ANALYTIC ASPECTS OF MULTILEVEL LARGE DEVIATIONS[1]

By D.A. Dawson and J. Gärtner

Fields Institute Toronto and Technical University Berlin

This is a continuation of our paper [4] on multilevel systems of diffusions. Here the focus is on multilevel large deviations for noninteracting diffusions. However, the tools developed in this paper may also be used to include mean field interactions. Given MN independent copies $\xi_{ij}(t)$, $i = 1, \dots, M$, $j = 1, \dots, N$, of a non-degenerate $\mathbb{R}^d$-valued diffusion, we consider the level II empirical processes $\Xi^{MN}(t) = M^{-1}\sum_{i=1}^{M}\delta_{\Xi_i^N(t)}$, where $\Xi_i^N(t) = N^{-1}\sum_{j=1}^{N}\delta_{\xi_{ij}(t)}$ denote the corresponding level I processes. Although the study of dynamical large deviations was initiated in [4], the method there was in fact only applied to obtain an integral representation of the rate function for a simple caricature of such a hierarchical model. The main objective of this paper is to provide an appropriate integral representation for the rate function describing the exponential decay of large deviation probabilities for the processes $\Xi^{MN}(\cdot)$ as $M, N \to \infty$. This requires the development of new tools which may also be of more general interest. In particular, we introduce an appropriate class of distributions on spaces of measures, provide an analogue of the Weierstrass polynomial approximation for functions of measures, and consider dual de Finetti approximations.

Contents

February 10, 1997.
[1] Supported by NSERC
AMS 1991 *subject classifications.* Primary 60F10, 60K35, 60J60
Key words and phrases. Large deviations, empirical processes, distributions of measures, hierarchical systems of diffusions.

1. Introduction and statement of main result.

1.1. Background and motivation. Multilevel systems of diffusions were introduced in Dawson and Gärtner [3] in order to model and study equilibrium and nonequilibrium phenomena. These models arise as a generalization of the well-known mean field models but reflect more closely than do mean field models the qualitative behavior of short and intermediate range systems. An important feature of this class of models is that for large, but finite, system size the corresponding effects are organized in multiple time scales, and in fact this provides a caricature of the behavior of short range systems at successively larger spatial scales. Dynamical large deviations play an important role in the study of these questions. A systematic study of multilevel dynamical large deviations was initiated in Dawson and Gärtner [4]. Although this was motivated by the problem of two-level empirical measure processes coming from a system of interacting diffusions, the method was in fact only applied to obtain an integral representation of the rate function for a simple caricature of such a hierarchical system of diffusions. This caricature involved the study of empirical measures of diffusions in $\mathbb{R}^d$, and we were able to use the theory of distributions on $\mathbb{R}^d$. In the present paper we now turn to the study of the multilevel system of diffusions which requires the development of new tools for the study of multilevel empirical measures. In particular, we introduce an appropriate class of distributions on spaces of measures and distribution-valued functions which may also be of more general interest. We also develop an analogue of Weierstrass polynomial approximation for functions of measures and consider dual de Finetti approximations. Here the focus is on multilevel large deviations for noninteracting diffusions. However, the tools developed in this paper may also be used to include mean field interactions as in Dawson and Gärtner [2].

1.2. Multilevel systems of independent diffusions. Let $(\xi(\cdot), P_{x,s})$ be a diffusion in $\mathbb{R}^d$ on a fixed time interval $[0,T]$ with time dependent generator

$$L_t := \frac{1}{2}\sum_{\alpha,\beta=1}^{d} a^{\alpha\beta}(\cdot,t)\frac{\partial^2}{\partial x^\alpha \partial x^\beta} + \sum_{\alpha=1}^{d} b^\alpha(\cdot,t)\frac{\partial}{\partial x^\alpha}.$$

$P_{x,s}$ is the probability law on the path space $C([0,T];\mathbb{R}^d)$ of the diffusion with trajectories $\xi(t)$ starting at x at time s. The family of probabilities $\{P_{x,s};\ (x,s) \in \mathbb{R}^d \times [0,T]\}$ will be considered as the solution to the martingale problem for $\{L_t;\ t \in [0,T]\}$. We will assume throughout that the coefficients

$a\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $b\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ are continuous, the diffusion matrix $a(x,t)$ is strictly positive definite for all (x,t), and the martingale problem is well-posed. We will write P_x instead of $P_{x,0}$.

Let $\mathcal{M}(E)$ and $C([0,T];E)$ denote the space of probability measures on a Polish space E equipped with the topology of weak convergence and the space of continuous paths $[0,T] \to E$ furnished with the topology of uniform convergence, respectively. Given a natural number N, $\mathcal{M}^N(E)$ will denote the closed subspace of $\mathcal{M}(E)$ consisting of N-point empirical measures, i.e. of measures of the form

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \qquad x_1, \dots, x_N \in E, \tag{1.1}$$

where δ_x denotes the Dirac measure at point x. The integral of a function ϕ with respect to a measure μ and the application of a distribution μ to a test function ϕ will both be denoted by $\langle \mu, \phi \rangle$.

Given N independent copies $\xi_1(\cdot), \dots, \xi_N(\cdot)$ of our diffusion process, the *level I empirical process* $(\Xi^N(\cdot), \mathcal{P}^N_\mu)$ is defined by

$$\Xi^N(t) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_j(t)}.$$

This process lives in the space $\mathcal{M}_I := \mathcal{M}(\mathbb{R}^d)$. Given an initial measure $\mu \in \mathcal{M}_I^N := \mathcal{M}^N(\mathbb{R}^d)$ of the form (1.1), $\mathcal{P}^N_\mu$ is the law of $\Xi^N(\cdot)$ on $\mathcal{C}_I := C([0,T];\mathcal{M}_I)$ under $P_{x_1} \otimes \cdots \otimes P_{x_N}$.

Now, given natural numbers M and N, consider M independent copies $\Xi^N_1(\cdot), \dots, \Xi^N_M(\cdot)$ of our level I empirical process $\Xi^N(\cdot)$. Then the *level II empirical process* $(\Xi^{MN}(\cdot), \mathcal{P}^{MN}_\nu)$ is defined by

$$\Xi^{MN}(t) := \frac{1}{M}\sum_{i=1}^{M} \delta_{\Xi^N_i(t)}.$$

This process lives in $\mathcal{M}_{II} := \mathcal{M}(\mathcal{M}_I)$. For $\nu \in \mathcal{M}_{II}^{MN} := \mathcal{M}^M(\mathcal{M}_I^N)$ of the form

$$\nu = \frac{1}{M}\sum_{i=1}^{M} \delta_{\mu_i}, \qquad \mu_1, \dots, \mu_M \in \mathcal{M}_I^N,$$

$\mathcal{P}^{MN}_\nu$ is the law of $\Xi^{MN}(\cdot)$ on $\mathcal{C}_{II} := C([0,T];\mathcal{M}_{II})$ under $\mathcal{P}^N_{\mu_1} \otimes \cdots \otimes \mathcal{P}^N_{\mu_M}$. We will identify $\mathcal{M}_{II}^{MN}$ with the subspace of $\mathcal{M}^M(\mathcal{M}_I)$ consisting of measures which are concentrated on $\mathcal{M}_I^N$.

Let us next review the behavior of the level I process $\Xi^N(\cdot)$. Assume that the initial measures $\mu_N = \Xi^N(0) \in \mathcal{M}_I^N$ satisfy $\mu_N \to \mu_0$ in $\mathcal{M}_I$ as $N \to \infty$. Then, by the dynamical law of large numbers, $\Xi^N(\cdot)$ converges in distribution to a deterministic measure-valued process $\mu(\cdot) = \mu(\cdot\,;\mu_0)$ given by the weak solution of the Fokker-Planck equation

$$\dot{\mu}(t) = L_t^* \mu(t),$$
$$\mu(0) = \mu_0,$$

where L_t^* is the formal adjoint of L_t.

In Dawson and Gärtner [2] it was shown that the level I family of probability measures $\{\mathcal{P}_\mu^N;\, \mu \in \mathcal{M}_I^N\}$ satisfies the large deviation principle as $N \to \infty$ with scale N and rate function S. The latter has the integral representation

$$S(\mu(\cdot)) = \frac{1}{2} \int_0^T \|\dot{\mu}(t) - L_t^* \mu(t)\|_{\mu(t),t}^2 \, dt \tag{1.2}$$

if $\mu(\cdot) \in \mathcal{C}_I$ is absolutely continuous as a $\mathcal{D}'(\mathbb{R}^d)$-valued function and $S(\mu(\cdot)) = \infty$ otherwise, where

$$\|\vartheta\|_{\mu,t}^2 := \sup_{\phi \in \mathcal{D}(\mathbb{R}^d)} \frac{|\langle \vartheta, \phi \rangle|^2}{\langle \mu, |\nabla \phi|_t^2 \rangle}, \qquad \vartheta \in \mathcal{D}'(\mathbb{R}^d), \tag{1.3}$$

and

$$|\nabla \phi|_t^2 := \sum_{\alpha,\beta=1}^d a^{\alpha\beta}(\cdot, t) \frac{\partial \phi}{\partial x^\alpha} \frac{\partial \phi}{\partial x^\beta}, \qquad \phi \in \mathcal{D}(\mathbb{R}^d). \tag{1.4}$$

We denote by $\mathcal{D}(\mathbb{R}^d)$ the Schwartz space of C^∞ functions on $\mathbb{R}^d$ with compact support and by $\mathcal{D}'(\mathbb{R}^d)$ the corresponding space of distributions. Note that the supremum in the definition of $\|\cdot\|_{\mu,t}^2$ has to be restricted to such ϕ for which $\langle \mu, |\nabla \phi|_t^2 \rangle \neq 0$. For simplicity of notation, we will not indicate this explicitly here and in similar expressions later on.

