N$. For $i\in\{1,\ldots,N\}$, again by (4.12), the process couple $((X_i^N(t),\mu^N(t)))_{t\in\{0,\ldots,T\}}$ is a Markov chain with state space $\mathcal{S}\times\mathcal{P}(\mathcal{S})$, and its law does not depend on the component i. Define the function ψ according to (4.7), and define Ψ according to (4.3). As in the more general situation of Example 4.3, Equation (4.4) (i.e., the fixed point equation $\Psi_\gamma(\mu) = \mu$) has a unique solution $\mu_*(\gamma)$ given any $\gamma\in\mathcal{D}\doteq\mathcal{P}(\mathcal{Y})$, representation (4.11) holds for $\mu_*(\gamma)$, and the mapping $\mathcal{P}(\mathcal{Y})\ni\gamma\mapsto\mu_*(\gamma)\in\mathcal{P}(\mathcal{X})$ is measurable. Let us assume that the maps $p\mapsto a_{ij}(t,p)$ are continuous for all $i,j\in\{1,\ldots,M\}$, $t\in\mathbb{N}_0$. In order to verify the hypotheses of Corollary 4.2, it then remains to check that $\mu_*(.)$ is continuous at any $\tilde{\gamma}\in\mathcal{P}(\mathcal{Y})$ such that $R(\tilde{\gamma}|\gamma_0)<\infty$. To do this, take $\tilde{\gamma}\in\mathcal{P}(\mathcal{Y})$ absolutely continuous with respect to γ_0, and let $(\tilde{\gamma}_n)\subset\mathcal{P}(\mathcal{Y})$ be such that $\tilde{\gamma}_n\to\tilde{\gamma}$ as $n\to\infty$. Recall (4.8), the definition of the functions $\bar{\varphi}_t$, and (4.9), the definition of the maps $\gamma\mapsto\alpha_t(\gamma)$. For $t\in\{0,\ldots,T\}$ set

$$D_t \doteq \big\{y\in\mathcal{Y} : \exists\,(y^n)_{n\in\mathbb{N}}\subset\mathcal{Y} \text{ such that, as } n\to\infty,\ y^n\to y \text{ but}$$
$$\bar{\varphi}_t\big((\alpha_0(\tilde{\gamma}_n),\ldots,\alpha_{t-1}(\tilde{\gamma}_n)),y^n\big)\nrightarrow\bar{\varphi}_t\big((\alpha_0(\tilde{\gamma}),\ldots,\alpha_{t-1}(\tilde{\gamma})),y\big)\big\}.$$

By definition of $\bar{\varphi}_0$ and φ_0, we have

$$D_0\subseteq\left\{y\in\mathcal{Y} : y_0\in\left\{\sum_{k=1}^{j}q_k : j\in\{0,\ldots,M\}\right\}\right\}.$$

It follows that $\gamma_0(D_0) = 0$ and, since $\tilde{\gamma}$ is absolutely continuous with respect to γ_0, $\tilde{\gamma}(D_0) = 0$. The extended mapping theorem implies that $\alpha_0(\tilde{\gamma}_n)\to\alpha_0(\tilde{\gamma})$ as $n\to\infty$. Using this convergence, the definition of $\bar{\varphi}_1$ in terms of φ, the continuity of $p\mapsto a_{ij}(t,p)$, and the fact that $\bar{\varphi}_0$ is continuous on $\mathcal{Y}\setminus D_0$, we find that

$$D_1\subseteq D_0\cup\left\{y\in\mathcal{Y} : y_1\in\left\{\sum_{k=1}^{j}a_{\iota(\bar{\varphi}_0(y))k}(0,\alpha_0(\tilde{\gamma})) : j\in\{0,\ldots,M\}\right\}\right\}.$$

Since $\bar{\varphi}_0(y)$ depends on y only through y_0 (in fact, $\bar{\varphi}_0(y) = \varphi_0(y_0)$), it follows that $\gamma_0(D_1) = 0$, hence $\tilde{\gamma}(D_1) = 0$. The extended mapping theorem in the version of Theorem 5.5 in Billingsley [1968, p. 34] implies that $\alpha_1(\tilde{\gamma}_n) \to \alpha_1(\tilde{\gamma})$ as $n \to \infty$. Proceeding by induction over t, one checks that

$$\begin{aligned} D_t \subseteq D_0 \cup \ldots \cup D_{t-1} \cup \Bigg\{ y \in \mathcal{Y} : \\ y_t \in \left\{ \sum_{k=1}^{j} a_{\iota(\bar{\varphi}_{t-1}((\alpha_0(\tilde{\gamma}),\ldots,\alpha_{t-1}(\tilde{\gamma})),y))k}(t-1, \alpha_{t-1}(\tilde{\gamma})) : j \in \{0, \ldots, M\} \right\} \Bigg\} \end{aligned}$$

and, since $\bar{\varphi}_{t-1}((\alpha_0(\tilde{\gamma}), \ldots, \alpha_{t-1}(\tilde{\gamma})), y)$ depends on y only through the components $(y_0, \ldots, y_{t-1})$, $\gamma_0(D_t) = 0 = \tilde{\gamma}(D_t)$, which implies that $\alpha_t(\tilde{\gamma}_n) \to \alpha_t(\tilde{\gamma})$ as $n \to \infty$. Set $D \doteq \bigcup_{t=0}^{T} D_t$ and recall (4.10), the definition of Φ. Let $y \in \mathcal{Y}$, $(y^n)_{n\in\mathbb{N}} \subset \mathcal{Y}$ be such that $y^n \to y$ as $n \to \infty$. Then

$$\Phi\big((\alpha_0(\tilde{\gamma}_n), \ldots, \alpha_{T-1}(\tilde{\gamma}_n)), y^n\big) \overset{n\to\infty}{\longrightarrow} \Phi\big((\alpha_0(\tilde{\gamma}), \ldots, \alpha_{T-1}(\tilde{\gamma})), y\big) \text{ if } y \notin D.$$

Since $\gamma_0(D) = 0 = \tilde{\gamma}(D)$, the extended mapping theorem yields

$$\tilde{\gamma}_n \circ \Phi\big((\alpha_0(\tilde{\gamma}_n), \ldots, \alpha_{T-1}(\tilde{\gamma}_n)), .\big)^{-1} \overset{n\to\infty}{\longrightarrow} \tilde{\gamma} \circ \Phi\big((\alpha_0(\tilde{\gamma}), \ldots, \alpha_{T-1}(\tilde{\gamma})), .\big)^{-1}.$$

Recalling representation (4.11) we conclude that

$$\mu_*(\tilde{\gamma}_n) \overset{n\to\infty}{\longrightarrow} \mu_*(\tilde{\gamma}),$$

which establishes continuity of $\mu_*(.)$ at any $\tilde{\gamma}$ with $R(\tilde{\gamma}\|\gamma_0) < \infty$ since any such $\tilde{\gamma}$ is absolutely continuous with respect to γ_0. Under the assumption that the maps $p \mapsto a_{ij}(t, p)$ are continuous, we have thus derived the large deviation principle for $(\mu^N)_{N\in\mathbb{N}}$ with rate function $\eta \mapsto R(\eta\|\Psi_{\gamma_0}(\eta))$; here $\Psi_{\gamma_0}(\eta)$ coincides with the law of a time-inhomogeneous $\mathcal{S}$-valued Markov chain with initial distribution q and transition matrices $A(t, \eta(t))$, $t \in \{0, \ldots, T-1\}$. The same arguments and a completely analogous construction work for weakly interacting Markov chains with countably infinite state space $\mathcal{S}$. Notice that we need not require the transition probabilities $a_{ij}(t, p)$ to be bounded away from zero; in particular, whether $a_{ij}(t, p)$ is equal to zero or strictly positive may depend on the measure variable p.

5 Weakly interacting Itô processes

In this section, we consider weakly interacting systems described by Itô processes as studied in Budhiraja et al. [2012]. We show that the Laplace principle rate function derived there in variational form can be expressed in non-variational form in terms of relative entropy. We do not give the most general conditions under which the results hold; in particular, we assume here that all particles obey the same deterministic initial condition.

Let $T > 0$ be a finite time horizon, let $d, d_1 \in \mathbb{N}$, and let $x_0 \in \mathbb{R}^d$. Set $\mathcal{X} \doteq \mathbf{C}([0,T], \mathbb{R}^d)$, $\mathcal{Y} \doteq \mathbf{C}([0,T], \mathbb{R}^{d_1})$, equipped with the maximum norm topology. Let b, σ be predictable functionals defined on $[0,T] \times \mathcal{X} \times \mathcal{P}(\mathbb{R}^d)$ with values in $\mathbb{R}^d$ and $\mathbb{R}^{d \times d_1}$, respectively. For $N \in \mathbb{N}$, let $((\Omega_N, \mathcal{F}^N, \mathbf{P}_N), (\mathcal{F}^N_t))$ be a stochastic basis satisfying the usual hypotheses and carrying N independent d_1-dimensional $(\mathcal{F}^N_t)$)-Wiener processes $W^N_1, \ldots, W^N_N$. The N-particle system is described by the solution to the system of stochastic differential equations

$$dX^N_i(t) = b\big(t, X^N_i, \mu^N(t)\big)dt + \sigma\big(t, X^N_i, \mu^N(t)\big)dW^N_i(t) \tag{5.1}$$

with initial condition $X^N_i(0) = x_0$, $i \in \{1, \ldots, N\}$, where $\mu^N(t)$ is the empirical measure of $X^N_1, \ldots, X^N_N$ at time $t \in [0,T]$, that is,

$$\mu^N(t, \omega) \doteq \frac{1}{N} \sum_{i=1}^{N} \delta_{X^N_i(t,\omega)}, \quad \omega \in \Omega_N.$$

The coefficients b, σ in Equation (5.1) may depend on the entire history of the solution trajectory, not only its current value as in the diffusion case. In the diffusion case, in fact, one has $b(t, \varphi, \nu) = \tilde{b}(t, \varphi(t), \nu)$, $\sigma(t, \varphi, \nu) = \tilde{\sigma}(t, \varphi(t), \nu)$ for some functions $\tilde{b}$, $\tilde{\sigma}$ defined on $[0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, and the solution process X^N is a Markov process with state space $\mathbb{R}^{N \times d}$.

Denote by μ^N the empirical measure of $(X^N_1, \ldots, X^N_N)$ over the time interval $[0,T]$, that is, μ^N is the $\mathcal{P}(\mathcal{X})$-valued random variable defined by

$$\mu^N_\omega \doteq \frac{1}{N} \sum_{i=1}^{N} \delta_{X^N_i(.,\omega)}, \quad \omega \in \Omega_N$$

The asymptotic behavior of μ^N as N tends to infinity can be characterized in terms of solutions to the “nonlinear” stochastic differential equation

$$dX(t) = b\big(t, X, \mathrm{Law}(X(t))\big)dt + \sigma\big(t, X, \mathrm{Law}(X(t))\big)dW(t) \tag{5.2}$$

with $\mathrm{Law}(X(0)) = \delta_{x_0}$, where W is a standard d_1-dimensional Wiener process defined on some stochastic basis. Notice that the law of the solution itself appears in the coefficients of Equation (5.2). In the diffusion case, the corresponding Kolmogorov forward equation is therefore a nonlinear parabolic partial differential equation, and it corresponds to the McKean-Vlasov equation of the weakly interacting system defined by (5.1).

