Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Multilevel large deviations

D.A. Dawson[1] and J. Gärtner[2]

3rd June 1992

[1] Carleton University Ottawa
Department of Mathematics
and Statistics
Ottawa, Canada
K1S 5B6

[2] Institut für Angewandte Analysis
und Stochastik
Hausvogteiplatz 5-7
D – O 1086 Berlin
Germany

Preprint No. 1
Berlin 1992

1991 Mathematics Subject Classification. Primary 60 F 10, 60 K 35, 60 J 60.
Key words and phrases. large deviations, hierarchical systems, interacting diffusions, empirical measures, randomly perturbed systems, McKean–Vlasov interaction.

Herausgegeben vom
Institut für Angewandte Analysis und Stochastik
Hausvogteiplatz 5-7
D – O 1086 Berlin

Fax: + 49 30 2004975
e-Mail (X.400): c=de;a=dbp;p=iaas-berlin;s=preprint
e-Mail (Internet): preprint@iaas-berlin.dbp.de

Multilevel large deviations*

D.A. Dawson[1] and J. Gärtner[2]

[1] Carleton University Ottawa, Dept. of Mathematics and Statistics, Canada K1S 5B6
[2] Institut für Angewandte Analysis und Stochastik, PF 1304, D-O 1086 Berlin, Germany

April 2, 1992

Summary. Let (ξ^N) be a sequence of random variables with values in a topological space which satisfy the large deviation principle. For each M and each N, let $\Xi^{M,N}$ denote the empirical measure associated with M independent copies of ξ^N. As a main result, we show that $(\Xi^{M,N})$ also satisfies the large deviation principle as $M, N \to \infty$. We derive several representations of the associated rate function. These results are then applied to empirical measure processes $\Xi^{M,N}(t) = M^{-1} \sum_{i=1}^{M} \delta_{\xi_i^N(t)}$, $0 \leq t \leq T$, where $(\xi_1^N(t), \ldots, \xi_M^N(t))$ is a system of weakly interacting diffusions with noise intensity $1/N$. This is a continuation of our previous work on the McKean-Vlasov limit and related hierarchical models ([4], [5]).

Contents

Introduction 2
Frequently used notation 8

1. Large deviation systems 8

2. Multilevel large deviations 15
2.1. Main result 16
2.2. Other representations of the rate function 30

3. Randomly perturbed dynamical systems 35
3.1. Non-interacting diffusions 36
3.2. McKean-Vlasov interaction 62

A. Appendix 67
A.1. $\mathcal{M}^{M,N}(X)$-sequences 67
A.2. Freidlin-Wentzell estimates 68

References 71

* Research partially supported by a Natural Science and Engineering Research Council of Canada operating grant

Introduction

In order to introduce the basic idea of multilevel large deviations, we begin with a sequence (ξ^N) of random variables on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a topological space Y. We assume that the sequence (ξ^N) satisfies the *large deviation principle* (as $N \to \infty$) with *scale* γ_N and *rate function* J:

(i) for each open subset G of Y,

$$\liminf_{N\to\infty} \gamma_N^{-1} \log P\left(\xi^N \in G\right) \geq -\inf_{y\in G} J(y);$$

(ii) for each closed subset F of Y,

$$\limsup_{N\to\infty} \gamma_N^{-1} \log P\left(\xi^N \in F\right) \leq -\inf_{y\in F} J(y);$$

(iii) the level sets $\{\, y \in Y : J(y) \leq \rho \,\}$, $\rho \geq 0$, are compact.

For each N, let $\xi_1^N, \xi_2^N, \ldots$ be independent copies of ξ^N, and denote by $\delta_{\xi_i^N}$ the Dirac measure at ξ_i^N. Now consider the empirical measures

$$\Xi^{M,N} := \frac{1}{M}\sum_{i=1}^{M} \delta_{\xi_i^N}$$

and regard them as random variables with values in $\mathcal{M}(Y)$, the space of Radon probability measures on Y furnished with the topology of weak convergence. The objective of this paper is to show that the empirical measures $\Xi^{M,N}$ satisfy the large deviation principle as $M, N \to \infty$ and to identify the rate function.

This question was partially motivated by our study of hierarchical systems of interacting diffusions (see [5], Section 5.2). However, in order to explain the relevant multilevel large deviation problem, we will describe the simpler non-interacting case. Let $\xi(t)$ be a diffusion process on $\mathbb{R}^d$ given by an Itô equation of the form

$$d\xi(t) = b(\xi(t), t)\, dt + dw(t),$$

where $w(t)$ denotes d-dimensional Brownian motion. For each N, let $\xi_1^N(t), \ldots, \xi_N^N(t)$ be independent copies of $\xi(t)$ with not necessarily coinciding non-random starting points $\xi_1^N(0), \ldots, \xi_N^N(0)$ such that $N^{-1}\sum_{j=1}^{N} \delta_{\xi_j^N(0)}$ converges to a measure ν in $\mathcal{M}^I := \mathcal{M}(\mathbb{R}^d)$ as $N \to \infty$. Consider the empirical measure process

$$\Xi^N(t) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_j^N(t)}, \qquad 0 \leq t \leq T.$$

In Dawson and Gärtner [4], Theorem 4.5, it was shown that, under mild conditions on the vector field b, the sequence $(\Xi^N(\cdot))$ of $C([0,T]; \mathcal{M}^I)$-valued random variables satisfies the large deviation principle with scale N and rate function S_ν given by

$$S_\nu(\mu(\cdot)) := \frac{1}{2}\int_0^T \|\dot{\mu}(t) - \mathcal{L}_t^* \mu(t)\|_{\mu(t)}^2 \, dt$$

if $\mu(\cdot) \in C([0,T];\mathcal{M}^I)$ is absolutely continuous and $\mu(0) = \nu$ and equal to $+\infty$ otherwise. Here $\mathcal{L}_t^*$ denotes the formal adjoint of the diffusion operator

$$\mathcal{L}_t := \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2}{(\partial x^i)^2} + \sum_{i=1}^{d} b^i(\cdot,t)\frac{\partial}{\partial x^i}$$

associated with $\xi(t)$. Further,

$$\|\theta\|_\mu^2 := \sup_{f\in\mathcal{D}} \frac{|\langle\theta, f\rangle|^2}{\langle\mu, |\nabla f|^2\rangle}, \qquad \theta \in \mathcal{D}', \tag{0.1}$$

where $\mathcal{D}$ and $\mathcal{D}'$ denote the Schwartz space of infinitely differentiable functions with compact support and the corresponding space of distributions, respectively, and $|\nabla f|^2 = \sum_{i=1}^d (\partial f/\partial x^i)^2$.

Now consider a two-parameter family $\{\xi_{i,j}^{M,N}(t);\ i = 1,\ldots,M,\ j = 1,\ldots,N\}$ of independent copies of the diffusion process $\xi(t)$. The *two-level empirical measure process* is defined by

$$\Xi^{M,N}(t) := \frac{1}{M}\sum_{i=1}^{M} \delta_{\Xi_i^N(t)}, \qquad 0 \le t \le T,$$

where, for each i,

$$\Xi_i^N(t) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_{i,j}^{M,N}(t)}, \qquad 0 \le t \le T.$$

The problem is to show that, if $\Xi^{M,N}(0)$ converges to some measure in $\mathcal{M}^{II} := \mathcal{M}(\mathcal{M}(\mathbb{R}^d))$ as $M, N \to \infty$, then the processes $\Xi^{M,N}(\cdot)$ considered as random variables in $C([0,T];\mathcal{M}^{II})$ satisfy the large deviation principle as $M, N \to \infty$ and to find a suitable representation of the rate function. This large deviation problem has a structure similar to that hypothesized above with the additional complication that the processes $\Xi_1^N(\cdot),\ldots,\Xi_N^N(\cdot)$ are *not identically distributed.* The law of the process $\Xi^{M,N}(\cdot)$ depends on the (non-random) initial measure $\Xi^{M,N}(0)$ which may be viewed as an additional parameter. For this reason, our general results on multilevel large deviations will be formulated in terms of *parametrized* families of probability laws.

Section 1 contains preliminary definitions and results on large deviation systems. In *Section 2* we will prove the general multilevel large deviation result and derive several representations of the associated rate function. Further, as an application of these results, in *Section 3* we will analyse a simple caricature of the hierarchical system of diffusions. The *Appendix* contains some auxiliary proofs which we separated from the main exposition of the material.

In order to provide an introduction to the ideas used in the proofs of the multilevel large deviation results in Section 2, we will now state the main result and sketch the proof in the simple case in which there is no parametrization and in which the space Y consists of a finite number of points, $Y = \{y_1, \ldots, y_r\}$. As before, let (ξ^N) be a sequence of Y-valued random variables. For each M and each N, let $\Xi^{M,N}$ denote the empirical measure of M independent copies $\xi_1^N, \ldots, \xi_M^N$ of ξ^N.

Theorem 0.1. *Assume that (ξ^N) satisfies the large deviation principle (as $N \to \infty$) with scale γ_N and rate function J. Then $(\Xi^{M,N})$ also satisfies the large deviation principle (as $M, N \to \infty$) having scale $M\gamma_N$ and rate function*

$$S(\nu) := \int_Y J(y)\, \nu(dy), \qquad \nu \in \mathcal{M}(Y).$$

Proof. a) *Lower large deviation bound.* Fix $\nu \in \mathcal{M}(Y)$ and an open neighborhood $U(\nu)$ of ν arbitrarily. It suffices to show that

$$\liminf_{M,N\to\infty} \frac{1}{M\gamma_N} \log P\left(\Xi^{M,N} \in U(\nu)\right) \geq -S(\nu).$$

We choose a partition of $\{1, \ldots, M\}$ into pairwise disjoint sets Λ_k^M of size $|\Lambda_k^M|$, $k = 1, \ldots, r$, such that

$$\lim_{M\to\infty} \frac{|\Lambda_k^M|}{M} = \nu(y_k), \qquad k = 1, \ldots, r. \tag{0.2}$$

There exists $\varepsilon > 0$ such that

$$\tilde{U}(\nu) := \{\, \tilde{\nu} \in \mathcal{M}(Y) : \tilde{\nu}(y_k) > \nu(y_k) - \varepsilon \text{ for all } k \,\} \subseteq U(\nu).$$

It is now easy to verify that for large M and all N,

$$\bigcap_{k=1}^{r} \{\xi_i^N = y_k \text{ for all } i \in \Lambda_k^M\} \subseteq \left\{\Xi^{M,N} \in \tilde{U}(\nu)\right\}.$$

For these M and N,

$$\begin{aligned} P\left(\Xi^{M,N} \in U(\nu)\right) &\geq P\left(\bigcap_{k=1}^{r} \{\xi_i^N = y_k \text{ for all } i \in \Lambda_k^M\}\right) \\ &= \prod_{k=1}^{r} \left[P\left(\xi^N = y_k\right)\right]^{|\Lambda_k^M|}. \end{aligned} \tag{0.3}$$

Since (ξ^N) satisfies the large deviation principle, we have

$$\liminf_{N\to\infty} \frac{1}{\gamma_N} \log P\left(\xi^N = y_k\right) \geq -J(y_k) \qquad \text{for } k = 1, \ldots, r. \tag{0.4}$$

Combining (0.3) with (0.2) and (0.4), we obtain

$$\begin{aligned}\liminf_{M,N\to\infty} \frac{1}{M\gamma_N} \log P\left(\Xi^{M,N} \in U(\nu)\right) \\ \geq \sum_{k=1}^{r} \liminf_{M,N\to\infty} \frac{|\Lambda_k^M|}{M} \frac{1}{\gamma_N} \log P\left(\xi^N = y_k\right) \\ \geq -\sum_{k=1}^{r} \nu(y_k) J(y_k) = -S(\nu).\end{aligned}$$

b) *Upper large deviation bound.* Fix $h > 0$ arbitrarily. Since $\mathcal{M}(Y)$ is compact, it will be sufficient to show that for each $\nu \in \mathcal{M}(Y)$ there exists a neighborhood $U(\nu)$ such that

$$\limsup_{M,N\to\infty} \frac{1}{M\gamma_N} \log P\left(\Xi^{M,N} \in U(\nu)\right) \leq -S(\nu) + h \tag{0.5}$$

if $S(\nu) < \infty$ and $\leq -h$ if $S(\nu) = \infty$. Assume that $J(y_k) < \infty$ for all k and, in particular, $S(\nu) < \infty$. Take $U(\nu) := \{\, \tilde{\nu} \in \mathcal{M}(Y) : S(\tilde{\nu}) > S(\nu) - h \,\}$. Note that $S(\Xi^{M,N}) = M^{-1} \sum_{i=1}^{M} J(\xi_i^N)$. Applying Chebyshev's exponential inequality, we obtain

$$\begin{aligned}P\left(\Xi^{M,N} \in U(\nu)\right) &= P\left(\frac{1}{M}\sum_{i=1}^{M} J(\xi_i^N) > S(\nu) - h\right) \\ &\leq \exp\left\{-\theta M\gamma_N(S(\nu) - h)\right\} E\exp\left\{\theta\gamma_N \sum_{i=1}^{M} J(\xi_i^N)\right\} \\ &\leq \exp\left\{-\theta M\gamma_N(S(\nu) - h)\right\} \left[E\exp\left\{\theta\gamma_N J(\xi^N)\right\}\right]^M\end{aligned}$$

for each $\theta > 0$. Thus, in order to prove (0.5), it will be enough to verify that the expectation on the right of the last inequality remains bounded as $N \to \infty$ for $0 < \theta < 1$. But, since

$$\limsup_{N\to\infty} \frac{1}{\gamma_N} \log P\left(\xi^N = y_k\right) \leq -J(y_k), \qquad \text{for } k = 1, \ldots, r,$$

we have

$$\begin{aligned}\limsup_{N\to\infty} E\exp\left\{\theta\gamma_N J(\xi^N)\right\} &= \limsup_{N\to\infty} \sum_{k=1}^{r} \exp\left\{\theta\gamma_N J(y_k)\right\} P\left(\xi^N = y_k\right) \\ &\leq 1 \qquad \text{for } 0 < \theta < 1.\end{aligned}$$

To handle the case when $J(y_k) = \infty$ for some k, one has to replace J by a function $\tilde{J}$ which coincides with J on $\{\, y : J(y) < \infty \,\}$ and which is 'arbitrarily large, but finite' on $\{\, y : J(y) = \infty \,\}$. Correspondingly, one has to replace S by $\tilde{S}(\mu) := \int \tilde{J}(y)\, \mu(dy)$, $\mu \in \mathcal{M}(Y)$.

c) *The compactness of the level sets* $\{\, \nu \in \mathcal{M}(Y) : S(\nu) \leq \rho \,\}$, $\rho \geq 0$, is obvious in this case. □

We now turn to the simple caricature of the hierarchical system referred to above. Consider the randomly perturbed dynamical system

$$d\,\xi^N(t) = b(\xi^N(t),t)\,dt + N^{-1/2}\,dw(t), \qquad \xi^N(0) = x, \tag{0.6}$$

in $\mathbb{R}^d$ with perturbation parameter $N^{-1/2}$. Then, under certain restrictions on the vector field b, according to Freidlin and Wentzell [7], Chap. 4, Theorem 1.1, the sequence $(\xi^N(\cdot))$ of $C([0,T];\mathbb{R}^d)$-valued random variables satisfies the large deviation principle with scale N and rate function I_x for each starting point $x \in \mathbb{R}^d$. The rate function has the representation

$$I_x(\varphi) := \frac{1}{2}\int_0^T |\dot\varphi(t) - b(\varphi(t),t)|^2\, dt$$

if $\varphi \in C([0,T];\mathbb{R}^d)$ is absolutely continuous and $\varphi(0) = x$ and $I_x(\varphi) = \infty$ otherwise. Given non-random starting points $\xi_1^N(0),\dots,\xi_M^N(0)$, let $\xi_1^N(t),\dots,\xi_M^N(t)$ be M independent processes of the form (0.6), and let

$$\Xi^{M,N}(t) := \frac{1}{M}\sum_{i=1}^{M}\delta_{\xi_i^N(t)}, \qquad 0 \le t \le T, \tag{0.7}$$

denote the associated empirical process. Assume that $\Xi^{M,N}(0) \to \nu$ in $\mathcal{M}(\mathbb{R}^d)$ as $M,N \to \infty$.

As an application of our multilevel large deviation results, we will prove in Section 3.1 that the family $(\Xi^{M,N}(\cdot))$ of $C([0,T];\mathcal{M}(\mathbb{R}^d))$-valued random variables satisfies the large deviation principle (as $M,N\to\infty$) with scale MN and rate function S_ν given by

$$S_\nu(\mu(\cdot)) := \frac{1}{2}\int_0^T \left\|\dot\mu(t) - (\mathcal{L}_t^0)^*\mu(t)\right\|_{\mu(t)}^2\, dt \tag{0.8}$$

if $\mu(\cdot)$ is absolutely continuous in $C([0,T];\mathcal{M}(\mathbb{R}^d))$ and $\mu(0)=\nu$ and $S_\nu(\mu(\cdot)) = \infty$ otherwise. Here the norm $\|\cdot\|_\mu$ is defined by (0.1) and

$$\mathcal{L}_t^0 := \sum_{i=1}^{d} b^i(\cdot,t)\frac{\partial}{\partial x^i}, \qquad 0 \le t \le T,$$

is the family of differential operators associated with the unperturbed dynamical system

$$\dot\varphi(t) = b(\varphi(t),t).$$

It may be noted that, for fixed N, the large deviation results of our previous paper [4] show that the family $(\Xi^{M,N}(\cdot))$ satisfies the large deviation principle as $M\to\infty$ with scale M and rate function NS_ν^N, where S_ν^N is defined by (0.8) except that $\mathcal{L}_t^0$ is replaced by the generator associated with (0.6). As a special case of Theorem 2.9 to be proved below, it will follow that these large deviation bounds are 'uniform' in N and that S_ν^N converges in some sense (but not pointwise!) to the rate function for $(\Xi^{M,N}(\cdot))$ as $M, N \to \infty$. However, in order to identify this rate function with (0.8) under a natural set of weak hypotheses involves a number of nontrivial technical steps carried out in Section 3.1.

Finally, in Section 3.2 we will extend the above result to the corresponding system with mean-field interaction

$$d\xi_i^N(t) = b(\xi_i^N(t); \Xi^{M,N}(t))\,dt + N^{-1/2}\,dw_i(t), \qquad i = 1, \dots, M,$$

where $\Xi^{M,N}(t)$ is again defined by (0.7) and $w_1(t), \dots, w_M(t)$ are independent d-dimensional Wiener processes. As in Dawson and Gärtner [4] and Gärtner [8], in order to treat unbounded drift coefficients b, we consider the processes $\Xi^{M,N}(\cdot)$ as random variables with values in the space $C([0,T]; \mathcal{M}_\infty)$, where $\mathcal{M}_\infty$ is a subset of $\mathcal{M}(\mathbb{R}^d)$ furnished with an 'inductive' topology. More precisely, we introduce a smooth function $\psi: \mathbb{R}^d \to [0,\infty)$ with $\psi(x) \to \infty$ as $|x| \to \infty$ depending on the 'growth of b at infinity', set $\mathcal{M}_R := \{\nu \in \mathcal{M}(\mathbb{R}^d) : \int \psi\, d\nu \le R\}$, $R > 0$, define

$$\mathcal{M}_\infty := \bigcup_{R>0} \mathcal{M}_R,$$

and equip this space with the strongest topology which induces the weak topology on $\mathcal{M}_R$ for each $R > 0$. The space $\mathcal{M}_\infty$ is not metrizable but satisfies a weak metrizability hypothesis formulated in Section 2. Refer to Appendix B in Gärtner [8] for a detailed discussion of the 'inductive' topology and the properties of $\mathcal{M}_\infty$ and $C([0,T]; \mathcal{M}_\infty)$.

Let the drift coefficient $b: \mathbb{R}^d \times \mathcal{M}_\infty \to \mathbb{R}^d$ be continuous and satisfy assumptions analogous to that in Dawson and Gärtner [4], Section 5. Suppose that the non-random initial measures $\Xi^{M,N}(0)$ converge to a measure ν in $\mathcal{M}_\infty$. Then we show that $(\Xi^{M,N}(\cdot))$ again satisfies the large deviation principle with scale MN and rate function S_ν, where S_ν is now defined by (0.8) with the operator $\mathcal{L}_t^0$ replaced by

$$\mathcal{L}^0(\mu(t)) := \sum_{i=1}^d b^i(\cdot\,; \mu(t)) \frac{\partial}{\partial x^i}.$$

The proof of this result is based on a reduction to a system of independent diffusions along the lines of Section 5 of Dawson and Gärtner [4]. At the end of this section we will briefly consider the corresponding McKean-Vlasov equations.

In the situation considered here, the process $\xi^N(t)$ lives in the metric space $\mathbb{R}^d$. But in the case of the interacting hierarchical model mentioned above, the role of $\xi^N(t)$ is played by an empirical measure process $\Xi^N(t)$ which lives in the non-metrizable space $\mathcal{M}_\infty$. Although the results will not be used in this generality in Section 3 of this paper, the latter fact has motivated the development of our main results for families of probability laws on a space Y which are parametrized by a space X, where X and Y are not necessarily metrizable. It should be noted that this introduces a number of technical complications which would not arise in the metrizable case.

Frequently used notation

By $\mathbb{N}$ and $\mathbb{R}^d$ we will denote the set of natural numbers and the d-dimensional Euclidean space, respectively.

Given a Hausdorff topological space X, we will denote by $C_b(X)$ and $\mathcal{M}(X)$ the space of real-valued bounded continuous functions on X with the supremum norm $\|\cdot\|$ and the space of Radon probability measures on the Borel σ-field $\mathcal{B}(X)$ of X furnished with the topology of weak convergence, respectively. By $\langle \nu, f\rangle$ we will abbreviate the integral of $f \in C_b(X)$ with respect to $\nu \in \mathcal{M}(X)$. By δ_x we will denote the Dirac measure at $x \in X$.

$C([0,T];X)$ will stand for the space of continuous functions $[0,T] \to X$. If X is a Polish space, then $C([0,T];X)$ will be endowed with the uniform topology corresponding to a complete separable metric on X.

We will use the abbreviations $\mathcal{M} := \mathcal{M}(\mathbb{R}^d)$, $C_b := C_b(\mathbb{R}^d)$, $C_{0,T} := C([0,T];\mathbb{R}^d)$, and $\mathcal{C}_{0,T} := C([0,T];\mathcal{M}(\mathbb{R}^d))$.

By $\mathcal{D}$ we will denote the Schwartz space of infinitely differentiable functions $\mathbb{R}^d \to \mathbb{R}$ with compact support equipped with the usual inductive topology. The corresponding space of real distributions will be denoted by $\mathcal{D}'$.

Finally, $\overline{A}$ and $\mathbb{1}_A$ will stand for the closure and the indicator function of a set A, respectively.

1. Large deviation systems

Let X and Y denote Hausdorff topological spaces, and let (X_N) be a sequence of subsets of X. A sequence (x_N) of points in X will be called an *X_N-sequence* if $x_N \in X_N$ for each N. We will assume throughout that each point in X is the limit of an X_N-sequence.

Let $\Pi = \{P_x^N; x \in X_N, N \in \mathbb{N}\}$ be a family of Radon probability measures on Y. (A probability measure ν on the Borel σ-field of a Hausdorff space is called a Radon measure if $\nu(A) = \sup\{\, \nu(K) : K \subseteq A, K \text{ compact}\,\}$ for each Borel set A.) Let I be a function on $X \times Y$ taking values in $[0,\infty]$, and introduce the notation

$$I(x;A) := \inf\{\, I(x;y) : y \in A\,\}, \qquad x \in X,\, A \subseteq Y.$$

Finally, let (γ_N) be a sequence of positive numbers tending to infinity as $N \to \infty$.

Definition 1.1. We will say that Π is a *large deviation system* with *rate function* I and *scale* γ_N if the following conditions are satisfied:

(i) *compactness of the level sets:* for each $x \in X$ and each $\rho \geq 0$ the set

$$\Phi(x;\rho) := \{\, y \in Y : I(x;y) \leq \rho\,\}$$

is compact (and, in particular, non-empty);

(ii) *lower large deviation bound:*

$$\liminf_{N\to\infty} \gamma_N^{-1} \log P_{x_N}^N(G) \geq -I(x;G)$$

for each open set G in Y, each $x \in X$, and each X_N-sequence (x_N) tending to x;

(iii) *upper large deviation bound:*

$$\limsup_{N\to\infty} \gamma_N^{-1} \log P^N_{x_N}(F) \le -I(x;F)$$

for each closed set F in Y, each $x \in X$, and each X_N-sequence (x_N) tending to x.

Given $A \subseteq X$ and $\rho \ge 0$, we define

$$\Phi(A;\rho) := \bigcup_{x\in A} \Phi(x;\rho).$$

Sometimes we will assume in addition that the level sets $\Phi(K;\rho)$ are compact for all compact subsets K of X and all $\rho \ge 0$.

Definition 1.2. Suppose that we are given in addition a surjective continuous map $\pi: Y \to X$ such that

$$P^N_x(\pi^{-1}\{x\}) = 1 \qquad \text{for each } N \in \mathbb{N} \text{ and all } x \in X_N.$$

Then we will say that the family Π forms a *special large deviation system* (with respect to π) having *rate function* $J: Y \to [0,\infty]$ and *scale* γ_N if Π is a large deviation system with scale γ_N and rate function

$$I(x;y) := \begin{cases} J(y) & \text{if } \pi(y) = x, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.1}$$

Note that the level sets $\Phi(K;\rho)$ associated with the rate function (1.1) are of the form

$$\Phi(K;\rho) = \{\, y \in Y : \pi(y) \in K, J(y) \le \rho \,\}.$$

A typical example we have in mind is the situation when X is a Polish space, $Y = C([0,T];X)$, $X_N = X$ for all N, $\pi(y(\cdot)) = y(0)$, and $\{P^N_x; x \in X\}$ is a Markov family of probability laws on Y for each N.

Formally, the notion of a large deviation system is more general than that of a special large deviation system, since in the first case the supports of the measures P^N_x are not assumed to be disjoint for different x. Nevertheless, each large deviation system may be regarded as a special large deviation system. To explain this, let $\Pi = \{P^N_x; x \in X_N, N \in \mathbb{N}\}$ be a family of Radon probability measures on Y. Given $N \in \mathbb{N}$ and $x \in X_N$, let us denote by $\tilde{P}^N_x$ the unique extension of the product measure $\delta_x \otimes P^N_x$ to a Radon probability measure on $\tilde{Y} := X \times Y$. (Here δ_x denotes the Dirac measure at x. For Radon extensions see Schwartz [13], Chap. 1, Theorem 17.) We further denote by $\tilde{\pi}$ the canonical projection of $X \times Y$ onto X. Clearly

$$\tilde{P}_x^N(\tilde{\pi}^{-1}\{x\}) = 1 \qquad \text{for each } N \in \mathbb{N} \text{ and all } x \in X_N.$$

Let $\tilde{\Pi}$ denote the family of measures $\tilde{P}_x^N$, $x \in X_N$, $N \in \mathbb{N}$.

Theorem 1.3. Π *is a large deviation system if and only if* $\tilde{\Pi}$ *is a special large deviation system (with respect to* $\tilde{\pi}$*) having the same scale and the same rate function.*

Proof. a) Assume that Π is a large deviation system with scale γ_N and rate function I. We must show that $\tilde{\Pi}$ is a large deviation system with the same scale and rate function $\tilde{I}: X \times \tilde{Y} \to [0, \infty]$ defined by

$$\tilde{I}(x_0; (x, y)) := \begin{cases} I(x; y) & \text{if } x = x_0, \\ +\infty & \text{otherwise.} \end{cases}$$

1^0 Let $\Phi(x; \rho)$ and $\tilde{\Phi}(x; \rho)$ denote the level sets associated with I and $\tilde{I}$, respectively. Since the sets $\Phi(x; \rho)$ are compact and $\tilde{\Phi}(x; \rho) = \{x\} \times \Phi(x; \rho)$, the level sets $\tilde{\Phi}(x; \rho)$, $x \in X$, $\rho \geq 0$, are also compact.

2^0 We next derive the lower large deviation bound for $\tilde{\Pi}$. Given $(x, y) \in X \times Y$ and open neighborhoods U and V of x and y, respectively, it suffices to check that

$$\liminf_{N \to \infty} \gamma_N^{-1} \log \tilde{P}_{x_N}^N(U \times V) \geq -I(x; y)$$

for each X_N-sequence (x_N) tending to x (cf. e.g. Freidlin and Wentzell [7], Chap. 3, Theorem 3.3). But this is immediate from the definition of $\tilde{P}_{x_N}^N$ and the lower large deviation bound for the measures $P_{x_N}^N$.

3^0 To derive the upper large deviation bound for $\tilde{\Pi}$, we fix $x \in X$, an X_N-sequence (x_N) with $x_N \to x$, and a closed subset F of $X \times Y$ arbitrarily. We must check that

$$\limsup_{N \to \infty} \gamma_N^{-1} \log \tilde{P}_{x_N}^N(F) \leq -\rho \tag{1.2}$$

for each $\rho \geq 0$ with $F \cap \tilde{\Phi}(x; \rho) = \emptyset$. Let therefore $\rho \geq 0$ be such that $F \cap \tilde{\Phi}(x; \rho) = \emptyset$ (provided that such ρ exists at all). Since $\Phi(x; \rho)$ is compact and $\tilde{\Phi}(x; \rho) = \{x\} \times \Phi(x; \rho)$, we find open neighborhoods U and W of x and $\Phi(x; \rho)$, respectively, such that $U \times W$ does not intersect F. Thus, for sufficiently large N,

$$\tilde{P}_{x_N}^N(F) \leq \tilde{P}_{x_N}^N((U \times W)^c) = P_{x_N}^N(W^c).$$

(The upper index c denotes the operation of taking the complement.) Hence, applying the upper large deviation bound for the probabilities on the right, we arrive at (1.2).

b) Suppose that $\tilde{\Pi}$ is a special large deviation system (with respect to $\tilde{\pi}$). Since, for each N and each $x \in X_N$, P_x^N is the image of the measure $\tilde{P}_x^N$ with respect to the canonical projection $X \times Y \to Y$, an application of the 'contraction principle' (see e.g. Varadhan [16], Theorem 2.4) yields that Π is a large deviation system having the same scale and the same rate function as $\tilde{\Pi}$. □

Fix a compact subset K of X and $\rho \geq 0$ arbitrarily, and consider the level sets

$$\Phi(K;\rho) = \{\, y \in Y : I(x;y) \leq \rho \text{ for some } x \in K \,\}$$

and

$$\tilde{\Phi}(K;\rho) = \{\, (x,y) \in X \times Y : x \in K, I(x;y) \leq \rho \,\}$$

associated with the large deviation systems Π and $\tilde{\Pi}$, respectively. Since $\Phi(K;\rho)$ is the continuous image of $\tilde{\Phi}(K;\rho)$ with respect to the projection $\tilde{\pi}$, the compactness of $\tilde{\Phi}(K;\rho)$ implies the compactness of $\Phi(K;\rho)$. On the other hand, $\tilde{\Phi}(K;\rho) \subseteq K \times \Phi(K;\rho)$. Therefore, the compactness of $\Phi(K;\rho)$ implies at least the relative compactness of $\tilde{\Phi}(K;\rho)$. In the next lemma we will see that $\tilde{\Phi}(K;\rho)$ is compact under the following additional assumption on X and (X_N).