Let us now switch to the level II picture. Assume that the initial measures $\nu_{MN} = \Xi^{MN}(0) \in \mathcal{M}_{II}^{MN}$ converge weakly to a measure ν in $\mathcal{M}_{II}$. Then the corresponding law of large numbers limit as both M and N go to infinity of the level II empirical process $\Xi^{MN}(\cdot)$ is given by the deterministic $\mathcal{M}_{II}$-valued process

$$\nu(t) = \int \delta_{\mu(t;\mu_0)} \, \nu_0(d\mu_0), \tag{1.5}$$

where $\mu(\cdot\,;\mu_0)$ is the level I law of large numbers limit. The measure-valued path $\nu(\cdot)$ is also a weak solution of a differential equation

$$\dot{\nu}(t) = \mathcal{L}_t^* \nu(t). \tag{1.6}$$

In order to give a precise meaning to this equation we have to introduce new spaces $\mathcal{D}(\mathcal{M}_I)$, $\mathcal{D}'(\mathcal{M}_I)$ of test functions and corresponding distributions on a space of probability measures. Then the operator $\mathcal{L}_t^*$ maps $\mathcal{M}_I$ into $\mathcal{D}'(\mathcal{M}_I)$

and is the formal adjoint of the operator $\mathcal{L}_t$ acting on functions $f \in \mathcal{D}(\mathcal{M}_I)$ given by

$$\mathcal{L}_t f(\mu) := \int L_t D f(\mu)(x)\, \mu(dx).$$

The derivative Df is defined by

$$\lim_{\gamma \downarrow 0} \gamma^{-1} \left[f((1-\gamma)\mu + \gamma\nu) - f(\mu) \right] = \langle \nu - \mu, Df(\mu) \rangle$$

for all $\mu, \nu \in \mathcal{M}_I$. Precise definitions of Df and corresponding higher order derivatives will be given in Section 3.1 where we will also introduce and systematically study the Schwartz spaces $\mathcal{D}(\mathcal{M}_I)$ and $\mathcal{D}'(\mathcal{M}_I)$. These spaces will be constructed in analogy with the 'classical' Schwartz spaces $\mathcal{D}(\mathbb{R}^d)$ and $\mathcal{D}'(\mathbb{R}^d)$, except that the notion of compact support has to be replaced by a different notion, namely, compact argument domain. A function f is said to have compact argument domain if $f(\mu)$ depends on the measure μ only via its mass distribution on a compact subset of $\mathbb{R}^d$.

1.3. Multilevel large deviations. Integral representation of the rate function. We now turn to the investigation of large deviations for the level II family $\{\mathcal{P}_\nu^{MN};\, \nu \in \mathcal{M}_{II}^{MN}\}$ of probability measures.

Theorem 1.1. *The family $\{\mathcal{P}_\nu^{MN};\, \nu \in \mathcal{M}_{II}^{MN}\}$ satisfies the large deviation principle as both M and N tend to infinity with scale MN and a certain rate function $S\colon \mathcal{C}_{II} \to [0,\infty]$. More precisely, given $\nu_{MN} \in \mathcal{M}_{II}^{MN}$ and $\nu \in \mathcal{M}_{II}$, suppose that $\nu_{MN} \to \nu$ in $\mathcal{M}_{II}$ as $M, N \to \infty$. Then*

(i) for each open subset G of $\mathcal{C}_{II}$,

$$\liminf_{M,N\to\infty} \frac{1}{MN} \log \mathcal{P}_{\nu_{MN}}^{MN}(G) \geq -\inf\{S(Q(\cdot))\colon Q(\cdot) \in G,\ Q(0) = \nu\};$$

(ii) for each closed subset F of $\mathcal{C}_{II}$,

$$\limsup_{M,N\to\infty} \frac{1}{MN} \log \mathcal{P}_{\nu_{MN}}^{MN}(F) \leq -\inf\{S(Q(\cdot))\colon Q(\cdot) \in F,\ Q(0) = \nu\};$$

(iii) for each compact subset $\mathcal{K}$ of $\mathcal{M}_{II}$, the level sets

$$\Psi(\mathcal{K}; s) := \{Q(\cdot) \in \mathcal{C}_{II}\colon S(Q(\cdot)) \leq s,\ Q(0) \in \mathcal{K}\}, \qquad s \geq 0,$$

are compact.

This is a particular case of the multilevel large deviation results in Dawson and Gärtner [4]. That paper also provides us with several descriptions of the rate function S which are either implicit or given in terms of complicated variational expressions and do not reflect properly the Markovian structure of the level II dynamics as does formula (1.2) for the level I empirical process.

The main objective of the present paper is to derive an analogous integral representation of the rate function S. To this end, for $Q \in \mathcal{M}_{II}$ and $t \in [0, T]$, we define

$$\|\theta\|^2_{Q,t} := \sup_{f \in \mathcal{D}(\mathcal{M}_I)} \frac{|\langle \theta, f \rangle|^2}{\left\langle Q, \langle \mu, |\nabla Df(\mu)|^2_t \rangle \right\rangle}, \qquad \theta \in \mathcal{D}'(\mathcal{M}_I), \tag{1.7}$$

where $\langle Q, \langle \mu, |\nabla Df(\mu)|^2_t \rangle\rangle$ is a shorthand for

$$\int Q(d\mu) \int \mu(dx) \, |\nabla Df(\mu)(x)|^2_t.$$

THEOREM 1.2. *The rate function S of the level II family $\{\mathcal{P}^{MN}_\nu;\, \nu \in \mathcal{M}^{MN}_{II}\}$ admits the representation*

$$S(Q(\cdot)) = \frac{1}{2} \int_0^T \left\| \dot{Q}(t) - \mathcal{L}^*_t Q(t) \right\|^2_{Q(t),t} dt \tag{1.8}$$

if $Q(\cdot) \in \mathcal{C}_{II}$ is absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued function. Otherwise $S(Q(\cdot)) = \infty$.

As we already mentioned, the Schwartz spaces $\mathcal{D}(\mathcal{M}_I)$ and $\mathcal{D}'(\mathcal{M}_I)$ will be introduced in Section 3.1. Absolute continuity of $\mathcal{D}'(\mathcal{M}_I)$-valued functions will be considered in Section 3.3.

The integral representation (1.8) expresses the rate function in terms of the drift operator and the Riemannian metric formally associated with the diffusion process $\Xi^{MN}(\cdot)$. To explain this, note that the generator $\mathcal{L}^N_t$ of the level I diffusion $\Xi^N(\cdot)$ has the form

$$\mathcal{L}^N_t f(\mu) = \frac{1}{2N} \langle \mu, \Sigma_t D^2 f(\mu) \rangle + \langle \mu, L_t Df(\mu) \rangle, \qquad f \in \mathcal{D}(\mathcal{M}_I),$$

where $D^2 f$ is the second order derivative of f and $\Sigma_t \colon \mathcal{D}((\mathbb{R}^d)^2) \to C_k(\mathbb{R}^d)$ is defined by

$$\Sigma_t \phi(x) := \sum_{\alpha,\beta=1}^d a^{\alpha\beta}(x,t) \frac{\partial^2 \phi}{\partial x^\alpha \partial y^\beta}(x,x).$$

The process $\Xi^{MN}(\cdot)$ then turns out to be a solution of the semimartingale equation

$$d\langle \Xi(t), f \rangle = \langle \Xi(t), \mathcal{L}^N_t f \rangle \, dt + \frac{1}{\sqrt{MN}} \, dM_t(f), \qquad f \in \mathcal{D}(\mathcal{M}_I),$$

where $M_t(f)$ is a bounded martingale with quadratic characteristic $\langle\langle M(f) \rangle\rangle_t$ given by

$$\begin{aligned} \frac{d}{dt} \langle\langle M(f) \rangle\rangle_t &= \int \Xi(t)(d\mu) \; \langle \mu, \Sigma_t (Df(\mu) \otimes Df(\mu)) \rangle \\ &= \left\langle \Xi(t), \langle \mu, |\nabla Df(\mu)|^2_t \rangle \right\rangle. \end{aligned} \tag{1.9}$$

This indicates that we are dealing with a random perturbation of the (infinite dimensional) dynamical system (1.6). The drift operator $(\mathcal{L}_t^N)^*$ is a singular perturbation of the vector field $\mathcal{L}_t^*$ governing the deterministic limiting dynamics (1.6), and the Riemannian norm associated with the quadratic diffusion form (1.9) coincides with $\|\cdot\|_{\Xi(t),t}$ which was defined in (1.7).

One of the obvious consequences of the Theorems 1.1 and 1.2 is the above mentioned law of large numbers for $\Xi^{MN}(\cdot)$ which we restate as a corollary.