For the statement of the Laplace principle, we need to consider controlled versions of Equations (5.1) and (5.2), respectively. For $N \in \mathbb{N}$, let $\mathcal{U}_N$ be the space of all $(\mathcal{F}_t^N)$-progressively measurable functions $u\colon [0,T]\times\Omega_N \to \mathbb{R}^{N\times d_1}$ such that

$$\mathbf{E}_N\left[\sum_{i=1}^{N}\int_0^T |u_i(t)|^2 dt\right] < \infty,$$

where $u = (u_1, \ldots, u_N)$ and $\mathbf{E}_N$ denotes expectation with respect to $\mathbf{P}_N$. Given $u \in \mathcal{U}_N$, the counterpart of Equation (5.1) is the system of controlled stochastic differential equations

$$\begin{aligned} d\bar{X}_i^N(t) = b\big(t, \bar{X}_i^N, \bar{\mu}^N(t)\big)dt + \sigma\big(t, \bar{X}_i^N, \bar{\mu}^N(t)\big)u_i(t)dt \\ + \sigma\big(t, \bar{X}_i^N, \bar{\mu}^N(t)\big)dW_i^N(t), \end{aligned} \tag{5.3}$$

with initial condition $\bar{X}_i^N(0) = x_0$, where $\bar{\mu}^N(t)$ denotes the empirical measure of $\bar{X}_1^N, \ldots, \bar{X}_N^N$ at time t.

Let $\mathcal{U}$ be the set of quadruples $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), u, W)$ such that the pair $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ forms a stochastic basis satisfying the usual hypotheses, W is a d_1-dimensional $(\mathcal{F}_t)$-Wiener process, and u is an $\mathbb{R}^{d_1}$-valued $(\mathcal{F}_t)$-progressively measurable process such that

$$\mathbf{E}\left[\int_0^T |u(t)|^2 dt\right] < \infty.$$

For simplicity, we may write $u \in \mathcal{U}$ instead of $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), u, W) \in \mathcal{U}$. Given $u \in \mathcal{U}$, the counterpart of Equation (5.2) is the controlled "nonlinear" stochastic differential equation

$$\begin{aligned} d\bar{X}(t) = b\big(t, \bar{X}, \mathrm{Law}(\bar{X}(t))\big)dt + \sigma\big(t, \bar{X}, \mathrm{Law}(\bar{X}(t))\big)u(t)dt \\ + \sigma\big(t, \bar{X}, \mathrm{Law}(\bar{X}(t))\big)dW(t) \end{aligned} \tag{5.4}$$

with initial condition $\mathrm{Law}(\bar{X}(0)) = \delta_{x_0}$. A solution of Equation (5.4) under $u \in \mathcal{U}$ is a continuous $\mathbb{R}^d$-valued process $\bar{X}$ defined on the given stochastic basis and adapted to the given filtration such that the integral version of Equation (5.4) holds with probability one. Denote by $\mathcal{R}_1$ the space of deterministic relaxed controls with finite first moments, that is, $\mathcal{R}_1$ is the set of all positive measures on $\mathcal{B}(\mathbb{R}^{d_1} \times [0,T])$ such that $r(\mathbb{R}^{d_1} \times [0,t]) = t$ for all $t \in [0,T]$ and $\int_{\mathbb{R}^{d_1}\times[0,T]} |y| r(dy \times dt) < \infty$. Equip $\mathcal{R}_1$ with the topology of weak convergence of measures plus convergence of first moments. Let $u \in \mathcal{U}$. The joint distribution of (u, W) can be identified with a probability measure on $\mathcal{B}(\mathcal{R}_1 \times \mathcal{Y})$. If $\bar{X}$ is a solution of Equation (5.4) under u, then the joint distribution of $(\bar{X}, u, W)$ can be identified with a probability measure on $\mathcal{B}(\mathcal{Z})$, where $\mathcal{Z} \doteq \mathcal{X} \times \mathcal{R}_1 \times \mathcal{Y}$.

Definition 5.1. *Weak uniqueness* of solutions is said to hold for Equation (5.4) if, whenever $u, \tilde{u} \in \mathcal{U}$ and $\bar{X}$, $\tilde{X}$ are two solutions of Equation (5.4) under u and $\tilde{u}$, respectively, such that $\mathbf{P} \circ \bar{X}(0)^{-1} = \tilde{\mathbf{P}} \circ \tilde{X}(0)^{-1}$, then $\mathbf{P} \circ (\bar{X}, u, W)^{-1} = \tilde{\mathbf{P}} \circ (\tilde{X}, \tilde{u}, \tilde{W})^{-1}$ as probability measures on $\mathcal{B}(\mathcal{X} \times \mathcal{R}_1 \times \mathcal{Y})$.

Notice that here we give a process version of what can be equivalently formulated in terms of probability measures on $\mathcal{B}(\mathcal{Z})$. Indeed, any integrable control process u corresponds to an $\mathcal{R}_1$-valued random variable. On the other hand, since the control appears linearly in Equations (5.3) and (5.4), given any adapted $\mathcal{R}_1$-valued random variable, one can find an integrable control process that produces the same solution process $\bar{X}$ [cf. Sections 2 & 6 in Budhiraja et al., 2012].

Remark 5.1. In Budhiraja et al. [2012], weak uniqueness for Equation (5.4) is required to hold over the class of all $\Theta \in \mathcal{P}(\mathcal{Z})$ that correspond to a weak solution of (5.4). This requirement is stronger than necessary. As can be seen from the definition of the rate function and the proof of Theorem 3.1 (and Theorem 7.1) there[1], it suffices to have weak uniqueness for Equation (5.4) over the class of all $\Theta \in \mathcal{P}(\mathcal{Z})$ that correspond to a weak solution of (5.4) and are such that

$$\int_{\mathcal{Z}} \int_{\mathbb{R}^{d_1} \times [0,T]} |y|^2 \, r(dy \times dt) \, \Theta(d\varphi \times dr \times dw) < \infty.$$

This is equivalent to requiring weak uniqueness of solutions for Equation (5.4) with respect to $\mathcal{U}$ as in Definition 5.1 above.

The Laplace principle given in Theorem 5.1 below is a version of Theorem 7.1 in Budhiraja et al. [2012]; also cf. Theorem 3.1 and Remark 3.2 there. The following assumptions are sufficient for the Laplace principle to hold:

(H1) The functions $b(t, ., .)$, $\sigma(t, ., .)$ are uniformly continuous and bounded on sets $B \times P$ whenever $B \subset \mathcal{X}$ is bounded and $P \subset \mathcal{P}(\mathbb{R}^d)$ is compact, uniformly in $t \in [0, T]$.

(H2) For all $N \in \mathbb{N}$, existence and uniqueness of solutions holds in the strong sense for the system of N equations given by (5.1).

[1] In the notation of Budhiraja et al. [2012], it follows from the proof of Lemma 5.1 there and a version of Fatou's lemma, that if Q is a limit point in the sense of convergence in distribution of the sequence of $\mathcal{P}(\mathcal{Z})$-valued random variables Q^N, then $\int_{\mathcal{Z}} \int_{\mathbb{R}^{d_1} \times [0,T]} |y|^2 r(dy \times dt) Q(d\varphi \times dr \times dw) < \infty$ with probability one. As to the rate function and the Laplace upper bound, notice that the class $\mathcal{P}_\infty$ only contains measures $\Theta \in \mathcal{P}(\mathcal{Z})$ such that $\int_{\mathcal{Z}} \int_{\mathbb{R}^{d_1} \times [0,T]} |y|^2 r(dy \times dt) \Theta(d\varphi \times dr \times dw) < \infty$.

(H3) Weak uniqueness of solutions holds for Equation (5.4).

(H4) If $u^N \in \mathcal{U}_N$, $N \in \mathbb{N}$, are such that

$$\sup_{N\in\mathbb{N}} \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N}\int_0^T |u_i^N(t)|^2\, dt\right] < \infty,$$

then $\{\bar{\mu}^N : N \in \mathbb{N}\}$ is tight as a family of $\mathcal{P}(\mathcal{X})$-valued random variables, where $\bar{\mu}^N$ is the empirical measure of the solution to the system of equations (5.3) under u^N.

Theorem 5.1 (Budhiraja et al. [2012])**.** *Grant (H1) – (H4). Then the sequence $(\mu^N)_{N\in\mathbb{N}}$ of $\mathcal{P}(\mathcal{X})$-valued random variables satisfies the Laplace principle with rate function $I\colon \mathcal{P}(\mathcal{X}) \to [0,\infty]$ given by*

$$I(\theta) = \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{X}^u)=\theta} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right],$$

where $\bar{X}^u$ is a solution of Equation (5.4) *over the time interval* $[0,T]$ *with* $\mathrm{Law}(X(0)) = \delta_{x_0}$*, and* $\inf\emptyset = \infty$ *by convention.*

Remark 5.2. The function I of Theorem 5.1 is indeed a rate function, that is, I is lower semicontinuous with values in $[0,\infty]$. The following hypothesis, which is analogous to the stability condition (H4), is sufficient to guarantee goodness of the rate function.

(H') If $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ is such that $\sup_{n\in\mathbb{N}} \mathbf{E}_n\left[\int_0^T |u_n(t)|^2 dt\right] < \infty$, then $\{\mathrm{Law}(\bar{X}^{u_n}) : n \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathcal{X})$.

Under this additional assumption, I is a good rate function and the Laplace principle implies the large deviation principle.