Countability Hypothesis. For each compact subset K of X there exists a set $X(K)$, $K \subseteq X(K) \subseteq X$, such that each point of K has a countable base of neighborhoods in $X(K)$ and is the limit of an X_N-sequence which belongs to $X(K)$ for all but finitely many terms.

Note that this hypothesis is fulfilled if X satisfies the first countability axiom.

Lemma 1.4. *Let the Countability Hypothesis be fulfilled. Assume that Π is a large deviation system with rate function I and scale γ_N. Given a compact subset K of X and $\rho \geq 0$, suppose that the level set $\Phi(K;\rho)$ is compact. Then $\tilde{\Phi}(K;\rho)$ is also compact. In particular, I is lower semi-continuous on $K \times Y$ for each compact subset K of X.*

Proof. The level set $\tilde{\Phi}(K;\rho)$ has the form

$$\tilde{\Phi}(K;\rho) = (K \times Y) \cap \{I \leq \rho\}. \tag{1.3}$$

It only remains to show that this set is closed. To this end we fix $\varepsilon > 0$ and $(x_0, y_0) \in K \times Y$ arbitrarily. It will be enough to check that there exist open neighborhoods U and V of x_0 and y_0, respectively, such that

$$\inf\{\, I(x;y) : x \in U \cap K, y \in V \,\} \geq I(x_0, y_0) - \varepsilon. \tag{1.4}$$

Set $\rho_0 := I(x_0, y_0) - \varepsilon/2$ and assume without loss of generality that $\rho_0 \geq 0$. Since $\Phi(x_0;\rho_0)$ is compact and does not contain y_0, there exist disjoint open neighborhoods V and W of y_0 and $\Phi(x_0;\rho_0)$, respectively. Applying the upper large deviation bound for Π to the complement of W, we find that

$$\limsup_{N\to\infty} \gamma_N^{-1} \log P^N_{x_N}(V) \leq -\rho_0$$

for each X_N-sequence (x_N) tending to x_0. From this and the Countability Hypothesis we conclude that there exists an open neighborhood U of x_0 such that

$$\limsup_{N\to\infty} \gamma_N^{-1} \log \sup_{x\in U\cap X(K)\cap X_N} P_x^N(V) \le -\rho_0 + \varepsilon/2. \tag{1.5}$$

Now choose $x \in U \cap K$ arbitrarily. Because of the Countability Hypothesis, we find an X_N-sequence (x_N) which tends to x and belongs to $U \cap X(K)$ for all but finitely many terms. Applying the lower large deviation bound for $P_{x_N}^N(V)$ and combining it with (1.5), we obtain

$$-I(x;V) \le -\rho_0 + \varepsilon/2 = -I(x_0;y_0) + \varepsilon \qquad \text{for all } x \in U \cap K.$$

This finally yields (1.4). □

In the rest of this section we collect some further properties of rate functions which will be used in the subsequent sections.

A subset of a Hausdorff space Z is called *universally measurable* if it belongs to the μ-completion of the Borel σ-field of Z for each Radon probability measure μ on Z. A function $f: Z \to \overline{\mathbb{R}}$ is called *sequentially lower semi-continuous* if $f(z) \le \liminf f(z_n)$ for each sequence (z_n) in Z with $z_n \to z$. If Z satisfies the first countability axiom, then a function $f: Z \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous iff it is sequentially lower semi-continuous.

Lemma 1.5. *a) Assume that Π is a special large deviation system with rate function $J: Y \to [0,\infty]$, and suppose that the associated level sets $\Phi(K;\rho)$ are compact for all compact subsets K of X and all $\rho \ge 0$. Then J is sequentially lower semi-continuous and universally measurable.*

b) Let the Countability Hypothesis be satisfied. Assume that Π is a large deviation system with rate function $I: X \times Y \to [0,\infty]$, and suppose that the associated level sets $\Phi(K;\rho)$ are compact for all compact subsets K of X and all $\rho \ge 0$. Then I is sequentially lower semi-continuous and universally measurable.

Proof. a) For each compact set $K \subseteq X$, the sets $\Phi(K;\rho)$, $\rho \ge 0$, are closed, i.e. the restriction of J to $\pi^{-1}(K)$ is lower semi-continuous. Since each converging sequence in Y is contained in $\pi^{-1}(K)$ for some compact subset K of X, this implies the sequential lower semi-continuity of J.

Fix a Radon probability measure ν on Y arbitrarily. Let (K_r) be an increasing sequence of compact subsets of X such that the Radon measure $\nu\circ\pi^{-1}$ is concentrated on $\bigcup K_r$. Then ν is concentrated on $\bigcup \pi^{-1}(K_r)$ and

$$\bigcup_{r=1}^{\infty} \Phi(K_r;\rho) \subseteq \{J \le \rho\} \subseteq \bigcup_{r=1}^{\infty} \Phi(K_r;\rho) \cup \left(Y \setminus \bigcup_{r=1}^{\infty} \pi^{-1}(K_r) \right)$$

for each $\rho \ge 0$. Since the sets $\Phi(K_r;\rho)$ are Borel measurable, this proves the universal measurability of J.

b) This is a consequence of a). Indeed, we know from Theorem 1.3 that I is the rate function of a special large deviation system. Moreover, Lemma 1.4 tells us that the associated level sets $\tilde{\Phi}(K;\rho)$ are compact. □

Lemma 1.6. *Assume that X satisfies the Countability Hypothesis and Y is regular. Let $\Pi = \{P_x^N; x \in X_N, N \in \mathbb{N}\}$ be a large deviation system with rate function I and scale γ_N, and suppose that the associated level sets $\Phi(K;\rho)$ are compact for all compact subsets K of X and all $\rho \geq 0$. Then the following assertions are valid.*

a) For each open subset G of Y the function $I(\,\cdot\,;G)$ is sequentially upper semi-continuous and universally measurable.

b) For each closed subset F of Y the function $I(\,\cdot\,;F)$ is sequentially lower semi-continuous and universally measurable.

c) For each bounded continuous function $g: Y \to \mathbb{R}$ the function

$$h_g(x) := \sup_{y \in Y} \left[g(y) - I(x;y)\right], \qquad x \in X,$$

is bounded, sequentially continuous, and universally measurable.

Proof. Let $f: X \to \overline{\mathbb{R}}$ be a sequentially lower semi-continuous function. By the Countability Hypothesis, each compact subspace of X satisfies the first countability axiom. Consequently, the set $\{f \leq \rho\} \cap K$ is closed (and, hence, Borel measurable) for each $\rho \in \mathbb{R}$ and each compact set $K \subseteq X$. This implies the universal measurability of f (cf. the proof of Lemma 1.5 a)). This also shows that each sequentially upper semi-continuous and each sequentially continuous function on X is universally measurable.

a) Let (x_n) be a sequence in X with $x_n \to x$, and let K be a compact subset of X containing (x_n). Given an open subset A of X and $h > 0$, we find an open neighborhood $U(x)$ of x such that

$$-I(x;A) - h \leq \liminf_{N\to\infty} \gamma_N^{-1} \log \inf_{\tilde{x} \in U(x)\cap K \cap X_N} P_{\tilde{x}}^N(A).$$

This is a consequence of the lower large deviation bound and the Countability Hypothesis. Applying the upper large deviation bound to $P_{\tilde{x}}^N(\overline{A})$, we see that the expression on the right does not exceed

$$-\sup_{\tilde{x} \in U(x)\cap K} I(\tilde{x};\overline{A}).$$

This shows that

$$\limsup_{n\to\infty} I(x_n;\overline{A}) \leq I(x;A) \tag{1.6}$$

for each open set $A \subseteq X$.

Now let G be an arbitrary open subset of X. Then (1.6) implies that

$$\limsup_{n\to\infty} I(x_n;G) \leq I(x;A)$$

for each open set A with $\overline{A} \subseteq G$. Because of the regularity of Y, this yields

$$\limsup_{n\to\infty} I(x_n;G) \leq I(x;G),$$

i.e. $I(\,\cdot\,;G)$ is sequentially lower semi-continuous.

b) Let (x_n) be a sequence in X with $x_n \to x$, and let K be a compact subset of X containing (x_n). We must show that

$$I(x;F) \le \liminf_{n\to\infty} I(x_n;F) \tag{1.7}$$

for each closed subset F of Y. It suffices to prove (1.7) for $F \cap \Phi(K;\rho)$ instead of F with $\rho > \liminf_{n\to\infty} I(x_n;F)$. We can and will therefore assume that F is compact.

Fix $\varepsilon > 0$ arbitrarily. Since I is lower semi-continuous on $K \times Y$ (see Lemma 1.4), we find for each point y in Y open neighborhoods U_y and V_y of x and y, respectively, such that

$$I(\tilde{x};\tilde{y}) > I(x;y) - \varepsilon \qquad \text{for all } (\tilde{x},\tilde{y}) \in (U_y \cap K) \times V_y.$$

Now select a finite covering of F by sets $V_k := V_{y_k}$ $(k = 1,\dots,r)$ with $y_1,\dots,y_r \in F$ and put $U := \bigcap_{k=1}^r U_{y_k}$. Then for all sufficiently large n, x_n belongs to $U \cap K$ and therefore

$$I(x_n;F) = \min_{1\le k\le r} I(x_n;F\cap V_k) \ge \min_{1\le k\le r} I(x;y_k) - \varepsilon \ge I(x;F) - \varepsilon.$$

This proves (1.7).

c) Again, let (x_n) be a sequence in X with $x_n \to x$, and let K be a compact subset of X containing (x_n). Since $I \ge 0$ and $I(x;Y) = 0$ for each $x \in X$, we have

$$\inf g \le h_g \le \sup g.$$

In particular, h_g is bounded. This also shows that

$$h_g(\tilde{x}) = \sup_{y\in\Phi(K;\rho)} [g(y) - I(\tilde{x};y)], \qquad \tilde{x} \in K,$$

for $\rho > \sup g - \inf g$.

Fix $\varepsilon > 0$ arbitrarily and choose a finite covering of $\Phi(K;\rho)$ by open neighborhoods $W(y_1),\dots,W(y_r)$ of $y_1,\dots,y_r$, respectively, such that

$$\sup_{y\in\overline{W(y_k)}} |g(y) - g(y_k)| < \varepsilon/2 \qquad \text{for } k = 1,\dots,r.$$

(Here we have used the regularity of Y.) Then

$$\begin{aligned}\liminf_{n\to\infty} h_g(x_n) &= \liminf_{n\to\infty} \max_{1\le k\le r} \sup_{y\in W(y_k)} [g(y) - I(x_n;y)] \\ &\ge \liminf_{n\to\infty} \max_{1\le k\le r} [g(y_k) - I(x_n;W(y_k))] - \varepsilon/2.\end{aligned}$$

It follows from assertion a) that the maximum on the right is sequentially lower semi-continuous. We can therefore continue as follows:

$$\begin{aligned} &\ge \max_{1\le k\le r} [g(y_k) - I(x;W(y_k))] - \varepsilon/2 \\ &\ge \max_{1\le k\le r} \sup_{y\in W(y_k)} [g(y) - I(x;y)] - \varepsilon \\ &= h_g(x).\end{aligned}$$

This proves the sequential lower semi-continuity of h_g.

It remains to show that h_g is sequentially upper semi-continuous. Using assertion b), we obtain

$$\begin{aligned}\limsup_{n\to\infty} h_g(x_n) &= \limsup_{n\to\infty} \max_{1\le k\le r} \sup_{y\in\overline{W(y_k)}} [g(y) - I(x_n; y)] \\ &\le \limsup_{n\to\infty} \max_{1\le k\le r} \left[g(y_k) - I\left(x_n; \overline{W(y_k)}\right)\right] + \varepsilon/2 \\ &= \max_{1\le k\le r} \left[g(y_k) - \liminf_{n\to\infty} I\left(x_n; \overline{W(y_k)}\right)\right] + \varepsilon/2 \\ &\le \max_{1\le k\le r} \left[g(y_k) - I\left(x; \overline{W(y_k)}\right)\right] + \varepsilon/2 \\ &\le h_g(x) + \varepsilon,\end{aligned}$$

and we are done. □

We remark that the results of this section are applicable to families of Radon probability measures indexed by an arbitrary directed set instead of $\mathbb{N}$.

2. Multilevel large deviations

The aim of this section is to study large deviations for empirical measures of independent copies of random variables which themselves satisfy the large deviation principle. Before formulating the precise results (Theorems 2.1, 2.2, 2.7, and 2.9 below), we introduce the necessary notation.

Throughout this section, X and Y are *completely regular* Hausdorff spaces, and (X_N) is a sequence of subsets of X such that each point in X is the limit of an X_N-sequence. By $\mathcal{M}(X)$ and $\mathcal{M}(Y)$ we denote the spaces of Radon probability measures on X and Y, respectively, furnished with the topology of weak convergence. $\mathcal{M}(X)$ and $\mathcal{M}(Y)$ are also completely regular Hausdorff spaces. Concerning this and further topological properties of the spaces $\mathcal{M}(X)$ and $\mathcal{M}(Y)$, the reader is referred to Topsøe [15].

Let $C_b(X)$ and $C_b(Y)$ denote the spaces of bounded continuous functions on X and Y, respectively, equipped with the sup-norm $\|\cdot\|$. Given $\mu \in \mathcal{M}(X)$ and $f \in C_b(X)$, $\langle \mu, f\rangle$ will stand for the integral of f with respect to μ. Correspondingly we define $\langle \nu, g\rangle$ for $\nu \in \mathcal{M}(Y)$ and $g \in C_b(Y)$. By $\overline{A}$ and $1\!\!1_A$ we will denote the closure and the indicator function of a set A, respectively.

Throughout this section we will assume that the following hypotheses are satisfied.

Metrizability Hypothesis. For each compact subset K of X there exists a metrizable set $X(K)$, $K \subseteq X(K) \subseteq X$, such that each point of K is the limit of an X_N-sequence which belongs to $X(K)$ for all but finitely many terms.

Tightness Hypothesis. Each converging sequence in $\mathcal{M}(X)$ is tight.

Note that the Metrizability Hypothesis implies the Countability Hypothesis of Section 1. The class of spaces X which fulfill the Tightness Hypothesis contains all metrizable spaces, all spaces which satisfy the second countability axiom, and all locally compact spaces (cf. Topsøe [15], Theorem 9.3). In particular, both hypotheses are satisfied in the case when X is a Polish space.

Given $M, N \in \mathbb{N}$, we denote by $\mathcal{M}^{M,N}(X)$ the subset of $\mathcal{M}(X)$ consisting of M-point empirical measures on X_N, i.e.

$$\mathcal{M}^{M,N}(X) := \left\{ M^{-1} \sum_{m=1}^{M} \delta_{x_m} : x_1, \ldots, x_M \in X_N \right\},$$

where δ_x is the Dirac measure at x. Each element of $\mathcal{M}(X)$ is the limit of an $\mathcal{M}^{M,N}(X)$-sequence as $M, N \to \infty$. The proof of this fact relies on the Metrizability Hypothesis and will be given in Appendix A.1.

2.1. Main result

Let $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ be a family of Radon probability measures on Y. By $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ we denote the family of Radon probability laws on $\mathcal{M}(Y)$ associated with the empirical measures of independent copies of Y-valued random variables with laws P_x^N. More precisely, given $\mu = M^{-1} \sum_{m=1}^{M} \delta_{x_m}$ $(x_1, \ldots, x_M \in X_N)$, $\mathcal{P}_\mu^{M,N}$ is the image of the Radon extension of the product measure $P_{x_1}^N \otimes \cdots \otimes P_{x_M}^N$ with respect to the continuous map

$$Y^M \in (y_1, \ldots, y_M) \longmapsto M^{-1} \sum_{m=1}^{M} \delta_{y_m} \in \mathcal{M}(Y).$$

Let $\pi: Y \to X$ be a surjective continuous map, and denote by $\hat{\pi}$ the induced map $\mathcal{M}(Y) \to \mathcal{M}(X)$ defined by $\hat{\pi}(\nu) := \nu \circ \pi^{-1}$, $\nu \in \mathcal{M}(Y)$, which is also continuous.

Theorem 2.1. *Assume that $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ is a special large deviation system (with respect to π) having rate function J and scale γ_N and that the associated level sets $\Phi(K;\rho)$ are compact for all compact subsets K of X and all $\rho \geq 0$. Then $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ is a special large deviation system (with respect to $\hat{\pi}$) with rate function*

$$S(\nu) := \int_Y J(y)\, \nu(dy), \qquad \nu \in \mathcal{M}(Y), \tag{2.1}$$

and scale $M\gamma_N$ as $M, N \to \infty$.

Given $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$, let $\mathcal{M}(\mu, \nu)$ denote the set of Radon probability measures on $X \times Y$ with left marginal μ and right marginal ν.

Theorem 2.2. *Assume that* $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ *is a large deviation system having rate function* I *and scale* γ_N *and that the associated level sets* $\Phi(K;\rho)$ *are compact for all compact subsets* K *of* X *and all* $\rho \geq 0$. *Then* $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ *is a large deviation system with rate function*

$$S(\mu;\nu) := \inf_{Q\in\mathcal{M}(\mu,\nu)} \int_{X\times Y} I(x;y)\, Q(dx,dy), \qquad \mu \in \mathcal{M}(X),\ \nu \in \mathcal{M}(Y), \quad (2.2)$$

and scale $M\gamma_N$ *as* $M, N \to \infty$.

By Lemma 1.5, the functions J and I are universally measurable. Therefore, the integrals in (2.1) and (2.2) are well-defined.

Before proving Theorem 2.1, we show how to derive Theorem 2.2 from Theorem 2.1 and Theorem 1.3.

Proof of Theorem 2.2. As in Section 1, let $\tilde{P}_x^N$ denote the Radon extension of $\delta_x \otimes P_x^N$. We know from Theorem 1.3 that $\{\tilde{P}_x^N; x \in X_N, N \in \mathbb{N}\}$ is a special large deviation system (with respect to the canonical projection $X \times Y \to X$) having rate function I and scale γ_N. Moreover, Lemma 1.4 tells us that the level sets $\tilde{\Phi}(K;\rho)$ are compact for all compact subsets K of X and all $\rho \geq 0$. Let $\{\tilde{\mathcal{P}}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ denote the family of Radon probability measures on $\mathcal{M}(X \times Y)$ associated with the empirical measures for $\{\tilde{P}_x^N; x \in X_N, N \in \mathbb{N}\}$. According to Theorem 2.1, this family forms a special large deviation system (with respect to the canonical projection $\mathcal{M}(X \times Y) \to \mathcal{M}(X)$) having rate function

$$\tilde{S}(Q) := \int_{X\times Y} I(x;y)\, Q(dx,dy), \qquad Q \in \mathcal{M}(X \times Y),$$

and scale $M\gamma_N$. But $\mathcal{P}_\mu^{M,N}$ is the image of $\tilde{\mathcal{P}}_\mu^{M,N}$ with respect to the canonical projection $\mathcal{M}(X \times Y) \to \mathcal{M}(Y)$ transforming Radon measures on $X \times Y$ into its marginals on Y. Therefore the assertion of Theorem 2.2 follows now by an application of the 'contraction principle' (see e.g. Varadhan [16], Theorem 2.4). □

The rest of this section is devoted to the proof of Theorem 2.1. To this end we assume that $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ is a special large deviation system (with respect to π) having rate function J and scale γ_N. Let I be defined by (1.1), and denote the associated level sets by $\Phi(K;\rho)$. We assume that $\Phi(K;\rho)$ is compact for each compact subset K of X and each $\rho \geq 0$.

The proof of Theorem 2.1 will be divided into several steps (Lemma 2.3 – Lemma 2.6 below).

Let

$$\Psi(\mu;\rho) := \left\{ \nu \in \mathcal{M}(Y) : \nu\circ\pi^{-1} = \mu,\ \int J\, d\nu \leq \rho \right\}, \qquad \mu \in \mathcal{M}(X),\ \rho \geq 0,$$

be the level sets associated with S, and define

$$\Psi(\mathcal{A};\rho) := \bigcup_{\mu\in\mathcal{A}} \Psi(\mu;\rho)$$

for $\mathcal{A} \subseteq \mathcal{M}(X)$.

Lemma 2.3. *The level sets $\Psi(\mu;0)$, $\mu \in \mathcal{M}(X)$, are non-empty. In particular, the map $\hat{\pi}$ is surjective.*

Proof. We first remark that the upper large deviation bound for $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ yields $I(x;Y) = 0$ and, hence, $\Phi(x;0) \neq \emptyset$ for all $x \in X$.

Choose $\mu \in \mathcal{M}(X)$ arbitrarily. There exists an increasing sequence (K_r) of compact sets such that μ is concentrated on $\bigcup_r K_r$. We write μ in the form

$$\mu = \sum_{r=1}^{\infty} p_r \mu_r\,,$$

where, for each r, μ_r is a probability measure concentrated on K_r, $p_r \geq 0$, and $\sum_r p_r = 1$.

Now fix $r \in \mathbb{N}$ arbitrarily. We claim that there exists a probability measure $\nu_r \in \mathcal{M}(Y)$ with $\nu_r \circ \pi^{-1} = \mu_r$ which is concentrated on the compact level set $\Phi(K_r;0)$. Indeed, one finds a sequence (x_{rn}) in K_r such that

$$\mu_{rn} := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_{ri}} \longrightarrow \mu_r$$

weakly as $n \to \infty$ (cf. the proof of Proposition A.1). For each n, choose a point y_{rn} in the non-empty set $\Phi(x_{rn};0)$ and define

$$\nu_{rn} := \frac{1}{n}\sum_{i=1}^{n} \delta_{y_{ri}}.$$

Then $\pi(y_{ri}) = x_{ri}$, $\nu_{rn} \circ \pi^{-1} = \mu_{rn}$, and the measures ν_{rn} are concentrated on $\Phi(K_r;0)$. In particular, the sequence (ν_{rn}) is tight. Selecting a converging subsequence, we find a probability measure ν which is concentrated on $\Phi(K_r;0)$ and satisfies $\nu_r \circ \pi^{-1} = \mu_r$. Hence the measure

$$\nu := \sum_{r=1}^{\infty} p_r \nu_r$$

belongs to $\mathcal{M}(Y)$ and satisfies $\nu \circ \pi^{-1} = \mu$. Since ν_r is concentrated on $\Phi(K_r;0)$ and $J = 0$ on $\Phi(K_r;0)$ for each $r \in \mathbb{N}$, we have $\int J\,d\nu = 0$ and, hence, $\nu \in \Psi(\mu;0)$. □

We next prove the compactness of the level sets.

Lemma 2.4. *a) For each $\mu \in \mathcal{M}(X)$ and each $\rho \geq 0$, the set $\Psi(\mu;\rho)$ is compact and tight.*

b) Suppose that $\mathcal{K}$ is a compact and tight subset of $\mathcal{M}(X)$. Then the sets $\Psi(\mathcal{K};\rho)$, $\rho \geq 0$, are compact and tight. In particular, if X is a Polish space, then $\Psi(\mathcal{K};\rho)$ is compact for each compact subset $\mathcal{K}$ of $\mathcal{M}(X)$ and each $\rho \geq 0$.

Proof. a) Let (K_r) be an increasing sequence of compact subsets of X such that $\mu(X \setminus K_r) < 1/r$ for each r. Define

$$Y^0 := \bigcup_{r=1}^{\infty} \Phi(K_r;r)$$

and equip Y^0 with the strongest topology which induces on $\Phi(K_r;r)$ the subspace topology of Y for each r. This topology on Y^0 is stronger than the subspace topology generated by Y. As free topological union of a countable number of compact spaces, the space Y^0 is normal (Postnikov [11], p. 30). In particular, it is a completely regular Hausdorff space. Let $\mathcal{B}(Y)$ and $\mathcal{B}(Y^0)$ denote the Borel σ-fields of Y and Y^0, respectively. One easily checks that $\mathcal{B}(Y^0) = \mathcal{B}(Y) \cap Y^0$.

Let $\mathcal{M}(Y^0)$ denote the space of Radon probability measures on Y^0 endowed with the topology of weak convergence. The continuous imbedding $\iota: Y^0 \to Y$ induces a continuous map $\hat{\iota}: \mathcal{M}(Y^0) \to \mathcal{M}(Y)$ which transforms each measure $\nu^0 \in \mathcal{M}(Y^0)$ into its image measure ν with respect to ι given by

$$\nu(A) = \nu^0(A \cap Y^0), \qquad A \in \mathcal{B}(Y).$$

Now define

$$\Psi^0(\mu;\rho) := \left\{ \nu^0 \in \mathcal{M}(Y^0) : \nu^0 \circ \iota^{-1} \circ \pi^{-1} = \mu, \ \int J \circ \iota \, d\nu^0 \leq \rho \right\}.$$

(The measurability of the map $J \circ \iota : Y^0 \to [0,\infty]$ follows from its lower semi-continuity proven below.) Since each measure $\nu \in \Psi(\mu;\rho)$ is concentrated on Y^0, the set $\Psi(\mu;\rho)$ is the image of $\Psi^0(\mu;\rho)$ with respect to $\hat{\iota}$. To prove the compactness and tightness of $\Psi(\mu;\rho)$, we have therefore only to show that $\Psi^0(\mu;\rho)$ is compact and tight in $\mathcal{M}(Y^0)$.

Applying Chebyshev's inequality, we obtain for each $\nu^0 \in \Psi^0(\mu;\rho)$ and each r the estimate

$$\begin{aligned} \nu^0\left(Y^0 \setminus \Phi(K_r;r)\right) &\leq \nu^0(J \circ \iota > r) + \nu^0\left(\iota^{-1} \circ \pi^{-1}(X \setminus K_r)\right) \\ &\leq \frac{1}{r} \int J \circ \iota \, d\nu^0 + \mu(X \setminus K_r) \leq \frac{\rho+1}{r}. \end{aligned}$$

Since the sets $\Phi(K_r;r)$ are compact in Y^0, this proves the tightness and, hence, the relative compactness of $\Psi^0(\mu;\rho)$.

It remains to show that $\Psi^0(\mu;\rho)$ is closed. To this end it suffices to check that the map $J \circ \iota$ is lower semi-continuous. Fix $\rho \geq 0$ arbitrarily. Then the set

$$\{J \circ \iota \leq \rho\} \cap \Phi(K_r;r) = \Phi(K_r; r \wedge \rho)$$

is closed in $\Phi(K_r;r)$ for each r. But this means that $\{J\circ\iota \le \rho\}$ is closed in the topology of Y^0.

b) Since $\mathcal{K}$ is tight, we find an increasing sequence (K_r) of compact subsets of X such that $\mu(X\setminus K_r) < 1/r$ for each $\mu\in\mathcal{K}$. We can therefore repeat the proof of part a) with $\Psi(\mu;\rho)$ and $\Psi^0(\mu;\rho)$ replaced by $\Psi(\mathcal{K};\rho)$ and $\Psi^0(\mathcal{K};\rho)$, respectively.

If X is Polish then, by Prokhorov's compactness criterion, the compactness of $\mathcal{K}$ implies the tightness of $\mathcal{K}$, and we can proceed as before. □

Now we turn to the proof of the lower large deviation bound. To this end we set

$$S(\mu;\nu) := \begin{cases} S(\nu) & \text{if } \nu\circ\pi^{-1} = \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

Lemma 2.5. *For each $\mathcal{M}^{M,N}(X)$-sequence $(\mu^{M,N})$ tending to $\mu\in\mathcal{M}(X)$ and each open subset G of $\mathcal{M}(Y)$ we have*

$$\liminf_{M,N\to\infty} \frac{1}{M\gamma_N} \log \mathcal{P}^{M,N}_{\mu^{M,N}}(G) \ge -S(\mu;G).$$

Proof. 1^0 Fix $\nu\in\mathcal{M}(Y)$ with $S(\nu)<\infty$ arbitrarily and choose an $\mathcal{M}^{M,N}(X)$-sequence $(\mu^{M,N})$ which converges weakly to $\mu := \nu\circ\pi^{-1}$ as $M,N\to\infty$. We write the measures $\mu^{M,N}$ in the form

$$\mu^{M,N} = \frac{1}{M}\sum_{i=1}^{M} \delta_{x_i^{M,N}} \qquad \text{with } x_1^{M,N},\dots,x_M^{M,N}\in X_N.$$

Given $M,N\in\mathbb{N}$, we consider a Y^M-valued random vector $(\xi_1^{M,N},\dots,\xi_M^{M,N})$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ the law of which coincides with the Radon extension of the product measure $P^N_{x_1^{M,N}}\otimes\cdots\otimes P^N_{x_M^{M,N}}$. I.e., $\xi_1^{M,N},\dots,\xi_M^{M,N}$ are independent Y-valued random variables with laws $P^N_{x_1^{M,N}},\dots,P^N_{x_M^{M,N}}$, respectively. We introduce the empirical measures

$$\Xi^{M,N} := \frac{1}{M}\sum_{i=1}^{M}\delta_{\xi_i^{M,N}}.$$

Let $U(\nu)$ be an open neighborhood of ν. Fix $h>0$ and sequences (M_n) and (N_n) of natural numbers with $M_n,N_n\to\infty$ as $n\to\infty$ arbitrarily. We must show that

$$\liminf_{n\to\infty}\frac{1}{M_n\gamma_{N_n}}\log\mathbb{P}\left(\Xi^{M_n,N_n}\in U(\nu)\right) \ge -S(\nu)-h \tag{2.3}$$

(cf. Freidlin and Wentzell [7], Chap. 3, Theorem 3.3).