Corollary 1.3. *Suppose that $\nu_{MN} \in \mathcal{M}_{II}^{MN}$ converges to a measure ν_0 in $\mathcal{M}_{II}$ as $M,N \to \infty$. Then*

$$\mathcal{P}_{\nu_{MN}}^{MN} \to \delta_{\nu(\cdot)}$$

weakly as $M,N \to \infty$, where $\nu(\cdot) \in \mathcal{C}_{II}$ is given by (1.5). The path $\nu(\cdot)$ is absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued function and the unique solution of equation (1.6) in $\mathcal{D}'(\mathcal{M}_I)$ with initial datum ν_0. As a function of ν_0, it maps $\mathcal{M}_{II}$ continuously into $\mathcal{C}_{II}$.

The only nontrivial aspect of the proof of the corollary is uniqueness which is established in Lemma 4.7.

1.4. Overview of the detailed development and proof. Let $\xi_{ij}(t)$, $i=1,\dots,M$, $j=1,\dots,N$, be MN independent copies of our diffusion process in $\mathbb{R}^d$ with generator L_t. We begin by fixing N and viewing $(\xi_{i1}(t),\dots,\xi_{iN}(t))$, $i=1,\dots,M$, as a system of M independent $(\mathbb{R}^d)^N$-valued diffusions. Then as $M\to\infty$, it follows from Dawson and Gärtner [2] that the $\mathcal{M}((\mathbb{R}^d)^N)$-valued processes $X^{MN}(\cdot)$ defined by

$$X^{MN}(t) := \frac{1}{M}\sum_{i=1}^{M} \delta_{(\xi_{i,1}(t),\dots,\xi_{i,N}(t))}$$

satisfy the large deviation principle with rate function

$$I^N(\mu(\cdot)) = \frac{1}{2}\int_0^T \left\|\dot\mu(t) - (L_t^N)^*\mu(t)\right\|_{\mu(t),t}^2 \, dt$$

if $\mu(\cdot) \in C([0,T];\mathcal{M}((\mathbb{R}^d)^N))$ is absolutely continuous and $I^N(\mu(\cdot)) = \infty$ otherwise. Here

$$\|\vartheta\|_{\mu,t}^2 := \sup_{\phi\in\mathcal{D}((\mathbb{R}^d)^N)} \frac{|\langle\vartheta,\phi\rangle|^2}{\langle\mu, |\nabla^N\phi|_t^2\rangle},$$

L_t^N denotes the generator of our $(\mathbb{R}^d)^N$-valued diffusion, and ∇^N stands for the Riemannian gradient with respect to the corresponding diffusion matrix. Thus, these quantities are defined in analogy with (1.3) and (1.4), but associated with the $(\mathbb{R}^d)^N$-valued diffusion.

In order to recast this in the level II setting, we introduce the mappings

$$\varepsilon^N : (\mathbb{R}^d)^N \to \mathcal{M}^N(\mathbb{R}^d)$$

defined by

$$\varepsilon^N(x_1,\dots,x_N) := \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$$

and the corresponding induced maps

$$\widehat{\varepsilon}^N\colon \mathcal{M}((\mathbb{R}^d)^N) \to \mathcal{M}(\mathcal{M}^N(\mathbb{R}^d)).$$

The next step is to obtain the rate function for large deviations of the level II empirical processes $\Xi^{MN}(\cdot) = \widehat{\varepsilon}^N(X^{MN}(\cdot))$ with N fixed and $M \to \infty$. An important step carried out in Proposition 2.1 is to establish that the restriction of $\widehat{\varepsilon}^N$ to $\mathcal{M}_s((\mathbb{R}^d)^N)$, the space of measures $\mu(dx_1,\dots,dx_N)$ on $(\mathbb{R}^d)^N$ which are invariant under permutations of the variables $x_1,\dots,x_N$, is a homeomorphism. In addition, in order to prescribe the rate function for $Q(\cdot) \in C([0,T];\mathcal{M}_{II})$, we must also introduce an appropriate class of distributions on $\mathcal{M}_I = \mathcal{M}(\mathbb{R}^d)$ and the appropriate notion of absolute continuity of distribution-valued functions, and this is carried out in Section 3. We prove in Lemma 4.4 that the corresponding rate function for $\widehat{\varepsilon}^N(X^{MN}(\cdot))$ is given by

$$S^N(Q(\cdot)) = \frac{N}{2}\int_0^T \left\| \dot{Q}(t) - (\mathcal{L}_t^N)^* Q(t)\right\|^2_{Q(t),t}\, dt$$

for $Q(\cdot) \in C([0,T];\mathcal{M}(\mathcal{M}_I^N))$ being absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued path.

The proof of Theorem 1.1 is then based on the result of Dawson and Gärtner [4], Theorem 2.9, which implies that $\Xi^{MN}(\cdot)$ satisfies the large deviation principle as M and N go to infinity with scale MN and rate function

$$\operatorname*{epi\,lim}_{N\to\infty} \frac{1}{N} S^N =: S^\infty.$$

By definition, this means that $\liminf_{N\to\infty} N^{-1}S^N(Q^N(\cdot)) \geq S^\infty(Q(\cdot))$ for each $Q(\cdot)$ and each sequence $Q^N(\cdot) \to Q(\cdot)$ and $\limsup_{N\to\infty} N^{-1}S^N(Q^N(\cdot)) \leq S^\infty(Q(\cdot))$ for each $Q(\cdot)$ and at least one sequence $Q^N(\cdot) \to Q(\cdot)$. The proof that this epilimit coincides with the integral representation of the rate function for $\Xi^{MN}(\cdot)$ given in Theorem 1.2 is carried out in the Lemmas 4.5 and 4.6.

2. Empirical measures of finite exchangeable random vectors.

Given a Polish space X, we will denote by $C_b(X)$ and $\mathcal{M}(X)$ the space of real-valued bounded continuous functions on X equipped with the uniform topology and the space of probability measures on X furnished with the topology of weak convergence, respectively. For $f \in C_b(X)$ and $\mu \in \mathcal{M}(X)$, we will write $\langle \mu, f\rangle$ for the integral of f with respect to μ.

Given a natural number N, let ε^N denote the map which transforms N-particle vectors into empirical measures:

$$\varepsilon^N(\underline{x}) := N^{-1}\sum_{j=1}^N \delta_{x_j}, \qquad \underline{x} \in (\mathbb{R}^d)^N. \tag{2.1}$$

Here and in the following $x_1, \dots, x_N$ denote the components of a vector $\underline{x} \in (\mathbb{R}^d)^N$, and δ_{x_j} is the Dirac measure at x_j. Let $\mathcal{M}^N(\mathbb{R}^d)$ denote the space of N-particle empirical measures, i.e. the closed subset of $\mathcal{M}(\mathbb{R}^d)$ consisting of the measures (2.1). We will consider ε^N as a map from $(\mathbb{R}^d)^N$ onto $\mathcal{M}^N(\mathbb{R}^d)$. Throughout this section, N will be held fixed, and we will write ε instead of ε^N.

Let $C_{b,s}((\mathbb{R}^d)^N)$ and $\mathcal{M}_s((\mathbb{R}^d)^N)$ denote the subspaces of $C_b((\mathbb{R}^d)^N)$ and $\mathcal{M}((\mathbb{R}^d)^N)$ consisting of symmetric functions and measures, respectively. A function $f(x_1, \dots, x_N)$ or a measure $\mu(dx_1, \dots, dx_N)$ on $(\mathbb{R}^d)^N$ are called symmetric if they are invariant under permutations of the variables $x_1, \dots, x_N$.

The map $\varepsilon\colon (\mathbb{R}^d)^N \to \mathcal{M}^N(\mathbb{R}^d)$ induces a map

$$\widehat{\varepsilon}\colon \mathcal{M}_s((\mathbb{R}^d)^N) \to \mathcal{M}(\mathcal{M}^N(\mathbb{R}^d))$$

which transforms each symmetric probability measure μ on $(\mathbb{R}^d)^N$ into its image $\mu \circ \varepsilon^{-1}$ with respect to ε. The main purpose of this section is to prove the following two propositions.

PROPOSITION 2.1. *The map* $\widehat{\varepsilon}\colon \mathcal{M}_s((\mathbb{R}^d)^N) \to \mathcal{M}(\mathcal{M}^N(\mathbb{R}^d))$ *is bijective and continuous in both directions.*

PROOF. The continuity of $\widehat{\varepsilon}$ is obvious from the continuity of ε. Choose $\mu, \nu \in \mathcal{M}_s((\mathbb{R}^d)^N)$ and assume that $\widehat{\varepsilon}(\mu) = \widehat{\varepsilon}(\nu)$, that is

$$\langle \mu, h \circ \varepsilon \rangle = \langle \nu, h \circ \varepsilon \rangle \qquad \text{for all } h \in C_b(\mathcal{M}^N(\mathbb{R}^d)).$$

To prove injectivity we have to show that this implies $\mu = \nu$. This will certainly be true if we show that the equation

$$h \circ \varepsilon = g$$

has at least one solution $h \in C_b(\mathcal{M}^N(\mathbb{R}^d))$ for each $g \in C_{b,s}((\mathbb{R}^d)^N)$. For functions h of the form $h(\mu) := \langle \mu^{\otimes N}, f \rangle$, the last equation turns into

$$\langle \varepsilon(\underline{x})^{\otimes N}, f \rangle = g(\underline{x}) \qquad \text{for all } \underline{x} \in (\mathbb{R}^d)^N. \tag{2.2}$$

We will see in Proposition 2.2 (which is the hard part of our proof) that this equation admits a (unique) solution $f \in C_{b,s}((\mathbb{R}^d)^N)$.