Consider the special case in which $d = d_1$, $x_0 = 0$, $b \equiv 0$, and $\sigma \equiv \mathrm{Id}_d$. In this case, $\mathcal{X} = \mathcal{Y}$ and μ^N is the empirical measure of N independent Wiener processes $W_1^N, \ldots, W_N^N$. Let γ_0 be Wiener measure on $\mathcal{B}(\mathcal{Y})$. Since $\mathrm{Law}(W_i^N) = \gamma_0$, Sanov's theorem implies that the sequence $(\mu^N)_{N\in\mathbb{N}}$ satisfies the large deviation / Laplace principle with good rate function $R(.\|\gamma_0)$. On the other hand, by Theorem 5.1, $(\mu^N)_{N\in\mathbb{N}}$ satisfies the Laplace principle with rate function

$$J(\gamma) \doteq \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{Y}^u)=\gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right], \quad \gamma \in \mathcal{P}(\mathcal{Y}),$$

where $\bar{Y}^u$ is the process given by

$$\bar{Y}^u(t) \doteq \int_0^t u(s)ds + W(t), \quad t \in [0,T]. \tag{5.5}$$

One checks that $J\colon \mathcal{P}(\mathcal{Y}) \to [0,\infty]$ has compact sublevel sets, hence is a good rate function. It follows that J coincides with the rate function obtained from Sanov's theorem. Consequently, for all $\gamma \in \mathcal{P}(\mathcal{Y})$,

$$R\bigl(\gamma\|\gamma_0\bigr) = \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{Y}^u)=\gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]. \tag{5.6}$$

Remark 5.3. Equation (5.6) provides a "weak" variational representation of relative entropy with respect to Wiener measure. In Appendix B, we give a direct proof of Equation (5.6). The variational representation is weak in the sense that the underlying stochastic basis may vary. In particular, the control process u may be adapted to a filtration that is strictly bigger than the natural filtration of the Wiener process. Notice that expectation in (5.6) is taken with respect to the probability measure of the stochastic basis that comes with the control process u.

Remark 5.4. Representation (5.6) may be compared to the following result obtained by Üstünel [2009]. Take as stochastic basis the canonical set-up; in our notation, $((\mathcal{Y},\mathcal{B}(\mathcal{Y}),\gamma_0),(\mathcal{B}_t))$, where $(\mathcal{B}_t)$ is the canonical filtration. Let W be the coordinate process. Thus W is a d_1-dimensional Wiener process under γ_0 with respect to $(\mathcal{B}_t)$. Let u be an $\mathbb{R}^{d_1}$-valued $(\mathcal{B}_t)$-progressively measurable process such that $\mathbf{E}_{\gamma_0}\left[\int_0^T |u(t)|^2 dt\right] < \infty$. Consider $\bar{Y}^u = \int_0^{\cdot} u(s)ds + W(.)$. Since $\bar{Y}^u(.,\omega) = \int_0^{\cdot} u(s,\omega)ds + \omega(.)$ for all $\omega \in \mathcal{Y}$, $\bar{Y}^u$ induces a Borel measurable mapping $\mathcal{Y} \to \mathcal{Y}$. Set $\gamma \doteq \gamma_0\circ(\bar{Y}^u)^{-1}$. By Theorem 8 in Üstünel [2009],

$$R\bigl(\gamma\|\gamma_0\bigr) \leq \mathbf{E}_{\gamma_0}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]. \tag{5.7}$$

Assume in addition that u is such that

$$\mathbf{E}\left[\exp\left(-\int_0^T u(t)\cdot dW(t) - \frac{1}{2}\int_0^T |u(t)|^2 dt\right)\right] = 1,$$

and that, for some $\mathbb{R}^{d_1}$-valued $(\mathcal{B}_t)$-progressively measurable process v,

$$\frac{d\gamma}{d\gamma_0} = \exp\left(-\int_0^T v(t)\cdot dW(t) - \frac{1}{2}\int_0^T |v(t)|^2 dt\right) \quad \gamma_0\text{-a.s.}$$

Theorem 7 in Üstünel [2009] then states that equality holds in (5.7) if and only if $\bar{Y}^u$ is γ_0-almost surely invertible as a mapping $\mathcal{Y} \to \mathcal{Y}$ with inverse $\bar{Y}^v = \int_0^{\cdot} v(s)ds + W(.)$. For similar results on abstract Wiener spaces see Lassalle [2012]; Corollary 8 and Remark 4 in Section 7 therein might be compared to Lemma B.1 in Appendix B here.

Let us return to the general case. Given $\theta \in \mathcal{P}(\mathcal{X})$, denote by $\theta(t)$ the marginal distribution of θ at time t and consider the stochastic differential equation

$$dX(t) = b\big(t, X, \theta(t)\big)dt + \sigma\big(t, X, \theta(t)\big)dW(t). \tag{5.8}$$

Equation (5.8) results from freezing the measure variable in Equation (5.2) at θ. We will assume existence and pathwise uniqueness for Equation (5.8).

(H5) Given any $\theta \in \mathcal{P}(\mathcal{X})$, weak existence and pathwise uniqueness hold for Equation (5.8).

Based on representation (5.6) and the contraction property of relative entropy, the rate function of Theorem 5.1 can be shown to be of relative entropy form.

Theorem 5.2. *Grant (H1) – (H5). Then the rate function I of Theorem 5.1 can be expressed in relative entropy form as*

$$I(\theta) = R\big(\theta \| \Psi(\theta)\big), \quad \theta \in \mathcal{P}(\mathcal{X}),$$

where $\Psi(\theta)$ is the law of the unique solution of Equation (5.8) under θ over the time interval $[0, T]$ with initial condition $X(0) = x_0$.

Remark 5.5. The hypotheses of Theorem 5.2 are satisfied if b, σ are locally Lipschitz continuous with σ uniformly bounded and b of sub-linear growth in the trajectory variable; see Appendix C. These sufficient conditions are at the same time more restrictive and more general than the assumptions made in Dawson and Gärtner [1987], where the large deviation principle is derived for weakly interacting Itô diffusions. There the coefficients are only required to be continuous, where continuity in the measure variable is with respect to an inductive topology that is stronger than the topology of weak convergence (but cf. Remark 1.1 above), and to satisfy a coercivity condition that allows for sub-linear growth of the dispersion coefficient and for super-linear growth of the drift vector in "stabilizing" directions. On the other hand, in Dawson and Gärtner [1987] the diffusion matrix has to be non-degenerate and independent of the measure variable, while here we can have degeneracy of $\sigma\sigma^{\mathsf{T}}$ as well as measure dependence. Lastly, since here both b and σ are functions of the entire trajectory history, one can capture systems with delay in the state dynamics.

Remark 5.6. Assumption (H5) can be weakened by requiring weak existence and pathwise uniqueness of solutions to Equation (5.8) only for $\theta \in \mathcal{P}(\mathcal{X})$ such that $I(\theta) < \infty$. Those measures θ are, by definition of I, distributions of Itô processes. The function Ψ introduced in Theorem 5.2 would then be defined only on the effective domain of I; for $\theta \in \mathcal{P}(\mathcal{X})$ with $I(\theta) = \infty$, one can then choose $\Psi(\theta)$ in such a way that θ is not absolutely continuous with respect to $\Psi(\theta)$ (for instance, by choosing between two Dirac measures).

Proof of Theorem 5.2. Let $\theta \in \mathcal{P}(\mathcal{X})$. By hypothesis, weak existence and pathwise uniqueness hold for Equation (5.8). By a result originally due to Yamada and Watanabe [1971] [also cf. Kallenberg, 1996], there is a Borel measurable mapping $\psi_\theta \colon \mathbb{R}^d \times \mathcal{Y} \to \mathcal{X}$ such that

$$\psi_\theta(x_0, W) = X \quad \mathbf{P}\text{-almost surely} \tag{5.9}$$

whenever X is a solution of Equation (5.8) under θ over time $[0, T]$ with initial condition $X(0) = x_0$ on some stochastic basis $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ carrying a d_1-dimensional Wiener process W. For such a solution, $\Psi(\theta) = \mathrm{Law}(X)$ by definition. Set $\psi_\theta(.) \doteq \psi_\theta(x_0, .)$, and let γ_0 be Wiener measure on $\mathcal{B}(\mathcal{Y})$. By Equation (5.9), $\Psi(\theta) = \psi_\theta(\gamma_0) = \gamma_0 \circ \psi_\theta^{-1}$. By Lemma A.1, the contraction property of relative entropy, and representation (5.6) it follows that

$$\begin{aligned}
R\big(\theta \| \Psi(\theta)\big) &= R\big(\theta \| \psi_\theta(\gamma_0)\big) \\
&= \inf_{\gamma \in \mathcal{P}(\mathcal{Y}) : \psi_\theta(\gamma) = \theta} R\big(\gamma \| \gamma_0\big) \\
&= \inf_{\gamma \in \mathcal{P}(\mathcal{Y}) : \psi_\theta(\gamma) = \theta} \; \inf_{u \in \mathcal{U} : \mathrm{Law}(\bar{Y}^u) = \gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right] \\
&= \inf_{u \in \mathcal{U} : \mathrm{Law}(\psi_\theta(\bar{Y}^u)) = \theta} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right],
\end{aligned}$$

where $\bar{Y}^u$ is defined by (5.5). Let $u \in \mathcal{U}$, and set $\tilde{X}^u \doteq \psi_\theta(\bar{Y}^u)$. Then, as a consequence of Equation (5.9), $\tilde{X}^u$ solves

$$dX(t) = b\big(t, X, \theta(t)\big)dt + \sigma\big(t, X, \theta(t)\big)u(t)dt + \sigma\big(t, X, \theta(t)\big)dW(t)$$

with initial distribution δ_{x_0}. If u is such that $\mathrm{Law}(\psi_\theta(\bar{Y}^u)) = \theta$, then $\tilde{X}^u$ is a solution of Equation (5.4) under u with initial distribution δ_{x_0}. By Assumption (H3), weak uniqueness holds for Equation (5.4), hence $\mathrm{Law}(\tilde{X}^u) = \mathrm{Law}(\bar{X}^u)$ whenever $\bar{X}^u$ is a solution of (5.4) under u with $\mathrm{Law}(\bar{X}^u(0)) = \delta_{x_0}$. It follows that

$$\begin{aligned}
R\big(\theta \| \Psi(\theta)\big) &= \inf_{u \in \mathcal{U} : \mathrm{Law}(\psi_\theta(\bar{Y}^u)) = \theta} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right] \\
&= \inf_{u \in \mathcal{U} : \mathrm{Law}(\bar{X}^u) = \theta} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right] \\
&= I(\theta),
\end{aligned}$$

where I is the rate function of Theorem 5.1. □

Remark 5.7. Assuming in addition to (H1) – (H5) hypothesis (H') of Remark 5.2, Theorem 5.2 can be proved by applying both Sanov's theorem and Theorem 5.1 to the weakly interacting system given by equations (5.1) with measure variable frozen at $\theta \in \mathcal{P}(\mathcal{X})$ and then evaluating the resulting rate functions at θ.

A Contraction property of relative entropy

Let $\mathcal{X}$, $\mathcal{Y}$ be Polish spaces. Denote by $\Pi_{\mathcal{X}}$, $\Pi_{\mathcal{Y}}$ the collection of all finite and measurable partitions of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Recall that relative entropy can be approximated in terms of finite sums; for $\eta, \nu \in \mathcal{P}(\mathcal{X})$,

$$R(\eta\|\nu) = \sup_{\pi\in\Pi_{\mathcal{X}}} \sum_{A\in\pi} \eta(A)\log\left(\frac{\eta(A)}{\nu(A)}\right), \tag{A.1}$$

see, for instance, Lemma 1.4.3(g) in Dupuis and Ellis [1997, p. 30]. For $\psi : \mathcal{Y} \to \mathcal{X}$ measurable, $\gamma \in \mathcal{P}(\mathcal{Y})$, denote by $\psi(\gamma) \doteq \gamma \circ \psi^{-1}$ the image measure of γ under ψ.