In the following we will write μ^n, x_i^n, ξ_i^n and Ξ^n instead of μ^{M_n,N_n}, $x_i^{M_n,N_n}$, $\xi_i^{M_n,N_n}$ and Ξ^{M_n,N_n}, respectively.

2^0 It is not hard to see that there exist pairwise disjoint open sets $G_1,\dots,G_r \subseteq Y$, compact sets $C_k \subseteq G_k$ ($k=1,\dots,r$), and $\varepsilon > 0$ such that

$$\tilde{U}(\nu) := \left\{ \tilde{\nu} \in \mathcal{M}(Y) : \tilde{\nu}(G_k) > \nu(C_k) - \varepsilon \text{ for } k=1,\dots,r \right\} \subseteq U(\nu) \tag{2.4}$$

(see e.g. Billingsley [2], Appendix III, Theorem 3 for a similar statement). We choose $\varepsilon' \in (0,1)$ so that

$$\varepsilon' + 2\sqrt{\varepsilon'} < \varepsilon. \tag{2.5}$$

Because of the Tightness Hypothesis, we find a compact set $K \supseteq \bigcup_{k=1}^r \pi(C_k)$ with

$$\mu^n(K) \geq 1 - \varepsilon'/3 \qquad \text{for all } n. \tag{2.6}$$

Since $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ is a large deviation system with rate function I, we find for each $x \in K$ an open neighborhood $U(x)$ such that

$$\liminf_{N\to\infty} \frac{1}{\gamma_N} \log \inf_{\tilde{x} \in U(x)\cap X(K)\cap X_N} P_{\tilde{x}}^N(G_k) \geq -I(x;G_k) - h/2 \tag{2.7}$$

for $k=1,\dots,r$, where $X(K)$ is taken from the Metrizability Hypothesis. Thereby we can assume that $U(x)$ is chosen so 'small' that

$$\inf_{\tilde{x} \in U(x)\cap K} I(\tilde{x};C_k) \geq I(x;C_k) - h/2 \tag{2.8}$$

for $k=1,\dots,r$. This follows from Lemma 1.6 b) and the Metrizability Hypothesis. Since $I(x;G_k) \leq I(x;C_k)$ for all k, we can combine (2.7) and (2.8) to arrive at

$$\liminf_{N\to\infty} \frac{1}{\gamma_N} \log \inf_{\tilde{x} \in U(x)\cap X(K)\cap X_N} P_{\tilde{x}}^N(G_k) \geq - \inf_{\tilde{x} \in U(x)\cap K} I(\tilde{x};C_k) - h$$

for $k=1,\dots,r$. We now choose a finite covering $\{U_1,\dots,U_q\}$ of K by open sets of the form $U_j = U(x_j)$ with $x_j \in K$ ($j=1,\dots,q$). Then

$$\liminf_{N\to\infty} \frac{1}{\gamma_N} \log \inf_{x \in U_j\cap X(K)\cap X_N} P_x^N(G_k) \geq - \inf_{x \in U_j\cap K} I(x;C_k) - h \tag{2.9}$$

for $j=1,\dots,q$ and $k=1,\dots,r$.

3^0 We find pairwise disjoint Borel sets $W_j \subseteq U_j$ ($j=1,\dots,q$) such that

$$\mu\left(K \setminus \bigcup_{j=1}^q W_j\right) < \varepsilon'/3 \qquad \text{and} \qquad \mu(\partial W_j) = 0 \qquad \text{for each } j, \tag{2.10}$$

where ∂W_j denotes the boundary of W_j. Given $n \in \mathbb{N}$, we introduce the pairwise disjoint sets of indices

$$\Lambda_j^n := \{\, i : x_i^n \in W_j \cap K \,\}, \qquad j=1,\dots,q.$$

For each j, we further select pairwise disjoint subsets $\Lambda_{j,1}^n, \ldots, \Lambda_{j,r}^n$ of Λ_j^n such that

$$\left|\Lambda_{j,k}^n\right| = \left[\frac{\nu(C_k \cap \pi^{-1}(W_j))}{\mu(W_j)} \left|\Lambda_j^n\right|\right], \qquad k = 1, \ldots, r,$$

where $|\Lambda|$ denotes the cardinality of the set Λ and $[a]$ is the integer part of $a \in \mathbb{R}$. If $\mu(W_j) = 0$, then we set $\Lambda_{j,k}^n = \emptyset$ for all k. Since $\left|\Lambda_j^n\right|/M_n = \mu^n(W_j \cap K)$, $\mu^n \to \mu$ weakly and $\mu(\partial W_j) = 0$, we have

$$\limsup_{n\to\infty} \left|\Lambda_j^n\right|/M_n \le \mu(W_j)$$

for all j, and, consequently,

$$\limsup_{n\to\infty} \left|\Lambda_{j,k}^n\right|/M_n \le \nu(C_k \cap \pi^{-1}(W_j)) \tag{2.11}$$

for all j and k.

4^0 We now claim that

$$\left\{\xi_i^n \in G_k \text{ for all } i \in \bigcup_{j=1}^q \Lambda_{j,k}^n \text{ and } k = 1, \ldots, r\right\} \subseteq \left\{\Xi^n \in \tilde{U}(\nu)\right\} \tag{2.12}$$

for all sufficiently large n. Remembering the definitions of $\tilde{U}(\nu)$ and $\Xi^{M,N}$ and taking into account that $\left|\Lambda_j^n\right|/M_n = \mu^n(W_j \cap K)$, we see that it will be enough to show that

$$\liminf_{n\to\infty} \sum_{j=1}^q \frac{\nu(C_k \cap \pi^{-1}(W_j))}{\mu(W_j)} \mu^n(W_j \cap K) > \nu(C_k) - \varepsilon \tag{2.13}$$

for each k. To this end, we fix k arbitrarily and assume without loss of generality that the sequences $(\mu^n(W_j \cap K))$ converge for each j. Otherwise the subsequent considerations must be done for an appropriate subsequence of (M_n, N_n). We introduce the index set

$$\Gamma := \left\{j : \lim_{n\to\infty} \mu^n(W_j \cap K) > (1 - \sqrt{\varepsilon'})\mu(W_j)\right\}.$$

Using (2.6) and (2.10), we obtain

$$\begin{aligned}\lim_{n\to\infty} \sum_{j=1}^q \mu^n(W_j \cap K) &= \lim_{n\to\infty} \mu^n\left(\bigcup_{j=1}^q W_j \cap K\right) \\ &\ge \mu\left(\bigcup_{j=1}^q W_j\right) - \varepsilon'/3 > 1 - \varepsilon'.\end{aligned}$$

Therefore

$$\begin{aligned}
1-\varepsilon' &\leq \lim_{n\to\infty}\sum_{j\in\Gamma}\mu^n(W_j)+\lim_{n\to\infty}\sum_{j\notin\Gamma}\mu^n(W_j\cap K)\\
&\leq \sum_{j\in\Gamma}\mu(W_j)+(1-\sqrt{\varepsilon'})\sum_{j\notin\Gamma}\mu(W_j)\\
&\leq 1-\sqrt{\varepsilon'}\sum_{j\notin\Gamma}\mu(W_j),
\end{aligned}$$

i.e.

$$\sum_{j\notin\Gamma}\mu(W_j)\leq\sqrt{\varepsilon'}.$$

Hence, using this, (2.6), (2.10), and (2.5), we obtain

$$\begin{aligned}
&\sum_{j\in\Gamma}\frac{\nu(C_k\cap\pi^{-1}(W_j))}{\mu(W_j)}\lim_{n\to\infty}\mu^n(W_j\cap K)\\
&\geq(1-\sqrt{\varepsilon'})\sum_{j\in\Gamma}\nu(C_k\cap\pi^{-1}(W_j))\\
&\geq(1-\sqrt{\varepsilon'})\left[\sum_{j}\nu(C_k\cap\pi^{-1}(W_j))-\sum_{j\notin\Gamma}\mu(W_j)\right]\\
&\geq(1-\sqrt{\varepsilon'})\left[\nu(C_k)-\varepsilon'-\sqrt{\varepsilon'}\right]\geq\nu(C_k)-\varepsilon'-2\sqrt{\varepsilon'}\\
&>\nu(C_k)-\varepsilon,
\end{aligned}$$

and we arrive at (2.13).

5^0 We have now collected all ingredients to prove (2.3). Using (2.4) and (2.12), we obtain for all sufficiently large n the inequality

$$\begin{aligned}
\mathbb{P}\left(\Xi^n\in U(\nu)\right)&\geq\mathbb{P}\left(\xi_i^n\in G_k \text{ for all } i\in\bigcup_{j=1}^{q}\Lambda_{j,k}^n \text{ and } k=1,\ldots,r\right)\\
&=\prod_{k=1}^{r}\prod_{j=1}^{q}\prod_{i\in\Lambda_{j,k}^n}P_{x_i^n}^{N_n}(G_k)\geq\prod_{k=1}^{r}\prod_{j=1}^{q}\left[\inf_{x\in U_j\cap X(K)\cap X_{N_n}}P_x^{N_n}(G_k)\right]^{|\Lambda_{j,k}^n|}.
\end{aligned}$$

Applying the large deviation bound (2.9) and taking into account (2.11), we find that

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{M_n\gamma_{N_n}}\log\mathbb{P}\left(\Xi^n\in U(\nu)\right)\\
&\geq-\sum_{k=1}^{r}\sum_{j=1}^{q}\left[\inf_{x\in W_j\cap K}I(x;C_k)+h\right]\nu(C_k\cap\pi^{-1}(W_j))
\end{aligned}$$

$$\geq -\sum_{k=1}^{r}\sum_{j=1}^{q}\int_{C_k\cap\pi^{-1}(W_j)}[J(y)+h]\,\nu(dy)$$
$$\geq -S(\nu)-h,$$

and we arrive at (2.3). Here we have also used that $\pi(C_k)\subseteq K$ and therefore

$$\inf_{x\in W_j\cap K} I(x;C_k)\leq J(y) \qquad \text{for each } y\in C_k\cap\pi^{-1}(W_j)$$

and all j and k. The above estimates work in the case when

$$\inf_{x\in W_j\cap K} I(x;C_k)<\infty \tag{2.14}$$

for all j and k. But they also work in the general situation with the conventions $0^0=1$ and $\infty\cdot 0=0$. To see this, one has to take into account that $\Lambda^n_{j,k}\neq\emptyset$ implies $\nu(C_k\cap\pi^{-1}(W_j))>0$ and this yields (2.14), since by assumption $S(\nu)=\int J\,d\nu<\infty$. □

It remains to derive the upper large deviation bound.

Lemma 2.6. *For each $\mathcal{M}^{M,N}(X)$-sequence $(\mu^{M,N})$ tending to $\mu\in\mathcal{M}(X)$ and each closed subset F of $\mathcal{M}(Y)$ we have*

$$\limsup_{M,N\to\infty}\frac{1}{M\gamma_N}\log\mathcal{P}^{M,N}_{\mu^{M,N}}(F)\leq -S(\mu;F).$$

Proof. 1^0 Fix $\mu\in\mathcal{M}(X)$ and an $\mathcal{M}^{M,N}(X)$-sequence $(\mu^{M,N})$ with $\mu^{M,N}\to\mu$ arbitrarily. As in the proof of the lower large deviation bound, we write $\mu^{M,N}$ in the form

$$\mu^{M,N}=\frac{1}{M}\sum_{i=1}^{M}\delta_{x_i^{M,N}} \qquad \text{with } x_1^{M,N},\ldots,x_M^{M,N}\in X_N.$$

Again, let $(\xi_1^{M,N},\ldots,\xi_M^{M,N})$ be a Y^M-valued random vector on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ the law of which coincides with the Radon extension of $P^N_{x_1^{M,N}}\otimes\cdots\otimes P^N_{x_M^{M,N}}$. Let

$$\Xi^{M,N}:=\frac{1}{M}\sum_{i=1}^{M}\delta_{\xi_i^{M,N}}$$

be the associated empirical measure. Fix $\rho\geq 0$ and $h>0$ arbitrarily. Let (M_n) and (N_n) be sequences of natural numbers tending to infinity. Let further $U(\Psi(\mu;\rho))$ be an open neighborhood of the (non-empty) compact set $\Psi(\mu;\rho)$ (see Lemma 2.3 and Lemma 2.4). We must show that

$$\limsup_{n\to\infty}\frac{1}{M_n\gamma_{N_n}}\log\mathbb{P}\left(\Xi^{M_n,N_n}\notin U(\Psi(\mu;\rho))\right)\leq -\rho+h \tag{2.15}$$

(cf. Freidlin and Wentzell [7], Chap. 3, Theorem 3.3).

In the following we will often write n instead of M_n, N_n. So we will use the notations μ^n, ξ_i^n and Ξ^n instead of μ^{M_n,N_n}, $\xi_i^{M_n,N_n}$ and Ξ^{M_n,N_n}, respectively.

To prove (2.15), we choose measures $\nu_1, \dots, \nu_m \in \Psi(\mu;\rho)$ and functions $g_{ij} \in C_b(Y)$ $(i = 1, \dots, m;\ j = 1, \dots, n_i)$ so that

$$U(\Psi(\mu;\rho)) \supseteq U_1(\Psi(\mu;\rho)) \supseteq U_{1/2}(\Psi(\mu;\rho)) \supseteq \Psi(\mu;\rho), \tag{2.16}$$

where

$$U_\alpha(\Psi(\mu;\rho)) := \bigcup_{i=1}^{m} \Big\{ \tilde\nu \in \mathcal{M}(Y) : |\langle \tilde\nu, g_{ij}\rangle - \langle \nu_i, g_{ij}\rangle| < \alpha \text{ for } j = 1, \dots, n_i \Big\}. \tag{2.17}$$

Because of the Tightness Hypothesis, we find an increasing sequence (K_r) of compact subsets of X such that

$$\mu^n(K_r) \geq 1 - 1/r \qquad \text{for all } n \text{ and } r. \tag{2.18}$$

Then

$$\mathcal{K} := \{ \tilde\mu \in \mathcal{M}(X) : \tilde\mu(K_r) \geq 1 - 1/r \text{ for all } r \}$$

is a compact and tight set of measures containing μ^n and μ. From Lemma 2.4 we know that $\Psi(\mathcal{K};\rho)$ is also compact. Because of this, we conclude from (2.16) that there exist functions $f_k \in C_b(X)$ $(k = 1, \dots, p)$ such that

$$U_{1/2}(\Psi(\mu;\rho)) \supseteq \Psi\left(\overline{V(\mu)} \cap \mathcal{K};\rho\right), \tag{2.19}$$

where

$$\overline{V(\mu)} := \Big\{ \tilde\mu \in \mathcal{M}(X) : |\langle \tilde\mu, f_k\rangle - \langle \mu, f_k\rangle| \leq 1 \text{ for } k = 1, \dots, p \Big\}. \tag{2.20}$$

2^0 We next construct mutually independent random variables $\tilde\xi_i^n$ which attain only finitely many values and are 'sufficiently close' to ξ_i^n for $i = 1, \dots, M_n$. To this end we fix a number

$$t > 8\rho \Big[\max\{ \|g_{ij}\| : 1 \leq i \leq m, 1 \leq j \leq n_i \} \vee \max\{ \|f_k\| : 1 \leq k \leq p \} \Big]. \tag{2.21}$$

Because of (2.18), we may choose $r_0 \in \mathbb{N}$ so that

$$\mu^n(X \setminus K_{r_0}) < \frac{h}{2t} \wedge \min_{i,j} \frac{1}{16\,\|g_{ij}\|} \wedge \min_k \frac{1}{16\,\|f_k\|} \qquad \text{for all } n,$$

i.e.

$$|\{ i : x_i^n \notin K_{r_0} \}| < M_n \left(\frac{h}{2t} \wedge \min_{i,j} \frac{1}{16\,\|g_{ij}\|} \wedge \min_k \frac{1}{16\,\|f_k\|} \right) \qquad \text{for all } n, \tag{2.22}$$

where $|\Lambda|$ denotes the cardinality of the set Λ. We select a finite covering of the compact set $\Phi(K_{r_0};t)$ by open sets $G_1,\ldots,G_q$ such that

$$\sup_{y,\tilde{y}\in\overline{G}_l} |g_{ij}(y) - g_{ij}(\tilde{y})| < 1/8 \qquad \text{for all } i,j \text{ and } 1 \le l \le q \tag{2.23}$$

and

$$\sup_{y,\tilde{y}\in\overline{G}_l} |f_k(\pi(y)) - f_k(\pi(\tilde{y}))| < 1/8 \qquad \text{for all } k \text{ and } 1 \le l \le q. \tag{2.24}$$

We choose pairwise disjoint measurable sets $A_1,\ldots,A_q$ such that $A_l \subseteq G_l$ for $l = 1,\ldots,q$ and $\bigcup_{l=1}^q A_l = \bigcup_{l=1}^q G_l$. We further set $A_0 := G_0 := Y \setminus \bigcup_{l=1}^q A_l$. Note that this set is closed. Given $0 \le l \le q$ and $1 \le r \le r_0$, we set

$$A_{l,r} := A_l \cap \pi^{-1}(K_r \setminus K_{r-1})$$

(with $K_0 := \emptyset$) and pick a point $y_{l,r} \in \overline{G}_l \cap \Phi(K_r;t)$ so that

$$J(y_{l,r}) = \min\{\, J(y) : y \in \overline{G}_l \cap \Phi(K_r;t) \,\}. \tag{2.25}$$

(If $\overline{G}_l \cap \Phi(K_r;t) = \emptyset$, then we choose $y_{l,r} \in \Phi(K_r;t)$ arbitrarily.) For $0 \le l \le q$ we further set

$$A_{l,0} := A_l \cap \pi^{-1}(X \setminus K_{r_0})$$

and $y_{l,0} := y_{l,r_0}$. Note that $\{A_{l,r}; 0 \le l \le q, 0 \le r \le r_0\}$ is a partition of Y into pairwise disjoint measurable sets. We introduce the index set

$$\begin{aligned} \Gamma \;:=\; & \{\, (l,r) : 0 \le l \le q, 1 \le r \le r_0, \overline{G}_l \cap \Phi(K_r;t) \ne \emptyset \,\} \\ & \cup \{\, (l,0) : 0 \le l \le q, \overline{G}_l \cap \Phi(K_{r_0};t) \ne \emptyset \,\}. \end{aligned}$$

Given $n \in \mathbb{N}$ and $1 \le i \le M_n$, we define

$$\tilde{\xi}_i^n := y_{l,r} \qquad \text{if } \xi_i^n \in A_{l,r}, \qquad 0 \le l \le q,\ 0 \le r \le r_0, \tag{2.26}$$

and introduce the associated empirical measure

$$\tilde{\Xi}^n := \frac{1}{M_n} \sum_{i=1}^{M_n} \delta_{\tilde{\xi}_i^n}.$$

The above constructed objects have the following properties:

(i) $y_{l,r} \in \overline{G}_l$ for all $(l,r) \in \Gamma$, $y_{l,r} \in \Phi(K_{r_0};t)$ for all l and r;

(ii) $J(y_{l,r}) \le \inf\{\, J(y) : y \in \overline{A}_{l,r} \,\}$ for $0 \le l \le q$ and $1 \le r \le r_0$;

(iii) $\tilde{\Xi}^n \circ \pi^{-1} \in \mathcal{K}$ a.s. for each n.

The last property follows from the observation that $\tilde{\Xi}^n \circ \pi^{-1}(K_{r_0}) = 1$ and $\tilde{\Xi}^n \circ \pi^{-1}(K_r) \geq \Xi^n \circ \pi^{-1}(K_r) = \mu^n(K_r)$ for $1 \leq r < r_0$ almost surely.

3^0 We next show that $\tilde{\Xi}^n$ satisfies an 'upper large deviation bound' with rate function

$$\tilde{S}(\nu) := \int_Y \tilde{J}(y)\,\nu(dy), \qquad \nu \in \mathcal{M}(Y),$$

where

$$\tilde{J}(y) := \begin{cases} J(y) & \text{for } y \in \{\, y_{l,r} : 0 \leq l \leq q, 0 \leq r \leq r_0 \,\}, \\ +\infty & \text{otherwise.} \end{cases}$$

More precisely, for each closed subset C of $\mathcal{M}(Y)$,

$$\limsup_{n\to\infty} \frac{1}{M_n \gamma_{N_n}} \log \mathbb{P}\left(\tilde{\Xi}^n \in C\right) \leq -\inf_{\nu\in C} \tilde{S}(\nu) + h. \tag{2.27}$$

The set $\tilde{\mathcal{M}}$ of all measures in $\mathcal{M}(Y)$ which are concentrated on $\{\, y_{l,r} : 0 \leq l \leq q, 0 \leq r \leq r_0 \,\}$ is compact. Since all realizations of the random measures $\tilde{\Xi}^n$ belong to $\tilde{\mathcal{M}}$ and $\tilde{S}(\nu) = +\infty$ for $\nu \notin \tilde{\mathcal{M}}$, it suffices to prove (2.27) for compact sets $C \subseteq \tilde{\mathcal{M}}$. But for this it will be enough to derive the following local large deviation bound: For each $\nu \in \tilde{\mathcal{M}}$ there exists an open neighborhood $U(\nu)$ of ν such that

$$\limsup_{n\to\infty} \frac{1}{M_n \gamma_{N_n}} \log \mathbb{P}\left(\tilde{\Xi}^n \in U(\nu)\right) \leq -S(\nu) + h. \tag{2.28}$$

To prove (2.28), we fix $\nu \in \tilde{\mathcal{M}}$ arbitrarily, choose $g \in C_b(Y)$ so that $g = J$ on $\{\, y_{l,r} : 0 \leq l \leq q, 0 \leq r \leq r_0 \,\}$ and put

$$U(\nu) := \{\, \tilde{\nu} : \langle \tilde{\nu}, g\rangle > S(\nu) - h/2 \,\}.$$

We introduce the index sets $\Lambda^n := \{\, i : x_i^n \in K_{r_0} \,\}$. Since $y_{l,r} \in \Phi(K_{r_0}; t)$, we have $g(y_{l,r}) \leq t$ for all l and r. Together with (2.22) this yields

$$\left\langle \tilde{\Xi}^n, g \right\rangle \leq \frac{1}{M_n} \sum_{i\in\Lambda^n} g(\tilde{\xi}_i^n) + \frac{h}{2}.$$

Using this and Chebyshev's exponential inequality, we obtain

$$\begin{aligned} \mathbb{P}\left(\tilde{\Xi}^n \in U(\nu)\right) &\leq \mathbb{P}\left(\sum_{i\in\Lambda^n} g(\tilde{\xi}_i^n) > M_n[S(\nu) - h]\right) \\ &\leq \exp\{-M_n \gamma_{N_n}[S(\nu) - h]\} \prod_{i\in\Lambda^n} \mathbb{E}\exp\left\{\gamma_{N_n} g(\tilde{\xi}_i^n)\right\}, \end{aligned} \tag{2.29}$$

where $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$. Here we have also used that the random variables $\tilde{\xi}_i^n$, $1 \leq i \leq M_n$, are mutually independent. For each

$i \in \Lambda^n$ we have

$$\mathbb{E}\exp\left\{\gamma_{N_n} g(\tilde{\xi}_i^n)\right\} = \sum_{l=0}^{q}\sum_{r=1}^{r_0} \exp\{\gamma_{N_n} J(y_{l,r})\}\, \mathbb{P}(\xi_i^n \in A_{l,r})$$

$$\leq \sum_{l=0}^{q}\sum_{r=1}^{r_0} \exp\{\gamma_{N_n} J(y_{l,r})\} \sup_{x\in K_r\cap X_{N_n}} P_x^{N_n}(\overline{A}_{l,r}). \tag{2.30}$$

(By convention, the supremum over the empty set is zero.) Applying the upper large deviation bound for $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ and taking into account the Metrizability Hypothesis and property (ii) of step 2^0, we obtain

$$\limsup_{n\to\infty} \frac{1}{\gamma_{N_n}} \log \sup_{x\in K_r\cap X_{N_n}} P_x^{N_n}(\overline{A}_{l,r}) \leq -J(y_{l,r}) \tag{2.31}$$

for $0 \leq l \leq q$ and $1 \leq r \leq r_0$. Combining (2.29), (2.30), and (2.31), we finally arrive at the desired bound (2.28).

4^0 We have now collected all ingredients to prove (2.15). Using (2.16), (2.17), (2.20), and property (iii) of step 2^0 and taking into account that $\Xi^n \circ \pi^{-1} = \mu^n \to \mu$ weakly as $n \to \infty$, we obtain for all sufficiently large n the estimate

$$\begin{aligned}
&\mathbb{P}(\Xi^n \notin U(\Psi(\mu;\rho))) \\
&\leq \mathbb{P}\left(\Xi^n \notin U_1(\Psi(\mu;\rho)),\, \tilde{\Xi}^n \in U_{1/2}(\Psi(\mu;\rho))\right) \\
&\quad + \mathbb{P}\left(\tilde{\Xi}^n \notin U_{1/2}(\Psi(\mu;\rho)),\, \tilde{\Xi}^n\circ\pi^{-1} \in \overline{V(\mu)}\cap\mathcal{K}\right) \\
&\quad + \mathbb{P}\left(\tilde{\Xi}^n\circ\pi^{-1} \notin \overline{V(\mu)}\right) \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{n_i} \mathbb{P}\left(\left|\left\langle \tilde{\Xi}^n - \Xi^n, g_{ij}\right\rangle\right| > \frac{1}{2}\right) \\
&\quad + \mathbb{P}\left(\tilde{\Xi}^n \notin U_{1/2}(\Psi(\mu;\rho)),\, \tilde{\Xi}^n\circ\pi^{-1} \in \overline{V(\mu)}\cap\mathcal{K}\right) \\
&\quad + \sum_{k=1}^{p} \mathbb{P}\left(\left|\left\langle \tilde{\Xi}^n - \Xi^n, f_k\circ\pi\right\rangle\right| > \frac{1}{2}\right).
\end{aligned}$$

To prove (2.15), it therefore suffices to show that

$$\limsup_{n\to\infty} \frac{1}{M_n\gamma_{N_n}} \log \mathbb{P}\left(\tilde{\Xi}^n \notin U_{1/2}(\Psi(\mu;\rho)), \tilde{\Xi}^n\circ\pi^{-1} \in \overline{V(\mu)}\cap\mathcal{K}\right) \leq -\rho + h, \tag{2.32}$$

$$\limsup_{n\to\infty} \frac{1}{M_n\gamma_{N_n}} \log \mathbb{P}\left(\left|\left\langle \tilde{\Xi}^n - \Xi^n, g_{ij}\right\rangle\right| > \frac{1}{2}\right) \leq -\rho \tag{2.33}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, and

$$\limsup_{n\to\infty} \frac{1}{M_n\gamma_{N_n}} \log \mathbb{P}\left(\left|\left\langle \tilde{\Xi}^n - \Xi^n, f_k\circ\pi\right\rangle\right| > \frac{1}{2}\right) \leq -\rho \tag{2.34}$$

for $k = 1, \ldots, p$.

Taking into account the inclusion (2.19) and remembering the definition of $\Psi\left(\overline{V(\mu)} \cap \mathcal{K}; \rho\right)$, we find that $\tilde{S}(\nu) \geq S(\nu) > \rho$ for each ν with $\nu \notin U_{1/2}(\Psi(\dot{\mu}; \rho))$ and $\nu \circ \pi^{-1} \in \overline{V(\mu)} \cap \mathcal{K}$. Therefore an application of the large deviation bound (2.27) yields (2.32).

Using property (i), (2.23), (2.22), and (2.21) of step 2^0, we get for all i, j:

$$\begin{aligned}
\left|\left\langle \tilde{\Xi}^n - \Xi^n, g_{ij} \right\rangle\right| &\leq \frac{1}{M_n} \sum_{k=1}^{M_n} \left| g_{ij}(\tilde{\xi}_k^n) - g_{ij}(\xi_k^n) \right| \\
&< \frac{1}{8} + 2\|g_{ij}\| \frac{1}{M_n} \sum_{k=1}^{M_n} \sum_{(l,r)\notin\Gamma} \mathbb{I}_{A_{l,r}}(\xi_k^n) \\
&\leq \frac{1}{8} + 2\|g_{ij}\| \frac{1}{M_n} \left[|\{k : x_k^n \notin K_{r_0}\}| + \sum_{k=1}^{M_n} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} \mathbb{I}_{A_{l,r}}(\xi_k^n) \right] \\
&\leq \frac{1}{4} + \frac{t}{4\rho} \frac{1}{M_n} \sum_{k=1}^{M_n} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} \mathbb{I}_{A_{l,r}}(\xi_k^n).
\end{aligned}$$

Using this estimate and applying Chebyshev's exponential inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left|\left\langle \tilde{\Xi}^n - \Xi^n, g_{ij} \right\rangle\right| > 1/2\right) \\
&\quad\leq \mathbb{P}\left(\sum_{k=1}^{M_n} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} t\, \mathbb{I}_{A_{l,r}}(\xi_k^n) > M_n\rho \right) \\
&\quad\leq \exp\{-M_n \gamma_{N_n} \rho\} \prod_{k=1}^{M_n} \mathbb{E} \exp\left\{ \gamma_{N_n} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} t\, \mathbb{I}_{A_{l,r}}(\xi_k^n) \right\}. \qquad (2.35)
\end{aligned}$$

Moreover, for each k,

$$\begin{aligned}
&\mathbb{E} \exp\left\{ \gamma_{N_n} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} t\, \mathbb{I}_{A_{l,r}}(\xi_k^n) \right\} \\
&\quad\leq 1 + \exp\{\gamma_{N_n} t\} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} \mathbb{P}(\xi_k^n \in A_{l,r}) \\
&\quad\leq 1 + \exp\{\gamma_{N_n} t\} \sum_{\substack{(l,r)\notin\Gamma \\ r\neq 0}} \sup_{x\in K_r \cap X_{N_n}} P_x^{N_n}(\overline{G}_l). \qquad (2.36)
\end{aligned}$$

But for $(l,r) \notin \Gamma$ and $r \neq 0$ we have $\overline{G}_l \cap \Phi(K_r;t) = \emptyset$ and therefore

$$\inf\{\, J(y) : y \in \overline{G}_l \cap \pi^{-1}(K_r)\,\} > t.$$

Because of this and the Metrizability Hypothesis, an application of the upper large deviation bound for $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$ yields

$$\limsup_{n\to\infty} \frac{1}{\gamma N_n} \log \sup_{x\in K_r\cap X_{N_n}} P_x^{N_n}(\overline{G}_l) < -t$$

for $(l,r) \notin \Gamma$ and $r \neq 0$. Thus, the expression on the right of (2.36) tends to 1 as $n \to \infty$. Combining this with (2.35), we arrive at (2.33). The proof of assertion (2.34) repeats that of (2.33) with g_{ij} replaced by $f_k \circ \pi$.