Let us next prove surjectivity. Given $g \in C_{b,s}((\mathbb{R}^d)^N)$, we define a continuous function $\Psi_g \in C_b(\mathcal{M}^N(\mathbb{R}^d))$ by

$$\Psi_g(\varepsilon(\underline{x})) = g(\underline{x}), \qquad \underline{x} \in (\mathbb{R}^d)^N.$$

By the Daniell-Stone Theorem (see e.g. Bauer [1]), for each $Q \in \mathcal{M}(\mathcal{M}^N(\mathbb{R}^d))$ the formula

$$\langle \mu, g \rangle := \int_{\mathcal{M}^N(\mathbb{R}^d)} Q(d\nu)\, \Psi_g(\nu), \qquad g \in C_{b,s}((\mathbb{R}^d)^N), \tag{2.3}$$

defines a measure $\mu \in \mathcal{M}_s((\mathbb{R}^d)^N)$. It turns out that μ is the preimage of Q with respect to $\widehat{\varepsilon}$. Indeed, for each $f \in C_b(\mathcal{M}^N(\mathbb{R}^d))$ the function $f \circ \varepsilon$ belongs to $C_{b,s}((\mathbb{R}^d)^N)$ and $\Psi_{f\circ\varepsilon} = f$. Therefore

$$\langle \widehat{\varepsilon}(\mu), f\rangle = \langle \mu, f \circ \varepsilon\rangle = \int Q(d\nu)\, \Psi_{f\circ\varepsilon}(\nu) = \langle Q, f\rangle$$

for all $f \in C_b(\mathcal{M}^N(\mathbb{R}^d))$.

Finally, the continuity of the inverse $\widehat{\varepsilon}^{-1}$ follows from (2.3). $\square$

We next introduce a linear operator $T = T^N$ on $C_{b,s}((\mathbb{R}^d)^N)$ by

$$Tf(\underline{x}) := \langle \varepsilon(\underline{x})^{\otimes N}, f\rangle, \qquad \underline{x} \in (\mathbb{R}^d)^N.$$

Note that the image of a function with compact support does not necessarily have compact support. Therefore, instead of considering spaces of symmetric continuous or C^∞ functions with compact support, we need to introduce slightly modified function spaces. For each compact K in $\mathbb{R}^d$, we will denote by $\mathcal{G}^0_{N,K}$ the linear space of symmetric continuous functions $g\colon (\mathbb{R}^d)^N \to \mathbb{R}$ such that $g(x_1, \ldots, x_N)$ does not change if x_1 varies outside of K. Let $\mathcal{G}^0_N$ be the union of these spaces. We equip $\mathcal{G}^0_{N,K}$ and $\mathcal{G}^0_N$ with the supremum norm. By $\mathcal{G}^\infty_{N,K}$ and $\mathcal{G}^\infty_N$ we will denote the subspace of $\mathcal{G}^0_{N,K}$ consisting of C^∞ functions and the union of the spaces $\mathcal{G}^\infty_{N,K}$ over all compacts K, respectively. We endow the spaces $\mathcal{G}^\infty_{N,K}$ with the metric

$$\rho_N(f,g) := \sum_{n=0}^{\infty} 2^{-n} \left(\|g-h\|_{N,n} \wedge 1\right),$$

where

$$\|g\|_{N,n} := \sup\left\{|\partial^\alpha g(\underline{x})|\colon \underline{x} \in (\mathbb{R}^d)^N,\ |\alpha| \le n\right\}$$

and ∂^α denote partial derivatives of order $|\alpha|$. This makes $\mathcal{G}^\infty_{N,K}$ into a separable Fréchet space. Finally, we furnish $\mathcal{G}^\infty_N$ with the strongest locally convex vector space topology which induces the original Fréchet topology on $\mathcal{G}^\infty_{N,K}$ for each compact K. For separability and representation by functions with compact support we refer to Lemma 3.7 below. We may now consider T also as a linear operator acting on $\mathcal{G}^0_{N,K}$, $\mathcal{G}^0_N$, $\mathcal{G}^\infty_{N,K}$, or $\mathcal{G}^\infty_N$.

Proposition 2.2. *The linear operator T is a homeomorphism on each of the spaces $C_{b,s}((\mathbb{R})^d)^N$, $\mathcal{G}^0_{N,K}$, $\mathcal{G}^0_N$, $\mathcal{G}^\infty_{N,K}$, and $\mathcal{G}^\infty_N$ for all compact sets K.*

Set $E := \{1, \ldots, N\}$ and fix r with $1 \le r \le N$ arbitrarily. Given $\underline{x} \in (\mathbb{R}^d)^N$ and $\underline{j} \in E^r$, we will denote by $\underline{x}_{\underline{j}}$ the vector $(x_{j_1}, \ldots, x_{j_r})$. We will prove Proposition 2.2 for the operator $S := N^N T$ instead of T. This operator has the form

$$Sf(\underline{x}) := \sum_{\underline{j}\in E^N} f(\underline{x}_{\underline{j}}), \qquad \underline{x} \in (\mathbb{R}^d)^N.$$

We will only check our assertion for the space $C_{b,s}((\mathbb{R}^d)^N)$. The necessary changes for handling the other function spaces will be evident from this. Let us therefore consider S as an operator on $C_{b,s}((\mathbb{R}^d)^N)$. The continuity of S is obvious. After showing that S is bijective, the continuity of S^{-1} will follow from the Open Mapping Theorem (or from the representation of S^{-1} given in Lemma 2.4 b) below). Let us first prove injectivity.

LEMMA 2.3. *Let $f\colon (\mathbb{R}^d)^N \to \mathbb{C}$ be a symmetric function which satisfies*

$$\sum_{\underline{j}\in E^N} f(\underline{x}_{\underline{j}}) = 0 \qquad \text{for all } \underline{x} \in (\mathbb{R}^d)^N. \tag{2.4}$$

Then f vanishes identically.

PROOF. 1^0 We first show that our lemma may be reduced to the following statement:

If $g\colon E^N \to \mathbb{C}$ is a symmetric function which satisfies

$$\sum_{\underline{j}\in E^N} g(\underline{k}_{\underline{j}}) = 0 \qquad \text{for all } \underline{k} \in E^N, \tag{2.5}$$

then g vanishes identically.

Indeed, suppose that $f\colon (\mathbb{R}^d)^N \to \mathbb{C}$ is symmetric and satisfies (2.4). Fix $\underline{z} \in (\mathbb{R}^d)^N$ arbitrarily. Given $\underline{k} \in E^N$, equation (2.4) holds in particular for $\underline{x} = \underline{z}_{\underline{k}}$. In other words, the symmetric function

$$g(\underline{k}) := f(\underline{z}_{\underline{k}}), \qquad \underline{k} \in E^N,$$

satisfies (2.5). Hence, g vanishes identically and therefore $f(\underline{z}) = 0$.

2^0 Before proving the above statement, we introduce some notation. Let $\mathcal{F}(E^N)$ denote the space of functions $E^N \to \mathbb{C}$. $\mathcal{F}(E^N)$ is a (finite dimensional) Hilbert space with inner product

$$(u, v) := \sum_{\underline{k}\in E^N} u(\underline{k})\overline{v(\underline{k})}.$$

We introduce functions $e_\ell\colon E \to \mathbb{C}$, $\ell \in E$, by

$$e_\ell(k) := N^{-1/2} \exp\left\{\frac{2\pi i}{N} k\ell\right\}, \qquad k \in E.$$

Given $\underline{\ell} = (\ell_1, \dots, \ell_N) \in E^N$, we set

$$e_{\underline{\ell}}(\underline{k}) := e_{\ell_1}(k_1) \dots e_{\ell_N}(k_N), \qquad \underline{k} = (k_1, \dots, k_N) \in E^N.$$

The functions $e_{\underline{\ell}}$, $\underline{\ell} \in E^N$, form an orthonormal basis in $\mathcal{F}(E^N)$.

3^0 Now let $g\colon E^N \to \mathbb{C}$ be a symmetric function satisfying (2.5). We represent g as the Fourier transform of a symmetric function $h\colon E^N \to \mathbb{C}$:

$$g = \sum_{\underline{\ell} \in E^N} h(\underline{\ell}) e_{\underline{\ell}}.$$

Substituting this in (2.5), we obtain

$$\sum_{\underline{\ell} \in E^N} h(\underline{\ell}) \phi_{\underline{\ell}} = 0 \tag{2.6}$$

with

$$\phi_{\underline{\ell}} := \sum_{\underline{j} \in E^N} \phi_{\underline{\ell}}^{\underline{j}} \tag{2.7}$$

and

$$\phi_{\underline{\ell}}^{\underline{j}}(\underline{k}) := e_{\underline{\ell}}(\underline{k}_{\underline{j}}) = e_{\ell_1}(k_{j_1}) \dots e_{\ell_N}(k_{j_N}).$$

4^0 It remains to deduce from (2.6) that the function h vanishes identically. To this end we prove the following facts:

(i) If the N-tuple $\underline{\ell}$ has more components equal to N than the N-tuple $\underline{m}$, then

$$(\phi_{\underline{\ell}}, e_{\underline{m}}) = 0.$$

(ii) Suppose that the N-tuples $\underline{\ell}$ and $\underline{m}$ have the same number of components equal to N. Then

$$(\phi_{\underline{\ell}}, e_{\underline{m}}) = \begin{cases} c_{\underline{m}}, & \text{if } \underline{\ell} \text{ is a permutation of } \underline{m}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.8}$$

where $c_{\underline{m}}$ is a positive constant depending on $\underline{m}$ only.