The following lemma extends the invariance property of relative entropy under bijective bi-measurable mappings as given by Lemma E.2.1 in Dupuis and Ellis [1997, p. 366] to arbitrary measurable transformations; also cf. Theorem 2.4.1 in Kullback [1978, pp. 19-20], where the inequality that is implied by Lemma A.1 is established.

Lemma A.1. *Let $\psi\colon \mathcal{Y} \to \mathcal{X}$ be a Borel measurable mapping. Let $\eta \in \mathcal{P}(\mathcal{X})$, $\gamma_0 \in \mathcal{P}(\mathcal{Y})$. Then*

$$R\big(\eta\|\psi(\gamma_0)\big) = \inf_{\gamma\in\mathcal{P}(\mathcal{Y}):\psi(\gamma)=\eta} R\big(\gamma\|\gamma_0\big), \tag{A.2}$$

where $\inf\emptyset = \infty$ *by convention.*

Proof. Suppose $\gamma \in \mathcal{P}(\mathcal{Y})$ is such that $\psi(\gamma) = \eta$. Then, by (A.1) and the definition of image measure,

$$\begin{aligned}
R\big(\eta\|\psi(\gamma_0)\big) &= \sup_{\pi\in\Pi_{\mathcal{X}}} \sum_{A\in\pi} \eta(A)\log\left(\frac{\eta(A)}{\psi(\gamma_0)(A)}\right)\\
&= \sup_{\pi\in\Pi_{\mathcal{X}}} \sum_{A\in\pi} \gamma(\psi^{-1}(A))\log\left(\frac{\gamma(\psi^{-1}(A))}{\gamma_0(\psi^{-1}(A))}\right)\\
&= \sup_{\pi\in\Pi_{\mathcal{X}}} \sum_{B\in\psi^{-1}(\pi)} \gamma(B)\log\left(\frac{\gamma(B)}{\gamma_0(B)}\right)\\
&\leq \sup_{\hat{\pi}\in\Pi_{\mathcal{Y}}} \sum_{B\in\hat{\pi}} \gamma(B)\log\left(\frac{\gamma(B)}{\gamma_0(B)}\right)
\end{aligned}$$

$$= R(\gamma\|\gamma_0),$$

where $\psi^{-1}(\pi)$ denotes the partition of $\mathcal{Y}$ induced by the inverse images of ψ. More precisely, $\psi^{-1}(\pi) \doteq \{\psi^{-1}(A) : A \in \pi\}$. Notice that $\psi^{-1}(\pi)$ is indeed a finite and measurable partition of $\mathcal{Y}$ since π is a finite and measurable partition of $\mathcal{X}$, inverse images under ψ are Borel measurable and $\psi^{-1}(A) \cap \psi^{-1}(\tilde{A}) = \emptyset$ whenever $A \cap \tilde{A} = \emptyset$. Since $\inf \emptyset = \infty$, it follows that

$$R\big(\eta\|\psi(\gamma_0)\big) \leq \inf_{\gamma \in \mathcal{P}(\mathcal{Y}):\psi(\gamma)=\eta} R\big(\gamma\|\gamma_0\big).$$

If $R\big(\eta\|\psi(\gamma_0)\big) = \infty$, then the above inequality is necessarily an equality, namely $\infty = \infty$. Thus in order to show the opposite inequality, we may assume that $R\big(\eta\|\psi(\gamma_0)\big) < \infty$. Now $R\big(\eta\|\psi(\gamma_0)\big) < \infty$ implies that η is absolutely continuous with respect to $\psi(\gamma_0)$, hence possesses a density $f \doteq \frac{d\eta}{d\psi(\gamma_0)}$. Set

$$\gamma(C) \doteq \int_C f(\psi(y))\gamma_0(dy), \quad C \in \mathcal{B}(\mathcal{Y}).$$

Then γ is a probability measure having density $f \circ \psi$ with respect to γ_0. Using the integral transformation formula and definition of f, we have for all $A \in \mathcal{B}(\mathcal{X})$,

$$\begin{aligned}
\psi(\gamma)(A) &= \int_{\mathcal{Y}} \mathbf{1}_{\psi^{-1}(A)}(y) \cdot f(\psi(y))\, \gamma_0(dy) \\
&= \int_{\mathcal{Y}} \mathbf{1}_A(\psi(y)) \cdot f(\psi(y))\, \gamma_0(dy) \\
&= \int_{\mathcal{X}} \mathbf{1}_A(x) \cdot f(x)\, \psi(\gamma_0)(dx) \\
&= \eta(A),
\end{aligned}$$

which means that $\psi(\gamma) = \eta$. Recalling that $f \circ \psi = \frac{d\gamma}{d\gamma_0}$, $f = \frac{d\eta}{d\psi(\gamma_0)}$,

$$\begin{aligned}
R\big(\gamma\|\gamma_0\big) &= \int_{\mathcal{Y}} f(\psi(y)) \log\big(f(\psi(y))\big)\gamma_0(dy) \\
&= \int_{\mathcal{X}} f(x) \log\big(f(x)\big)\psi(\gamma_0)(dx) \\
&= R\big(\eta\|\psi(\gamma_0)\big),
\end{aligned}$$

which proves inequality "$\geq$" in (A.2). □

The proof of Lemma A.1 shows that the probability measure γ defined by $\gamma(dy) \doteq \frac{d\eta}{d\psi(\gamma_0)}(\psi(y))\gamma_0(dy)$ attains the infimum in (A.2) whenever that infimum is finite.

B Relative entropy with respect to Wiener measure

Let $\mathcal{Y}$ be the Polish space $\mathbf{C}([0,T],\mathbb{R}^d)$ equipped with the maximum norm topology. Let $\mathcal{U}$ be defined as in Section 5 with $d_1 = d$. Thus $\mathcal{U}$ is the set of quadruples $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),u,W)$ such that the pair $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t))$ forms a stochastic basis satisfying the usual hypotheses, W is a d-dimensional $(\mathcal{F}_t)$-Wiener process, and u is an $\mathbb{R}^d$-valued $(\mathcal{F}_t)$-progressively measurable process with $\mathbf{E}\left[\int_0^T |u(t)|^2 dt\right] < \infty$. Given $u \in \mathcal{U}$, define $\bar{Y}^u$ according to (5.5), that is,

$$\bar{Y}^u(t) \doteq W(t) + \int_0^t u(s)ds, \quad t \in [0,T].$$

The following result provides a variational representation of relative entropy with respect to Wiener measure.

Lemma B.1. *Let γ_0 be Wiener measure on $\mathcal{B}(\mathcal{Y})$. Then for all $\gamma \in \mathcal{P}(\mathcal{Y})$,*

$$R(\gamma\|\gamma_0) = \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{Y}^u)=\gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right], \tag{B.1}$$

where $\inf\emptyset = \infty$ *by convention.*

The proof of inequality "$\leq$" in (B.1) relies on the lower semicontinuity of relative entropy and the Donsker-Varadhan variational formula; it may be confronted to the first part of the proof of Theorem 3.1 in Boué and Dupuis [1998]. The proof of inequality "$\geq$" exploits the variational formulation and uses arguments contained in Föllmer [1985, 1986].

Proof of Lemma B.1. In order to prove inequality "$\leq$" in (B.1), it suffices to show that, for all $u \in \mathcal{U}$,

$$R\big(\mathrm{Law}(\bar{Y}^u)\|\gamma_0\big) \leq \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]. \tag{B.2}$$

Let $u \in \mathcal{U}$, and set $\gamma \doteq \mathrm{Law}(\bar{Y}^u) = \mathbf{P}\circ(\bar{Y}^u)^{-1}$. In accordance with Definition 3.2.3 in Karatzas and Shreve [1991, p. 132], a process v defined on $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t))$ is called simple if there are $N \in \mathbb{N}$, $0 = t_0 < \ldots < t_N = T$, and uniformly bounded $\mathbb{R}^d$-valued random variables $\xi_0,\ldots,\xi_N$ such that ξ_i is $\mathcal{F}_{t_i}$-measurable and

$$v(t,\omega) = \xi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=0}^N \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t).$$

By Proposition 3.2.6 in Karatzas and Shreve [1991, p. 134], there exists a sequence $(v_n)_{n\in\mathbb{N}}$ of simple processes such that $\mathbf{E}\left[\int_0^T |u(t)-v_n(t)|^2 dt\right] \to 0$

as $n \to \infty$. Let $(v_n)_{n\in\mathbb{N}}$ be such a sequence. For $n \in \mathbb{N}$, set $\gamma_n \doteq \mathrm{Law}(\bar{Y}^{v_n})$. Then $\gamma_n \to \gamma$ in $\mathcal{P}(\mathcal{Y})$ since

$$\mathbf{E}\left[\sup_{t\in[0,T]} |\bar{Y}^u(t) - \bar{Y}^{v_n}(t)|^2\right] \leq T \cdot \mathbf{E}\left[\int_0^T |u(t) - v_n(t)|^2 dt\right] \stackrel{n\to\infty}{\longrightarrow} 0.$$

Therefore, by the lower semicontinuity of $R(.\|\gamma_0)$,

$$R\big(\mathrm{Law}(\bar{Y}^u)\|\gamma_0\big) = R\big(\gamma\|\gamma_0\big) \leq \liminf_{n\to\infty} R\big(\gamma_n\|\gamma_0\big) = \liminf_{n\to\infty} R\big(\mathrm{Law}(\bar{Y}^{v_n})\|\gamma_0\big).$$

On the other hand, $\mathbf{E}\left[\frac{1}{2}\int_0^T |v_n(t)|^2 dt\right] \to \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]$ as $n \to \infty$. It is therefore enough to show that (B.2) holds whenever u is a simple process. Thus assume that u is simple. Let Z be the $\mathcal{F}_T$-measurable $(0,\infty)$-valued random variable given by

$$Z \doteq \exp\left(-\int_0^T u(s)\cdot dW(s) - \frac{1}{2}\int_0^T |u(s)|^2 ds\right).$$

Notice that $\mathbf{E}[Z] = 1$ since u is uniformly bounded. Define a probability measure $\tilde{\mathbf{P}}$ on $(\Omega, \mathcal{F}_T)$ by

$$\frac{d\tilde{\mathbf{P}}}{d\,\mathbf{P}} \doteq Z.$$

By Girsanov's theorem [Theorem 3.5.1 in Karatzas and Shreve, 1991, p. 191], $\bar{Y}^u$ is an $(\mathcal{F}_t)$-Wiener process with respect to $\tilde{\mathbf{P}}$. By the Donsker-Varadhan variational formula for relative entropy [Lemma 1.4.3(a) in Dupuis and Ellis, 1997, p. 29],

$$R\big(\gamma\|\gamma_0\big) = \sup_{g\in \boldsymbol{C}_b(\mathcal{Y})}\left\{\int_{\mathcal{Y}} g(y)\,\gamma(dy) - \log\int_{\mathcal{Y}} e^{g(y)}\,\gamma_0(dy)\right\}. \tag{B.3}$$