This completes the proof of the upper large deviation bound (2.15). □

2.2. Other representations of the rate function

In this subsection we derive two more representations of the rate function (2.2). Let us begin with the derivation of a dual expression for the marginal problem (2.2).

Theorem 2.7. *Let the assumptions of Theorem 2.2 be satisfied. Then*

$$S(\mu;\nu) = \sup_{g\in C_b(Y)} \left[\langle \nu, g\rangle - \left\langle \mu, \sup_{y\in Y}[g(y) - I(\,\cdot\,;y)] \right\rangle \right] \tag{2.37}$$

for all $\mu \in \mathcal{M}(X)$ *and* $\nu \in \mathcal{M}(Y)$.

Remark 2.8. From Lemma 1.6 c) we know that the last supremum on the right of (2.37) is sequentially continuous. Therefore

$$S(\mu;\nu) = \sup_{\substack{f\oplus g \le I \\ f\in C_{b,s}(X),\, g\in C_b(Y)}} [\langle \mu, f\rangle + \langle \nu, g\rangle],$$

where $C_{b,s}(X)$ denotes the space of bounded sequentially continuous functions on X. If X is Polish, then $C_{b,s}(X) = C_b(X)$. The representation (2.37) of the rate function S may therefore be regarded as a version of the dual representation for marginal problems, see Kellerer [9]. Unfortunately, the (rather general) assumptions in Kellerer [9] do not exactly fit our needs. The proof given below employs the large deviation background of the functional S.

Proof of Theorem 2.7. Given $Q \in \mathcal{M}(\mu,\nu)$ and $g \in C_b(Y)$, we have

$$\begin{aligned} &\int Q(dx,dy)\, I(x;y) \\ &\quad \ge \int Q(dx,dy) \left(g(y) - \sup_{\tilde y\in Y} [g(\tilde y) - I(x;\tilde y)] \right) \\ &\quad = \langle \nu, g\rangle - \left\langle \mu, \sup_{y\in Y} [g(y) - I(\,\cdot\,;y)] \right\rangle . \end{aligned}$$

This shows that

$$S(\mu;\nu) \geq \sup_{g\in C_b(Y)} \left[\langle \nu, g\rangle - \left\langle \mu, \sup_{y\in Y} [g(y) - I(\,\cdot\,;y)] \right\rangle \right] \tag{2.38}$$

for all $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$.

Fix $\mu \in \mathcal{M}(X)$ arbitrarily and set $\tilde{S}(\mu;\nu) := S(\mu;\nu)$ for $\nu \in \mathcal{M}(Y)$ and $\tilde{S}(\mu;\nu) := +\infty$ for $\nu \in C_b(Y)^* \setminus \mathcal{M}(Y)$. Note that $S(\mu;\,\cdot\,)$ is convex and the level sets $\Psi(\mu;\rho)$, $\rho \geq 0$, are compact in $C_b(Y)^*$ (cf. Lemma 2.4 a)). Therefore the function $\tilde{S}(\mu;\,\cdot\,)$ is convex and lower semi-continuous on $C_b(Y)^*$. But this means that $\tilde{S}(\mu;\,\cdot\,)$ coincides with its bipolar, i.e.

$$S(\mu;\nu) = \sup_{g\in C_b(Y)} [\langle \nu, g\rangle - L(\mu;g)], \qquad \nu \in \mathcal{M}(Y),$$

where

$$L(\mu;g) := \sup_{\nu\in\mathcal{M}(Y)} [\langle \nu, g\rangle - S(\mu;\nu)], \qquad g \in C_b(Y)$$

(see e.g. Ekeland and Temam [6], Chap. 1, Proposition 4.1).

To prove the inequality opposite to (2.38), it will therefore be enough to show that

$$L(\mu;g) \geq \left\langle \mu, \sup_{y\in Y}[g(y) - I(\,\cdot\,;y)] \right\rangle, \qquad g \in C_b(Y). \tag{2.39}$$

Note that this inequality remains valid if one replaces g by $g + \text{const}$.

To prove (2.39), we fix $g \in C_b(Y)$ and assume without loss of generality that $g \leq 0$. We choose an $\mathcal{M}^{M,N}(X)$-sequence $(\mu^{M,N})$ with $\mu^{M,N} \to \mu$ as $M,N \to \infty$. By Theorem 2.2, $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ is a large deviation system with rate function S and scale $M\gamma_N$. We can therefore apply the Laplace-Varadhan method to obtain

$$L(\mu;g) = \lim_{M,N\to\infty} \frac{1}{M\gamma_N} \log \int_{\mathcal{M}(Y)} \mathcal{P}_{\mu^{M,N}}^{M,N}(d\nu)\, \exp\{M\gamma_N\langle \nu, g\rangle\} \tag{2.40}$$

(see e.g. Varadhan [16], Theorem 2.2). It follows from the definition of the measures $\mathcal{P}_{\mu^{M,N}}^{M,N}$ that

$$\begin{aligned} &\frac{1}{M\gamma_N} \log \int_{\mathcal{M}(Y)} \mathcal{P}_{\mu^{M,N}}^{M,N}(d\nu)\, \exp\{M\gamma_N\langle \nu, g\rangle\} \\ &\qquad = \int_X \mu^{M,N}(dx)\, \frac{1}{\gamma_N} \log \int_Y P_x^N(dy) \exp\{\gamma_N g(y)\}. \end{aligned} \tag{2.41}$$

Now fix $\varepsilon > 0$ arbitrarily. Let (M_n) and (N_n) be sequences of natural numbers tending to infinity. We will write μ^n instead of μ^{M_n,N_n}. Because of the Tightness Hypothesis, there exists a compact set $K \subseteq X$ such that

$$\mu^n(K) \geq 1 - \varepsilon \qquad \text{for all } n. \tag{2.42}$$

Applying the Laplace-Varadhan method to the large deviation system $\{P_x^N; x \in X_N, N \in \mathbb{N}\}$, we get

$$\lim_{N\to\infty} \frac{1}{\gamma_N} \log \int_Y P_{x_N}^N(dy) \exp\{\gamma_N g(y)\} = \sup_{y\in Y} [g(y) - I(x;y)]$$

for each X_N-sequence (x_N) tending to $x \in X$. Therefore, taking into account the Metrizability Hypothesis and Lemma 1.6 c), we find for each $x \in K$ an open neighborhood $U(x)$ of x such that

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{\gamma_{N_n}} \log \inf_{\tilde{x}\in U(x)\cap K\cap X_{N_n}} \int P_{\tilde{x}}^{N_n}(dy) \exp\{\gamma_{N_n} g(y)\} \\ \geq \sup_{\tilde{x}\in U(x)\cap K} \sup_{y\in Y} [g(y) - I(\tilde{x};y)] - \varepsilon. \end{aligned} \tag{2.43}$$

We select a finite covering $\{G_1, \ldots, G_r\}$ of K by such open neighborhoods. Taking into account (2.42), we find pairwise disjoint measurable sets $A_i \subseteq G_i$ with $\mu(\partial A_i) = 0$ $(1 \leq i \leq r)$ and

$$\mu\left(X \setminus \bigcup_{i=1}^r (A_i \cap K)\right) < 2\varepsilon. \tag{2.44}$$

From this and (2.42) we conclude that

$$\mu^n\left(X \setminus \bigcup_{i=1}^r (A_i \cap K)\right) < 4\varepsilon \tag{2.45}$$

for all sufficiently large n. Combining (2.40) with (2.41) and taking into account (2.45), we obtain

$$\begin{aligned} L(\mu;g) \geq & \sum_{i=1}^r \liminf_{n\to\infty} \int_{A_i\cap K} \mu^n(dx) \frac{1}{\gamma_{N_n}} \log \int_Y P_x^{N_n}(dy) \exp\{\gamma_{N_n} g(y)\} \\ & - 4\varepsilon \|g\| . \end{aligned} \tag{2.46}$$

Since (2.43) holds for $U(x)$ replaced by A_i, we get

$$\begin{aligned} & \liminf_{n\to\infty} \int_{A_i\cap K} \mu^n(dx) \frac{1}{\gamma_{N_n}} \log \int P_x^{N_n}(dy) \exp\{\gamma_{N_n} g(y)\} \\ & \geq \liminf_{n\to\infty} \mu^n(A_i) \frac{1}{\gamma_{N_n}} \log \inf_{x\in A_i\cap K\cap X_{N_n}} \int P_x^{N_n}(dy) \exp\{\gamma_{N_n} g(y)\} \\ & \geq \mu(A_i) \left(\sup_{x\in A_i\cap K} \sup_y [g(y) - I(x;y)] - \varepsilon \right) \\ & \geq \mu(A_i \cap K) \left(\sup_{x\in A_i\cap K} \sup_y [g(y) - I(x;y)] - \varepsilon \right) - (\|g\| + \varepsilon)\, \mu(A_i \setminus K) \\ & \geq \int_{A_i\cap K} \mu(dx) \left(\sup_y [g(y) - I(x;y)] - \varepsilon \right) - (\|g\| + \varepsilon)\, \mu(A_i \setminus K) \end{aligned} \tag{2.47}$$

for $1 \leq i \leq r$. Here we have used that, as a consequence of $g \leq 0$, the expression under the first integral is nonpositive. We have also used the bound $\sup_y[g(y) - I(x;y)] \geq -\|g\|$. Substituting (2.47) in (2.46) and taking into account (2.44), we arrive at

$$L(\mu;g) \geq \int_X \mu(dx) \sup_y [g(y) - I(x;y)] - \varepsilon\,(7\,\|g\| + 1 + \varepsilon).$$

Since ε may be chosen arbitrarily small, this proves (2.39), and we are done. □

We are now going to derive a further useful representation of the rate function S (formulas (2.48) and (2.49) below) under the restriction that X and Y are Polish spaces. Then, in particular, X satisfies the Metrizability and the Tightness Hypotheses and $\mathcal{M}(X)$ and $\mathcal{M}(Y)$ are also Polish spaces. For each $N \in \mathbb{N}$, we will denote by $\mathcal{M}(X_N)$ the space of Radon probability measures on X_N and by $\mathcal{M}^M(X_N)$ the subspace of M-point empirical measures. We will consider $\mathcal{M}(X_N)$ as a subspace of $\mathcal{M}(X)$ and identify $\mathcal{M}^M(X_N)$ with $\mathcal{M}^{M,N}(X)$, the subspace of $\mathcal{M}(X)$ consisting of M-point empirical measures concentrated on X_N. Let E_x^N denote expectation with respect to P_x^N.

Before formulating our result, we need to introduce the notion of e-convergence ('convergence in terms of the epigraph', cf. Wets [18]). Let Z be a Hausdorff space, and let f_n, $n \in \mathbb{N}$, and f be functions from Z into $\mathbb{R} \cup \{+\infty\}$. We will say that the sequence (f_n) is *e-convergent* to f,

$$f = \operatorname*{e-lim}_{n\to\infty} f_n,$$

if for each $z \in Z$ the following conditions are satisfied:

(i) $$f(z) \leq \sup_{V \in \mathcal{V}(z)} \liminf_{n\to\infty} \inf_{\tilde{z}\in V} f_n(\tilde{z});$$

(ii) $$f(z) \geq \sup_{V \in \mathcal{V}(z)} \limsup_{n\to\infty} \inf_{\tilde{z}\in V} f_n(\tilde{z}).$$

Here $\mathcal{V}(z)$ denotes the system of neighborhoods of z. If Z satisfies the first countability axiom, then (i) and (ii) are equivalent to the following conditions:

(i′) for each sequence (z_n) tending to z it holds

$$f(z) \leq \liminf_{n\to\infty} f_n(z_n);$$

(ii′) there exists a sequence (z_n^0) tending to z such that

$$f(z) \geq \limsup_{n\to\infty} f_n(z_n^0).$$

This notion of convergence is useful in studying sequences of lower semi-continuous convex functions; in particular, a sequence of lower semi-continuous convex functions converges in this sense if and only if their convex conjugates converge (cf. Wets [18]).

Theorem 2.9. *Assume that X and Y are Polish spaces, and let the assumptions of Theorem 2.2 be satisfied. Suppose further that the map $x \mapsto P_x^N$ from X_N into $\mathcal{M}(Y)$ is continuous for each $N \in \mathbb{N}$. Then the following assertions are valid.*

a) For each $N \in \mathbb{N}$, $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^M(X_N), M \in \mathbb{N}\}$ is a large deviation system with scale M and rate function

$$S^N(\mu;\nu) := \sup_{f \in C_b(Y)} \left[\langle \nu, f\rangle - \langle \mu, \log E_\cdot^N e^f\rangle\right], \quad \mu \in \mathcal{M}(X_N),\ \nu \in \mathcal{M}(Y).$$

b) $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ is a large deviation system with scale $M\gamma_N$ (as $M, N \to \infty$). The corresponding rate function S satisfies

$$S(\mu;\cdot) = \operatorname*{e-lim}_{N\to\infty} \gamma_N^{-1} S^N(\mu^N;\cdot) \tag{2.48}$$

for each $\mu \in \mathcal{M}(X)$ and each sequence of measures $\mu^N \in \mathcal{M}(X_N)$ tending to μ.

c) Let Z be a regular Hausdorff space and π a continuous map from $\mathcal{M}(Y)$ into Z. Denote by $\mathcal{Q}_\mu^{M,N}$ the image of the measure $\mathcal{P}_\mu^{M,N}$ with respect to π ($\mu \in \mathcal{M}^{M,N}(X)$, $M \in \mathbb{N}$, $N \in \mathbb{N}$). Then, for each $N \in \mathbb{N}$, $\{\mathcal{Q}_\mu^{M,N}; \mu \in \mathcal{M}^M(X_N), M \in \mathbb{N}\}$ is a large deviation system with scale M and rate function

$$S_\pi^N(\mu;z) := \inf_{\pi(\nu)=z} S^N(\mu;\nu), \qquad \mu \in \mathcal{M}(X_N),\ z \in Z.$$

$\{\mathcal{Q}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ is a large deviation system with scale $M\gamma_N$ (as $M, N \to \infty$) and rate function

$$S_\pi(\mu;z) := \inf_{\pi(\nu)=z} S(\mu;\nu), \qquad \mu \in \mathcal{M}(X),\ z \in Z.$$

Moreover,

$$S_\pi(\mu;\cdot) = \operatorname*{e-lim}_{N\to\infty} \gamma_N^{-1} S_\pi^N(\mu^N;\cdot) \tag{2.49}$$

for each $\mu \in \mathcal{M}(X)$ and each sequence of measures $\mu^N \in \mathcal{M}(X_N)$ tending to μ.

Proof. a) Assertion a) is a Sanov type theorem. Its proof can be found in Dawson and Gärtner [4], Section 3.5.

b) That $\{\mathcal{P}_\mu^{M,N}; \mu \in \mathcal{M}^{M,N}(X), M \in \mathbb{N}, N \in \mathbb{N}\}$ is a large deviation system is a restatement of Theorem 2.2. It only remains to prove (2.48). To this end, we fix $\mu \in \mathcal{M}(X)$ and a sequence of measures $\mu^N \in \mathcal{M}(X_N)$ with $\mu^N \to \mu$. Since X is Polish, the subspaces X_N are metrizable. We may therefore apply Proposition A.1 to find for each $N \in \mathbb{N}$ measures $\mu^{M,N} \in \mathcal{M}^M(X_N)$, $M \in \mathbb{N}$, such that

$$\mu^{M,N} \to \mu^N \qquad \text{in } \mathcal{M}(X_N) \text{ as } M \to \infty. \tag{2.50}$$

Since $\mathcal{M}(X)$ is metrizable, we conclude from this that for each $N \in \mathbb{N}$ there exists $M_0(N) \in \mathbb{N}$ such that

$$\mu^{M,N} \to \mu \quad \text{in } \mathcal{M}(X) \text{ as } M, N \to \infty \text{ and } M \geq M_0(N). \tag{2.51}$$

Given $\nu \in \mathcal{M}(Y)$, let V be an arbitrary neighborhood of ν. From assertion a) and (2.50) we conclude that

$$-\inf_{\tilde{\nu}\in V} S^N(\mu^N;\tilde{\nu}) \leq \liminf_{M\to\infty} M^{-1} \log \mathcal{P}^{M,N}_{\mu^{M,N}}(V) \tag{2.52}$$

for each N. On the other hand, because of Theorem 2.2 and (2.51), we have

$$-\inf_{\tilde{\nu}\in \overline{V}} S(\mu;\tilde{\nu}) \geq \limsup_{\substack{M,N\to\infty \\ M\geq M_0(N)}} M^{-1}\gamma_N^{-1} \log \mathcal{P}^{M,N}_{\mu^{M,N}}(\overline{V}). \tag{2.53}$$

Combining both estimates, we arrive at

$$\inf_{\tilde{\nu}\in \overline{V}} S(\mu;\tilde{\nu}) \leq \liminf_{N\to\infty} \inf_{\tilde{\nu}\in V} \gamma_N^{-1} S^N(\mu^N;\tilde{\nu}).$$

Taking into account that $S(\mu;\cdot)$ is lower semi-continuous and $\mathcal{M}(Y)$ is regular, we conclude from this that

$$S(\mu;\nu) \leq \sup_{V\in\mathcal{V}(\nu)} \liminf_{N\to\infty} \inf_{\tilde{\nu}\in V} \gamma_N^{-1} S^N(\mu^N;\tilde{\nu}). \tag{2.54}$$

Using the large deviation bounds

$$-\inf_{\tilde{\nu}\in \overline{V}} S^N(\mu^N;\tilde{\nu}) \geq \limsup_{M\to\infty} M^{-1} \log \mathcal{P}^{M,N}_{\mu^{M,N}}(\overline{V})$$

and

$$-\inf_{\tilde{\nu}\in V} S(\mu;\tilde{\nu}) \leq \liminf_{\substack{M,N\to\infty \\ M\geq M_0(N)}} M^{-1}\gamma_N^{-1} \log \mathcal{P}^{M,N}_{\mu^{M,N}}(V)$$

opposite to (2.52) and (2.53), respectively, we find in a similar manner that

$$S(\mu;\nu) \geq \sup_{V\in\mathcal{V}(\nu)} \limsup_{N\to\infty} \inf_{\tilde{\nu}\in V} \gamma_N^{-1} S^N(\mu^N;\tilde{\nu}). \tag{2.55}$$

(2.54) and (2.55) together imply the e-convergence of $S^N(\mu^N;\cdot)$ to $S(\mu;\cdot)$.

c) The first half of assertion c) is a consequence of the contraction principle. The proof of the e-convergence (2.49) follows the proof of (2.48) with $\mathcal{P}^{M,N}_\mu$, S^N, and S replaced by $\mathcal{Q}^{M,N}_\mu$, S^N_π, and S_π, respectively. □

3. Randomly perturbed dynamical systems

3.1. Non-interacting diffusions

In this section we deal with diffusion processes in $\mathbb{R}^d$ with generator

$$\mathcal{L}_t^\varepsilon := \frac{\varepsilon^2}{2}\sum_{i,j=1}^d a^{ij}(\,\cdot\,,t)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^d b^i(\,\cdot\,,t)\frac{\partial}{\partial x^i} \tag{3.1}$$

depending on a small parameter $\varepsilon > 0$. More precisely, given $T > 0$, we consider for each $\varepsilon > 0$ the solution $\{P_{x,t}^\varepsilon; (x,t) \in \mathbb{R}^d \times [0,T]\}$ to the martingale problem for $\{\mathcal{L}_t^\varepsilon; t \in [0,T]\}$. Here $P_{x,t}^\varepsilon$ is the law on $C_{0,T} := C([0,T];\mathbb{R}^d)$ of the diffusion process governed by $\{\mathcal{L}_t^\varepsilon; t \in [0,T]\}$ which starts at time t at point x. For details see e.g. Stroock and Varadhan [14]. We will often write P_x^ε instead of $P_{x,0}^\varepsilon$. We impose the following assumption on the diffusion matrix $a(x,t) = \{a^{ij}(x,t)\}$ and the drift vector $b(x,t) = \{b^i(x,t)\}$.

Assumption (D1). The diffusion matrix $a\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^d$ and the drift vector $b\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ are continuous. For each $(x,t) \in \mathbb{R}^d \times [0,T]$, the matrix $a(x,t)$ is symmetric and strictly positive definite.

This condition guarantees uniqueness of the solution to our martingale problem for each $\varepsilon > 0$ (Stroock and Varadhan [14], Theorem 7.2.1 and Corollary 10.1.2), but it does not exclude explosion. In the next assumption we require that our processes do not explode. Non-explosion criteria can be found e.g. in Stroock and Varadhan [14], Chap. 10.

Assumption (D2). The martingale problem for $\{\mathcal{L}_t^\varepsilon; t \in [0,T]\}$ is well-posed for each $\varepsilon > 0$.

We next want to formulate the Freidlin-Wentzell result on large deviations for the family of probability measures $\{P_x^\varepsilon; x \in \mathbb{R}^d, \varepsilon > 0\}$ (Theorem 3.1 below). To this end we need some further notation.

Given $(x,t) \in \mathbb{R}^d \times [0,T]$, we denote by $|\cdot|_{x,t}$ and $\nabla_{x,t}$, respectively, the Riemannian norm and the Riemannian gradient in the tangent space at x for the Riemannian structure on $\mathbb{R}^d$ associated with the diffusion matrix $a(\,\cdot\,,t)$. In particular,

$$|z|_{x,t}^2 = \sum_{i,j=1}^d a_{ij}(x,t) z^i z^j, \qquad z = (z^1,\dots,z^d) \in \mathbb{R}^d,$$

and

$$|\nabla_{x,t} f|_{x,t}^2 = \sum_{i,j=1}^d a^{ij}(x,t) \frac{\partial f(x)}{\partial x^i}\frac{\partial f(x)}{\partial x^j}.$$

Here $\{a_{ij}(x,t)\}$ denotes the inverse of the matrix $\{a^{ij}(x,t)\}$. (Of course, if $a(\,\cdot\,,t)$ is not sufficiently smooth, then there is not really a Riemannian structure associated with $a(\,\cdot\,,t)$, but the above formulas still make sense.) Suppressing the dependence on x, we will often write $|\cdot|_t$ and ∇_t instead of $|\cdot|_{\cdot,t}$ and $\nabla_{\cdot,t}$, respectively.

We define a functional $I: C_{0,T} \to [0,\infty]$ by setting

$$I(\varphi) := \frac{1}{2} \int_0^T |\dot{\varphi}(t) - b(\varphi(t),t)|^2_{\varphi(t),t}\, dt \tag{3.2}$$

if $\varphi \in C_{0,T}$ is absolutely continuous and $I(\varphi) := +\infty$ otherwise. Let

$$\Phi(A;\rho) := \{\, \varphi \in C_{0,T} : \varphi(0) \in A, I(\varphi) \le \rho \,\}, \qquad A \subseteq \mathbb{R}^d,\ \rho \ge 0,$$

denote the associated level sets.

$C_{0,T}$, I, and $\Phi(A;\rho)$ are defined with respect to the time interval $[0,T]$. Given an arbitrary time interval $[s,t] \subseteq [0,T]$, the associated objects will be denoted by $C_{s,t}$, $I_{s,t}$, and $\Phi_{s,t}(A;\rho)$, respectively.

Assumption (D3). (i) For each compact set $K \subset \mathbb{R}^d$ and each $\rho \ge 0$, the set $\Phi(K;\rho)$ is bounded in $C_{0,T}$.

(ii) For each $t \in [0,T)$ and each $x \in \mathbb{R}^d$, the equation

$$\dot{\varphi}(u) = b(\varphi(u),u), \qquad u \in [t,T], \tag{3.3}$$

has at least one solution $\varphi \in C_{t,T}$ with $\varphi(t) = x$.

We are now ready to state the large deviation result of Freidlin and Wentzell in a form which is convenient for our purposes.

Theorem 3.1. *Let the Assumptions (D1)-(D3) be fulfilled. Then $\{P_x^\varepsilon; x \in \mathbb{R}^d, \varepsilon > 0\}$ is a special large deviation system (with respect to the map $\varphi(\cdot) \mapsto \varphi(0)$) with rate function I and scale ε^{-2} as $\varepsilon \to 0$. Moreover, the level sets $\Phi(K;\rho)$ are compact for all compact sets $K \subset \mathbb{R}^d$ and all $\rho \ge 0$.*

For bounded and uniformly continuous drift and diffusion coefficients with uniformly non-degenerate diffusion matrix the proof can be found in Freidlin and Wentzell [7], Chap. 5, Theorem 3.1, in the time homogeneous situation and in Wentzell [17], Theorem 4.3.3, for time inhomogeneous coefficients. Azencott [1], Chap. III, Theorem 2.13, allowed explosion and degeneracy of the diffusion matrix but assumed local Lipschitz continuity of the drift and diffusion coefficients.

In Appendix A.2 it will be shown how Theorem 3.1 may be derived from the results in Wentzell [17] by use of localization techniques.

Remark 3.2. a) Let Assumption (D1) be satisfied. Then Assumption (D3) is equivalent to the condition that the sets $\Phi_{s,t}(K;\rho)$ are *bounded* in $C_{s,t}$ and *non-empty* for $0 \le s < t \le T$, each compact set $K \subset \mathbb{R}^d$, and all $\rho \ge 0$.

Indeed, part (ii) of Assumption (D3) is obviously equivalent to the condition that the sets $\Phi_{s,t}(K;\rho)$ are non-empty. Now let (D1) and (D3) be satisfied. To check that $\Phi_{s,t}(K;\rho)$ is bounded, we define a map $\iota: C_{s,t} \to C_{0,T}$ by setting $\iota(\varphi)(u) := \varphi(s)$ for $u \in [0,s]$, $\iota(\varphi)(u) := \varphi(u)$ for $u \in [s,t]$, and by choosing $\iota(\varphi)(u)$, $u \in [t,T]$, to be a path of the dynamical system (3.3) with $\iota(\varphi)(t) = \varphi(t)$ which exists according to part (ii) of Assumption (D3). It follows that $\iota(\Phi_{s,t}(K;\rho)) \subseteq \Phi(K;\hat{\rho})$ for some $\hat{\rho} > \rho$. Since $\Phi(K;\hat{\rho})$ is bounded by part (i) of Assumption (D3), this yields the boundedness of $\Phi_{s,t}(K;\rho)$.

b) If the drift coefficient b is time-independent, then Assumption (D1) and part (i) of Assumption (D3) together imply part (ii) of Assumption (D3). Otherwise one would find $t \in [0,T)$, $x \in \mathbb{R}^d$, and an unbounded continuously differentiable function $\varphi: [t,T) \to \mathbb{R}^d$ satisfying $\varphi(t) = x$ and $\dot{\varphi}(u) = b(\varphi(u))$ for $u \in [t,T)$. Set $\varphi_n(u) := x$ for $u \in [0, t+1/n]$ and $\varphi_n(u) := \varphi(u - 1/n)$ for $u \in [t+1/n, T]$. Then (φ_n) is an unbounded sequence in $C_{0,T}$ which belongs to $\Phi(x;\rho)$ for some $\rho \geq 0$. But this contradicts part (i) of Assumption (D3).

c) In general, the Assumptions (D1) and (D2) do not imply Assumption (D3). To see this, let F be a bounded smooth real function with $F'(x) > 0$ for all x. Then the Assumptions (D1) and (D2) are satisfied for $d = 1$, $a(x,t) = a(x) = (F'(x))^{-2}$ and $b(x,t) \equiv 0$. But each solution of the equation

$$\dot{\varphi}(t) = \Big(\frac{\rho}{T}\, a(\varphi(t))\Big)^{1/2}, \qquad t \geq 0,$$

explodes before time T for $\rho > (F(+\infty) - F(-\infty))^2/T$. Consequently, for each such ρ, the level sets $\Phi(K;\rho)$ are not bounded in $C_{0,T}$.

d) Let Assumption (D1) be fulfilled. Suppose that there exist a continuously differentiable function $U: \mathbb{R}^d \to \mathbb{R}$ with

$$\lim_{|x|\to\infty} U(x) = +\infty \tag{3.4}$$

and $\lambda \geq 0$ such that

$$\mathcal{L}_t^0 U + \frac{1}{2}|\nabla_t U|_t^2 \leq \lambda U \qquad \text{for all } t \in [0,T]. \tag{3.5}$$

Then Assumption (D3) is satisfied. Before turning to the proof, let us remark that condition (3.5) is certainly fulfilled in the case when $b(\cdot\,,t) = -\nabla_t U$, $t \in [0,T]$.