Note that $(\phi_{\underline{\ell}}, e_{\underline{m}})$ is invariant under (separate) permutations of the N-tuples $\underline{\ell}$ and $\underline{m}$. Therefore, in order to prove (i) and (ii), it will be enough to consider functions $\phi_{\underline{\ell}}$ and $e_{\underline{m}}$ with

$$\underline{\ell} = (\ell_1, \dots, \ell_r, N, \dots, N) \quad \text{and} \quad \underline{m} = (m_1, \dots, m_s, N, \dots, N),$$

where $0 \le r \le s \le N$ and $\ell_1, \dots, \ell_r$ and $m_1, \dots, m_s$ are not equal to N. We remark that $e_N \equiv N^{-1/2}$.

To prove (i), we note that

$$\phi_{\underline{\ell}}^{\underline{j}}(\underline{k}) = N^{-(N-r)/2} e_{\ell_1}(k_{j_1}) \dots e_{\ell_r}(k_{j_r}) \tag{2.9}$$

and

$$e_{\underline{m}}(\underline{k}) = N^{-(N-s)/2} e_{m_1}(k_1) \dots e_{m_s}(k_s). \tag{2.10}$$

By assumption, $r < s$. Hence, for each $\underline{j} \in E^N$ we find $\alpha \in \{1, \dots, s\}$ such that the function (2.9) does not depend on the coordinate k_α. But, since $m_\alpha \neq N$, we conclude from this that $\phi^{\underline{j}}_{\underline{\ell}}$ is orthogonal to $e_{\underline{m}}$. Because of (2.7), this shows that $\phi_{\underline{\ell}}$ is orthogonal to $e_{\underline{m}}$.

To prove (ii), suppose that $r = s$. Then, by similar reasoning, we conclude from (2.9) and (2.10) that $\phi^{\underline{j}}_{\underline{\ell}}$ is orthogonal to $e_{\underline{m}}$ if $(j_1, \dots, j_r)$ is not a permutation of $(1, \dots, r)$. Using this, we conclude from (2.7) that

$$
\begin{aligned}
(\phi_{\underline{\ell}}, e_{\underline{m}}) &= \sum_\pi \sum_{j_{r+1}=1}^{N} \cdots \sum_{j_N=1}^{N} \left(\phi^{\pi(1),\dots,\pi(r),j_{r+1},\dots,j_N}_{\ell_1,\dots,\ell_r,N,\dots,N}, e_{m_1,\dots,m_r,N,\dots,N} \right) \\
&= N^{N-r} \sum_\pi \left(e_{\ell_1,\dots,\ell_r,N,\dots,N}, e_{m_{\pi(1)},\dots,m_{\pi(r)},N,\dots,N} \right),
\end{aligned}
$$

where π runs over all permutations of $(1, \dots, r)$. This yields (2.8) with $c_{\underline{m}}$ equal to N^{N-r} times the number of permutations π for which $(m_{\pi(1)}, \dots, m_{\pi(r)})$ coincides with $(m_1, \dots, m_r)$.

5^0 Now let $\underline{m} \in E^N$ be an arbitrary N-tuple with no component equal to N. Then, evaluating the inner product with $e_{\underline{m}}$ on both sides of equation (2.6) and taking into account the assertions (i) and (ii) from step 4^0 as well as the symmetry of h, we see that $h(\underline{m}) = 0$. Repeating this argument, we successively find that $h(\underline{m}) = 0$ for all N-tuples $\underline{m}$ with one component equal to N, two components equal to N, and so on. Thus h vanishes identically. $\square$

Given r with $1 \leq r \leq N$, let $E^r_{\neq}$ denote the set of all tuples $(j_1, \dots, j_r) \in E^r$ with $j_\alpha \neq j_\beta$ for $\alpha \neq \beta$.

LEMMA 2.4. *a) For each r, $1 \leq r \leq N$, and each function $g_r \in C_{b,s}((\mathbb{R}^d)^r)$ there exists a function $f_r \in C_{b,s}((\mathbb{R}^d)^r)$ such that*

$$
\sum_{\underline{j} \in E^r} f_r(\underline{x}_{\underline{j}}) = \sum_{\underline{j} \in E^r_{\neq}} g_r(\underline{x}_{\underline{j}}) \tag{2.11}
$$

for all $\underline{x} \in (\mathbb{R}^d)^N$. In particular, the operator S on $C_{b,s}((\mathbb{R}^d)^N)$ is surjective.

b) The inverse of the operator S has the form

$$
S^{-1} g(x_1, \dots, x_N) = \sum_\pi c_\pi g(x_{\pi(1)}, \dots, x_{\pi(N)}) \tag{2.12}
$$

with certain coefficients $c_\pi \in \mathbb{R}$, where π runs over all maps of E into itself.

PROOF. a) We prove the solvability of (2.11) by induction with respect to r. Clearly $f_1 = g_1$ is a solution for $r = 1$. To accomplish the induction step from $r - 1$ to r, let us fix r with $1 < r \leq N$ and $g_r \in C_{b,s}((\mathbb{R}^d)^r)$ arbitrarily. We introduce the functions

$$
q_s(x_1, \dots, x_s) := \sum_{\{j_1,\dots,j_r\} = \{1,\dots,s\}} g_r(x_{j_1}, \dots, x_{j_r}), \tag{2.13}
$$

$1 \le s < r$. The sum on the right is taken over all r-tuples $(j_1, \dots, j_r) \in E^r$ such that the set $\{j_1, \dots, j_r\}$ coincides with $\{1, \dots, s\}$. Clearly q_s belongs to $C_{b,s}((\mathbb{R}^d)^s)$ for each s. According to our induction assumption, we find functions $p_s \in C_{b,s}((\mathbb{R}^d)^s)$ such that

$$\sum_{\underline{j} \in E^s} p_s(\underline{x}_{\underline{j}}) = \sum_{\underline{j} \in E^s_{\neq}} q_s(\underline{x}_{\underline{j}}), \qquad \underline{x} \in (\mathbb{R}^d)^N, \tag{2.14}$$

for $1 \le s < r$. Let

$$f_r(\underline{x}) := g_r(\underline{x}) - \sum_{s=1}^{r-1} \frac{1}{N^{r-s}\binom{r}{s}(s!)^2} \sum_{\underline{k} \in F^s_{\neq}} p_s(\underline{x}_{\underline{k}}), \qquad \underline{x} \in (\mathbb{R}^d)^r, \tag{2.15}$$

where $F := \{1, \dots, r\}$ and $F^s_{\neq}$ consists of all s-tuples $(k_1, \dots, k_s) \in F^s$ with $k_\alpha \neq k_\beta$ for $\alpha \neq \beta$. We claim that the function (2.15) solves (2.11). Indeed, for this function we obtain

$$\begin{aligned}
\sum_{\underline{j} \in E^r} f_r(\underline{x}_{\underline{j}}) &= \sum_{s=1}^{r} \frac{1}{s!} \sum_{\underline{k} \in E^s_{\neq}} \sum_{\{j_1, \dots, j_r\} = \{k_1, \dots, k_s\}} g_r(\underline{x}_{\underline{j}}) \\
&\quad - \sum_{s=1}^{r-1} \frac{1}{N^{r-s}\binom{r}{s}(s!)^2} \sum_{\underline{k} \in F^s_{\neq}} \sum_{\underline{j} \in E^r} p_s((\underline{x}_{\underline{j}})_{\underline{k}}) \\
&= \sum_{\underline{j} \in E^r_{\neq}} g_r(\underline{x}_{\underline{j}}) + \sum_{s=1}^{r-1} \frac{1}{s!} \left[\sum_{\underline{k} \in E^s_{\neq}} q_s(\underline{x}_{\underline{k}}) - \sum_{\underline{j} \in E^s} p_s(\underline{x}_{\underline{j}}) \right] \\
&= \sum_{\underline{j} \in E^r_{\neq}} g_r(\underline{x}_{\underline{j}})
\end{aligned}$$

for all $\underline{x} \in (\mathbb{R}^d)^N$. Here we have used (2.13), the fact that

$$\sum_{\underline{j} \in E^r} p_s((\underline{x}_{\underline{j}})_{\underline{k}})$$

does not depend on $\underline{k} \in F^s_{\neq}$, and (2.14).

Take $r = N$ in (2.11) to see that the operator S is surjective.

b) Formula (2.12) follows from the inductive construction of the solutions to (2.11) given in a). $\square$

3. Preliminaries on distributions and distribution-valued functions.

3.1. Spaces of distributions on $\mathcal{M}$. Given a compact set $K \subset \mathbb{R}^d$, let $\mathcal{D}_K(\mathbb{R}^d)$ denote the space of real-valued C^∞-functions ϕ on $\mathbb{R}^d$ with $\operatorname{supp} \phi \subseteq K$

endowed with the usual Fréchet topology. Denote by $\mathcal{D}(\mathbb{R}^d)$ the union of the spaces $\mathcal{D}_K(\mathbb{R}^d)$ furnished with the corresponding inductive topology. The topologically dual space $\mathcal{D}'(\mathbb{R}^d)$ is the space of Schwartz distributions on $\mathbb{R}^d$. Abbreviate $\mathcal{M} := \mathcal{M}(\mathbb{R}^d)$. The objective of this section is to introduce a Schwartz space $\mathcal{D}(\mathcal{M})$ of test functions on $\mathcal{M}$ and a corresponding space $\mathcal{D}'(\mathcal{M})$ of distributions and to derive some properties of these spaces.