Recall that $\gamma = \mathbf{P}\circ(\bar{Y}^u)^{-1}$ and $\gamma_0 = \mathbf{P}\circ W^{-1}$, but also $\gamma_0 = \tilde{\mathbf{P}}\circ(\bar{Y}^u)^{-1}$ since $\bar{Y}^u$ is a Wiener process under $\tilde{P}$. Let $\tilde{\mathbf{E}}$ denote expectation with respect to $\tilde{\mathbf{P}}$. By the convexity of $-\log$ and Jensen's inequality, for all $g \in \boldsymbol{C}_b(\mathcal{Y})$,

$$\begin{aligned}
&\int_{\mathcal{Y}} g(y)\,\gamma(dy) - \log\int_{\mathcal{Y}} e^{g(y)}\,\gamma_0(dy)\\
&= \mathbf{E}\left[g(\bar{Y}^u)\right] - \log\mathbf{E}\left[\exp\left(g(W)\right)\right]\\
&= \mathbf{E}\left[g(\bar{Y}^u)\right] - \log\tilde{\mathbf{E}}\left[\exp\left(g(\bar{Y}^u)\right)\right]\\
&= \mathbf{E}\left[g(\bar{Y}^u)\right] - \log\mathbf{E}\left[\exp\left(g(\bar{Y}^u)\right)\cdot Z\right]\\
&= \mathbf{E}\left[g(\bar{Y}^u)\right] - \log\mathbf{E}\left[\exp\left(g(\bar{Y}^u) - \int_0^T u(t)\cdot dW(t) - \frac{1}{2}\int_0^T |u(t)|^2 dt\right)\right]\\
&\leq \mathbf{E}\left[g(\bar{Y}^u)\right] - \mathbf{E}\left[g(\bar{Y}^u) - \int_0^T u(t)\cdot dW(t) - \frac{1}{2}\int_0^T |u(t)|^2 dt\right]\\
&= \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]
\end{aligned}$$

since $\mathbf{E}\left[\int_0^T u(t)\cdot dW(t)\right] = 0$ as u is square integrable. In view of (B.3), inequality (B.2) follows.

In order to prove inequality "$\geq$" in (B.1), it suffices to consider probability measures with finite relative entropy with respect to Wiener measure. Let $\gamma \in \mathcal{P}(\mathcal{Y})$ be such that $R(\gamma\|\gamma_0) < \infty$. In particular, γ is absolutely continuous with respect to γ_0. We have to show that there exists $u \in \mathcal{U}$ such that $\mathrm{Law}(\bar{Y}^u) = \gamma$ and $R(\gamma\|\gamma_0) \geq \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right]$. Let Y be the coordinate process on the canonical space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, and let $(\mathcal{B}_t)_{t\in[0,T]}$ be the canonical filtration (the natural filtration of Y). Denote by $(\hat{\mathcal{B}}_t)$ the γ_0-augmentation of $(\mathcal{B}_t)$. Both γ_0 and γ extend naturally to $\hat{\mathcal{B}}_T \supset \mathcal{B}(\mathcal{Y})$. Clearly, Y is a $(\hat{\mathcal{B}}_t)$-Wiener process under γ_0. Since $R(\gamma\|\gamma_0) < \infty$, there is a $[0,\infty)$-valued $\hat{\mathcal{B}}_T$-measurable random variable ξ such that

$$\frac{d\gamma}{d\gamma_0} = \xi, \qquad \mathbf{E}_{\gamma_0}[\xi] = 1, \qquad \mathbf{E}_{\gamma}[|\log(\xi)|] = \mathbf{E}_{\gamma_0}[|\log(\xi)|\xi] < \infty.$$

Set $Z(t) \doteq \mathbf{E}_{\gamma_0}[\xi|\hat{\mathcal{B}}_t]$, $t \in [0,T]$. By a version of Itô's martingale representation theorem [Theorem III.4.33 in Jacod and Shiryaev, 2003, p. 189], there exists an $\mathbb{R}^d$-valued $(\hat{\mathcal{B}}_t)$-progressively measurable process v such that $\gamma_0(\int_0^T |v(t)|^2 dt < \infty) = 1$ and

(B.4) $$Z(t) = 1 + \int_0^t v(s)\cdot dY(s) \text{ for all } t \in [0,T], \quad \gamma_0\text{-a.s.}$$

In particular, Z is a continuous process. By the continuity and martingale property of Z, and since $Z(T) = \xi$,

$$\gamma\left(\inf_{t\in[0,T]} Z(t) > 0\right) = 1.$$

Define an $\mathbb{R}^d$-valued $(\hat{\mathcal{B}}_t)$-progressively measurable process u by

(B.5) $$u(t) \doteq \frac{1}{Z(t)}\cdot v(t)\cdot \mathbf{1}_{\{\inf_{s\in[0,t]} Z(s)>0\}}, \quad t \in [0,T].$$

Thus $u(t) = v(t)/Z(t)$ γ-almost surely. Applying Itô's formula to calculate $\log(Z(t))$ (more precisely, Itô's formula is applied to $\varphi_\varepsilon(Z(t))$ with $\varphi_\varepsilon \in \mathbf{C}^2(\mathbb{R})$ such that $\varphi_\varepsilon(x) = \log(x)$ for all $x \geq \varepsilon > 0$), one checks that

(B.6)
$$Z(t) = \exp\left(\int_0^t u(s)\cdot dY(s) - \frac{1}{2}\int_0^t |u(s)|^2 dt\right) \text{ for all } t \in [0,T], \quad \gamma\text{-a.s.}$$

Set $\tilde{Y}(t) \doteq Y(t) - \int_0^t u(s)ds$, $t \in [0,T]$. Then $\tilde{Y}$ is a $(\hat{\mathcal{B}}_t)$-Wiener process with respect to γ. Clearly, $\tilde{Y}$ is continuous and $(\hat{\mathcal{B}}_t)$-adapted. Since γ is absolutely continuous with respect to γ_0, the quadratic covariation processes of Y are the same with respect to γ_0 as with respect to γ. Since $\int_0^{\cdot} u(t)dt$ is a process of finite total variation with γ-probability one, it follows that $\tilde{Y}$ has the same quadratic covariations under γ as Y under γ_0. In view of Lévy's characterization of the Wiener process [Theorem 3.3.16 in Karatzas and Shreve, 1991, p. 157], it suffices to check that $\tilde{Y}$ is a local martingale with respect to $(\hat{\mathcal{B}}_t)$ and γ. But this follows from the version of Girsanov's theorem provided by Theorem III.3.11 in Jacod and Shiryaev [2003, pp. 168-169] and the fact that, thanks to (B.4), the quadratic covariations of the continuous processes Y_i, $i \in \{1, \dots, d\}$, and Z are given by

$$[Y_i, Z](t) = \langle Y_i, Z \rangle (t) = \int_0^t v_i(s)ds \text{ for all } t \in [0,T], \quad \gamma_0\text{-a.s.},$$

and $v(t) = u(t) \cdot Z(t)$ γ-almost surely. For $n \in \mathbb{N}$, define a $(\hat{\mathcal{B}}_t)$-stopping time τ_n by

$$\tau_n \doteq \inf \left\{ t \geq 0 : \int_0^t |u(s)|^2 ds > n \right\} \wedge T.$$

Set

$$\xi_n \doteq \exp\left(\int_0^{\tau_n} u(s) \cdot dY(s) - \frac{1}{2} \int_0^{\tau_n} |u(s)|^2 ds \right).$$

Then ξ_n is well-defined with $\xi_n > 0$ γ_0-almost surely (hence also γ-almost surely). By Novikov's criterion [Corollary 3.5.13 in Karatzas and Shreve, 1991, p. 199] and the version of Girsanov's theorem cited in the first part of the proof,

$$\frac{d\gamma_n}{\gamma_0} \doteq \xi_n$$

defines a probability measure γ_n which is equivalent to γ_0. As a consequence, γ is absolutely continuous with respect to γ_n with density given by ξ / ξ_n. It follows that

$$\begin{aligned}
R(\gamma \| \gamma_0) &= \mathbf{E}_\gamma \left[\log(\xi) \right] \\
&= \mathbf{E}_\gamma \left[\log\left(\frac{\xi}{\xi_n} \right) \right] + \mathbf{E}_\gamma \left[\log(\xi_n) \right] \\
&= R(\gamma \| \gamma_n) + \mathbf{E}_\gamma \left[\int_0^{\tau_n} u(s) \cdot dY(s) - \frac{1}{2} \int_0^{\tau_n} |u(s)|^2 ds \right] \\
&= R(\gamma \| \gamma_n) + \mathbf{E}_\gamma \left[\int_0^{\tau_n} u(s) \cdot d\tilde{Y}(s) \right] + \mathbf{E}_\gamma \left[\frac{1}{2} \int_0^{\tau_n} |u(s)|^2 ds \right] \\
&= R(\gamma \| \gamma_n) + \mathbf{E}_\gamma \left[\frac{1}{2} \int_0^{\tau_n} |u(s)|^2 ds \right]
\end{aligned}$$

since $\tilde{Y}$ is a γ-Wiener process and $\int_0^T \mathbf{1}_{[0,\tau_n]}(s) \cdot |u(s)|^2 ds \leq n$ by construction of τ_n. Since relative entropy is nonnegative and $\mathbf{E}_\gamma \left[\frac{1}{2}\int_0^{\tau_n} |u(s)|^2 ds\right] \to \mathbf{E}_\gamma \left[\frac{1}{2}\int_0^{T} |u(s)|^2 ds\right]$ in $[0,\infty]$ as $n \to \infty$ by monotone convergence, we obtain

$$R(\gamma\|\gamma_0) \geq \mathbf{E}_\gamma \left[\frac{1}{2}\int_0^T |u(s)|^2 ds\right]. \tag{B.7}$$

Since $R(\gamma\|\gamma_0) < \infty$ by assumption, also $\mathbf{E}_\gamma \left[\frac{1}{2}\int_0^T |u(s)|^2 ds\right] < \infty$, which together with (B.6) actually implies equality in (B.7).