Using (3.5), we find that

$$\begin{aligned}
\frac{d}{dt}\Big(e^{-\lambda t} U(\varphi(t))\Big) &= e^{-\lambda t}\Big[(\dot{\varphi}(t), \nabla_t U(\varphi(t)))_t - \lambda U(\varphi(t))\Big] \\
&\leq e^{-\lambda t}\Big[(\dot{\varphi}(t) - b(\varphi(t),t), \nabla_t U(\varphi(t)))_t - \frac{1}{2}\,|\nabla_t U(\varphi(t))|_t^2\Big] \\
&\leq e^{-\lambda t}\,\frac{1}{2}\,|\dot{\varphi}(t) - b(\varphi(t),t)|_t^2
\end{aligned}$$

for all absolutely continuous paths $\varphi \in C_{0,T}$ and Lebesgue-almost all $t \in [0,T]$. Here $(\cdot\,,\cdot)_t$ denotes the Riemannian inner product with respect to the diffusion matrix $a(\cdot\,,t)$. Thus, integration yields

$$e^{-\lambda t}U(\varphi(t)) \leq U(\varphi(0)) + I(\varphi), \qquad t \in [0,T].$$

Together with (3.4), this implies the boundedness of the level sets, i.e. part (i) of Assumption (D3). Now let φ be a path of the dynamical system (3.3) in a right-open time interval $[t,t') \subset [0,T]$. Then, analogous to the above, we obtain

$$e^{-\lambda u}U(\varphi(u)) \leq e^{-\lambda t}U(\varphi(t)) \qquad \text{for } u \in [t,t').$$

Hence, φ is bounded on $[t,t')$. In other words, the paths of our dynamical system do not explode. Together with Assumption (D1), this implies part (ii) of Assumption (D3).

We are now going to study large deviations for empirical processes of N independent copies of our diffusion processes in the limit as $N \to \infty$ and $\varepsilon \to 0$.

Before formulating our result (Theorem 3.3 below), we introduce some further notation. By $\mathcal{M} := \mathcal{M}(\mathbb{R}^d)$ we denote the space of probability measures on $\mathbb{R}^d$ endowed with Prokhorov's metric. Given $N \in \mathbb{N}$, let $\mathcal{M}^N$ be the subset of $\mathcal{M}$ consisting of N-point empirical measures, i.e. of measures μ of the form

$$\mu = N^{-1}\sum_{i=1}^{N}\delta_{x_i} \qquad \text{with } x_1,\ldots,x_N \in \mathbb{R}^d. \tag{3.6}$$

Also let $\mathcal{C}_{0,T} := C([0,T];\mathcal{M})$ be the space of continuous functions from $[0,T]$ into $\mathcal{M}$ furnished with the topology of uniform convergence. Both $\mathcal{M}$ and $\mathcal{C}_{0,T}$ are Polish spaces.

Given $N \in \mathbb{N}$, $\varepsilon > 0$, and a measure $\mu \in \mathcal{M}^N$ of the form (3.6), we denote by $\mathcal{P}_\mu^{N,\varepsilon}$ the law on $\mathcal{C}_{0,T}$ of the *empirical process* associated with N independent diffusions having law $P_{x_1}^\varepsilon,\ldots,P_{x_N}^\varepsilon$, respectively. More precisely, $\mathcal{P}_\mu^{N,\varepsilon}$ is the image of the product measure $P_{x_1}^\varepsilon \otimes \cdots \otimes P_{x_N}^\varepsilon$ with respect to the continuous map

$$(C_{0,T})^N \ni (y_1(\cdot),\cdots,y_N(\cdot)) \longmapsto \left(t \mapsto N^{-1}\sum_{i=1}^{N}\delta_{y_i(t)}\right) \in \mathcal{C}_{0,T}. \tag{3.7}$$

We denote by $\mathcal{D}$ the Schwartz space of test functions $\mathbb{R}^d \to \mathbb{R}$ having compact support and possessing continuous derivatives of all orders. We endow $\mathcal{D}$ with the usual inductive topology. Let $\mathcal{D}'$ be the corresponding space of real distributions. For each compact set $K \subset \mathbb{R}^d$, $\mathcal{D}_K$ will denote the subspace of $\mathcal{D}$ consisting of all test functions the support of which is contained in K. Given $\vartheta \in \mathcal{D}'$ and $f \in \mathcal{D}$, let $\langle \vartheta, f\rangle$ denote the application of the test function f to the distribution ϑ. A distribution-valued function $\vartheta(\cdot)\colon [0,T] \to \mathcal{D}'$ will be called *absolutely continuous* if for each compact set $K \subset \mathbb{R}^d$ there exist a neighborhood U_K of 0 in $\mathcal{D}_K$ and an absolutely continuous function $H_K: [0,T] \to \mathbb{R}$ such that

$$|\langle \vartheta(s), f\rangle - \langle \vartheta(t), f\rangle| \le |H_K(s) - H_K(t)|$$

for all $s,t \in [0,T]$ and $f \in U_K$. If $\vartheta(\cdot)$ is absolutely continuous, then the derivative in the distribution sense $\dot\vartheta(t)$ exists for Lebesgue-almost all $t \in [0,T]$, see Dawson and Gärtner [4], Lemma 4.2.

Given $\mu \in \mathcal{M}$ and $t \in [0,T]$, we introduce a normed linear space $T_{\mu,t} := \{ \vartheta \in \mathcal{D}' : \|\vartheta\|_{\mu,t} < \infty \}$ with norm $\|\cdot\|_{\mu,t}$ defined by

$$\|\vartheta\|^2_{\mu,t} := \sup_{f \in \mathcal{D}_{\mu,t}} \frac{|\langle \vartheta, f\rangle|^2}{\langle \mu, |\nabla_t f|^2_t\rangle}, \qquad \vartheta \in \mathcal{D}'. \tag{3.8}$$

Here $\mathcal{D}_{\mu,t} := \{ f \in \mathcal{D} : \langle \mu, |\nabla_t f|^2_t\rangle \neq 0 \}$. Heuristically speaking this means that, for each $t \in [0,T]$, we consider $\mathcal{M}$ as an infinite dimensional 'Riemannian manifold' with 'tangent spaces' $T_{\mu,t}$ and 'Riemannian norm' $\|\cdot\|_{\mu,t}$, $\mu \in \mathcal{M}$.

We define a functional $S^0: \mathcal{C}_{0,T} \to [0,\infty]$ by setting

$$S^0(\mu(\cdot)) := \frac{1}{2}\int_0^T \|\dot\mu(t) - (\mathcal{L}^0_t)^*\mu(t)\|^2_{\mu(t),t}\, dt \tag{3.9}$$

if $\mu(\cdot)$ is absolutely continuous and $S^0(\mu(\cdot)) := +\infty$ otherwise. Here

$$\mathcal{L}^0_t := \sum_{i=1}^d b^i(\,\cdot\,,t)\frac{\partial}{\partial x^i}$$

denotes the operator (3.1) for $\varepsilon = 0$ corresponding to the unperturbed motion $\dot x = b(x,t)$, and $(\mathcal{L}^0_t)^*$ is the formal adjoint of $\mathcal{L}^0_t$ acting on $\mathcal{D}'$. Let

$$\Psi^0(\mathcal{A};\rho) := \{ \mu(\cdot) \in \mathcal{C}_{0,T} : \mu(0) \in \mathcal{A}, S^0(\mu(\cdot)) \le \rho \}, \qquad \mathcal{A} \subseteq \mathcal{M},\ \rho \ge 0,$$

be the level sets associated with S^0.

Theorem 3.3. *Let the Assumptions (D1)–(D3) be satisfied. Then* $\{\mathcal{P}^{N,\varepsilon}_\mu; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ *is a special large deviation system (with respect to the map* $\mu(\cdot) \mapsto \mu(0)$*) with rate function* S^0 *and scale* $N\varepsilon^{-2}$ *as* $N \to \infty$ *and* $\varepsilon \to 0$*. The level sets* $\Psi^0(\mathcal{K};\rho)$ *are compact for all compact subsets* $\mathcal{K}$ *of* $\mathcal{M}$ *and all* $\rho \ge 0$.

The rest of the present section is devoted to the proof of Theorem 3.3. We will assume throughout that the Assumptions (D1)–(D3) are satisfied. We will first show that $\{\mathcal{P}^{N,\varepsilon}_\mu; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ is a special large deviation system and then identify the rate function.

Given $t \in [0,T]$, we will denote by π_t the canonical projection $C_{0,T} \to \mathbb{R}^d$ defined by $\pi_t(\varphi) := \varphi(t)$, $\varphi \in C_{0,T}$. Let $\hat\pi_t$ be the induced map $\mathcal{M}(C_{0,T}) \to \mathcal{M}(\mathbb{R}^d)$, i.e. $\hat\pi_t(Q) := Q \circ \pi_t^{-1}$, $Q \in \mathcal{M}(C_{0,T})$. Since each measure $Q \in \mathcal{M}(C_{0,T})$ may be regarded as the probability law of a stochastic process with continuous paths, the measures $\hat\pi_t(Q)$, $t \in [0,T]$, may be interpreted as the one-dimensional distributions of this process. Recall that I denotes the rate function for $\{P^\varepsilon_x; x \in \mathbb{R}^d, \varepsilon > 0\}$ given by (3.2).

Lemma 3.4. *The family* $\{\mathcal{P}_\mu^{N,\varepsilon}; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ *forms a special large deviation system (with respect to the map* $\mu(\cdot) \mapsto \mu(0)$*) having scale* $N\varepsilon^{-2}$ *and rate function*

$$S(\mu(\cdot)) := \inf\Big\{ \langle Q, I\rangle : Q \in \mathcal{M}(C_{0,T}),\ \hat{\pi}_t(Q) = \mu(t) \text{ for all } t \in [0,T] \Big\}, \quad (3.10)$$

$\mu(\cdot) \in \mathcal{C}_{0,T}$. *The associated level sets*

$$\Psi(\mathcal{K};\rho) := \{\, \mu(\cdot) \in \mathcal{C}_{0,T} : \mu(0) \in \mathcal{K}, S(\mu(\cdot)) \le \rho \,\}$$

are compact for all compact sets $\mathcal{K} \subset \mathcal{M}$ *and all* $\rho \ge 0$.

Proof. Given $N \in \mathbb{N}$, $\varepsilon > 0$, and a measure $\mu \in \mathcal{M}^N$ of the form

$$\mu = N^{-1} \sum_{i=1}^{N} \delta_{x_i}, \qquad x_1, \ldots, x_N \in \mathbb{R}^d,$$

let $\widehat{\mathcal{P}}_\mu^{N,\varepsilon}$ denote the image of the product measure $P_{x_1}^\varepsilon \otimes \cdots \otimes P_{x_N}^\varepsilon$ with respect to the map

$$(C_{0,T})^N \ni (y_1(\cdot), \ldots, y_N(\cdot)) \longmapsto N^{-1} \sum_{i=1}^{N} \delta_{y_i(\cdot)} \in \mathcal{M}(C_{0,T}).$$

Since $\{P_x^\varepsilon; x \in \mathbb{R}^d, \varepsilon > 0\}$ is a special large deviation system (with respect to π_0) with scale ε^{-2} and rate function I and because of the compactness of the associated level sets (Theorem 3.1), we may apply Theorem 2.1 to see that the measures $\widehat{\mathcal{P}}_\mu^{N,\varepsilon}$ form a special large deviation system (with respect to $\hat{\pi}_0$) having scale $N\varepsilon^{-2}$ and rate function

$$\widehat{S}(Q) := \int_{C_{0,T}} I(\varphi)\, Q(d\varphi), \qquad Q \in \mathcal{M}(C_{0,T}).$$

Moreover, according to Lemma 2.4 b), the level sets

$$\widehat{\Psi}(\mathcal{K};\rho) := \Big\{ Q \in \mathcal{M}(C_{0,T}) : \hat{\pi}_0(Q) \in \mathcal{K}, \widehat{S}(Q) \le \rho \Big\}$$

are compact for all compact sets $\mathcal{K} \subset \mathcal{M}$ and all $\rho \ge 0$. But $\mathcal{P}_\mu^{N,\varepsilon}$ is the image of $\widehat{\mathcal{P}}_\mu^{N,\varepsilon}$ with respect to the continuous map

$$\mathcal{M}(C_{0,T}) \ni Q \longmapsto (t \mapsto \hat{\pi}_t(Q)) \in \mathcal{C}_{0,T}. \quad (3.11)$$

(Concerning continuity, see the proof of Lemma 4.6 in Dawson and Gärtner [4].) Therefore the assertions of our lemma now follow by an application of the 'contraction principle'. □

In the rest of this section we show that the functionals S^0 and S defined by (3.9) and (3.10), respectively, coincide. We will first show that $S \geq S^0$ (Lemma 3.6). After that we will prove the opposite inequality for bounded smooth drift and diffusion coefficients (Lemma 3.7). The final (and most difficult) part of this section is devoted to the derivation of the inequality $S \leq S^0$ for arbitrary unbounded coefficients satisfying the Assumptions (D1)–(D3).

Given a function $g\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $t \in [0,T]$, we will denote by $g(t)$ the function $g(t)(x) := g(x,t)$, $x \in \mathbb{R}^d$. We will denote by $C_k^{2,1} = C_k^{2,1}(\mathbb{R}^d \times [0,T])$ the set of continuous real-valued functions on $\mathbb{R}^d \times [0,T]$ having compact support and possessing continuous spatial derivatives of first and second order and a continuous time derivative of first order. We begin with the following lemma which was proved in Dawson and Gärtner [4], Lemma 4.8, for $\mathcal{L}_t^\varepsilon$, $\varepsilon > 0$, instead of $\mathcal{L}_t^0$. It is also valid for $\mathcal{L}_t^0$, since the presence of second order derivatives in $\mathcal{L}_t^\varepsilon$ played no role in the proof.

Lemma 3.5. *For each $\mu(\cdot) \in \mathcal{C}_{0,T}$,*

$$S^0(\mu(\cdot)) = \sup_{g \in C_k^{2,1}} J(\mu(\cdot); g), \tag{3.12}$$

where

$$\begin{aligned} J(\mu(\cdot); g) \quad := \quad & \langle \mu(T), g(T)\rangle - \langle \mu(0), g(0)\rangle \\ & - \int_0^T \left\langle \mu(t), \left(\frac{\partial}{\partial t} + \mathcal{L}_t^0\right) g(t) + \frac{1}{2} |\nabla_t\, g(t)|_t^2 \right\rangle dt. \end{aligned} \tag{3.13}$$

Lemma 3.6. $S \geq S^0$.

Proof. Given $g \in C_k^{2,1}$ and $\varepsilon \geq 0$, we introduce the bounded continuous functional

$$\begin{aligned} F_g^\varepsilon(x(\cdot)) \quad := \quad & g(x(T)) - g(x(0)) \\ & - \int_0^T \left[\left(\frac{\partial}{\partial t} + \mathcal{L}_t^\varepsilon\right) g(x(t),t) + \frac{1}{2} |\nabla_t\, g(x(t),t)|_t^2 \right] dt, \end{aligned}$$

$x(\cdot) \in C_{0,T}$. For $\varepsilon > 0$ this functional has the form

$$F_g^\varepsilon = M_T^\varepsilon - \frac{1}{2\varepsilon^2} \ll M^\varepsilon \gg_T, \tag{3.14}$$

where

$$M_t^\varepsilon(x(\cdot)) := g(x(t)) - g(x(0)) - \int_0^t \left(\frac{\partial}{\partial s} + \mathcal{L}_s^\varepsilon\right) g(x(s),s)\, ds,$$

$t \in [0,T]$, is a bounded continuous P_x^ε-martingale with quadratic characteristic

$$\ll M^\varepsilon \gg_t (x(\cdot)) := \varepsilon^2 \int_0^t |\nabla_s\, g(x(s),s)|_s^2\, ds, \qquad t \in [0,T],$$

for each $x \in \mathbb{R}^d$. Therefore,

$$\exp\left\{\frac{1}{\varepsilon^2}M_t^\varepsilon - \frac{1}{2\varepsilon^4} \ll M^\varepsilon \gg_t\right\}$$

is an exponential P_x^ε-martingale. Because of (3.14), this implies that

$$\varepsilon^2 \log E_x^\varepsilon \exp\left\{\varepsilon^{-2} F_g^\varepsilon\right\} = 0 \tag{3.15}$$

for all $g \in C_k^{2,1}$, $x \in \mathbb{R}^d$, and $\varepsilon > 0$. Here E_x^ε denotes expectation with respect to P_x^ε. Note that F_g^ε converges to F_g^0 uniformly as $\varepsilon \to 0$. Applying the Laplace-Varadhan method (Varadhan [16], Theorem 2.2) for the large deviation system $\{P_x^\varepsilon; x \in \mathbb{R}^d, \varepsilon > 0\}$, we may therefore pass on the left-hand side of (3.15) to the limit as $\varepsilon \to 0$ to obtain

$$\sup_{\varphi(0)=x} \left[F_g^0(\varphi) - I(\varphi)\right] = 0, \qquad x \in \mathbb{R}^d.$$

Hence, $I \geq F_g^0$ and, in particular,

$$\langle Q, I\rangle \geq \langle Q, F_g^0\rangle$$

for each measure $Q \in \mathcal{M}(C_{0,T})$. But

$$\langle Q, F_g^0\rangle = J(\hat{\pi}.(Q); g),$$

where J is defined by (3.13). Combining these facts with (3.10) and (3.12), we finally obtain for each $\mu(\cdot) \in \mathcal{C}_{0,T}$:

$$S(\mu(\cdot)) = \inf_{\hat{\pi}.(Q)=\mu(\cdot)} \langle Q, I\rangle \geq \sup_{g \in C_k^{2,1}} J(\mu(\cdot); g) = S^0(\mu(\cdot)).$$

□

Lemma 3.7. *Assume that the diffusion and drift coefficients (a, b) are bounded and uniformly continuous and that the diffusion matrix a is uniformly non-degenerate and possesses bounded continuous spatial derivatives of first order. Then $S \leq S^0$.*

To prove this lemma, we introduce functionals $S^\varepsilon: \mathcal{C}_{0,T} \to [0, \infty]$, $\varepsilon \geq 0$, by setting

$$S^\varepsilon(\mu(\cdot)) := \frac{1}{2}\int_0^T \|\dot{\mu}(u) - (\mathcal{L}_u^\varepsilon)^* \mu(u)\|_{\mu(u),u}^2 \, du \tag{3.16}$$

if $\mu(\cdot)$ is absolutely continuous and $S^\varepsilon(\mu(\cdot)) := +\infty$ otherwise. Note that for $\varepsilon = 0$ this coincides with our previous definition of S^0.

Let $k: \mathbb{R}^d \to \mathbb{R}$ be a symmetric C^∞ function such that $k(x) > 0$ for $|x| < 1$, $k(x) = 0$ for $|x| \geq 1$, $\int k(x)\, dx = 1$, and

$$\int_{|x|<1} \frac{|\nabla k(x)|^2}{k(x)} \, dx =: \kappa < \infty, \tag{3.17}$$

where ∇ denotes the 'usual' gradient with respect to the Euclidean norm $|\cdot|$. We introduce the smoothing kernels

$$k_\delta(x) := \delta^{-d} k(\delta^{-1} x), \qquad x \in \mathbb{R}^d, \quad \delta > 0.$$

Given a measure $\mu \in \mathcal{M}$ and a function $f \in C_b$, we will denote by μ_δ and f_δ the convolution of μ and f with the kernel k_δ, respectively.

Together with Theorem 2.9 c), the following statement provides the key for the proof of Lemma 3.7.

Lemma 3.8. *Assume that the diffusion and drift coefficients (a, b) are bounded and uniformly continuous and that the diffusion matrix a is uniformly non-degenerate and possesses bounded continuous spatial derivatives of first order. Then*

$$\limsup_{\delta\to 0} \limsup_{\varepsilon\to 0} S^\varepsilon(\mu_\delta(\cdot)) \le S^0(\mu(\cdot))$$

for each $\mu(\cdot) \in \mathcal{C}_{0,T}$.

Proof. We fix $\mu(\cdot) \in \mathcal{C}_{0,T}$ arbitrarily and assume without loss of generality that $\mu(\cdot)$ is absolutely continuous.

1^0 Because of our assumptions on the diffusion matrix a, we find constants C'_a and $0 < \underline{\gamma} < \overline{\gamma} < \infty$ such that

$$\sum_{j=1}^d \left(\sum_{i=1}^d \frac{\partial}{\partial x^i} a^{ij}(x,t) \right)^2 \le C'_a \tag{3.18}$$

and

$$\underline{\gamma} \sum_{i=1}^d \lambda_i^2 \le \sum_{i,j=1}^d a^{ij}(x,t)\lambda_i\lambda_j \le \overline{\gamma} \sum_{i=1}^d \lambda_i^2 \tag{3.19}$$

for all $(x,t) \in \mathbb{R}^d \times [0,T]$ and $\lambda_1, \dots, \lambda_d \in \mathbb{R}$. Note that the diffusion matrix a may be written in the form $\sigma\sigma^*$, i.e.

$$a^{ij}(x,t) = \sum_{k=1}^d \sigma_k^i(x,t)\sigma_k^j(x,t). \tag{3.20}$$

Consider the operators

$$\mathcal{H}_t := \sum_{i,j=1}^d a^{ij}(\cdot\,,t) \frac{\partial^2}{\partial x^i \partial x^j}, \qquad t \in [0,T].$$

We first want to show that

$$\int_0^T \|\mathcal{H}_t^* \mu_\delta(t)\|^2_{\mu_\delta(t),t}\, dt < \infty \qquad \text{for each } \delta > 0. \tag{3.21}$$

Note that, in general, this fails to be true for $\mu_\delta(\cdot)$ replaced by $\mu(\cdot)$. By definition (3.8), we have

$$\|\mathcal{H}_t^*\mu_\delta(t)\|^2_{\mu_\delta(t),t} = \sup_{f\in\mathcal{D}_{\mu_\delta(t),t}} \frac{|\langle \mu(t), k_\delta * \mathcal{H}_t f\rangle|^2}{\left\langle \mu(t), k_\delta * |\nabla_t f|_t^2 \right\rangle}. \tag{3.22}$$

Writing the convolution $k_\delta * \mathcal{H}_t f$ as an integral, integrating by parts, and using (3.20), we find that

$$k_\delta * \mathcal{H}_t f = I_\delta^{(1)} + I_\delta^{(2)}, \tag{3.23}$$

where

$$\begin{aligned} I_\delta^{(1)}(x,t) \quad := \quad & -\int_{|x-y|<\delta} dy \sum_{k=1}^d \left(\sum_{i=1}^d \sigma_k^i(y,t) \frac{\frac{\partial k_\delta(x-y)}{\partial y^i}}{\sqrt{k_\delta(x-y)}} \right) \\ & \times \left(\sum_{j=1}^d \sigma_k^j(y,t) \sqrt{k_\delta(x-y)}\, \frac{\partial f(y)}{\partial y^j} \right) \end{aligned}$$

and

$$I_\delta^{(2)}(x,t) := -\int dy\, k_\delta(x-y) \sum_{j=1}^d \left(\sum_{i=1}^d \frac{\partial}{\partial y^i} a^{ij}(y,t) \right) \frac{\partial f(y)}{\partial y^j}.$$

Applying the Cauchy-Schwarz inequality and again using (3.20), we obtain

$$\begin{aligned} \left| I_\delta^{(1)}(x,t) \right|^2 \quad \le \quad & \int_{|x-y|<\delta} dy \sum_{k=1}^d \left(\sum_{i=1}^d \sigma_k^i(y,t) \frac{\frac{\partial k_\delta(x-y)}{\partial y^i}}{\sqrt{k_\delta(x-y)}} \right)^2 \\ & \times \int dy \sum_{k=1}^d \left(\sum_{j=1}^d \sigma_k^j(y,t) \sqrt{k_\delta(x-y)}\, \frac{\partial f(y)}{\partial y^j} \right)^2 \\ = \quad & \int_{|x-y|<\delta} dy \sum_{i,j=1}^d a^{ij}(y,t) \frac{\frac{\partial k_\delta(x-y)}{\partial y^i}}{\sqrt{k_\delta(x-y)}} \frac{\frac{\partial k_\delta(x-y)}{\partial y^j}}{\sqrt{k_\delta(x-y)}} \\ & \times \int dy\, k_\delta(x-y) \sum_{i,j=1}^d a^{ij}(y,t) \frac{\partial f(y)}{\partial y^i} \frac{\partial f(y)}{\partial y^j}. \end{aligned}$$

Together with (3.19) and (3.17), this implies that

$$|I_\delta^{(1)}|^2 \le \delta^{-2} \kappa \overline{\gamma}\, k_\delta * |\nabla f|^2. \tag{3.24}$$

Similarly, using (3.17)-(3.19), one gets

$$|I_\delta^{(2)}|^2 \le C'_a \underline{\gamma}^{-1}\, k_\delta * |\nabla f|^2. \tag{3.25}$$

Combining (3.23) with (3.24) and (3.25), we conclude that

$$\begin{aligned} |\langle \mu(t), k_\delta * \mathcal{H}_t f\rangle|^2 &\le \left\langle \mu(t), |k_\delta * \mathcal{H}_t f|^2 \right\rangle \\ &\le C_\delta \left\langle \mu(t), k_\delta * |\nabla_t f|_t^2 \right\rangle, \end{aligned}$$

where the constant C_δ does not depend on f or t. Hence, the supremum on the right of (3.22) does not exceed C_δ for all $t \in [0,T]$, and we arrive at (3.21).

2^0 Note that the absolute continuity of $\mu(\cdot)$ implies the absolute continuity of $\mu_\delta(\cdot)$ for each $\delta > 0$. Since $(a+b)^2 \le a^2/\theta + b^2/(1-\theta)$ for $0<\theta<1$ and

$$\mathcal{L}_t^\varepsilon - \mathcal{L}_t^0 = \frac{\varepsilon^2}{2}\mathcal{H}_t,$$

we get

$$\begin{aligned} &\|\dot\mu_\delta(t) - (\mathcal{L}_t^\varepsilon)^*\mu_\delta(t)\|^2_{\mu_\delta(t),t} \\ &\quad\le \frac{1}{\theta}\left\|\dot\mu_\delta(t) - (\mathcal{L}_t^0)^*\mu_\delta(t)\right\|^2_{\mu_\delta(t),t} + \frac{\varepsilon^4}{4(1-\theta)}\|\mathcal{H}_t^*\mu_\delta(t)\|^2_{\mu_\delta(t),t} \end{aligned}$$

for arbitrary $\theta \in (0,1)$ and all $t \in [0,T]$. Thus,

$$S^\varepsilon(\mu_\delta(\cdot)) \le \frac{1}{\theta} S^0(\mu_\delta(\cdot)) + \frac{\varepsilon^4}{8(1-\theta)}\int_0^T \|\mathcal{H}_t^*\mu_\delta(t)\|^2_{\mu_\delta(t),t}\,dt.$$

Together with (3.21), this implies that

$$\limsup_{\varepsilon\to 0} S^\varepsilon(\mu_\delta(\cdot)) \le S^0(\mu_\delta(\cdot)) \qquad \text{for each } \delta > 0.$$

This reduces the proof of our lemma to the verification of the inequality

$$\limsup_{\delta\to 0} S^0(\mu_\delta(\cdot)) \le S^0(\mu(\cdot)). \tag{3.26}$$

3^0 To prove (3.26), we fix $\theta \in (0,1)$ arbitrarily. We first show that

$$\left\langle \mu_\delta(t), |\nabla_t f|_t^2 \right\rangle \ge \theta \left\langle \mu(t), |\nabla_t (k_\delta * f)|_t^2 \right\rangle \tag{3.27}$$

for all $t \in [0,T]$ and $f \in \mathcal{D}$ provided that δ is sufficiently small. Because of (3.19) and the uniform continuity of the diffusion matrix,

$$\sum_{i,j=1}^d a^{ij}(y,t)\lambda_i\lambda_j \ge \theta \sum_{i,j=1}^d a^{ij}(x,t)\lambda_i\lambda_j$$

for all $x, y \in \mathbb{R}^d$ with $|x-y| < \delta$ and all $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$ provided that δ is sufficiently small. Taking into account (3.20), we obtain for such δ:

$$\begin{aligned}
k_\delta * |\nabla_t f|_t^2(x) &= \int dy\, k_\delta(x-y) \sum_{i,j=1}^d a^{ij}(y,t) \frac{\partial f(y)}{\partial y^i} \frac{\partial f(y)}{\partial y^j} \\
&\geq \theta \int dy\, k_\delta(x-y) \sum_{i,j=1}^d a^{ij}(x,t) \frac{\partial f(y)}{\partial y^i} \frac{\partial f(y)}{\partial y^j} \\
&= \theta \sum_{k=1}^d \int dy\, k_\delta(x-y) \left(\sum_{i=1}^d \sigma_k^i(x,t) \frac{\partial f(y)}{\partial y^i} \right)^2 \\
&\geq \theta \sum_{k=1}^d \left(\sum_{i=1}^d \sigma_k^i(x,t) \int dy\, k_\delta(x-y) \frac{\partial f(y)}{\partial y^i} \right)^2 \\
&= \theta \sum_{k=1}^d \left(\sum_{i=1}^d \sigma_k^i(x,t) \frac{\partial}{\partial x^i} (k_\delta * f)(x) \right)^2 \\
&= \theta |\nabla_t (k_\delta * f)|_t^2(x).
\end{aligned}$$

This implies (3.27).