Given a measure $\mu \in \mathcal{M}$ and a Borel set $B \subseteq \mathbb{R}^d$, we will denote by $\mu|_B$ the restriction of μ to B defined by $\mu|_B(A) := \mu(A \cap B)$. In general, functions of the form $f(\mu) = g(\langle \mu, \phi \rangle)$, $\mu \in \mathcal{M}$, do not have compact support even if g and ϕ have compact support. This forces us to replace the notion of compact support by the following notion which will turn out to be more adequate for our purposes.

DEFINITION 3.1. A function $f\colon \mathcal{M} \to \mathbb{R}$ is said to have *compact argument domain* if there exists a compact set $K \subset \mathbb{R}^d$ such that $\mu|_K = \nu|_K$ implies $f(\mu) = f(\nu)$.

Let $\overline{\mathbb{R}^d} := \mathbb{R}^d \cup \{\infty\}$ denote Alexandroff's one point compactification of $\mathbb{R}^d$, and let $\bar{\mathcal{M}} := \mathcal{M}(\overline{\mathbb{R}^d})$ be the space of probability measures on $\overline{\mathbb{R}^d}$ endowed with the topology of weak convergence. Given $\bar\mu \in \bar{\mathcal{M}}$ and $z \in \mathbb{R}^d$, define $\mu_z \in \mathcal{M}$ by $\mu_z(A) := \bar\mu(A) + \bar\mu(\{\infty\})\,\delta_z(A)$, i.e. by shifting the mass at ∞ to point z. Suppose now that $f\colon \mathcal{M} \to \mathbb{R}$ has compact argument domain. Then, for sufficiently large $|z|$, the function $\bar f(\bar\mu) := f(\mu_z)$, $\bar\mu \in \bar{\mathcal{M}}$, does not depend on z and is a natural *extension* of f onto $\bar{\mathcal{M}}$. Moreover, f is continuous if and only if $\bar f$ is continuous. In the following we will denote the extension $\bar f$ again by f.

LEMMA 3.2. *Assume that f belongs to $C_b(\mathcal{M})$ and has compact argument domain. Then there exists a smallest compact set $K \subset \mathbb{R}^d$ with the properties stated in Definition 2.1.*

This smallest compact set will be called *argument domain* of f and will be denoted by $\operatorname{arg\,dom} f$. We remark that one finds functions $f \in C_b(\mathcal{M})$ for which there is no smallest closed set $F \subseteq \mathbb{R}^d$ with the property that $\mu|_F = \nu|_F$ implies $f(\mu) = f(\nu)$.

PROOF OF LEMMA 3.2. Fix $f \in C_b(\mathcal{M})$ arbitrarily, and let $\mathcal{K}$ denote the system of compact sets $K \subset \mathbb{R}^d$ such that $f(\mu) = f(\nu)$ for $\mu|_K = \nu|_K$. We have to show that the intersection of all sets $K \in \mathcal{K}$ also belongs to $\mathcal{K}$.

1^0 We first check that $\mathcal{K}$ is closed under finite intersections. Choose $K_1, K_2 \in \mathcal{K}$ and $\mu, \nu \in \mathcal{M}$ with $\mu|_{K_1\cap K_2} = \nu|_{K_1 \cap K_2}$ arbitrarily. We want to show that this implies $f(\mu) = f(\nu)$. To this end, we define probability measures $\mu_1, \mu_2 \in \bar{\mathcal{M}}$ as follows:

$$
\begin{aligned}
\mu_1(A) &:= \mu(A \cap K_1) + \mu(K_1^c)\,\delta_\infty(A),\\
\mu_2(A) &:= \mu_1(A \cap K_2) + \mu_1(K_2^c)\,\delta_\infty(A)\\
&= \mu(A \cap K_1 \cap K_2) + \mu((K_1 \cap K_2)^c)\,\delta_\infty(A).
\end{aligned}
$$

In exactly the same way we construct probability measures $\nu_1, \nu_2 \in \bar{\mathcal{M}}$ from ν by first 'sweeping' the mass outside of K_1 to ∞ and then also 'sweeping' the mass outside of K_2 to ∞. Since $\mu|_{K_1} = \mu_1|_{K_1}$ and $\mu_1|_{K_2} = \mu_2|_{K_2}$, we have $f(\mu) = f(\mu_1) = f(\mu_2)$. Correspondingly, $f(\nu) = f(\nu_1) = f(\nu_2)$. But $\mu_2 = \nu_2$, and therefore $f(\mu) = f(\nu)$.

2^0 We show that the compact set

$$\underline{K} := \bigcap_{K \in \mathcal{K}} K$$

belongs to $\mathcal{K}$. Let us first assume that $\underline{K} = \emptyset$. Then we find a finite number of compact sets $K_1, \ldots, K_r \in \mathcal{K}$ such that $K_1 \cap \cdots \cap K_r = \emptyset$. By step 1^0, this implies that the empty set belongs to $\mathcal{K}$. Now suppose that $\underline{K} \neq \emptyset$. Let G be an arbitrary open neighborhood of $\underline{K}$. Then there exist finitely many sets $K_1, \ldots, K_r \in \mathcal{K}$ such that $K_1 \cap \cdots \cap K_r \subset G$. According to step 1^0, $K_1 \cap \cdots \cap K_r$ belongs to $\mathcal{K}$. Hence, $\mu|_G = \nu|_G$ implies $f(\mu) = f(\nu)$. Now let μ and ν be such that $\mu|_{\underline{K}} = \nu|_{\underline{K}}$. We want to show that this yields $f(\mu) = f(\nu)$. To this end, let (G_n) be a sequence of bounded open sets such that $G_n \downarrow \underline{K}$. We construct probabilities $\mu_n, \nu_n \in \bar{\mathcal{M}}$ by 'sweeping', respectively, the masses of μ and ν in $G_n \backslash \underline{K}$ to ∞:

$$\begin{aligned}\mu_n(A) &:= \mu(A \cap (\underline{K} \cup G_n^c)) + \mu(G_n \backslash \underline{K})\, \delta_\infty(A),\\ \nu_n(A) &:= \nu(A \cap (\underline{K} \cup G_n^c)) + \nu(G_n \backslash \underline{K})\, \delta_\infty(A).\end{aligned}$$

Since $\mu_n|_{G_n} = \nu_n|_{G_n}$, we get $f(\mu_n) = f(\nu_n)$. Our assertion now follows from the observation that $\mu_n \to \mu$ and $\nu_n \to \nu$ in $\bar{\mathcal{M}}$ and the continuity of f on $\bar{\mathcal{M}}$. Therefore $\underline{K}$ indeed belongs to $\mathcal{K}$. $\square$

Given a natural number m, we denote by $C_{b,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ the space of bounded continuous functions $f\colon \mathcal{M} \times (\mathbb{R}^d)^m \to \mathbb{R}$ with the property that $f(\mu; x_1, \ldots, x_m)$ is symmetric in the variables $x_1, \ldots, x_m \in \mathbb{R}^d$ for every $\mu \in \mathcal{M}$. For each $f \in C_{b,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ and each $\mu \in \mathcal{M}$, we will denote by $f(\mu)$ the function in $C_{b,s}((\mathbb{R}^d)^m)$ given by $f(\mu)(\underline{x}) := f(\mu; \underline{x})$, $\underline{x} \in (\mathbb{R}^d)^m$.

Definition 3.3. A function $f \in C_{b,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ is said to have *compact argument domain* if $f(\cdot\,; \underline{x})$ has compact argument domain for each $\underline{x} \in (\mathbb{R}^d)^m$ and there exists a compact set $K \subset \mathbb{R}^d$ such that $\arg\operatorname{dom} f(\cdot\,; \underline{x}) \subseteq K$ for each $\underline{x} \in (\mathbb{R}^d)^m$ and $\operatorname{supp} f(\mu; \cdot) \subseteq K^m$ for each $\mu \in \mathcal{M}$. The smallest compact set K with this property will be called *argument domain* of f and denoted by $\arg\operatorname{dom} f$.

Note that each function $f \in C_{b,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ with compact argument domain allows a natural extension $\bar{f} \in C_{b,s}(\bar{\mathcal{M}} \times (\overline{\mathbb{R}^d})^m)$ which again will be denoted by f.

Let $C_k(\mathcal{M})$ and $C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ denote, respectively, the vector spaces of functions in $C_b(\mathcal{M})$ and $C_{b,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ with compact argument domain. We furnish $C_k(\mathcal{M})$ and $C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ with the supremum norm. These spaces may be considered as the analogues of the spaces $C_k(\mathbb{R}^d)$ and $C_{k,s}((\mathbb{R}^d)^m)$ of continuous functions $\mathbb{R}^d \to \mathbb{R}$ and symmetric continuous functions $(\mathbb{R}^d)^m \to \mathbb{R}$ with compact support, except that these functions have compact argument domain rather than compact support.