Now we are in a position to choose $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),u,W) \in \mathcal{U}$ such that

$$\mathbf{P} \circ \left(W + \int_0^\cdot u(s)ds\right)^{-1} = \gamma \quad \text{and} \quad R(\gamma\|\gamma_0) \geq \mathbf{E}\left[\frac{1}{2}\int_0^T |u(s)|^2 ds\right].$$

Take $\Omega \doteq \mathcal{Y}$, let $\mathcal{F}$ be the γ-completion of $\mathcal{B}_T$, and take $\mathbf{P}$ equal to γ, extended to the additional null sets. Let $(\mathcal{F}_t)$ be the γ-augmentation of $(\mathcal{B}_t)$. Notice that $\hat{\mathcal{B}}_t \subseteq \mathcal{F}_t$, $t \in [0,T]$, and that $(\mathcal{F}_t)$ satisfies the usual hypotheses. Define the control process u according to (B.5), and set $W \doteq \tilde{Y}$. Then W is an $(\mathcal{F}_t)$-Wiener process under $\mathbf{P}$ and

$$\mathbf{P} \circ \left(W + \int_0^\cdot u(s)ds\right)^{-1} = \gamma \circ \left(\tilde{Y} + \int_0^\cdot u(s)ds\right)^{-1} = \gamma \circ Y^{-1} = \gamma$$

since Y is the identity on $\mathcal{Y} = \Omega$. Finally, by (B.7),

$$R(\gamma\|\gamma_0) \geq \mathbf{E}\left[\frac{1}{2}\int_0^T |u(s)|^2 ds\right],$$

where expectation is taken with respect to $\mathbf{P} = \gamma$. □

Remark B.1. Lemma B.1 allows to derive a version of Theorem 3.1 in Boué and Dupuis [1998], the representation theorem for Laplace functionals with respect to a Wiener process. The starting point here as there is the following abstract representation formula for Laplace functionals [Proposition 1.4.2 in Dupuis and Ellis, 1997, p. 27]. Let $\mathcal{S}$ be a Polish space, $\nu \in \mathcal{P}(\mathcal{S})$. Then for all $f\colon \mathcal{S} \to \mathbb{R}$ bounded and measurable,

$$-\log \int_{\mathcal{S}} e^{-f(x)}\nu(dx) = \inf_{\mu\in\mathcal{P}(\mathcal{S})} \left\{R(\mu\|\nu) + \int_{\mathcal{S}} f(x)\mu(dx)\right\}. \tag{B.8}$$

With $\mathcal{S} = \mathcal{Y}$, $\nu = \gamma_0$ Wiener measure as above, Equation (B.8) and Lemma B.1

imply that

$$\begin{aligned}
&-\log \int_{\mathcal{Y}} e^{-f(y)}\gamma_0(dy) \\
&= \inf_{\gamma\in\mathcal{P}(\mathcal{Y})} \left\{ \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{Y}^u)=\gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt\right] + \int_{\mathcal{Y}} f(y)\gamma(dy) \right\} \\
&= \inf_{\gamma\in\mathcal{P}(\mathcal{Y})} \inf_{u\in\mathcal{U}:\mathrm{Law}(\bar{Y}^u)=\gamma} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt + f\left(\bar{Y}^u\right)\right] \\
&= \inf_{u\in\mathcal{U}} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt + f\left(\bar{Y}^u\right)\right].
\end{aligned}$$

Let $\hat{W}$ be a standard d-dimensional Wiener process over time $[0,T]$ defined on some probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{P}})$. Since $\int_{\mathcal{Y}} e^{-f(y)}\gamma_0(dy) = \mathbf{E}_{\hat{\mathbf{P}}}\left[e^{-f(W)}\right]$, it follows that for all $f\colon \mathcal{S}\to\mathbb{R}$ bounded and measurable,

$$-\log \mathbf{E}_{\hat{\mathbf{P}}}\left[e^{-f(W)}\right] = \inf_{u\in\mathcal{U}} \mathbf{E}\left[\frac{1}{2}\int_0^T |u(t)|^2 dt + f\left(\bar{Y}^u\right)\right]. \tag{B.9}$$

The difference with the formula as stated in Boué and Dupuis [1998] lies in the fact that the control processes there all live on the canonical space and are adapted to the canonical filtration, while here the stochastic bases for the control processes may vary; also cf. the related representation formula in Budhiraja and Dupuis [2000], where the control processes are allowed to be adapted to filtrations larger than that induced by the driving Wiener process.

C Sufficient conditions for Hypotheses (H1) – (H5)

As in Section 5, let b, σ be predictable functionals on $[0,T]\times\mathcal{X}\times\mathcal{P}(\mathbb{R}^d)$ with values in $\mathbb{R}^d$ and $\mathbb{R}^{d\times d_1}$, respectively. Let d_{bL} be the bounded Lipschitz metric on $\mathcal{P}(\mathbb{R}^d)$, that is,

$$d_{bL}(\nu,\tilde{\nu}) \doteq \sup\left\{ \int_{\mathbb{R}^d} f(x)\nu(dx) - \int_{\mathbb{R}^d} f(x)\tilde{\nu}(dx) : \|f\|_{bL} \leq 1 \right\},$$

where $\|.\|_{bL}$ is defined for functions $f\colon \mathbb{R}^d\to\mathbb{R}$ by

$$\|f\|_{bL} \doteq \sup_{x\in\mathbb{R}^d} |f(x)| + \sup_{x,y\in\mathbb{R}^d: x\neq y} \frac{|f(x)-f(y)|}{|x-y|}.$$

If X, Y are two $\mathbb{R}^d$-valued random variables defined on the same probability space, then

$$d_{bL}\big(\mathrm{Law}(X), \mathrm{Law}(Y)\big) \leq \mathbf{E}\left[|X-Y|\right].$$

Consider the following local Lipschitz and growth conditions on b, σ.

(L) For every $M \in \mathbb{N}$ there exists $L_M > 0$ such that for all $t \in [0,T]$, all $\varphi, \tilde{\varphi} \in \mathcal{X}$, all $\nu, \tilde{\nu} \in \mathcal{P}(\mathbb{R}^d)$,

$$
\begin{aligned}
&|b(t,\varphi,\nu) - b(t,\tilde{\varphi},\tilde{\nu})| + |\sigma(t,\varphi,\nu) - \sigma(t,\tilde{\varphi},\tilde{\nu})| \\
&\leq L_M \left(\sup_{s\in[0,t]} |\varphi(s) - \tilde{\varphi}(s)| + d_{bL}(\nu,\tilde{\nu}) \right)
\end{aligned}
$$

whenever $\sup_{s\in[0,t]} |\varphi(s)| \vee |\tilde{\varphi}(s)| \leq M$.

(G) There exist a constant $K > 0$ such that for all $t \in [0,T]$, all $\varphi \in \mathcal{X}$, all $\nu \in \mathcal{P}(\mathbb{R}^d)$,

$$
|b(t,\varphi,\nu)| \leq K \left(1 + \sup_{s\in[0,t]} |\varphi(s)| \right), \qquad |\sigma(t,\varphi,\nu)| \leq K.
$$

The boundedness condition on σ is used only in the verification of Hypothesis (H4).

Proposition C.1. *Grant condition (L). Let* $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),u,W) \in \mathcal{U}$*. Suppose that* X*,* $\tilde{X}$ *are solutions of Equation* (5.4) *over the time interval* $[0,T]$ *under control* u *with initial condition* $X(0) = \tilde{X}(0)$ $\mathbf{P}$*-almost surely. Then* X*,* $\tilde{X}$ *are indistinguishable, that is,*

$$
\mathbf{P}\left(X(t) = \tilde{X}(t) \text{ for all } t \in [0,T] \right) = 1.
$$

Proof. For $M \in \mathbb{N}$ define an $(\mathcal{F}_t)$-stopping time τ_M by

$$
\tau_M(\omega) \doteq \inf \left\{ t \in [0,T] : |X(t,\omega)| \vee |\tilde{X}(t,\omega)| \vee \int_0^t |u(s,\omega)|^2 ds \geq M \right\}
$$

with $\inf \emptyset = \infty$. Observe that $\mathbf{P}(\tau_M \leq T) \to 0$ as $M \to \infty$ since X, $\tilde{X}$ are continuous processes and $\mathbf{E}\left[\int_0^T |u(s)|^2 ds\right] < \infty$. Set $\theta(t) \doteq \mathrm{Law}(X(t))$, $\tilde{\theta}(t) \doteq \mathrm{Law}(\tilde{X}(t))$, $t \in [0,T]$. Using Hölder's inequality, Doob's maximal inequality, the Itô isometry, and condition (L), we obtain for $M \in \mathbb{N}$, all

$t \in [0,T]$,

$$
\begin{aligned}
&\mathbf{E}\left[\sup_{s\in[0,t]}\left|X(s\wedge\tau_M)-\tilde{X}(s\wedge\tau_M)\right|^2\right]\\
&\leq 4T\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left|b\big(r,X,\theta(r)\big)-b\big(r,\tilde{X},\tilde{\theta}(r)\big)\right|^2 dr\right]\\
&\quad+4\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left|\sigma\big(r,X,\theta(r)\big)-\sigma\big(r,\tilde{X},\tilde{\theta}(r)\big)\right|^2 dr\cdot\int_0^{t\wedge\tau_M}|u(r)|^2dr\right]\\
&\quad+16\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left|\sigma\big(r,X,\theta(r)\big)-\sigma\big(r,\tilde{X},\tilde{\theta}(r)\big)\right|^2 dr\right]\\
&\leq 4T\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left|b\big(r,X,\theta(r)\big)-b\big(r,\tilde{X},\tilde{\theta}(r)\big)\right|^2 dr\right]\\
&\quad+(4M+16)\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left|\sigma\big(r,X,\theta(r)\big)-\sigma\big(r,\tilde{X},\tilde{\theta}(r)\big)\right|^2 dr\right]\\
&\leq 8L_M^2\,(T+M+4)\,\mathbf{E}\left[\int_0^{t\wedge\tau_M}\left(\sup_{s\in[0,r]}|X(s)-\tilde{X}(s)|^2+d_{bL}(\theta(r),\tilde{\theta}(r))^2\right)dr\right]\\
&\leq 16L_M^2\,(T+M+4)\int_0^t\mathbf{E}\left[\sup_{s\in[0,r]}\left|X(s\wedge\tau_M)-\tilde{X}(s\wedge\tau_M)\right|^2\right]dr.
\end{aligned}
$$

An application of Gronwall's lemma yields that

$$
\mathbf{E}\left[\sup_{s\in[0,T]}\left|X(s\wedge\tau_M)-\tilde{X}(s\wedge\tau_M)\right|^2\right]=0,
$$

hence $\mathbf{P}(X(t)=\tilde{X}(t)$ for all $t\leq\tau_M)=1$ for all $M\in\mathbb{N}$. This implies the assertion since $\tau_M\nearrow\infty$ as $M\to\infty$ $\mathbf{P}$-almost surely. □

Proposition C.1 says that under condition (L) pathwise uniqueness holds for Equation (5.4) with respect to $\mathcal{U}$. As in the classical case of uncontrolled Itô diffusions, pathwise uniqueness implies uniqueness in law. The proof of Proposition C.2 below is in fact analogous to that of Proposition 1 in Yamada and Watanabe [1971]; also cf. Proposition 5.3.20 in Karatzas and Shreve [1991, p. 309]).