On the other hand, we have

$$\begin{aligned}
&\left|\left\langle \dot{\mu}_\delta(t) - (\mathcal{L}_t^0)^* \mu_\delta(t), f \right\rangle\right|^2 \\
&\quad = \left|\left\langle \dot{\mu}(t) - (\mathcal{L}_t^0)^* \mu(t), k_\delta * f \right\rangle + \left\langle \mu(t), \mathcal{L}_t^0(k_\delta * f) - k_\delta * \mathcal{L}_t^0 f \right\rangle\right|^2 \\
&\quad \leq \frac{1}{\theta} \left|\left\langle \dot{\mu}(t) - (\mathcal{L}_t^0)^* \mu(t), k_\delta * f \right\rangle\right|^2 + \frac{1}{1-\theta} \left\langle \mu(t), \left|\mathcal{L}_t^0(k_\delta * f) - k_\delta * \mathcal{L}_t^0 f\right|^2 \right\rangle.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality and using (3.19), we see that

$$\begin{aligned}
&\left|\mathcal{L}_t^0(k_\delta * f) - k_\delta * \mathcal{L}_t^0 f\right|^2(x) \\
&\quad = \left| \int dy\, k_\delta(x-y) \sum_{i=1}^d \left[b^i(x,t) - b^i(y,t) \right] \frac{\partial f(y)}{\partial y^i} \right|^2 \\
&\quad \leq \int dy\, k_\delta(x-y) \sum_{i=1}^d \left| b^i(x,t) - b^i(y,t) \right|^2 \int dy\, k_\delta(x-y) \sum_{i=1}^d \left(\frac{\partial f(y)}{\partial y^i} \right)^2 \\
&\quad \leq B(\delta) \underline{\gamma}^{-1} k_\delta * |\nabla_t f|_t^2(x),
\end{aligned}$$

where

$$B(\delta) := \sup_{\substack{|x-y|<\delta \\ t\in[0,T]}} \sum_{i=1}^d \left| b^i(x,t) - b^i(y,t) \right|^2 .$$

Therefore,

$$\left|\left\langle \dot{\mu}_\delta(t) - (\mathcal{L}_t^0)^* \mu_\delta(t), f \right\rangle\right|^2 \quad \leq \quad \frac{1}{\theta} \left|\left\langle \dot{\mu}(t) - (\mathcal{L}_t^0)^* \mu(t), k_\delta * f \right\rangle\right|^2$$

$$+\frac{B(\delta)}{\underline{\gamma}(1-\theta)}\left\langle\mu_\delta(t),|\nabla_t f|_t^2\right\rangle. \tag{3.28}$$

Using the definition (3.8) of the norm $\|\cdot\|_{\mu,t}$, (3.28), and (3.27), we find that

$$\left\|\dot\mu_\delta(t)-(\mathcal{L}_t^0)^*\mu_\delta(t)\right\|^2_{\mu_\delta(t),t}\le\frac{1}{\theta^2}\left\|\dot\mu(t)-(\mathcal{L}_t^0)^*\mu(t)\right\|^2_{\mu(t),t}+\frac{B(\delta)}{\underline{\gamma}(1-\theta)},$$

i.e.

$$S^0(\mu_\delta(\cdot))\le\frac{1}{\theta^2}S^0(\mu(\cdot))+\frac{T}{2\underline{\gamma}(1-\theta)}B(\delta) \tag{3.29}$$

for all sufficiently small $\delta>0$. Since the drift vector b is uniformly continuous, $B(\delta)$ tends to zero as $\delta\to 0$. Therefore, letting in (3.29) first $\delta\to 0$ and then $\theta\uparrow 1$, we finally arrive at assertion (3.26). □

Proof of Lemma 3.7. According to Theorem 3.1, $\{P_x^\varepsilon; x\in\mathbb{R}^d,\varepsilon>0\}$ is a special large deviation system with rate function I and scale ε^{-2}. Moreover, the associated level sets $\Phi(K;\rho)$ are compact. Note also that the map $x\mapsto P_x^\varepsilon$ from $\mathbb{R}^d$ into $\mathcal{M}(C_{0,T})$ is continuous for each $\varepsilon>0$ (Stroock and Varadhan [14], Corollary 10.1.4). From Lemma 3.4 and its proof we know that $\mathcal{P}_\mu^{N,\varepsilon}$ is the image of the measure $\widehat{\mathcal{P}}_\mu^{N,\varepsilon}$ with respect to the continuous map (3.11) and that $\{\mathcal{P}_\mu^{N,\varepsilon};\mu\in\mathcal{M}^N,N\in\mathbb{N},\varepsilon>0\}$ is a special large deviation system with rate function S and scale $N\varepsilon^{-2}$. In Dawson and Gärtner [4], Theorem 4.5, it was shown that the family $\{\mathcal{P}_\mu^{N,\varepsilon};\mu\in\mathcal{M}^N,N\in\mathbb{N}\}$ forms a special large deviation system having rate function $\varepsilon^{-2}S^\varepsilon$ and scale N for each $\varepsilon>0$, where S^ε is given by (3.16). We may therefore apply assertion c) of Theorem 2.9 (with respect to the measures P_x^ε and the map (3.11)) to obtain

$$S=\operatorname*{e-lim}_{\varepsilon\to 0}S^\varepsilon.$$

In particular, we have

$$S(\mu_\delta(\cdot))\le\liminf_{\varepsilon\to 0}S^\varepsilon(\mu_\delta(\cdot))$$

for each $\mu(\cdot)\in\mathcal{C}_{0,T}$ and each $\delta>0$. Since the functional S is lower semicontinuous and $\mu_\delta(\cdot)\to\mu(\cdot)$ in $\mathcal{C}_{0,T}$ as $\delta\to 0$, we conclude from this that

$$S(\mu(\cdot))\le\liminf_{\delta\to 0}\liminf_{\varepsilon\to 0}S^\varepsilon(\mu_\delta(\cdot)).$$

Combining this with Lemma 3.8, we arrive at $S\le S^0$. □

So far, we have shown that the inequality $S\le S^0$ and therefore also the assertion of Theorem 3.3 are valid for bounded uniformly continuous drift and diffusion coefficients with a uniformly non-degenerate diffusion matrix possessing bounded continuous spatial derivatives of first order. In the rest of this subsection we show how to prove the inequality $S\le S^0$ for merely continuous unbounded drift and diffusion coefficients. The crucial condition which we imposed on the coefficients to control their growth at infinity is the compactness of the level sets, cf. Assumption (D3). Our approach is to derive an appropriate inequality for bounded coefficients which implies $S \leq S^0$ and to make explicit use of the compactness of the level sets to extend the mentioned inequality to the case of unbounded coefficients.

Note that the Assumptions (D1)–(D3) for the time interval $[0,T]$ imply the analogous assumptions for each subinterval $[s,t] \subseteq [0,T]$. Given $0 \leq s < t \leq T$, we may therefore apply Theorem 3.1 to conclude that the family $\{P^\varepsilon_{x,s}; x \in \mathbb{R}^d, \varepsilon > 0\}$ of probabilities on $C_{s,t}$ forms a special large deviation system with rate function $I_{s,t}$ and compact level sets $\Phi_{s,t}(K;\rho)$. By the 'contraction principle', the family $\{P^\varepsilon_{x,s} \circ \pi_t^{-1}; x \in \mathbb{R}^d, \varepsilon > 0\}$ of probability laws on $\mathbb{R}^d$ also forms a large deviation system having rate function

$$I^{s,t}(x;y) := \inf\{\, I_{s,t}(\varphi) : \varphi \in C_{s,t},\ \varphi(s) = x,\ \varphi(t) = y\,\}, \qquad x, y \in \mathbb{R}^d. \tag{3.30}$$

Moreover, the level sets

$$\Phi^{s,t}(K;\rho) := \Big\{\, y \in \mathbb{R}^d : I^{s,t}(x;y) \leq \rho \text{ for some } x \in K \,\Big\}$$

are compact for all compact sets $K \subset \mathbb{R}^d$ and all $\rho \geq 0$. Given $\nu_1, \dots, \nu_r \in \mathcal{M} = \mathcal{M}(\mathbb{R}^d)$, we will denote by $\mathcal{M}(\nu_1, \dots, \nu_r)$ the set of probability laws on $(\mathbb{R}^d)^r$ with marginals $\nu_1, \dots, \nu_r$.

Lemma 3.9. *The rate function S of the special large deviation system $\{\mathcal{P}^{N,\varepsilon}_\mu; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ has the form*

$$S(\mu(\cdot)) = \sup_{0=t_0<t_1<\cdots<t_r\leq T} \sum_{k=1}^{r} S^{t_{k-1},t_k}(\mu(t_{k-1});\mu(t_k)), \qquad \mu(\cdot) \in \mathcal{C}_{0,T},$$

where

$$S^{s,t}(\mu;\nu) := \inf_{Q \in \mathcal{M}(\mu,\nu)} \int_{(\mathbb{R}^d)^2} I^{s,t}(x;y)\, Q(dx,dy)$$

for $0 \leq s < t \leq T$ and $\mu, \nu \in \mathcal{M}$.

Proof. As a first step we show that

$$S(\mu(\cdot)) = \sup_{0<t_1<\cdots<t_r\leq T} S^{t_1,\dots,t_r}(\mu(0);\mu(t_1),\dots,\mu(t_r)), \tag{3.31}$$

where

$$S^{t_1,\dots,t_r}(\nu_0;\nu_1,\dots,\nu_r) := \inf_{Q \in \mathcal{M}(\nu_0,\dots,\nu_r)} \int_{(\mathbb{R}^d)^r} \sum_{k=1}^{r} I^{t_{k-1},t_k}(y_{k-1};y_k)\, Q(dy_0,\dots,dy_r) \tag{3.32}$$

with the convention $t_0 := 0$.

Let $\mathcal{M}^{[0,T]}$ be the space of functions $[0,T] \to \mathcal{M}$ equipped with the product topology. Given $N \in \mathbb{N}$, $\varepsilon > 0$, and $\mu \in \mathcal{M}^N$, we denote by $\widetilde{\mathcal{P}}_\mu^{N,\varepsilon}$ the image of the measure $\mathcal{P}_\mu^{N,\varepsilon}$ with respect to the continuous imbedding $\mathcal{C}_{0,T} \to \mathcal{M}^{[0,T]}$. Let $\mathcal{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$, $0 < t_1 < \cdots < t_r \le T$, denote the finite dimensional distributions of $\mathcal{P}_\mu^{N,\varepsilon}$. In other words, $\mathcal{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$ is the image of $\mathcal{P}_\mu^{N,\varepsilon}$ with respect to the map

$$\mathcal{C}_{0,T} \ni \mu(\cdot) \longmapsto (\mu(t_1), \ldots, \mu(t_r)) \in \mathcal{M}^r.$$

$\{\widetilde{\mathcal{P}}_\mu^{N,\varepsilon}; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ is a special large deviation system with rate function $\widetilde{S}$ equal to S on $\mathcal{C}_{0,T}$ and equal to $+\infty$ on $\mathcal{M}^{[0,T]} \setminus \mathcal{C}_{0,T}$. By the 'contraction principle', for each partition $0 < t_1 < \cdots < t_r \le T$ of $[0,T]$, the measures $\mathcal{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$ also form a large deviation system as $N \to \infty$ and $\varepsilon \to 0$. Let us denote the associated rate function by $S^{t_1,\ldots,t_r}$. Note that $\widetilde{\mathcal{P}}_\mu^{N,\varepsilon}$ is the projective limit of the measures $\mathcal{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$, $0 \le t_1 < \cdots < t_r \le T$. Projective limits of large deviation systems have been considered in Dawson and Gärtner [4], Theorem 3.3. That result yields the formula (3.31).

We next verify (3.32). To this end, let $\hat{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$ denote the law of the empirical measure associated with N 'independent copies' of the measures $P_x^{\varepsilon;t_1,\ldots,t_r}$, $x \in \mathbb{R}^d$, where $P_x^{\varepsilon;t_1,\ldots,t_r}$, $0 < t_1 < \cdots < t_r \le T$, are the finite dimensional distributions of the diffusion P_x^ε. The measure $P_\mu^{N,\varepsilon;t_1,\ldots,t_r}$ is the image of $\hat{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}$ with respect to the continuous map $\mathcal{M}((\mathbb{R}^d)^r) \to (\mathcal{M}(\mathbb{R}^d))^r$ which transforms each probability measure on $(\mathbb{R}^d)^r$ into its r marginals on $\mathbb{R}^d$. Applying Theorem 3.1 and the 'contraction principle', we find that the family $\{P_x^{\varepsilon;t_1,\ldots,t_r}; x \in \mathbb{R}^d, \varepsilon > 0\}$ forms a large deviation system with scale ε^{-2} and rate function

$$\begin{aligned} I^{t_1,\ldots,t_r}(y_0; y_1, \ldots, y_r) &:= \inf\{\, I(\varphi) : \varphi \in C_{0,T}, \varphi(t_0) = y_0, \ldots, \varphi(t_r) = y_r \,\} \\ &= \sum_{k=1}^r I^{t_{k-1},t_k}(y_{k-1}; y_k), \qquad y_0, \ldots, y_r \in \mathbb{R}^d. \end{aligned}$$

Now, in order to identify the rate function $S^{t_1,\ldots,t_r}$ of $\{P_\mu^{N,\varepsilon;t_1,\ldots,t_r}; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$, we may apply Theorem 2.2 to compute the rate function of $\{\hat{P}_\mu^{N,\varepsilon;t_1,\ldots,t_r}; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon > 0\}$ and then apply the 'contraction principle'. In this way we arrive at (3.32).

To complete the proof, it remains to check that

$$\begin{aligned} &\inf_{Q \in \mathcal{M}(\nu_0,\ldots,\nu_r)} \int_{(\mathbb{R}^d)^r} \sum_{k=1}^r I^{t_{k-1},t_k}(y_{k-1}; y_k)\, Q(dy_0, \ldots, dy_r) \\ &= \sum_{k=1}^r \inf_{Q \in \mathcal{M}(\nu_{k-1},\nu_k)} \int_{(\mathbb{R}^d)^2} I^{t_{k-1},t_k}(y_{k-1}; y_k)\, Q(dy_{k-1}, dy_k) \end{aligned}$$

for $0 = t_0 < t_1 < \cdots < t_r \le T$ and $\nu_0, \ldots, \nu_r \in \mathcal{M}$.

Given $Q \in \mathcal{M}(\nu_0, \ldots, \nu_r)$, let $Q_1, \ldots, Q_r$ denote the 'two-point' marginals of Q corresponding to the variables $(y_0, y_1), \ldots, (y_{r-1}, y_r)$, respectively. Then

$$\int_{(\mathbb{R}^d)^r} \sum_{k=1}^{r} I^{t_{k-1},t_k}(y_{k-1}; y_k)\, Q(dy_0, \ldots, dy_r)$$
$$= \sum_{k=1}^{r} \int_{(\mathbb{R}^d)^2} I^{t_{k-1},t_k}(y_{k-1}; y_k)\, Q_k(dy_{k-1}, dy_k) \tag{3.33}$$

which yields the inequality '$\geq$'. To prove the opposite inequality, fix $Q_k \in \mathcal{M}(\nu_{k-1}, \nu_k)$, $k = 1, \ldots, r$, arbitrarily. Each of these measures Q_k may be written in the form

$$Q_k(dy_{k-1}, dy_k) = q_k(y_{k-1}, dy_k)\nu_{k-1}(dy_{k-1}),$$

where $q_k(y_{k-1}, \cdot\,)$ is the regular conditional probability distribution of $Q_k(dy_{k-1}, dy_k)$ given y_{k-1}. Let Q denote the law of the (time inhomogeneous) Markov chain on $\mathbb{R}^d$ with initial distribution ν_0 and transition kernels q_k:

$$Q(dy_0, \ldots, dy_r) := \nu_0(dy_0) q_1(y_0, dy_1) \cdots q_r(y_{r-1}, dy_r).$$

One easily checks that the 'two-point' marginals of Q coincide with $Q_1, \ldots, Q_r$ and, in particular, Q belongs to $\mathcal{M}(\nu_0, \ldots, \nu_r)$. Thus, equation (3.33) is also valid in this case, and we obtain the inequality '$\leq$'. □

For the remainder of this subsection, we fix s, t with $0 \leq s < t \leq T$ arbitrarily. Lemma 3.9 tells us that, in order to prove the inequality $S \leq S^0$, it will be sufficient to show that

$$\inf_{Q \in \mathcal{M}(\mu(s), \mu(t))} \int_{(\mathbb{R}^d)^2} I^{s,t}(x; y)\, Q(dx, dy) \leq S^0_{s,t}(\mu(\cdot)) \tag{3.34}$$

for all paths $\mu(\cdot) \in \mathcal{C}_{s,t}$, where $S^0_{s,t}$ is defined by (3.9) except that the time interval $[0, T]$ is replaced by $[s, t]$.

We now switch to 'time-reversed' objects. More precisely, we set

$$\bar{I}^{u,v}(x; y) := I^{s+t-v, s+t-u}(y; x), \qquad s \leq u < v \leq t, \quad x, y \in \mathbb{R}^d, \tag{3.35}$$

and

$$\bar{\mathcal{L}}^0_u := -\mathcal{L}^0_{s+t-u}, \qquad u \in [s, t].$$

Moreover, we define a functional $\bar{S}^0_{s,t} \colon \mathcal{C}_{s,t} \to [0, \infty]$ by setting

$$\bar{S}^0_{s,t}(\mu(\cdot)) := \frac{1}{2} \int_s^t \left\| \dot{\mu}(u) - (\bar{\mathcal{L}}^0_u)^* \mu(u) \right\|^2_{\mu(u), s+t-u} du$$

if $\mu(\cdot) \in \mathcal{C}_{s,t}$ is absolutely continuous and $\bar{S}^0_{s,t}(\mu(\cdot)) := +\infty$ otherwise. Note that $\bar{I}^{s,t}(x; y) = I^{s,t}(y; x)$ and that the diffusion operators $\bar{\mathcal{L}}^0_u$ correspond to the time-reversed dynamics $\dot{\varphi}(u) = -b(\varphi(u), s+t-u)$. Replacing $\mu(\cdot)$ by the time-reversed path $\bar{\mu}(u) := \mu(s+t-u)$, $u \in [s,t]$, we see that (3.34) is equivalent to

$$\inf_{Q \in \mathcal{M}(\mu(s),\mu(t))} \int_{(\mathbb{R}^d)^2} \bar{I}^{s,t}(x;y)\, Q(dx,dy) \leq \bar{S}^0_{s,t}(\mu(\cdot)) \tag{3.36}$$

for all $\mu(\cdot) \in \mathcal{C}_{s,t}$.

Let (χ_m) be a sequence of continuous functions on $\mathbb{R}^d$ with compact support such that $0 \leq \chi_m \uparrow 1$ pointwise. Sometimes we will consider χ_m as function on $\mathbb{R}^d \times \mathbb{R}^d$ by setting $\chi_m(x,y) := \chi_m(y)$. We claim that

$$\inf_{Q \in \mathcal{M}(\mu,\nu)} \left\langle Q, \chi_m \bar{I}^{s,t} \right\rangle \longrightarrow \inf_{Q \in \mathcal{M}(\mu,\nu)} \left\langle Q, \bar{I}^{s,t} \right\rangle \qquad \text{as } m \to \infty \tag{3.37}$$

for arbitrary $\mu, \nu \in \mathcal{M}$. Since $\mathcal{M}(\mu,\nu)$ is compact in $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ and the functions $\chi_m \bar{I}^{s,t}$ are nonnegative and continuous, the infimum on the left of (3.37) is attained for some $Q_m \in \mathcal{M}(\mu,\nu)$. We may assume without loss of generality that (Q_m) converges to a measure $Q \in \mathcal{M}(\mu,\nu)$ weakly as $m \to \infty$. But then

$$\liminf_{m\to\infty} \left\langle Q_m, \chi_m \bar{I}^{s,t} \right\rangle \geq \liminf_{m\to\infty} \left\langle Q_m, \chi_n \bar{I}^{s,t} \right\rangle \geq \left\langle Q, \chi_n \bar{I}^{s,t} \right\rangle$$

for each n. Hence, letting $n \to \infty$, we obtain

$$\liminf_{m\to\infty} \left\langle Q_m, \chi_m \bar{I}^{s,t} \right\rangle \geq \left\langle Q, \bar{I}^{s,t} \right\rangle,$$

and this proves (3.37).

According to a duality theorem for marginal problems,

$$\inf_{Q \in \mathcal{M}(\mu,\nu)} \left\langle Q, \chi_m \bar{I}^{s,t} \right\rangle = \sup_{\substack{f \oplus g \leq \chi_m \bar{I}^{s,t} \\ f,g \in C_b}} \left[\langle \mu, f \rangle + \langle \nu, g \rangle \right]$$

(Kellerer [9], Theorem 2.6 and Proposition 1.33; cf. also Theorem 2.7 and Remark 2.8 above). The expression on the right does not exceed

$$\sup_{g \in C_b} \left[\langle \nu, g \rangle - \left\langle \mu, \sup_y \left[g(y) - \chi_m(y) \bar{I}^{s,t}(\,\cdot\,;y) \right] \right\rangle \right].$$

We next show that this supremum may be restricted to *nonnegative* functions $g \in \mathcal{D}$, $g \not\equiv 0$. Denote the expression under the supremum by $H(g)$. Since $H(g) = H(g + \text{const})$, it suffices to take the supremum over strictly positive $g \in C_b$. Now fix $g \in C_b$ with $g \geq \text{const} > 0$ arbitrarily and choose a sequence (g_n) of nonnegative continuous functions with compact support so that $g_n \uparrow g$ pointwise. Then

$$H(g) \leq \liminf_{n\to\infty} H(g_n).$$

This shows that we may restrict ourselves to nonnegative functions $g \in C_b$, $g \not\equiv 0$, with compact support. But each such function can be approached uniformly by nonnegative functions from $\mathcal{D}$, and we arrive at the desired result. Hence, we have shown that

$$\inf_{Q\in\mathcal{M}(\mu,\nu)} \langle Q, \chi_m \bar{I}^{s,t}\rangle \le \sup_{\substack{g\ge 0, g\not\equiv 0\\ g\in\mathcal{D}}} \left[\langle \nu, g\rangle - \left\langle \mu, \sup_y \left[g(y) - \chi_m(y)\bar{I}^{s,t}(\,\cdot\,;y)\right]\right\rangle\right]$$

for each $m \in \mathbb{N}$ and all $\mu, \nu \in \mathcal{M}$.

Putting all things together, we see that the proof of inequality (3.36) may be reduced to the verification of

$$\langle \mu(t), f\rangle - \left\langle \mu(s), \sup_y \left[f(y) - \chi_m(y)\bar{I}^{s,t}(\,\cdot\,;y)\right]\right\rangle \le \bar{S}^0_{s,t}(\mu(\cdot)) \tag{3.38}$$

for $\mu(\cdot) \in \mathcal{C}_{s,t}$, $m \in \mathbb{N}$, and all nonnegative $f \in \mathcal{D}$, $f \not\equiv 0$. Given a non-empty open set $G \subseteq \mathbb{R}^d$, we introduce semi-norms $\|\cdot\|^G_{\mu,u}$, $(\mu, u) \in \mathcal{M}\times[s,t]$, by setting

$$\left(\|\vartheta\|^G_{\mu,u}\right)^2 := \sup_{\substack{f\in\mathcal{D}_{\mu,t}\\ \text{supp } f\subset G}} \frac{|\langle \vartheta, f\rangle|^2}{\langle \mu, |\nabla_u f|^2_u\rangle}, \qquad \vartheta \in \mathcal{D}',$$

and define a functional $\bar{S}^{0,G}_{s,t}\colon \mathcal{C}_{s,t} \to [0,\infty]$ by

$$\bar{S}^{0,G}_{s,t}(\mu(\cdot)) := \frac{1}{2}\int_s^t \left(\left\|\dot{\mu}(u) - (\bar{\mathcal{L}}^0_u)^*\mu(u)\right\|^G_{\mu(u),s+t-u}\right)^2 du \tag{3.39}$$

if $\mu(\cdot)$ is absolutely continuous and $\bar{S}^{0,G}_{s,t}(\mu(\cdot)) := +\infty$ otherwise. Note that, for $G = \mathbb{R}^d$, $\|\cdot\|^G_{\mu,u}$ and $\bar{S}^{0,G}_{s,t}$ coincide with $\|\cdot\|_{\mu,u}$ and $\bar{S}^0_{s,t}$, respectively.

It will turn out later that inequality (3.38) is satisfied even if the functional $\bar{S}^0_{s,t}$ is replaced by the smaller functional $\bar{S}^{0,U(K)}_{s,t}$, where $U(K)$ denotes an arbitrarily small neighborhood of the set $K := K_{s,t}(f)$ which is defined as follows:

$$K_{s,t}(f) := \left\{ x \in \mathbb{R}^d : \min_{\substack{u\in[s,t]\\ y\in\text{supp } f}} \bar{I}^{u,t}(x;y) \le \|f\| \right\}. \tag{3.40}$$

Here supp f and $\|f\|$ denote the support and the sup-norm of f, respectively. By convention, $\bar{I}^{t,t}(x;y) := 0$ for $x = y$ and $\bar{I}^{t,t}(x;y) := \infty$ otherwise. In this form, the desired inequality can be proved first for bounded smooth drift and diffusion coefficients and then be extended to unbounded coefficients by approaching them by bounded ones. The point is that the set K is compact and 'behaves well' under such approximations (see Lemma 3.11 below). The introduction of this set is the reason that we switched to 'time-reversed' objects. Otherwise we would have to consider the set (3.40) with $\bar{I}$ replaced by I which, in general, is not compact.

Let us introduce the functions

$$h(x,u) := \sup_y \left[f(y) - \bar{I}^{u,t}(x;y)\right] \tag{3.41}$$

and

$$h^{(m)}(x,u) := \sup_y \left[f(y) - \chi_m(y)\bar{I}^{u,t}(x;y)\right], \tag{3.42}$$

$(x,u) \in \mathbb{R}^d \times [s,t]$. By convention, $0 \cdot \infty = 0$ on the right of (3.42). Note that $h(t) = h^{(m)}(t) = f$ for large m. Thus, for such m, the expression on the left of (3.38) may be written in the form $\langle \mu(t), h^{(m)}(t)\rangle - \langle \mu(s), h^{(m)}(s)\rangle$.

Summarizing the above considerations, we have found that the proof of the inequality $S \leq S^0$ may be reduced to the following lemma.

Lemma 3.10. *For $\mu(\cdot) \in \mathcal{C}_{s,t}$, all non-vanishing nonnegative functions $f \in \mathcal{D}$, and all sufficiently large m, we have*

$$\left\langle \mu(t), h^{(m)}(t)\right\rangle - \left\langle \mu(s), h^{(m)}(s)\right\rangle \leq \bar{S}^{0,U(K)}_{s,t}(\mu(\cdot)),$$

where $U(K)$ denotes an arbitrary open neighborhood of the set $K = K_{s,t}(f)$ given by (3.40) and $h^{(m)}$ is defined by (3.42).

The proof of this lemma will be broken down into several steps. To this end, let us fix $\mu(\cdot) \in \mathcal{C}_{s,t}$, $m \in \mathbb{N}$, and a non-vanishing nonnegative function $f \in \mathcal{D}$ arbitrarily. We will assume without loss of generality that $\chi_m = 1$ on supp f and, in particular, $h^{(m)}(t) = 1$. We first collect some properties of the set K and the functions h and $h^{(m)}$ defined by (3.40), (3.41), and (3.42), respectively. Let B_R denote the open ball in $\mathbb{R}^d$ with center 0 and radius R.

Lemma 3.11. *a) The set K is compact.*

b) Assume that the diffusion and drift coefficients (a,b) are bounded and uniformly continuous and that the diffusion matrix a is uniformly non-degenerate. Then the function h is nonnegative and continuous, and

$$\text{supp } h \subseteq K \times [s,t].$$

c) Assume that the diffusion and drift coefficients (a,b) are bounded and uniformly continuous and that the diffusion matrix a is uniformly non-degenerate. Let $\{(a_n,b_n)\}$ be a sequence of diffusion and drift coefficients having the same properties. Label each object associated with (a_n,b_n) with the subscript n. Suppose that $a_n \to a$ and $b_n \to b$ uniformly on $\mathbb{R}^d \times [s,t]$. Then

$$h_n \to h \text{ uniformly on } \mathbb{R}^d \times [s,t]$$

and

$$K_n \subset U(K)$$

for each neighborhood $U(K)$ of K and all sufficiently large n.

d) Let (a,b) denote arbitrary diffusion and drift coefficients satisfying the Assumptions (D1)-(D3). Then there exists $R > 0$ such that the following holds true. If $(\tilde{a},\tilde{b})$ is a pair of bounded and uniformly continuous diffusion and drift *coefficients with uniformly non-degenerate diffusion matrix* $\tilde{a}$ *such that* $(\tilde{a}, \tilde{b})$ *coincides with* (a, b) *on* $B_R \times [s,t]$, *then*

$$K = \tilde{K} \text{ and } h^{(m)} \geq \tilde{h}.$$

Here $\tilde{K}$ *and* $\tilde{h}$ *are defined in the same way as* K *and* h, *respectively, but with respect to the coefficients* $(\tilde{a}, \tilde{b})$ *instead of* (a, b).

Proof. a) Under the Assumptions (D1) and (D3), the function $\bar{I}^{u,t}(x;y)$ is continuous in the variables $(u, x, y) \in [s,t) \times \mathbb{R}^d \times \mathbb{R}^d$. Moreover, if $u_n \uparrow t$, $x_n \to x$, $y_n \to y$, and $x \neq y$, then $\bar{I}^{u_n,t}(x_n; y_n) \to \infty$. Recall that $\bar{I}^{t,t}(x;y) = 0$ for $x = y$ and $\bar{I}^{t,t}(x;y) = \infty$ otherwise. From these properties and the compactness of supp f one easily concludes that the minimum in (3.40) is attained and the set K is closed. Suppose that K is not compact. Then we find points x_n with $|x_n| \to \infty$, $y_n \in \operatorname{supp} f$, and $u_n \in [s,t)$ such that

$$I^{s,s+t-u_n}(y_n; x_n) = \bar{I}^{u_n,t}(x_n; y_n) \leq \|f\|.$$

Remembering the definition (3.30) of $I^{s,s+t-u_n}(y_n; x_n)$, we find functions $\varphi_n \in C_{s,t}$ with $\varphi_n(s) = y_n$, $\varphi_n(s+t-u_n) = x_n$, and $I_{s,t}(\varphi_n) = I^{s,s+t-u_n}(y_n; x_n)$. Hence, $I_{s,t}(\varphi_n)$ remains bounded as $n \to \infty$. Thus, since the level sets corresponding to the rate function $I_{s,t}$ are compact (cf. Remark 3.2 a)), the sequence (φ_n) is bounded in $C_{s,t}$. But this contradicts our assumption that $|x_n| \to \infty$.

b) Since the drift coefficient b is bounded and continuous, there exists a solution of $\dot{\varphi}(v) = b(\varphi(v), v)$, $v \in [s, s+t-u]$, with $\varphi(s+t-u) = x$ for each $(x, u) \in \mathbb{R}^d \times [s,t]$. Hence, for each $(x,u) \in \mathbb{R}^d \times [s,t]$, we find some y with $\bar{I}^{u,t}(x;y) = I^{s,s+t-u}(y;x) = 0$. This implies the nonnegativity of h. If $h(x,u) > 0$, then there exists $y \in \operatorname{supp} f$ such that

$$f(y) - \bar{I}^{u,t}(x;y) > 0,$$

i.e. x belongs to K. This shows that the support of h is contained in $K \times [s,t]$. Taking into account the above mentioned properties of $\bar{I}^{u,t}(x;y)$, one also easily checks that h is continuous.

c) We first remark that, under our assumptions on (a_n, b_n) and (a, b),

$$\bar{I}_n^{u_n,t}(x_n; y_n) \to \bar{I}^{u,t}(x;y) \tag{3.43}$$

for $u_n \to u$, $x_n \to x$, and $y_n \to y$ except in the case when $u = t$ and $x = y$. Set

$$M_n(x) := \min_{\substack{u \in [s,t] \\ y \in \operatorname{supp} f}} \bar{I}_n^{u,t}(x;y), \qquad x \in \mathbb{R}^d.$$

Using the definition of $\bar{I}_n^{u,t}(x;y) = I_n^{s,s+t-u}(y;x)$ as a minimum over path integrals and taking into account the assumptions on (a_n, b_n) and (a, b), one easily checks that $M_n(x) \to \infty$ as $|x| \to \infty$ uniformly in n. This means that the sets $K_n = \{ x \in \mathbb{R}^d : M_n(x) \le \|f\| \}$ are bounded uniformly in n. Therefore, in order to show that $K_n \subset U(K)$ for each neighborhood $U(K)$ of K and all sufficiently large n, it will be enough to check that $x_n \in K_n$ and $x_n \to x$ together imply $x \in K$. For each n, we find $u_n \in [s,t]$ and $y_n \in \operatorname{supp} f$ such that

$$\bar{I}_n^{u_n,t}(x_n; y_n) \le \|f\|.$$

We assume without loss of generality that $u_n \to u$ and $y_n \to y$. Because of (3.43), this yields

$$\bar{I}^{u,t}(x; y) \le \|f\|.$$

Thus, since $u \in [s,t]$ and $y \in \operatorname{supp} f$, the point x indeed belongs to K.