The system of sets

$$\{(\mu,\nu) \in \mathcal{M} \times \mathcal{M} \colon |\langle \mu, \phi_k\rangle - \langle \nu, \phi_k\rangle| < 1 \text{ for } k = 1,\dots,r\},$$

$r \in \mathbb{N}$, $\phi_1,\dots,\phi_r \in C_b(\mathbb{R}^d)$, forms the base of a uniform structure on $\mathcal{M}$ which is compatible with the topology of weak convergence. We will consider $\mathcal{M}$ to be equipped with this uniform structure. We furnish $\mathcal{M} \times (\mathbb{R}^d)^m$ with the product of the uniform structures on $\mathcal{M}$ and $\mathbb{R}^d$. In the same way, using functions $\phi_1,\dots,\phi_r \in C_b(\overline{\mathbb{R}^d})$, one may define uniform structures on $\bar{\mathcal{M}}$ and $\bar{\mathcal{M}} \times (\overline{\mathbb{R}^d})^m$.

LEMMA 3.4. *All functions in $C_k(\mathcal{M})$ and $C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$, $m \in \mathbb{N}$, are uniformly continuous.*

PROOF. If f belongs to $C_k(\mathcal{M})$, then its extension $\bar{f} \colon \bar{\mathcal{M}} \to \mathbb{R}$ is continuous and, since $\bar{\mathcal{M}}$ is compact, even uniformly continuous on $\bar{\mathcal{M}}$. Consequently, f is uniformly continuous on $\mathcal{M}$. The same argument applies to functions in $C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$. $\square$

DEFINITION 3.5. A function $f \in C_k(\mathcal{M})$ will be called *differentiable* if there exists a function $Df \in C_{k,s}(\mathcal{M} \times \mathbb{R}^d)$ such that

(3.1)
$$\lim_{\gamma \downarrow 0} \gamma^{-1} [f((1-\gamma)\mu + \gamma\nu) - f(\mu)] = \langle \nu - \mu, Df(\mu)\rangle \qquad \text{for all } \mu,\nu \in \mathcal{M}.$$

The function Df will be called (first order) *derivative* of f. Higher order derivatives $D^m f \in C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$, $m = 2,3,\dots$, are defined recursively by

(3.2)
$$\begin{aligned}\lim_{\gamma \downarrow 0} \gamma^{-1} \left[D^{m-1} f((1-\gamma)\mu + \gamma\nu)(x_1,\dots,x_{m-1}) - D^{m-1} f(\mu)(x_1,\dots,x_{m-1})\right]& \\ = \langle \nu - \mu, D^m f(\mu)(x_1,\dots,x_{m-1},\cdot\,)\rangle&\end{aligned}$$

for all $\mu,\nu \in \mathcal{M}$ and $x_1,\dots,x_{m-1} \in \mathbb{R}^d$.

Putting $\nu = \delta_x$, $x \in \mathbb{R}^d$, and remembering that $Df(\mu)$ has compact support, we see that (3.1) defines Df uniquely. Moreover,

$$\frac{d}{d\gamma} f((1-\gamma)\mu + \gamma\nu) = \langle \nu - \mu, Df((1-\gamma)\mu + \gamma\nu)\rangle \qquad \text{for } 0 \le \gamma \le 1.$$

Hence

$$f(\nu) - f(\mu) = \int_0^1 \langle \nu - \mu, Df((1-\theta)\mu + \theta\nu)\rangle \, d\theta \tag{3.3}$$

for all $\mu, \nu \in \mathcal{M}$. Replacing ν by $(1-\gamma)\mu + \gamma\nu$, we find that

$$\gamma^{-1}\left[f((1-\gamma)\mu + \gamma\nu) - f(\mu)\right] = \int_0^1 \langle \nu - \mu, Df((1-\theta\gamma)\mu + \theta\gamma\nu)\rangle \, d\theta.$$

Since Df is uniformly continuous (Lemma 3.4), we conclude from this that the convergence in (3.1) is uniform in $\mu, \nu \in \mathcal{M}$. If the derivative $D^m f$ exists for some $m \in \{2, 3, \dots\}$, then it is unique and the convergence in (3.2) also turns out to be uniform in $\mu, \nu \in \mathcal{M}$ and $x_1, \dots, x_{m-1} \in \mathbb{R}^d$.

LEMMA 3.6. *Given $f \in C_k(\mathcal{M})$, suppose that the derivatives $D^m f \in C_{k,s}(\mathcal{M} \times (\mathbb{R}^d)^m)$ exist for $m = 1, 2, \dots, M$. Then*

$$\operatorname{arg\,dom} f \supseteq \operatorname{arg\,dom} Df \supseteq \cdots \supseteq \operatorname{arg\,dom} D^M f.$$

PROOF. Set $K := \operatorname{arg\,dom} D^{m-1} f$ and let us show that $\operatorname{arg\,dom} D^m f \subseteq K$. If $\mu_1|_K = \mu_2|_K$, then we conclude from (3.2) that

$$\langle \delta_{x_m} - \mu_1, D^m f(\mu_1)(x_1, \dots, x_{m-1}, \cdot\,)\rangle = \langle \delta_{x_m} - \mu_2, D^m f(\mu_2)(x_1, \dots, x_{m-1}, \cdot\,)\rangle$$

for all $x_1, \dots, x_m \in \mathbb{R}^d$. Since both $D^m f(\mu_1)$ and $D^m f(\mu_2)$ have compact support, this implies $D^m f(\mu_1) = D^m f(\mu_2)$. If $x_i \notin K$ for some $i \in \{1, \dots, m-1\}$, then (3.2) yields

$$\langle \delta_{x_m} - \mu, D^m f(\mu)(x_1, \dots, x_{m-1}, \cdot\,)\rangle = 0 \tag{3.4}$$

for all $x_m \in \mathbb{R}^d$, and we obtain $D^m f(\mu)(x_1, \dots, x_m) = 0$. For arbitrary $x_1, \dots, x_{m-1}$ we also find that the left hand side of (3.4) does not depend on x_m outside of K. Hence, $D^m f(\mu)(x_1, \dots, x_m) = 0$ for $x_m \notin K$, and we are done. $\square$

Let us now turn to the introduction of the Schwartz space $\mathcal{D}(\mathcal{M})$ of test functions and the corresponding space $\mathcal{D}'(\mathcal{M})$ of distributions.

A variable $\underline{x} \in (\mathbb{R}^d)^m$ is a vector $\underline{x} = (x_1, \dots, x_m)$ with $x_i = (x_i^1, \dots, x_i^d) \in \mathbb{R}^d$ for $i = 1, \dots, m$. Therefore, $\underline{x}$ may be considered as a $m \times d$ matrix $\{x_i^j\}$ of real variables. A $m \times d$ multi-index $\alpha = \{\alpha_i^j\}$ is a $m \times d$ matrix of nonnegative integers α_i^j $(i = 1, \dots, m;\ j = 1, \dots, d)$. With each such multi-index α is associated the differential operator

$$\partial^\alpha := \frac{\partial^{\alpha_1^1}}{\partial x_1^1} \cdots \frac{\partial^{\alpha_1^d}}{\partial x_1^d} \cdots \frac{\partial^{\alpha_m^1}}{\partial x_m^1} \cdots \frac{\partial^{\alpha_m^d}}{\partial x_m^d}$$

whose order is $|\alpha| := \sum_{i,j} \alpha_i^j$. In particular, $\partial^\alpha f = f$ for $|\alpha| = 0$.

Let $\mathcal{D}(\mathcal{M})$ denote the set of functions $f \in C_k(\mathcal{M})$ for which the derivatives $\partial^\alpha D^m f$ exist and are continuous on $\mathcal{M} \times (\mathbb{R}^d)^m$ for all $m = 0, 1, 2, \dots$ and all $m \times d$ multi-indices α. On $\mathcal{D}(\mathcal{M})$ we define seminorms $\|\cdot\|_n$, $n = 0, 1, 2, \dots$, by

$$\|f\|_n := \sup\left\{ |\partial^\alpha D^m f(\mu)(\underline{x})| : (\mu, \underline{x}) \in \mathcal{M} \times (\mathbb{R}^d)^m,\ |\alpha| + m \le n \right\}.$$

These seminorms generate a metric

$$\rho(f,g) := \sum_{n=0}^{\infty} 2^{-n}(\|f-g\|_n \wedge 1), \qquad f,g \in \mathcal{D}(\mathcal{M}).$$

For compact sets $K \subset \mathbb{R}^d$, we consider the subspaces

$$\mathcal{D}_K(\mathcal{M}) := \{f \in \mathcal{D}(\mathcal{M})\colon \operatorname{arg\,dom} f \subseteq K\}.$$

With respect to the metric ρ, the spaces $\mathcal{D}_K(\mathcal{M})$ are (locally convex) Fréchet spaces. To verify the completeness of $\mathcal{D}_K(\mathcal{M})$, let (f_n) be a Cauchy sequence in $\mathcal{D}_K(\mathcal{M})$. Then there exist functions $f^{(\alpha,m)}$ such that $\partial^\alpha D^m f_n \to f^{(\alpha,m)}$ as $n \to \infty$ uniformly on $\mathcal{M} \times (\mathbb{R}^d)^m$ for all m and all multi-indices α. Clearly $f^{(\alpha,m)} \in C_k(\mathcal{M} \times (\mathbb{R}^d)^m)$ and $\operatorname{arg\,dom} f^{(\alpha,m)} \subseteq K$. Set $f := f^{(0,0)}$ and $f^{(m)} := f^{(0,m)}$. It only remains to show that $f^{(\alpha,m)} = \partial^\alpha D^m f$.