Proposition C.2. *Assume that pathwise uniqueness holds for Equation* (5.4) *given any deterministic initial condition. Let* $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),u,W)\in\mathcal{U}$, $((\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbf{P}}),(\tilde{\mathcal{F}}_t),\tilde{u},\tilde{W})\in\mathcal{U}$ *be such that* $\mathbf{P}\circ(u,W)^{-1}=\tilde{\mathbf{P}}\circ(\tilde{u},\tilde{W})^{-1}$ *as probability measures on* $\mathcal{B}(\mathcal{R}_1\times\mathcal{Y})$. *Suppose that* X, $\tilde{X}$ *are solutions of Equation* (5.4) *over the time interval* $[0,T]$ *under control* u *and* $\tilde{u}$, *respectively, with initial condition* x_0 $\mathbf{P}/\tilde{\mathbf{P}}$*-almost surely. Then* $\mathbf{P}\circ(X,u,W)^{-1}=\tilde{\mathbf{P}}\circ(\tilde{X},\tilde{u},\tilde{W})^{-1}$ *as probability measures on* $\mathcal{B}(\mathcal{Z})$.

Proof (sketch). Set $\hat{\mathcal{Z}} \doteq \mathcal{X}\times\mathcal{X}\times\mathcal{R}_1\times\mathcal{Y}$ and $\mathcal{G} \doteq \mathcal{B}(\hat{\mathcal{Z}})$. Let $\hat{Z} = (Z, \tilde{Z}, \rho, \hat{W})$ be the canonical process on $\hat{\mathcal{Z}}$, and let $(\mathcal{G}_t)_{t\in[0,T]}$ be the canonical filtration (i.e., the natural filtration of Z). Let $\mathbf{R}$ be the probability measure on $\mathcal{B}(\mathcal{R}_1 \times \mathcal{Y})$ given by

$$\mathbf{R} \doteq \mathbf{P}\circ(u, W)^{-1} = \tilde{\mathbf{P}} \circ (\tilde{u}, \tilde{W})^{-1}.$$

Let $\mathbf{Q}\colon \mathcal{R}_1\times\mathcal{Y}\times\mathcal{B}(\mathcal{X})$ be a regular conditional distribution of $\mathrm{Law}(X, u, W)$ given (u, W); thus for all $A \in \mathcal{B}(\mathcal{R}_1 \times \mathcal{Y})$, all $B \in \mathcal{B}(\mathcal{X})$,

$$\mathbf{P}\left(X \in B,\ (u, W) \in A\right) = \int_A \mathbf{Q}(r, w; B)\, \mathbf{R}\left(d(r, w)\right).$$

Analogously, let $\tilde{\mathbf{Q}}\colon \mathcal{R}_1 \times \mathcal{Y} \times \mathcal{B}(\mathcal{X})$ be a regular conditional distribution of $\mathrm{Law}(\tilde{X}, \tilde{u}, \tilde{W})$ given $(\tilde{u}, \tilde{W})$. Define $\hat{\mathbf{P}} \in \mathcal{P}(\hat{\mathcal{Z}})$ by setting, for $B, \tilde{B} \in \mathcal{B}(\mathcal{X})$, $A \in \mathcal{B}(\mathcal{R}_1 \times \mathcal{Y})$,

$$\hat{\mathbf{P}}\left(B \times \tilde{B} \times A\right) \doteq \int_A \mathbf{Q}(r, w; B) \cdot \tilde{\mathbf{Q}}(r, w; \tilde{B})\, \mathbf{R}\left(d(r, w)\right).$$

Let $\hat{\mathcal{G}}$ be the $\hat{\mathbf{P}}$-completion of $\mathcal{G}$, and denote by $(\hat{\mathcal{G}}_t)$ the right continuous filtration induced by the $\hat{\mathbf{P}}$-augmentation of $(\mathcal{G}_t)$. Then $((\hat{\mathcal{Z}}, \hat{\mathcal{G}}, \hat{\mathbf{P}}), (\hat{\mathcal{G}}_t), \rho, \hat{W}) \in \mathcal{U}$, where $(\rho, \hat{W})$ are the last two components of the canonical process $\hat{Z}$. One checks that

$$\hat{\mathbf{P}} \circ (Z, \rho, \hat{W})^{-1} = \mathbf{P}\circ(X, u, W)^{-1}, \quad \hat{\mathbf{P}} \circ (\tilde{Z}, \rho, \hat{W})^{-1} = \tilde{\mathbf{P}} \circ (\tilde{X}, \tilde{u}, \tilde{W})^{-1},$$

and that Z, $\tilde{Z}$ are solutions of Equation (5.4) over the time interval $[0, T]$ under control $((\hat{\mathcal{Z}}, \hat{\mathcal{G}}, \hat{\mathbf{P}}), (\hat{\mathcal{G}}_t), \rho, \hat{W}) \in \mathcal{U}$ with initial condition x_0 $\hat{\mathbf{P}}$-almost surely, where ρ is being identified with the control process $v(t) \doteq \int_{\mathbb{R}^{d_1}} y\rho_t(dy)$. By hypothesis, pathwise uniqueness holds for Equation (5.4) with deterministic initial condition; it follows that

$$\hat{\mathbf{P}}\left(Z(t) = \tilde{Z}(t) \text{ for all } t \in [0, T]\right) = 1,$$

which implies $\mathbf{P}\circ(X, u, W)^{-1} = \tilde{\mathbf{P}} \circ (\tilde{X}, \tilde{u}, \tilde{W})^{-1}$. □

The following lemma is used in the verification of Hypothesis (H4).

Lemma C.1. *Let $(\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ be a stochastic basis satisfying the usual hypotheses, and let M be a continuous local martingale with respect to $(\mathcal{F}_t)$ with quadratic variation $\langle M\rangle$. Suppose there exists a finite constant $C > 0$ such that for $\mathbf{P}$-almost all $\omega \in \Omega$, all $t, s \in [0, T]$,*

$$|\langle M\rangle(t, \omega) - \langle M\rangle(s, \omega)| \leq C \cdot |t - s|.$$

Then for every $\delta_0 \in (0, T]$,

$$\mathbf{E}\left[\sup_{\delta\in(0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} |M(t) - M(s)|\right] \leq 192 \cdot \sqrt{C} \cdot (e \cdot T)^{1/4}.$$

Proof. Since the assertion is about the behavior of M only up to time T, we may assume that $\lim_{t\to\infty}\langle M\rangle(t) = \infty$ $\mathbf{P}$-almost surely. For $s \geq 0$ set $\tau_s \doteq \inf\{t \geq 0 : \langle M\rangle(t) > s\}$. Then τ_s is an $(\mathcal{F}_t)$-stopping time for every $s \geq 0$. By the Dambis-Dubins-Schwarz theorem [for instance, Theorem 3.4.6 in Karatzas and Shreve, 1991, p. 174], setting $W(t, \omega) \doteq M(\tau_t(\omega), \omega)$, $t \geq 0$, $\omega \in \Omega$, defines a standard Wiener process with respect to the filtration $(\mathcal{F}_{\tau_t})$ and for $\mathbf{P}$-almost all $\omega \in \Omega$, all $t \geq 0$,

$$M(t, \omega) = W\big(\langle M\rangle(t, \omega), \omega\big).$$

Using the Garsia-Rodemich-Rumsey inequality one can show [cf. Appendix in Fischer and Nappo, 2010] that for every $p \geq 1$, every $\tilde{T} > 0$, there exists a p-integrable random variable $\xi_{p,\tilde{T}}$ such that $\mathbf{E}\left[|\xi_{p,\tilde{T}}|^p\right] \leq 192^p \cdot p^{p/2}$ and for $\mathbf{P}$-almost all $\omega \in \Omega$, all $t, s \in [0, T]$ such that $|t - s| \leq \tilde{T}/e$,

$$|W(t, \omega) - W(s, \omega)| \leq \xi_{p,\tilde{T}}(\omega) \cdot \sqrt{|t-s| \log\left(\tfrac{\tilde{T}}{|t-s|}\right)}.$$

Clearly, $x \mapsto x\log(\tilde{T}/x)$ is increasing on $(0, \tilde{T}/e]$, $\lim_{x\to 0+} x\log(\tilde{T}/x) = 0$, and $(x\log(\tilde{T}/x))^{1/2} \leq (\tilde{T} \cdot x)^{1/4}$ for all $x \in (0, \tilde{T}/e]$. Since $\langle M\rangle$ is non-decreasing with $\langle M\rangle(0) = 0$ and, by hypothesis, $|\langle M\rangle(t) - \langle M\rangle(s)| \leq C \cdot |t-s|$, it follows that for $\mathbf{P}$-almost all $\omega \in \Omega$, every $\delta \in (0, T]$,

$$\begin{aligned}
&\sup_{t,s\in[0,T]:|t-s|\leq\delta} |M(t,\omega) - M(s,\omega)| \\
&= \sup_{t,s\in[0,T]:|t-s|\leq\delta} \left|W\big(\langle M\rangle(t,\omega),\omega\big) - W\big(\langle M\rangle(s,\omega),\omega\big)\right| \\
&\leq \sup_{t,s\in[0,T]:|t-s|\leq\delta} \xi_{p,e\cdot C\cdot T}(\omega) \\
&\qquad \cdot \sqrt{|\langle M\rangle(t,\omega) - \langle M\rangle(s,\omega)| \log\left(\tfrac{e\cdot C\cdot T}{|\langle M\rangle(t,\omega)-\langle M\rangle(s,\omega)|}\right)} \\
&\leq \sup_{t,s\in[0,T]:|t-s|\leq\delta} \xi_{p,e\cdot C\cdot T}(\omega) \cdot \sqrt{C} \cdot \sqrt{\delta \log\left(\tfrac{e\cdot T}{\delta}\right)} \\
&\leq \sqrt{C} \cdot \xi_{p,e\cdot C\cdot T}(\omega) \cdot (e \cdot T \cdot \delta)^{1/4}.
\end{aligned}$$

The assertion follows by choosing p equal to one, inserting the term containing the supremum over $\delta \in (0, \delta_0]$, and taking expectations. □

Proposition C.3. *Conditions (L) and (G) entail hypotheses (H1) – (H5).*

Proof. Hypothesis (H1) is an immediate consequence of conditions (L) and (G). To verify Hypothesis (H2), let $N \in \mathbb{N}$ and define functions $b_N : [0,T] \times \mathcal{X}^N \to \mathbb{R}^{N\times d}$, $\sigma_N : [0,T] \times \mathcal{X}^N \to \mathbb{R}^{N\times d\times N\times d_1}$ according to

$$\begin{aligned} b_N(t,\boldsymbol{\varphi}) &\doteq \Big(b\big(t,\varphi_1,\mu^N_{\boldsymbol{\varphi}(t)}\big),\ldots,b\big(t,\varphi_N,\mu^N_{\boldsymbol{\varphi}(t)}\big)\Big)^{\mathsf{T}}, \\ \sigma_N(t,\boldsymbol{\varphi}) &\doteq \operatorname{diag}\Big(\sigma\big(t,\varphi_1,\mu^N_{\boldsymbol{\varphi}(t)}\big),\ldots,\sigma\big(t,\varphi_N,\mu^N_{\boldsymbol{\varphi}(t)}\big)\Big), \end{aligned}$$

where $\mu^N_{\boldsymbol{\varphi}(t)} \doteq \frac{1}{N}\sum_{i=1}^N \delta_{\varphi_i(t)}$. Then b_N, σ_N are the coefficients for the system of N stochastic differential equations given by (5.1). Thanks to conditions (L) and (G), b_N, σ_N are locally Lipschitz continuous and of sublinear growth. The Itô existence and uniqueness theorem [for instance, Theorem V.12.1 in Rogers and Williams, 2000, p. 132] thus yields pathwise uniqueness and existence of strong solutions for the system of equations (5.1). By Proposition C.1 in conjunction with condition (L), pathwise uniqueness holds for Equation (5.4). By Proposition C.2, it follows that weak uniqueness holds for Equation (5.4); hence Hypothesis (H3) is satisfied.