We next prove that h_n converges to h uniformly. By the assertions a) and b), h is continuous, $\operatorname{supp} h \subseteq K \times [s,t]$, and $\operatorname{supp} h_n \subseteq K_n \times [s,t]$. Recall that the sets K_n are bounded uniformly in n. Hence, it will be enough to check that

$$h_n(x_n, u_n) \to h(x,u)$$

for $x_n \to x$ and $u_n \to u$. If $u \ne t$, then we may use (3.43) to obtain

$$h_n(x_n, u_n) \ge f(y) - \bar{I}_n^{u_n,t}(x_n; y) \to f(y) - \bar{I}^{u,t}(x; y)$$

for each $y \in \mathbb{R}^d$, i.e.

$$\liminf_{n\to\infty} h_n(x_n, u_n) \ge h(x,u).$$

This inequality is also true for $u = t$. Indeed, in this case $\bar{I}_n^{u_n,t}(x_n; x_n) \to 0$, and therefore

$$h_n(x_n, u_n) \ge f(x_n) - \bar{I}_n^{u_n,t}(x_n; x_n) \to f(x) = h(x,t).$$

Let us now prove the opposite inequality

$$\limsup_{n\to\infty} h_n(x_n, u_n) \le h(x,u).$$

Since h is nonnegative, we assume without loss of generality that $h_n(x_n, u_n) > 0$ for all n. Then we find $y_n \in \operatorname{supp} f$ such that

$$h_n(x_n, u_n) = f(y_n) - \bar{I}_n^{u_n,t}(x_n; y_n).$$

Since f has compact support, we may also assume that $y_n \to y$ for some $y \in \operatorname{supp} f$. Then, using (3.43) once more, we obtain

$$\limsup_{n\to\infty} h_n(x_n, u_n) \le f(y) - \bar{I}^{u,t}(x; y) \le h(x,u).$$

d) According to Lemma A.3, we may choose R so large that for any pair $(\tilde{a}, \tilde{b})$ of diffusion and drift coefficients with the mentioned properties

$$\Phi_{s,s+t-u}(\operatorname{supp} f; \|f\|) = \tilde{\Phi}_{s,s+t-u}(\operatorname{supp} f; \|f\|) \subseteq B_R^{[s,s+t-u]}, \tag{3.44}$$

$s \leq u < t$. Using the representations of $I^{s,s+t-u}(y;x)$ and $\tilde{I}^{s,s+t-u}(y;x)$ as minima over path integrals, we conclude from (3.44) that $K = \tilde{K}$.

Let us now show that $h^{(m)} \geq \tilde{h}$. Since χ_m has compact support, the function $h^{(m)}$ is nonnegative. Because of this and

$$h^{(m)}(x,t) \geq f(x) = \tilde{h}(x,t), \qquad x \in \mathbb{R}^d,$$

it suffices to check that

$$h^{(m)}(x,u) \geq \tilde{h}(x,u)$$

for all $(x,u) \in \mathbb{R}^d \times [s,t)$ with $\tilde{h}(x,u) > 0$. For such (x,u), we find $y \in \operatorname{supp} f$ with

$$\tilde{h}(x,u) = f(y) - \tilde{I}^{s,s+t-u}(y;x) > 0, \tag{3.45}$$

and there is a path $\varphi \in C_{s,s+t-u}$ with $\varphi(s) = y$, $\varphi(s+t-u) = x$, and

$$\tilde{I}^{s,s+t-u}(y;x) = \tilde{I}_{s,s+t-u}(\varphi) < \|f\|.$$

Hence, φ belongs to $\tilde{\Phi}_{s,s+t-u}(\operatorname{supp} f; \|f\|)$. By (3.44), this implies that the path φ is entirely contained in the ball B_R. Therefore, since $(\tilde{a},\tilde{b}) = (a,b)$ on $B_R \times [s, s+t-u]$, $\tilde{I}_{s,s+t-u}(\varphi)$ coincides with $I_{s,s+t-u}(\varphi)$, and we arrive at

$$\tilde{I}^{s,s+t-u}(y;x) \geq I^{s,s+t-u}(y;x).$$

Substituting this in (3.45), we finally obtain

$$\tilde{h}(x,u) \leq f(y) - \chi_m(y) I^{s,s+t-u}(y;x) \leq h^{(m)}(x,u).$$

□

Next we show that the assertion of Lemma 3.10 is valid for bounded smooth drift and diffusion coefficients. More precisely, we have the following lemma.

Lemma 3.12. *Assume that the diffusion and drift coefficients (a,b) are bounded and uniformly continuous and possess bounded continuous derivatives of first order. Assume further that the diffusion matrix a is uniformly non-degenerate. Then*

$$\langle \mu(t), h(t) \rangle - \langle \mu(s), h(s) \rangle \leq \bar{S}^{0,U(K)}_{s,t}(\mu(\cdot)) \tag{3.46}$$

for each open neighborhood $U(K)$ of the set K.

Before proving this lemma, let us remark that the function h defined by the variational expression (3.41) turns out to be a viscosity solution of the Hamilton-Jacobi equation

$$\left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}^0_u \right) h(u) + \frac{1}{2} |\nabla_{s+t-u} h(u)|^2_{s+t-u} = 0, \qquad u \in [s,t], \tag{3.47}$$

with 'initial datum' $h(t) = f$, cf. Crandall and Lions [3] and Lions [10]. Thus, if h were smooth, then we would obtain

$$\begin{aligned}
&\langle \mu(t), h(t)\rangle - \langle \mu(s), h(s)\rangle \\
&\quad = \langle \mu(t), h(t)\rangle - \langle \mu(s), h(s)\rangle \\
&\qquad - \int_s^t du \left\langle \mu(u), \left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}_u^0\right) h(u) + \frac{1}{2}|\nabla_{s+t-u} h(u)|^2_{s+t-u}\right\rangle \\
&\quad = \int_s^t du \left[\langle \dot{\mu}(u) - (\bar{\mathcal{L}}_u^0)^* \mu(u), h(u)\rangle - \frac{1}{2}\left\langle \mu(u), |\nabla_{s+t-u} h(u)|^2_{s+t-u}\right\rangle\right] \\
&\quad \le \frac{1}{2}\int_s^t du \, \frac{|\langle \dot{\mu}(u) - (\bar{\mathcal{L}}_u^0)^* \mu(u), h(u)\rangle|^2}{\left\langle \mu(u), |\nabla_{s+t-u} h(u)|^2_{s+t-u}\right\rangle}.
\end{aligned}$$

Since supp $h \subseteq K \times [s,t]$ (Lemma 3.11 b)), the last expression would then not exceed $\bar{S}_{s,t}^{0,U(K)}(\mu(\cdot))$, and this would imply (3.46). Unfortunately, these arguments fail to be rigorous because of the non-differentiability of h. Note also that the uniqueness conditions of Crandall and Lions [3] are not fulfilled for (3.47). In the following steps the above approach will be made precise by adapting the 'vanishing viscosity method' to our situation, cf. e.g. Lions [10].

Given $u \in [s,t]$ and $\varepsilon \ge 0$, we define the operators

$$\bar{\mathcal{L}}_u^\varepsilon := \frac{\varepsilon^2}{2}\sum_{i,j=1}^d \bar{a}^{ij}(\,\cdot\,,u)\frac{\partial^2}{\partial x^i \partial x^j} - \sum_{i=1}^d \bar{b}^i(\,\cdot\,,u)\frac{\partial}{\partial x^i}, \qquad u \in [s,t],$$

where $\bar{a}^{ij}(\,\cdot\,,u) := a^{ij}(\,\cdot\,,s+t-u)$ and $\bar{b}^i(\,\cdot\,,u) := b^i(\,\cdot\,,s+t-u)$. Note that for $\varepsilon = 0$ this coincides with our previous definition of $\bar{\mathcal{L}}_u^0$. We further consider functionals $\bar{S}_{s,t}^{\varepsilon,G}: \mathcal{C}_{s,t} \to [0,\infty]$, $\varepsilon \ge 0$, which are defined in the same way as $\bar{S}_{s,t}^{0,G}$ but with $\bar{\mathcal{L}}_u^0$ replaced by $\bar{\mathcal{L}}_u^\varepsilon$, see (3.39). To prove Lemma 3.12, we need the following statement which is a slight modification of Lemma 3.8.

Lemma 3.13. *Assume that the diffusion and drift coefficients (a,b) are bounded and uniformly continuous and that the diffusion matrix a is uniformly non-degenerate and possesses bounded continuous spatial derivatives of first order. Let G and H be non-empty open subsets of $\mathbb{R}^d$ such that $\overline{G}$ is compact and $\overline{G} \subset H$. Then*

$$\limsup_{\delta\to 0}\limsup_{\varepsilon\to 0} \bar{S}_{s,t}^{\varepsilon,G}(\mu_\delta(\cdot)) \le \bar{S}_{s,t}^{0,H}(\mu(\cdot)).$$

Proof of Lemma 3.12. 1^0 Let $U(K)$ and $V(K)$ denote open neighborhoods of K such that $\overline{V(K)}$ is compact and $\overline{V(K)} \subset U(K)$. Let $g: \mathbb{R}^d \times [s,t] \to \mathbb{R}$ be a C^∞ function with supp $g \subset V(K) \times [s,t]$. Then, similar to the computations immediately after the statement of Lemma 3.12, we find that

$$\langle \mu_\delta(t), g(t)\rangle - \langle \mu_\delta(s), g(s)\rangle$$

$$-\frac{1}{2}\int_s^t du \left\langle \mu_\delta(u), \left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}_u^\varepsilon\right) g(u) + \frac{1}{2}|\nabla_{s+t-u} g(u)|^2_{s+t-u} \right\rangle$$

$$\leq \frac{1}{2}\int_s^t du \frac{|\langle \dot{\mu}_\delta(u) - (\bar{\mathcal{L}}_u^\varepsilon)^* \mu_\delta(u), g(u)\rangle|^2}{\left\langle \mu_\delta(u), |\nabla_{s+t-u} g(u)|^2_{s+t-u} \right\rangle}$$

for all $\delta > 0$ and $\varepsilon > 0$, where as before $\mu_\delta(u) = k_\delta * \mu(u)$. Hence, remembering the definition of $\bar{S}_{s,t}^{\varepsilon,V(K)}$, we obtain

$$\begin{aligned}\langle \mu_\delta(t), g(t)\rangle - \langle \mu_\delta(s), g(s)\rangle &\leq \bar{S}_{s,t}^{\varepsilon,V(K)}(\mu_\delta(\cdot)) \\ &+ \frac{1}{2}\int_s^t du \left\langle \mu_\delta(u), \left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}_u^\varepsilon\right) g(u) + \frac{1}{2}|\nabla_{s+t-u} g(u)|^2_{s+t-u} \right\rangle. \quad (3.48)\end{aligned}$$

In fact, this inequality is valid for all $g \in C_k^{2,1}(\mathbb{R}^d \times [s,t])$ with supp $g \subset V(K) \times [s,t]$.

2^0 Because of our assumptions on (a, b), the martingale problem for $\{\bar{\mathcal{L}}_u^\varepsilon; u \in [s,t]\}$ admits a unique solution $\{\bar{P}_{x,u}^\varepsilon; (x,u) \in \mathbb{R}^d \times [s,t]\}$ on $C_{s,t}$ for each $\varepsilon > 0$. Let $\bar{E}_{x,u}^\varepsilon$ denote expectation with respect to $\bar{P}_{x,u}^\varepsilon$. For each $\varepsilon > 0$, the function

$$h^\varepsilon(x,u) := \varepsilon^2 \log \bar{E}_{x,u}^\varepsilon \exp\left\{\varepsilon^{-2} f(x(t))\right\}, \qquad (x,u) \in \mathbb{R}^d \times [s,t], \qquad (3.49)$$

is the unique bounded solution of the Cauchy problem

$$\begin{aligned}&\left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}_u^\varepsilon\right) h^\varepsilon(u) + \frac{1}{2}|\nabla_{s+t-u} h^\varepsilon(u)|^2_{s+t-u} = 0, \qquad u \in [s,t], \\ &h^\varepsilon(t) = f.\end{aligned} \qquad (3.50)$$

Let $\pi_t: C_{s,t} \to \mathbb{R}^d$ denote the projection defined by $\pi_t(\varphi) := \varphi(t)$, $\varphi \in C_{s,t}$. The measures $\bar{P}_{x,u}^\varepsilon$ satisfy the Freidlin-Wentzell large deviation principle (Theorem 3.1). Hence, by the 'contraction principle', for each $u \in [s,t)$ the family $\{\bar{P}_{x,u}^\varepsilon \circ \pi_t^{-1}; x \in \mathbb{R}^d, \varepsilon > 0\}$ of probabilities on $\mathbb{R}^d$ forms a large deviation system with rate function $\bar{I}^{u,t}$ and scale ε^{-2}. Therefore, applying the Laplace-Varadhan method and remembering the definition (3.41) of the function h, we derive from (3.49) that

$$h^\varepsilon \to h \qquad \text{boundedly and pointwise on } \mathbb{R}^d \times [s,t] \qquad (3.51)$$

as $\varepsilon \to 0$.

3^0 We now apply inequality (3.48) for

$$g(x,t) = \theta\zeta(x) h^\varepsilon(x,t),$$

where $0 < \theta < 1$ and ζ is a C^∞ function such that $0 \leq \zeta \leq 1$, $\zeta = 1$ in a neighborhood of K, and supp $\zeta \subset V(K)$. As a result, we obtain

$$\begin{aligned}\langle \mu_\delta(t), \zeta h^\varepsilon(t)\rangle - \langle \mu_\delta(s), \zeta h^\varepsilon(s)\rangle &\leq \theta^{-1}\bar{S}_{s,t}^{\varepsilon,V(K)}(\mu_\delta(\cdot)) \\ &+ \frac{1}{2}\int_s^t du \left\langle \mu_\delta(u), \left(\frac{\partial}{\partial u} + \bar{\mathcal{L}}_u^\varepsilon\right)(\zeta h^\varepsilon(u)) + \frac{\theta}{2}|\nabla_{s+t-u}(\zeta h^\varepsilon(u))|^2_{s+t-u} \right\rangle.\end{aligned}$$

For each $u \in [s,t]$, $h^\varepsilon(u) \longrightarrow h(u)$ boundedly and pointwise as $\varepsilon \longrightarrow 0$ by (3.51), and $h(u)$ is continuous and supp $h(u) \subseteq K$ by Lemma 3.11 b). Since $\zeta = 1$ on K, this implies that $\zeta h^\varepsilon(u) \longrightarrow h(u)$ boundedly and pointwise as $\varepsilon \longrightarrow 0$. Moreover, $\mu_\delta(u) \longrightarrow \mu(u)$ weakly as $\delta \longrightarrow 0$. Therefore

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\Big[\langle\mu_\delta(t),\zeta h^\varepsilon(t)\rangle - \langle\mu_\delta(s),\zeta h^\varepsilon(s)\rangle\Big] = \langle\mu(t),h(t)\rangle - \langle\mu(s),h(s)\rangle,$$

and this limit coincides with the expression on the left of (3.46). On the other hand, according to Lemma 3.13,

$$\limsup_{\delta\to 0}\limsup_{\varepsilon\to 0}\bar{S}^{\varepsilon,V(K)}_{s,t}(\mu_\delta(\cdot)) \le \bar{S}^{0,U(K)}_{s,t}(\mu(\cdot)).$$

Thus, in order to finish the proof of Lemma 3.12, it only remains to check that

$$\limsup_{\varepsilon\to 0}\int_s^t du\,\left\langle\mu_\delta(u),\left(\frac{\partial}{\partial u}+\bar{\mathcal{L}}^\varepsilon_u\right)(\zeta h^\varepsilon(u)) + \frac{\theta}{2}\left|\nabla_{s+t-u}(\zeta h^\varepsilon(u))\right|^2_{s+t-u}\right\rangle \le 0 \tag{3.52}$$

for all $\theta \in (0,1)$ and $\delta > 0$. We therefore fix $\theta \in (0,1)$ and $\delta > 0$ arbitrarily. We have

$$\begin{aligned}&\left|\nabla_{s+t-u}(\zeta h^\varepsilon(u))\right|^2_{s+t-u}\\ &\qquad \le \frac{1}{\theta}\zeta\left|\nabla_{s+t-u}h(u)\right|^2_{s+t-u} + \frac{1}{1-\theta}\left|\nabla_{s+t-u}\zeta\right|^2_{s+t-u}(h^\varepsilon(u))^2.\end{aligned}$$

Taking also into account that h^ε satisfies equation (3.50), we see that the verification of (3.52) reduces to the proof of

$$\lim_{\varepsilon\to 0}\int_s^t du\,\left\langle\mu_\delta(u),\bar{\mathcal{L}}^\varepsilon_u(\zeta h^\varepsilon(u)) - \zeta\bar{\mathcal{L}}^\varepsilon_u h^\varepsilon(u)\right\rangle = 0 \tag{3.53}$$

and

$$\lim_{\varepsilon\to 0}\int_s^t du\,\left\langle\mu_\delta(u),\left|\nabla_{s+t-u}\zeta\right|^2_{s+t-u}(h^\varepsilon(u))^2\right\rangle = 0. \tag{3.54}$$

To prove (3.53), we remark that

$$\left\langle\mu_\delta(u),\bar{\mathcal{L}}^\varepsilon_u(\zeta h^\varepsilon(u)) - \zeta\bar{\mathcal{L}}^\varepsilon_u h^\varepsilon(u)\right\rangle = \langle\mu(u),R^\varepsilon_\delta(u)\rangle,$$

where

$$\begin{aligned}R^\varepsilon_\delta(x,u) \;&:=\; \int dy\,k_\delta(x-y)\,(\bar{\mathcal{L}}^\varepsilon_u\zeta)(y)\,h^\varepsilon(y,u)\\ &\qquad -\varepsilon^2\int dy\sum_{i,j=1}^d\frac{\partial}{\partial y^i}\left[k_\delta(x-y)\bar{a}^{ij}(y,u)\frac{\partial\zeta(y)}{\partial y^j}\right]h^\varepsilon(y,u).\end{aligned}$$

This is a result of integration by parts under the assumption that the diffusion matrix a is continuously differentiable with respect to all spatial variables. Because of this, (3.53) now turns out to be straightforward from the fact that $h^\varepsilon \longrightarrow h$ boundedly and pointwise as $\varepsilon \longrightarrow 0$ and the observation that the supports of $h(u)$ and the derivatives of ζ are disjoint. The same argument yields (3.54). □

We now want to remove the differentiability assumptions imposed on the drift and diffusion coefficients by means of approximation.

Lemma 3.14. *Let (a,b) be a pair of bounded and uniformly continuous diffusion and drift coefficients with uniformly non-degenerate diffusion matrix a. Denote by $\{(a_n,b_n)\}$ a sequence of diffusion and drift coefficients with the same properties. For each n, let the functional $\bar{S}_n^{0,G}$ be defined in the same way as $\bar{S}^{0,G} := \bar{S}_{s,t}^{0,G}$ but with respect to (a_n,b_n) instead of (a,b). Assume that $a_n \to a$ and $b_n \to b$ uniformly. Then*

$$\limsup_{n\to\infty} \bar{S}_n^{0,G}(\mu(\cdot)) \le \bar{S}^{0,G}(\mu(\cdot))$$

for each non-empty open set $G \subseteq \mathbb{R}^d$.

Proof. This is a modification of step 3^0 in the proof of Lemma 3.8. Roughly speaking, instead of comparing $k_\delta * \mathcal{L}_t^0 f$ with $\mathcal{L}_t^0(k_\delta * f)$ and $k_\delta * |\nabla_t f|_t^2$ with $|\nabla_t(k_\delta * f)|_t^2$, one has to compare $\bar{\mathcal{L}}_u^{0,n} f$ with $\bar{\mathcal{L}}_u^0 f$ and $(|\nabla_{s+t-u}^{(n)} f|_{s+t-u}^{(n)})^2$ with $|\nabla_{s+t-u} f|_{s+t-u}^2$, respectively. Here $\nabla^{(n)}$, $|\cdot|^{(n)}$, and $\bar{\mathcal{L}}^{0,n}$ are defined in the same way as ∇, $|\cdot|$, and $\bar{\mathcal{L}}^0$, respectively, but with (a,b) replaced by (a_n,b_n). Moreover, one only considers functions $f \in \mathcal{D}$ with the additional property that supp $f \subset G$. The details are left to the reader. □

Lemma 3.15. *The assertion of Lemma 3.12 is valid without the differentiability assumption on the coefficients a and b.*

Proof. Let (a,b) be a pair of bounded and uniformly continuous diffusion and drift coefficients with uniformly non-degenerate diffusion matrix a. Select a sequence $\{(a_n,b_n)\}$ of diffusion and drift coefficients such that $a_n \to a$ and $b_n \to b$ uniformly and, for each n, the diffusion matrix a_n is uniformly non-degenerate and the coefficients (a_n,b_n) are bounded and uniformly continuous and possess bounded continuous derivatives of first order. Label each object associated with (a_n,b_n) with the subscript n.

Let $U(K)$ denote an arbitrary open neighborhood of K. We may apply Lemma 3.12 for (a_n,b_n) instead of (a,b) to obtain

$$\langle \mu(t), h_n(t)\rangle - \langle \mu(s), h_n(s)\rangle \le \bar{S}_n^{0,U(K)}(\mu(\cdot))$$

for large n. Here we have also used that, as a consequence of Lemma 3.11 c), $U(K)$ is a neighborhood of K_n for large n. Letting $n \to \infty$ and applying Lemma 3.11 c) and Lemma 3.14, we arrive at

$$\langle \mu(t), h(t)\rangle - \langle \mu(s), h(s)\rangle \le \bar{S}^{0,U(K)}(\mu(\cdot)),$$

and we are done. □

We are now in a position to switch from bounded to unbounded coefficients and therefore to finish the proof of Lemma 3.10.

Proof of Lemma 3.10. Let (a,b) be arbitrary diffusion and drift coefficients satisfying the Assumptions (D1)–(D3). Let $U(K)$ be a bounded open neighborhood of K. We choose R and bounded coefficients $(\tilde{a},\tilde{b})$ as in assertion d) of Lemma 3.11. We thereby assume without loss of generality that $U(K)$ is contained in the ball B_R. Let $\tilde{h}$ and $\tilde{K}$ be also as in Lemma 3.11 d). Since $(a,b) = (\tilde{a},\tilde{b})$ on $U(K)\times[s,t]$ and $K = \tilde{K}$, the functional $\bar{S}^{0,U(K)}_{s,t}$ will not change if we replace the coefficients (a,b) by $(\tilde{a},\tilde{b})$ (and K by $\tilde{K}$). Hence, we may apply Lemma 3.15 with respect to the coefficients $(\tilde{a},\tilde{b})$ to obtain

$$\left\langle \mu(t), \tilde{h}(t) \right\rangle - \left\langle \mu(s), \tilde{h}(s) \right\rangle \leq \bar{S}^{0,U(K)}(\mu(\cdot)). \tag{3.55}$$

Recall that $\tilde{h}(t) = h^{(m)}(t) = f$ and that $\tilde{h}(s) \leq h^{(m)}(s)$ by Lemma 3.11 d). Substituting this in (3.55), we arrive at the assertion of Lemma 3.10. □

The proof of Theorem 3.3 is now complete.

3.2. McKean-Vlasov interaction

In this subsection it will be shown that our large deviation result for empirical processes of independent diffusions (Theorem 3.3) carries over to diffusions with mean field interaction. We will deal with large systems of coupled diffusions which interact via the empirical measure continuously entering the drift vector.

Let $U: \mathbb{R}^d \to \mathbb{R}$ be a nonnegative twice continuously differentiable function such that $U(x) \to \infty$ as $|x| \to \infty$. Given $R > 0$, let $\mathcal{M}_R$ denote the subspace of $\mathcal{M}$ consisting of all μ for which $\langle \mu, U\rangle \leq R$, and let $\mathcal{C}_R$ denote the space $C([0,T];\mathcal{M}_R)$ furnished with the uniform topology. We introduce a space $\mathcal{M}_\infty$ of admissible probability measures and a corresponding space $\mathcal{C}_\infty$ of measure-valued paths by setting

$$\mathcal{M}_\infty := \bigcup_{R>0} \mathcal{M}_R \quad \text{and} \quad \mathcal{C}_\infty := \bigcup_{R>0} \mathcal{C}_R.$$

We equip both spaces with the strongest topology which induces on $\mathcal{M}_R$ and $\mathcal{C}_R$, respectively, the given topology for each $R > 0$. Concerning the topological properties of these non-metrizable spaces, the reader is referred to Appendix B in Gärtner [8].

We consider diffusion operators

$$\mathcal{L}^\varepsilon(\mu) := \frac{\varepsilon^2}{2} \sum_{i,j=1}^{d} a^{ij}(\,\cdot\,)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^i(\,\cdot\,;\mu)\frac{\partial}{\partial x^i},$$

$\mu \in \mathcal{M}_\infty$, $\varepsilon \geq 0$. By $|\cdot|_x$ and ∇_x we will denote the Riemannian norm and the Riemannian gradient in the tangent space at $x \in \mathbb{R}^d$ associated with the diffusion matrix $\{a^{ij}(x)\}$. We impose the following conditions on the diffusion matrix $a(x) = \{a^{ij}(x)\}$ and the drift vector $b(x;\mu) = \{b^i(x;\mu)\}$.

Assumption (M1). The maps $a\colon \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $b\colon \mathbb{R}^d \times \mathcal{M}_\infty \to \mathbb{R}^d$ are continuous. For each $x \in \mathbb{R}^d$, the matrix $a(x)$ is symmetric and strictly positive definite.

Assumption (M2). There exists a constant $\lambda \geq 0$ such that

$$\left\langle \mu, \mathcal{L}^\varepsilon(\mu)U + \frac{1}{2}|\nabla U|^2 \right\rangle \leq \lambda \langle \mu, U\rangle$$

for all probability measures μ on $\mathbb{R}^d$ with compact topological support and all $\varepsilon \in (0,1)$.

Assumption (M3). For each $\bar{\mu}(\cdot) \in \mathcal{C}_\infty$, there exists a constant $\bar{\lambda} \geq 0$ such that

$$\mathcal{L}^\varepsilon(\bar{\mu}(t))U + \frac{1}{2}|\nabla U|^2 \leq \bar{\lambda}\, U$$

for all $t \in [0,T]$ and all $\varepsilon \in (0,1)$.

Assumption (M4). For each $\bar{\mu}(\cdot) \in \mathcal{C}_\infty$, the function

$$\mathcal{C}_\infty \ni \mu(\cdot) \longmapsto \int_0^T \left\langle \mu(t), |b(\,\cdot\,;\mu(t)) - b(\,\cdot\,;\bar{\mu}(t))|^2 \right\rangle dt \in [0,\infty]$$

is sequentially continuous at point $\mu(\cdot) = \bar{\mu}(\cdot)$.

For each $N \in \mathbb{N}$ and each $\varepsilon \in (0,1)$, we consider an N-particle system of interacting diffusions which is given by the solution $\{P^{N,\varepsilon}_{\mathbf{x}}; \mathbf{x} \in (\mathbb{R}^d)^N\}$ of the martingale problem for the diffusion operator $\mathcal{L}^{N,\varepsilon}$ acting on functions f on $(\mathbb{R}^d)^N$ according to

$$\mathcal{L}^{N,\varepsilon} f(\mathbf{x}) := \sum_{k=1}^N \mathcal{L}^\varepsilon_k\Big(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\Big) f(\mathbf{x}), \qquad \mathbf{x} = (x_1,\ldots,x_N) \in (\mathbb{R}^d)^N.$$

Here $P^{N,\varepsilon}_{\mathbf{x}}$, $\mathbf{x} \in (\mathbb{R}^d)^N$, are probability laws on $C([0,T];(\mathbb{R}^d)^N)$ and $\mathcal{L}^\varepsilon_k(\mu)$ is the operator $\mathcal{L}^\varepsilon(\mu)$ acting on the variable x_k. It was pointed out in Dawson and Gärtner [4], Section 5.1, that, as a consequence of the Assumptions (M1) and (M2), the martingale problem for $\mathcal{L}^{N,\varepsilon}$ is well-posed for each $N \in \mathbb{N}$ and each $\varepsilon \in (0,1)$.

Given $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, and an N-particle empirical measure

$$\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}, \qquad x_1,\ldots,x_N \in \mathbb{R}^d, \tag{3.56}$$

we denote by $\mathcal{P}^{N,\varepsilon}_\mu$ the law of the empirical process associated with our N-particle system starting at μ. It is defined as the image of the measure $P^{N,\varepsilon}_{(x_1,\ldots,x_N)}$ with respect to the continuous map

$$C([0,T];(\mathbb{R}^d)^N) \ni (x_1(\cdot),\ldots,x_N(\cdot)) \longmapsto \Big(t \mapsto \frac{1}{N}\sum_{i=1}^N \delta_{x_i(t)}\Big) \in \mathcal{C}_\infty.$$

Let further $\mathcal{M}^N$ stand for the subset of $\mathcal{M}_\infty$ consisting of measures of the form (3.56).

To formulate our large deviation result, we introduce functionals $S^\varepsilon: \mathcal{C}_\infty \to [0,\infty]$, $0 \le \varepsilon < 1$, by setting

$$S^\varepsilon(\mu(\cdot)) := \frac{1}{2}\int_0^T \|\dot\mu(t) - \mathcal{L}^\varepsilon(\mu(t))^*\mu(t)\|^2_{\mu(t)}\, dt \tag{3.57}$$

if $\mu(\cdot) \in \mathcal{C}_\infty$ is absolutely continuous and $S^\varepsilon(\mu(\cdot)) := +\infty$ otherwise. Here the norms $\|\cdot\|_\mu$ are defined as in (3.8) but now with respect to our time-homogeneous diffusion matrix a. We also define the level sets

$$\Psi^\varepsilon(\mathcal{A};\rho) := \{\, \mu(\cdot) \in \mathcal{C}_\infty : \mu(0) \in \mathcal{A}, S^\varepsilon(\mu(\cdot)) \le \rho \,\}, \qquad \mathcal{A} \subseteq \mathcal{M}_\infty,\ \rho \ge 0.$$

We are now ready to state our result.

Theorem 3.16. *Let the Assumptions (M1)-(M4) be satisfied. Then* $\{\mathcal{P}^{N,\varepsilon}_\mu; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon \in (0,1)\}$ *is a special large deviation system (with respect to the map* $\mu(\cdot) \mapsto \mu(0)$*) with rate function* S^0 *and scale* $N\varepsilon^{-2}$ *as* $N \to \infty$ *and* $\varepsilon \to 0$. *The level sets* $\Psi^0(\mathcal{K};\rho)$ *are compact in* $\mathcal{C}_\infty$ *for all compact subsets* $\mathcal{K}$ *of* $\mathcal{M}_\infty$ *and all* $\rho \ge 0$.

In Dawson and Gärtner [4], Section 5, it was shown that, for fixed $\varepsilon \in (0,1)$, the family $\{\mathcal{P}^{N,\varepsilon}_\mu; \mu \in \mathcal{M}^N, N \in \mathbb{N}\}$ forms a special large deviation system with rate function $\varepsilon^{-2}S^\varepsilon$ and scale N. This assertion was proved by 'freezing' the interaction $\bar\mu(\cdot)$ in the drift vector which made it possible to reduce the 'local' large deviation bounds to that for non-interacting diffusions governed by the 'frozen' operators

$$\bar{\mathcal{L}}^\varepsilon_t := \mathcal{L}^\varepsilon(\bar\mu(t)), \qquad t \in [0,T]. \tag{3.58}$$

This idea also works well in studying large deviations for $N \to \infty$ and $\varepsilon \to 0$ simultaneously. Since the changes consist in obvious modifications only, we will not present the details here. Instead, for the orientation of the reader, we will state the corresponding lemmas without proof.

The first step consists in proving the following lemma.

Lemma 3.17. *For all positive numbers* r *and* ρ *there exists a compact set* $\mathcal{K}$ *in* $\mathcal{C}_\infty$ *such that*

$$\limsup_{\substack{N\to\infty\\ \varepsilon\to 0}} N^{-1}\varepsilon^2 \log \sup_{\mu\in\mathcal{M}_r\cap\mathcal{M}^N} \mathcal{P}^{N,\varepsilon}_\mu(\mathcal{C}_\infty \setminus \mathcal{K}) \le -\rho.$$

The proof of this lemma relies on the fact that sets of the form

$$\mathcal{K} = \mathcal{C}_R \cap \bigcap_n \mathcal{K}_n$$

are compact in $\mathcal{C}_\infty$ for any $R > 0$ and all sets $\mathcal{K}_n$ of the form

$$\mathcal{K}_n = \{\, \mu(\cdot) \in \mathcal{C}_\infty : \langle \mu(\cdot), f_n \rangle \in K_n \,\},$$

where $\{f_n; n \in \mathbb{N}\}$ is a countable dense subset of $\mathcal{D}$ in the sup-norm and K_n, $n \in \mathbb{N}$, denote compact subsets of $C([0,T];\mathbb{R})$. Lemma 3.17 can therefore be derived from the next two lemmas which make essential use of Assumption (M2).

Lemma 3.18. *Given positive numbers r and R, we have*

$$\sup_{\mu \in \mathcal{M}_r \cap \mathcal{M}^N} \mathcal{P}_\mu^{N,\varepsilon}(\mathcal{C}_\infty \setminus \mathcal{C}_R) \le \exp\{-N\varepsilon^{-2} R_T\}$$

for all $N \in \mathbb{N}$ and $\varepsilon \in (0,1)$, where $R_T := R\exp\{-\lambda T\} - r$ and λ is taken from Assumption (M2).

Lemma 3.19. *Given $R > 0$, $\rho > 0$, and a function $f \in \mathcal{D}$, we find a compact subset K of $C([0,T];\mathbb{R})$ such that*

$$\mathcal{P}_\mu^{N,\varepsilon}(\mathcal{C}_R \setminus \mathcal{K}_f) \le \exp\{-N\varepsilon^{-2}\rho\}$$

for all $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\mu \in \mathcal{M}^N$, where

$$\mathcal{K}_f := \{\, \mu(\cdot) \in \mathcal{C}_\infty : \langle \mu(\cdot), f \rangle \in K \,\}.$$

Lemma 3.17 allows us to reduce the proof of Theorem 3.16 to the consideration of 'local' large deviation bounds. To obtain these bounds we fix $\bar{\mu}(\cdot) \in \mathcal{C}_\infty$ arbitrarily and consider the operators $\bar{\mathcal{L}}_t^\varepsilon$ defined by (3.58). The Assumptions (M1) and (M3) guarantee that the martingale problem for $\{\bar{\mathcal{L}}_t^\varepsilon; t \in [0,T]\}$ admits a unique solution $\{\bar{P}_{x,t}^\varepsilon; x \in \mathbb{R}^d \times [0,T]\}$ on $C([0,T];\mathbb{R}^d)$ for each $\varepsilon \in (0,1)$. Given $N \in \mathbb{N}$ and $\varepsilon \in (0,1)$, let $\bar{\mathcal{P}}_\mu^{N,\varepsilon}$, $\mu \in \mathcal{M}^N$, denote the laws on $C([0,T];\mathcal{M})$ of the empirical processes of N independent diffusions governed by the operators $\bar{\mathcal{L}}_t^\varepsilon$. As a consequence of Assumption (M3), condition (3.5) in Remark 3.2 d) is fulfilled in the situation considered here. Hence, we may apply Theorem 3.3 to conclude that $\{\bar{\mathcal{P}}_\mu^{N,\varepsilon}; \mu \in \mathcal{M}^N, N \in \mathbb{N}, \varepsilon \in (0,1)\}$ is a special large deviation system as $N \to \infty$ and $\varepsilon \to 0$ with scale $N\varepsilon^{-2}$ and rate function $\bar{S}^0$ given by (3.57) for $\varepsilon = 0$ except that $\mathcal{L}_t^0$ is replaced by $\bar{\mathcal{L}}_t^0$. Note that $S^0(\bar{\mu}(\cdot)) = \bar{S}^0(\bar{\mu}(\cdot))$. Fix $N \in \mathbb{N}$, $\varepsilon \in (0,1)$, and $\mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ arbitrarily. Then, by the Cameron-Martin-Girsanov Theorem, the measure $P_{\mathbf{x}}^{N,\varepsilon}$ is absolutely continuous with respect to $\bar{P}_{\mathbf{x}}^{N,\varepsilon} := \bar{P}_{x_1,0}^\varepsilon \otimes \cdots \otimes \bar{P}_{x_N,0}^\varepsilon$, and

$$\frac{dP_{\mathbf{x}}^{N,\varepsilon}}{d\bar{P}_{\mathbf{x}}^{N,\varepsilon}} = \exp\left\{ M_T^{N,\varepsilon} - \frac{1}{2} \langle\!\langle M^{N,\varepsilon} \rangle\!\rangle_T \right\},$$

where $M^{N,\varepsilon}$ is a continuous local $\bar{P}_{\mathbf{x}}^{N,\varepsilon}$-martingale with quadratic characteristic

$$\langle\!\langle M^{N,\varepsilon} \rangle\!\rangle_t (\mathbf{x}(\cdot)) = N\varepsilon^{-2} \int_0^t \left\langle \nu_{\mathbf{x}(u)}, |b(\,\cdot\,; \nu_{\mathbf{x}(u)}) - b(\,\cdot\,; \bar{\mu}(u))|^2 \right\rangle du.$$

Here $\nu_{\mathbf{x}} := N^{-1} \sum_{i=1}^{N} \delta_{x_i}$ denotes the empirical measure of the configuration $\mathbf{x} = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$. Because of Assumption (M4), this allows to obtain the following 'local' large deviation bounds from the corresponding bounds for the 'frozen' probabilities $\bar{\mathcal{P}}_{\mu}^{N,\varepsilon}$.

Lemma 3.20. *Given $\mu_N \in \mathcal{M}^N$ and $\mu \in \mathcal{M}_\infty$, suppose that $\mu_N \to \mu$ in $\mathcal{M}_\infty$. Then the following assertions are valid for each $\bar{\mu}(\cdot) \in \mathcal{C}_\infty$ with $\bar{\mu}(0) = \mu$.*

a) For each open neighborhood $\mathcal{V}$ of $\bar{\mu}(\cdot)$ in $\mathcal{C}_\infty$,

$$\liminf_{\substack{N\to\infty \\ \varepsilon\to 0}} N^{-1}\varepsilon^2 \log \mathcal{P}_{\mu_N}^{N,\varepsilon}(\mathcal{V}) \geq -S^0(\bar{\mu}(\cdot)).$$

b) For each $\gamma > 0$ there exists an open neighborhood $\mathcal{V}$ of $\bar{\mu}(\cdot)$ in $\mathcal{C}_\infty$ such that

$$\limsup_{\substack{N\to\infty \\ \varepsilon\to 0}} N^{-1}\varepsilon^2 \log \mathcal{P}_{\mu_N}^{N,\varepsilon}(\mathcal{V}) \leq -S^0(\bar{\mu}(\cdot)) + \gamma \tag{3.59}$$

provided that $S^0(\bar{\mu}(\cdot)) < \infty$. If $S^0(\bar{\mu}(\cdot)) = \infty$, then this assertion holds with the expression on the right of (3.59) replaced by $-\gamma$.

We remark that in the proof of assertion a), in order to switch from the topology on $C([0,T];\mathcal{M})$ to the topology on $\mathcal{C}_\infty = C([0,T];\mathcal{M}_\infty)$, we have also applied Lemma 3.18 with $\mathcal{P}_{\mu}^{N,\varepsilon}$ replaced by $\bar{\mathcal{P}}_{\mu}^{N,\varepsilon}$ and Assumption (M2) replaced by Assumption (M3).

Finally, the relative compactness of the level sets $\Psi^0(\mathcal{K};\rho)$ follows by a combination of Lemma 3.17 and Lemma 3.20 a). That these sets are closed can be deduced from the representation of $S^0(\mu(\cdot))$ in the form (3.12)–(3.13) with $\mathcal{L}_t^0$ replaced by $\mathcal{L}^0(\mu(t))$.

We close this section with a few remarks on the McKean-Vlasov equations related to our empirical processes. For each ε, $0 \leq \varepsilon < 1$, the weak solutions $\mu(\cdot) \in \mathcal{C}_\infty$ of the McKean-Vlasov equation

$$\dot{\mu}(t) = \mathcal{L}^\varepsilon(\mu(t))^* \mu(t), \qquad t \in [0,T],$$

coincide with the zeros of the corresponding rate function S^ε. The Assumptions (M1) and (M2) imply that there is at least one solution for each initial datum $\mu(0) \in \mathcal{M}_\infty$ and each $\varepsilon \in [0,1)$, see Gärtner [8]. But our assumptions do not ensure uniqueness. We refer to Scheutzow [12] for a discussion of uniqueness and non-uniqueness in the degenerate case $\varepsilon = 0$. Adequate uniqueness conditions for $\varepsilon \neq 0$ can be found e.g. in [8], Section 2.3. These conditions also ensure uniqueness for $\varepsilon = 0$ under the additional assumption that the degenerate Fokker-Planck equation

$$\dot{\mu}(t) = \mathcal{L}^0(\bar{\mu}(t))^* \mu(t), \qquad t \in [0,T],$$

admits a unique weak solution $\mu(\cdot) \in C([0,T];\mathcal{M})$ for each initial datum $\mu(0) \in \mathcal{M}$ and each $\bar{\mu}(\cdot) \in \mathcal{C}_\infty$. This is certainly true if the vector field $b(x;\mu)$ is continuously differentiable in x.

Let $\mu^\varepsilon(\cdot)$ and $\mu^0(\cdot)$ denote weak solutions of the McKean-Vlasov equation for the operators $\mathcal{L}^\varepsilon(\cdot)$ and $\mathcal{L}^0(\cdot)$, respectively. Assuming uniqueness for $0 \leq \varepsilon < 1$ and using results from [8], one also readily checks that $\mu^\varepsilon(\cdot) \to \mu^0(\cdot)$ in $\mathcal{C}_\infty$ as $\varepsilon \to 0$ provided that $\mu^\varepsilon(0) \to \mu^0(0)$ in $\mathcal{M}_\infty$.

A. Appendix

A.1. $\mathcal{M}^{M,N}(X)$-sequences

Let X be a completely regular Hausdorff space, and let (X_N) be a sequence of subsets of X such that each point in X is the limit of an X_N-sequence. Denote by $\mathcal{M}(X)$ the space of Radon probability measures on X equipped with the topology of weak convergence. Given $M, N \in \mathbb{N}$, denote by $\mathcal{M}^{M,N}(X)$ the subset of M-point empirical measures on X_N. In the following we assume that X and (X_N) satisfy the Metrizability Hypothesis of Section 2.

Proposition A.1. *Each measure in $\mathcal{M}(X)$ is the weak limit of an $\mathcal{M}^{M,N}(X)$-sequence as $M, N \to \infty$.*

Proof. Fix $\mu \in \mathcal{M}(X)$ arbitrarily. We must find measures $\mu^{M,N} \in \mathcal{M}^{M,N}(X)$ with $\mu^{M,N} \to \mu$ in $\mathcal{M}(X)$. Since μ is a Radon measure, there exists a sequence of compact subsets $K_1 \subseteq K_2 \subseteq \ldots$ of X such that μ is concentrated on $\bigcup_r K_r$. We can therefore write μ in the form

$$\mu = \sum_{r=1}^{\infty} \mu_r\,, \tag{A.1}$$

where, for each r, μ_r is a measure which is concentrated on K_r.

Now fix $r \in \mathbb{N}$ arbitrarily. Let (ξ_m) be a sequence of independent K_r-valued random variables with joint law $\mu_r(\cdot)/\mu_r(X)$. According to the strong law of large numbers,

$$\frac{1}{M} \sum_{m=1}^{[M\mu_r(X)]} f(\xi_m) \longrightarrow \int_{K_r} f\, d\mu_r \qquad \text{a.s.} \tag{A.2}$$

as $M \to \infty$ for each $f \in C(K_r)$, where $[x]$ denotes the integer part of $x \in \mathbb{R}$ and $C(K_r)$ is the space of continuous functions on K_r. Since K_r is metrizable, the space $C(K_r)$ is separable. Because of this, (A.2) implies the weak convergence

$$\frac{1}{M} \sum_{m=1}^{[M\mu_r(X)]} \delta_{\xi_m} \longrightarrow \mu_r \qquad \text{a.s.}$$

In particular, there exists a sequence (x_{rm}) in K_r with

$$\frac{1}{M} \sum_{m=1}^{[M\mu_r(X)]} \delta_{x_{rm}} \longrightarrow \mu_r \tag{A.3}$$

weakly as $M \to \infty$.

Let $X(K_r)$ be the metrizable set introduced in the Metrizability Hypothesis, and let ρ_r be a metric on $X(K_r)$ which generates the subspace topology of X. Using the Metrizability Hypothesis and the compactness of K_r, we see that $X_N \cap X(K_r)$ is non-empty for all sufficiently large N and $\rho_r(x, X_N \cap X(K_r)) \to 0$ as $N \to \infty$ uniformly in $x \in K_r$. For each m and each N, we can therefore select a point $x_{rm}^N \in X_N$ such that $x_{rm}^N \in X(K_r)$ for all sufficiently large N and all m and

$$\rho_r(x_{rm}^N, x_{rm}) \to 0 \qquad \text{uniformly in } m$$

as $N \to \infty$. Together with (A.3) this implies that

$$\mu_r^{M,N} := \frac{1}{M} \sum_{m=1}^{[M\mu_r(X)]} \delta_{x_{rm}^N} \longrightarrow \mu_r \tag{A.4}$$

weakly as $M, N \to \infty$. Note that $\mu_r^{M,N}(X) \leq \mu_r(X)$ and $\mu_r^{M,N}(X) \to \mu_r(X)$ as $M, N \to \infty$ for each r. Therefore $\sum_r \mu_r^{M,N}(X) \leq 1$ and $\sum_r \mu_r^{M,N}(X) \to 1$ as $M, N \to \infty$.

Now we define

$$\mu^{M,N} := \sum_{r=1}^{\infty} \mu_r^{M,N} + \frac{k_{M,N}}{M} \delta_{x_0^N}, \tag{A.5}$$

where x_0^N is an arbitrary point in X_N and $k_{M,N}$ is a nonnegative integer making $\mu^{M,N}$ into a probability measure. By construction, $\mu^{M,N} \in \mathcal{M}^{M,N}(X)$ for all $M, N \in \mathbb{N}$ and $k_{M,N}/M \to 0$ as $M, N \to \infty$. Passing in (A.5) to the limit as $M, N \to \infty$ and using thereby (A.4) and (A.1), we find that

$$\mu^{M,N} \to \mu \qquad \text{in } \mathcal{M}(X) \text{ as } M, N \to \infty,$$

and we are done. □

A.2. Freidlin-Wentzell estimates

The objective of this appendix is to prove Theorem 3.1. From Wentzell [17], Theorem 4.3.3, we know that the assertion of Theorem 3.1 is valid under the following hypothesis.

Assumption (W). The diffusion matrix $a: \mathbb{R}^d \times [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^d$ and the drift vector $b: \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ are bounded and uniformly continuous. The matrix a is symmetric, positive definite, and uniformly non-degenerate.

Our idea consists in reducing the general case (Assumptions (D1)–(D3) of Section 3.1) to that of Wentzell by changing the drift and diffusion coefficients outside of a ball with center 0 in such a way that the new coefficients satisfy Assumption (W). If the ball is sufficiently large, then this change will not influence the considered large deviation quantities.

In the sequel, B_R will denote the open ball in $\mathbb{R}^d$ with center 0 and radius R, and $B_R^{[u,v]}$ will stand for the set of functions on the interval $[u,v]$ with values in B_R.

We first prove the compactness of the level sets.

Lemma A.2. *Let the Assumptions (D1) and (D3) be satisfied. Then the sets* $\Phi_{s,t}(K;\rho)$ *are compact for* $0 \le s < t \le T$, *all compact sets* $K \subset \mathbb{R}^d$, *and all* $\rho \ge 0$.

Proof. 1^0 The Assumptions (D1) and (D3) imply the corresponding assumptions for the time interval $[s,t]$ instead of $[0,T]$, cf. Remark 3.2 a). Therefore it will be sufficient to consider $s = 0$ and $t = T$ only.

2^0 We show that the function $I: C_{0,T} \to [0,\infty]$ is lower semi-continuous. Suppose that $\varphi_n \to \varphi$ in $C_{0,T}$. We choose R so large that the paths φ_n, $n \in \mathbb{N}$, and φ belong to the ball B_R. We replace the coefficients a and b by new coefficients $\tilde{a}$ and $\tilde{b}$, respectively, so that $a = \tilde{a}$ and $b = \tilde{b}$ on $\overline{B_R} \times [0,T]$ and $\tilde{a}$ and $\tilde{b}$ satisfy Assumption (W). Then the associated rate function $\tilde{I}$ is lower semi-continuous. But $\tilde{I}(\varphi_n) = I(\varphi_n)$, $n \in \mathbb{N}$, and $\tilde{I}(\varphi) = I(\varphi)$. Hence $I(\varphi) \le \liminf I(\varphi_n)$.

3^0 Now fix a compact set $K \subset \mathbb{R}^d$ and $\rho \ge 0$ arbitrarily. By Assumption (D3), the set $\Phi(K;\rho)$ is bounded and non-empty. Thus, $\Phi(K;\rho) \subseteq B_R^{[0,T]}$ for some $R > 0$. Replacing a and b by $\tilde{a}$ and $\tilde{b}$, respectively, as in step 2^0 and denoting the associated level set by $\tilde{\Phi}(K;\rho)$, we find that $\Phi(K;\rho) \subseteq \tilde{\Phi}(K;\rho)$. Since $\tilde{\Phi}(K;\rho)$ is compact, $\Phi(K;\rho)$ is relatively compact. From step 2^0 we know that $\Phi(K;\rho)$ is closed. Hence $\Phi(K;\rho)$ is compact. □

We next show that $\Phi_{s,t}(K;\rho)$ coincides with $\tilde{\Phi}_{s,t}(K;\rho)$ for sufficiently large R.

Lemma A.3. *Let the Assumptions (D1) and (D3) be satisfied. Let a compact subset* K *of* $\mathbb{R}^d$ *and* $\rho \ge 0$ *be given. Then there exists* $R > 0$ *such that the following holds true. For any diffusion and drift coefficients* $(\tilde{a},\tilde{b})$ *satisfying Assumption (W) and coinciding with* (a,b) *on* $B_R \times [0,T]$, *we have*

$$\tilde{\Phi}_{s,t}(K;\rho) = \Phi_{s,t}(K;\rho) \subseteq B_R^{[s,t]}, \qquad 0 \le s < t \le T,$$

where $\Phi_{s,t}(K;\rho)$ *and* $\tilde{\Phi}_{s,t}(K;\rho)$ *denote the level sets associated with* (a,b) *and* $(\tilde{a},\tilde{b})$, *respectively.*

Proof. We fix a compact set $K \subset \mathbb{R}^d$ and $\rho \ge 0$ arbitrarily.

1^0 We show that

$$\Phi_{s,t}(K;\rho) \subseteq B_R^{[s,t]} \qquad \text{for } 0 \le s < t \le T \tag{A.6}$$

provided that R is sufficiently large. Suppose the contrary. Then we find a sequence of functions $\varphi_n: [s_n,t_n] \to \mathbb{R}^d$ with $0 \le s_n < t_n \le T$, $\varphi_n(s_n) \in K$,

$I_{s_n,t_n}(\varphi_n) \le \rho$ and $|\varphi_n(t_n)| \to \infty$. In accordance with part (ii) of Assumption (D3), we may continue φ_n to a function $\psi_n \in C_{0,T}$ by setting $\psi_n(u) = \varphi_n(s_n)$ for $u \in [0, s_n]$, $\psi_n = \varphi_n$ on $[s_n, t_n]$, and choosing ψ_n on $[t_n, T]$ as a solution of $\dot{\psi}_n(u) = b(\psi_n(u), u)$, $u \in [t_n, T]$. Then, on the one hand, $I(\psi_n)$ remains bounded for $n \to \infty$, i.e. the sequence (ψ_n) belongs to $\Phi(K; \tilde{\rho})$ for some $\tilde{\rho} > \rho$. But, on the other hand, the sequence (ψ_n) is unbounded in $C_{0,T}$, and this contradicts part (i) of Assumption (D3).

2^0 We choose R so large that (A.6) is fulfilled. Let $(\tilde{a}, \tilde{b})$ be diffusion and drift coefficients which satisfy Assumption (W) and coincide with (a, b) on $B_R \times [0, T]$. Fix $0 \le s < t \le T$ arbitrarily. Recall that $I_{s,t}$ and $\tilde{I}_{s,t}$ denote the rate functions associated with (a, b) and $(\tilde{a}, \tilde{b})$, respectively. It remains to check that $\tilde{\Phi}_{s,t}(K; \rho)$ coincides with $\Phi_{s,t}(K; \rho)$. Since $I_{s,t}$ and $\tilde{I}_{s,t}$ coincide on $B_R^{[s,t]}$, it will be sufficient to verify that not only $\Phi_{s,t}(K; \rho)$ but also $\tilde{\Phi}_{s,t}(K; \rho)$ is contained in $B_R^{[s,t]}$. Suppose that $\tilde{\Phi}_{s,t}(K; \rho) \not\subseteq B_R^{[s,t]}$. Then we find a function $\varphi \in \tilde{\Phi}_{s,t}(K; \rho)$ and $u \in (s, t]$ such that $|\varphi(v)| < R$ for $v \in [s, u)$ and $|\varphi(u)| = R$, i.e. $I_{s,u}(\varphi) = \tilde{I}_{s,u}(\varphi) \le \rho$ and $\varphi \notin B_R^{[s,t]}$. Therefore $\varphi \in \Phi_{s,u}(K; \rho) \setminus B_R^{[s,u]}$ which contradicts (A.6). □

Proof of Theorem 3.1. The compactness of the level sets was shown in Lemma A.2. Fix $x \in \mathbb{R}^d$, $\rho \ge 0$, and $\varphi \in C_{0,T}$ with $\varphi(0) = x$ arbitrarily. Let $x^\varepsilon \in \mathbb{R}^d$ be such that $x^\varepsilon \to x$ as $\varepsilon \to 0$. Denote by $U(\varphi)$ and $U(\Phi(x; \rho))$ bounded open neighborhoods of φ and $\Phi(x; \rho)$, respectively (cf. Assumption (D3)). It suffices to check that

$$\liminf_{\varepsilon \to 0} \varepsilon^2 \log P^\varepsilon_{x^\varepsilon}(U(\varphi)) \ge -I(\varphi) \tag{A.7}$$

and

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \log P^\varepsilon_{x^\varepsilon}\Big(C_{0,T} \setminus U(\Phi(x; \rho))\Big) \le \rho \tag{A.8}$$

(cf. Freidlin and Wentzell [7], Chap. 3, Theorem 3.3).

We choose $R > 0$ so that $U(\varphi)$ and $U(\Phi(x; \rho))$ are contained in $B_R^{[0,T]}$ and the assertions of Lemma A.3 are valid for $K = \{x\}$ and certain coefficients $\tilde{a}$ and $\tilde{b}$. This means that $I(\varphi)$, $\Phi(x; \rho)$, and the probabilities on the left of (A.7) and (A.8) will not change if we replace a and b by $\tilde{a}$ and $\tilde{b}$, respectively. But, according to Wentzell [17], Theorem 4.3.3, the bounds (A.7) and (A.8) hold for the diffusion processes with diffusion matrix $\tilde{a}$ and drift vector $\tilde{b}$ instead of a and b, respectively. This proves (A.7) and (A.8), and we are done. □

References

1. Azencott, R.: Grandes deviations et applications. In: Ecole d'Eté de Probabilités de Saint-Flour VIII-1978. Lect. Notes Math. 774, pp. 2–176. Berlin, Heidelberg, New York: Springer 1980
2. Billingsley, P.: Convergence of probability measures. New York: Wiley 1968
3. Crandall, M. G., Lions, P.: Viscosity solutions of Hamilton-Jacobi equations. Trans. Am. Math. Soc. **277**, 1–42 (1983)
4. Dawson, D. A., Gärtner, J.: Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics **20**, 247–308 (1987)
5. Dawson, D. A., Gärtner, J.: Long time behaviour of interacting diffusions. In: Norris, J. R. (ed.) Stochastic calculus in application. Proc. Cambridge Symp., 1987, pp. 29–54. Longman 1988
6. Ekeland, I., Temam, R.: Convex analysis and variational problems. Amsterdam, Oxford: North Holland 1976
7. Freidlin, M. I., Wentzell, A. D.: Random perturbations of dynamical systems. New York: Springer 1984
8. Gärtner, J.: On the McKean-Vlasov limit for interacting diffusions. Math. Nachr. **137**, 197–248 (1988)
9. Kellerer, H. G.: Duality theorems for marginal problems. Z. Wahrscheinlichkeitstheor. Verw. Geb. **67**, 399–432 (1984)
10. Lions, P.: Generalized solutions of Hamilton-Jacobi equations. Boston, London, Melbourne: Pitman 1982
11. Postnikov, M. M.: Introduction to Morse theory (in Russian). Moscow: Nauka 1971
12. Scheutzow, M.: Uniqueness and non-uniqueness of solutions of Vlasov-McKean equations. J. Austral. Math. Soc. (Series A) **43**, 246–256 (1987)
13. Schwartz, L.: Radon measures on arbitrary topological spaces and cylindrical measures. Oxford Univ. Press 1973
14. Stroock, D. W., Varadhan, S. R. S.: Multidimensional diffusion processes. New York: Springer 1979
15. Topsøe, F.: Topology and measure. Lect. Notes Math., vol. 133. Berlin, Heidelberg, New York: Springer 1970
16. Varadhan, S. R. S.: Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46. Philadelphia: SIAM 1984
17. Wentzell, A. D.: Limit theorems on large deviations for Markov stochastic processes. Dordrecht: Kluwer Acad. Publ. 1990
18. Wets, R. J. B.: Convergence of convex functions, variational inequalities and convex optimization problems. In: Cottle, R. W., Giannessi, F., Lions, J. L. (eds.) Variational inequalities and complementarity problems, pp. 375–403. Chichester: Wiley 1980