In order to verify Hypothesis (H4), let $u^N \in \mathcal{U}_N$, $N \in \mathbb{N}$, be such that $\sup_{N\in\mathbb{N}} \frac{1}{N}\sum_{i=1}^N \mathbf{E}\left[\int_0^T |u_i^N(t)|^2 dt\right] < \infty$. For $N \in \mathbb{N}$, let $\bar{\mu}^N$ be the empirical measure of the solution to the system of equations (5.3) under u^N. We have to show that $\{\mathbf{P}_N \circ (\bar{\mu}^N)^{-1} : N \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathcal{P}(\mathcal{X}))$. Choose $\delta_0 \in (0, 1\wedge T]$, and define a function $G : \mathcal{P}(\mathcal{X}) \to [0,\infty]$ by

$$G(\theta) \doteq \int_{\mathcal{X}} \left(|\varphi(0)| + \sup_{\delta\in(0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} |\varphi(t)-\varphi(s)| \right) \theta(d\varphi).$$

Then G is a tightness function, that is, G is measurable and the sublevel sets $\{\theta : G(\theta) \leq c\}$, $c \in [0,\infty)$, are compact in $\mathcal{P}(\mathcal{X})$. This latter property is a consequence of the Ascoli-Arzelà characterization of relatively compact sets in $\mathcal{P}(\mathcal{X})$ [for instance, Theorem 2.4.9 in Karatzas and Shreve, 1991, p. 62], the Markov inequality and Fatou's lemma. We are going to show that $\sup_{N\in\mathbb{N}} \mathbf{E}_N\left[G(\bar{\mu}^N)\right] < \infty$, which implies that $\{\mathbf{P}_N \circ (\bar{\mu}^N)^{-1} : N \in \mathbb{N}\}$ is tight. By construction, for $N \in \mathbb{N}$,

$$\begin{aligned} G(\bar{\mu}^N) &= \frac{1}{N}\sum_{i=1}^N \left(\left|\bar{X}_i^N(0)\right| + \sup_{\delta\in(0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} \left|\bar{X}_i^N(t) - \bar{X}_i^N(s)\right| \right) \\ &= |x_0| + \frac{1}{N}\sum_{i=1}^N \sup_{\delta\in(0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} \bigg| \int_s^t b\big(r,\bar{X}_i^N,\bar{\mu}^N(r)\big)dr \\ &\quad + \int_s^t \sigma\big(r,\bar{X}_i^N,\bar{\mu}^N(r)\big)u_i^N(r)dr + \int_s^t \sigma\big(r,\bar{X}_i^N,\bar{\mu}^N(r)\big)dW_i^N(r) \bigg|. \end{aligned}$$

Thanks to condition (G), for every $i \in \{1, \ldots, N\}$,

$$\sup_{\delta \in (0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} \left| \int_s^t b\big(r, \bar{X}_i^N, \bar{\mu}^N(r)\big) dr \right| \leq K \left(1 + \|X_i^N\|_\infty\right)$$

and, by Hölder's inequality,

$$\begin{aligned} &\sup_{\delta \in (0,\delta_0]} \delta^{-1/4} \cdot \sup_{t,s\in[0,T]:|t-s|\leq\delta} \left| \int_s^t \sigma\big(r, \bar{X}_i^N, \bar{\mu}^N(r)\big) u_i^N(r) dr \right| \\ &\qquad \leq \sqrt{T} K \cdot \sqrt{\int_0^T |u_i^N(t)|^2 dt} \;\leq\; \sqrt{T} K \left(\frac{1}{2} + \frac{1}{2} \int_0^T |u_i^N(t)|^2 dt \right). \end{aligned}$$

The process $\int_0^{\cdot} \sigma(r, \bar{X}_i^N, \bar{\mu}^N(r)) dW_i^N(r)$ is a vector of continuous local martingales which, thanks to condition (G), satisfy the hypothesis of Lemma C.1 with $C = K^2$. It follows that there exists a finite constant $K_T > 0$ depending only on K and T such that

$$\begin{aligned} &\mathbf{E}_N \left[G(\bar{\mu}^N) \right] \\ &\leq |x_0| + K_T \left(1 + \frac{1}{N} \sum_{i=1}^N \mathbf{E}_N \left[\|X_i^N\|_\infty \right] + \frac{1}{N} \sum_{i=1}^N \mathbf{E}_N \left[\int_0^T |u_i^N(t)|^2 dt \right] \right). \end{aligned}$$

Since $\sup_{N\in\mathbb{N}} \frac{1}{N} \sum_{i=1}^N \mathbf{E} \left[\int_0^T |u_i^N(t)|^2 dt \right] < \infty$ by hypothesis, it remains to check that, for some finite constant $\hat{K}_T > 0$ depending only on K and T,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{E}_N \left[\|X_i^N\|_\infty \right] \leq \hat{K}_T \left(1 + \frac{1}{N} \sum_{i=1}^N \mathbf{E}_N \left[\int_0^T |u_i^N(t)|^2 dt \right] \right).$$

But this follows by standard arguments involving localization along the stopping times $\tau_M^N \doteq \inf\{t \in [0,T] : \max_{i\in\{1,\ldots,N\}} \sup_{s\leq t} |X_i^N(s)| \geq M\}$, $M \in \mathbb{N}$, Hölder's inequality, Doob's maximal inequality, Itô's isometry, condition (G), and Gronwall's lemma.

Hypothesis (H5) is again a consequence of the Itô existence and uniqueness theorem since under conditions (L) and (G), given any $\theta \in \mathcal{P}(\mathcal{X})$, the mappings $(t, \varphi) \mapsto b(t, \varphi, \theta(t))$, $(t, \varphi) \mapsto \sigma(t, \varphi, \theta(t))$ are predictable, locally Lipschitz continuous, and of sub-linear growth. □

References

P. Billingsley. *Convergence of Probability Measures.* Wiley series in Probability and Statistics. John Wiley & Sons, New York, 1968.

M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.*, 26(4):1641–1659, 1998.

A. Budhiraja and P. Dupuis. A variational representation for positive functionals of infinite dimensional Brownian motion. *Probab. Math. Statist.*, 20(1):39–61, 2000.

A. Budhiraja, P. Dupuis, and M. Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.*, 40(1):74–102, 2012.

P. Dai Pra and F. den Hollander. McKean-Vlasov limit for interacting random processes in random media. *J. Stat. Phys.*, 84(3/4):735–772, 1996.

D. A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

P. Del Moral and A. Guionnet. Large deviations for interacting particle systems: Applications to non linear filtering. *Stochastic Processes Appl.*, 78(1):69–95, 1998.

P. Del Moral and T. Zajic. A note on the Laplace-Varadhan integral lemma. *Bernoulli*, 9(1):49–65, 2003.

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics*. Springer, New York, 2nd edition, 1998.

B. Djehiche and I. Kaj. The rate function for some measure-valued jump processes. *Ann. Probab.*, 23(3):1414–1438, 1995.

K. R. Duffy. Mean field Markov models of Wireless Local Area Networks. *Markov Process. Related Fields*, 16(2):295–328, 2010.

P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1997.

R. S. Ellis. *Entropy, Large Deviations and Statistical Mechanics*, volume 271 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 1985.

S. Feng. Large deviations for Markov processes with mean field interaction and unbounded jumps. *Probab. Theory Relat. Fields*, 100(2):227–252, 1994a.

S. Feng. Large deviations for empirical process of mean field interacting particle system with unbounded jumps. *Ann. Probab.*, 22(4):2122–2151, 1994b.

M. Fischer and G. Nappo. On the moments of the modulus of continuity of Itô processes. *Stochastic Anal. Appl.*, 28(1):103–122, 2010.

H. Föllmer. An entropy approach to the time reversal of diffusion processes. In M. Metivier and E. Pardoux, editors, *Stochastic differential systems: Proc. IFIP-WG 7/1 Conf., Marseille-Luminy, 1984*, volume 69 of *Lecture Notes in Control and Inform. Sci.*, pages 156–163. Springer-Verlag, Berlin, 1985.

H. Föllmer. Time reversal on Wiener space. In S. A. Albeverio, P. Blanchard, and L. Streit, editors, *Stochastic Processes – Mathematics and Physics: Proc. 1st BiBoS-Symposium, Bielefeld, 1984*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer-Verlag, Berlin, 1986.

J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2nd edition, 2003.

O. Kallenberg. On the existence of universal functional solutions to classical SDEs. *Ann. Probab.*, 24(1):196–205, 1996.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition, 1991.

S. Kullback. *Information Theory and Statistics*. Peter Smith, Gloucester, Massachusetts, reprint of the 1968 Dover edition, 1978.

R. Lassalle. Invertibility of adapted perturbations of the identity on abstract Wiener space. *J. Funct. Anal.*, 262(6):2734–2776, 2012.

C. Léonard. Large deviations for long range interacting particle systems with jumps. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 31(2):289–323, 1995a.

C. Léonard. On large deviations for particle systems associated with spatially homogenous Boltzmann type equations. *Probab. Theory Relat. Fields*, 101 (1):1–44, 1995b.

H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56(6):1907–1911, 1966.

L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Volume 2: Itô Calculus.* Cambridge University Press, Cambridge, 2000.

H. Tanaka. Limit theorems for certain diffusion processes with interaction. In K. Itô, editor, *Stochastic Analysis: Proc. Taniguchi Symposium, Katata & Kyoto, 1982*, volume 32 of *North-Holland Mathematical Library*, pages 469–488. Elsevier, Amsterdam, 1984.

A. S. Üstünel. Entropy, invertibility and variational calculus of adapted shifts on Wiener space. *J. Funct. Anal.*, 257(11):3655–3689, 2009.

S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.

T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.