Passing to the limit as $n \to \infty$ in

$$\begin{aligned} D^{m-1} f_n(\nu)(x_1,\dots,x_{m-1}) &- D^{m-1} f_n(\mu)(x_1,\dots,x_{m-1}) \\ &= \int_0^1 \langle \nu-\mu, D^m f_n((1-\theta)\mu+\theta\nu)(x_1,\dots,x_{m-1},\cdot)\rangle \, d\theta, \end{aligned}$$

we find that

$$\begin{aligned} f^{(m-1)}(\nu)(x_1,\dots,x_{m-1}) &- f^{(m-1)}(\mu)(x_1,\dots,x_{m-1}) \\ &= \int_0^1 \left\langle \nu-\mu, f^{(m)}((1-\theta)\mu+\theta\nu)(x_1,\dots,x_{m-1},\cdot)\right\rangle d\theta. \end{aligned}$$

From this we successively conclude that $f^{(m)} = D^m f$ for $m = 1,2,\dots$. Given $m \in \{0,1,2,\dots\}$ and $\mu \in \mathcal{M}$, $(D^m f_n(\mu))$ is a Cauchy sequence in the Fréchet space $\mathcal{D}_K((\mathbb{R}^d)^m)$ with limit $D^m f(\mu)$. This finally yields $f^{(\alpha,m)} = \partial^\alpha D^m f$ for all α.

We furnish $\mathcal{D}(\mathcal{M})$ with the strongest *locally convex* vector space topology which induces on $\mathcal{D}_K(\mathcal{M})$ the original Fréchet topology for each compact set $K \subset \mathbb{R}^d$. From now on we will consider $\mathcal{D}(\mathcal{M})$ to be equipped with this topology and refer to $\mathcal{D}(\mathcal{M})$ as the *Schwartz space* of test functions on $\mathcal{M}$.

The space $\mathcal{D}(\mathcal{M})$ has the following properties:

(i) A convex balanced set W is open in $\mathcal{D}(\mathcal{M})$ if and only if $W \cap \mathcal{D}_K(\mathcal{M})$ is open in $\mathcal{D}_K(\mathcal{M})$ for each compact set $K \subset \mathbb{R}^d$.

(ii) If E is a bounded subset of $\mathcal{D}(\mathcal{M})$, then $E \subseteq \mathcal{D}_K(\mathcal{M})$ for some compact set $K \subset \mathbb{R}^d$.

(iii) If (f_n) is a Cauchy sequence in $\mathcal{D}(\mathcal{M})$, then (f_n) is a converging sequence in $\mathcal{D}_K(\mathcal{M})$ for some compact set $K \subset \mathbb{R}^d$.

(iv) $\mathcal{D}(\mathcal{M})$ is complete.

The proof of (i)–(iv) is essentially the same as that of Theorem 6.5 in Rudin [7]. Only part (ii) needs an explanation. We proceed indirectly. Let E be a subset of $\mathcal{D}(\mathcal{M})$ and suppose that E lies in no $\mathcal{D}_K(\mathcal{M})$. Abbreviate $K_m := [-m, m]^d$, $m \in \mathbb{N}$. Then, for each m, we find a function $f_m \in E$ which does not belong to $\mathcal{D}_{K_m}(\mathcal{M})$. Hence, there exist measures $\mu_m, \nu_m \in \mathcal{M}$ with $\mu_m|_{K_m} = \nu_m|_{K_m}$ and $f_m(\mu_m) \neq f_m(\nu_m)$. The sets

$$V_m := \left\{ f \in \mathcal{D}(\mathcal{M})\colon |f(\mu_m) - f(\nu_m)| < \frac{1}{m} |f_m(\mu_m) - f_m(\nu_m)| \right\}$$

are convex balanced open neighborhoods of 0 in $\mathcal{D}(\mathcal{M})$. Moreover, $V_m \supseteq \mathcal{D}_K(\mathcal{M})$ for each compact set K and all m with $K_m \supseteq K$. Hence, for each compact set K, the intersection of the sets $V_m \cap \mathcal{D}_K(\mathcal{M})$ coincides with a finite intersection which is open in $\mathcal{D}_K(\mathcal{M})$. This shows that

$$V := \bigcap_m V_m$$

is a convex balanced open neighborhood of 0 in $\mathcal{D}(\mathcal{M})$. Since $f_m \in E$ and $f_m \notin mV$, $m = 1, 2, \ldots$, we conclude that E is not bounded. This proves (ii).

Let $\mathcal{D}'(\mathcal{M})$ denote the space of real-valued linear continuous functionals on $\mathcal{D}(\mathcal{M})$ equipped with the weak* topology. We will refer to $\mathcal{D}'(\mathcal{M})$ as the space of *Schwartz distributions* on $\mathcal{M}$. $\langle \Lambda, f \rangle$ will denote the application of the distribution $\Lambda \in \mathcal{D}'(\mathcal{M})$ to the test function $f \in \mathcal{D}(\mathcal{M})$. Let us quote some basic properties of $\mathcal{D}'(\mathcal{M})$ which are straightforward adaptations from the 'classical' situation:

(v) A linear functional $\Lambda\colon \mathcal{D}(\mathcal{M}) \to \mathbb{R}$ belongs to $\mathcal{D}'(\mathcal{M})$ if and only if $\langle \Lambda, f_n \rangle \to 0$ whenever $f_n \to 0$ in $\mathcal{D}(\mathcal{M})$.

(vi) Given $\Lambda_n \in \mathcal{D}'(\mathcal{M})$, $n \in \mathbb{N}$, suppose that the finite limit

$$\lambda(f) := \lim_{n \to \infty} \langle \Lambda_n, f \rangle$$

exists for each $f \in \mathcal{D}(\mathcal{M})$. Then there exists $\Lambda \in \mathcal{D}'(\mathcal{M})$ such that

$$\langle \Lambda, f \rangle = \lambda(f) \qquad \text{for all } f \in \mathcal{D}(\mathcal{M}).$$

3.2. Bernstein and de Finetti approximations. The aim of this subsection is to prove a version of Weierstrass' Approximation Theorem by showing that each function in $\mathcal{D}(\mathcal{M})$ can be approached by Bernstein polynomials. This will then be used to prove the separability of the Fréchet spaces $\mathcal{D}_K(\mathcal{M})$ which is crucial for the proof of Lemma 3.14 in Section 3.3. Our Bernstein approximation also leads to a dual de Finetti approximation for distributions in $\mathcal{D}'(\mathcal{M})$ and probability measures in $\mathcal{M}(\mathcal{M})$.

Let us begin with two technical lemmas. Recall that the maps $\varepsilon = \varepsilon^N$, the spaces $\mathcal{G}^0_{N,K}$ and $\mathcal{G}^\infty_{N,K}$, and the metrics ρ_N were introduced in Section 2. Each function $g \in \mathcal{G}^0_{N,K}$ has a natural extension onto $(\overline{\mathbb{R}^d})^N$ which will be denoted by the same symbol g. For fixed $N \in \mathbb{N}$, we set $E := \{1, \dots, N\}$ and denote by $E^\ell_<$ the set of all indices $\underline{j} = (j_1, \dots, j_\ell) \in E^\ell$ with $j_1 < j_2 < \cdots < j_\ell$, $\ell = 1, \dots, N$. For each such $\underline{j}$ and each $\underline{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$, we set $\underline{x}_{\underline{j}} := (x_{j_1}, \dots, x_{j_\ell})$.

LEMMA 3.7. *Fix $N \in \mathbb{N}$ and a compact set $K \subset \mathbb{R}^d$ arbitrarily.*

a) Each function $g \in \mathcal{G}^0_{N,K}$ admits a unique representation of the form

$$g(\underline{x}) = g_0 + \sum_{\ell=1}^{N} \sum_{\underline{j} \in E^\ell_<} g_\ell(\underline{x}_{\underline{j}}), \qquad \underline{x} \in (\mathbb{R}^d)^N, \tag{3.5}$$

where g_0 is a constant, $g_\ell \in C_{k,s}((\mathbb{R}^d)^\ell)$ and $\operatorname{supp} g_\ell \subseteq K^\ell$ for $\ell = 1, \dots, N$. Conversely, for any such functions $g_1, \dots, g_N$, the function g defined by (3.5) belongs to $\mathcal{G}^0_{N,K}$. Moreover, the map $g \mapsto g_\ell$ from $\mathcal{G}^0_{N,K}$ into $C_{k,s}((\mathbb{R}^d)^\ell)$ is linear and continuous for $\ell = 1, 2, \dots, N$.

b) The same holds true with $\mathcal{G}^0_{N,K}$ and $C_{k,s}((\mathbb{R}^d)^\ell)$ replaced by $\mathcal{G}^\infty_{N,K}$ and $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$, respectively.

c) The space $\mathcal{G}^\infty_{N,K}$ is separable with respect to ρ_N. Moreover, $\mathcal{G}^\infty_{N,K}$ is dense in $\mathcal{G}^0_{N,K}$ in the uniform topology.

PROOF. a) Note that g_0 in (3.5) coincides with the constant value of g on $(K^c)^N$. We therefore will assume without loss of generality that $g = 0$ on $(K^c)^N$ and prove the decomposition (3.5) for $g_0 = 0$. Given $g \in \mathcal{G}^0_{N,K}$ with this property, take

$$\begin{aligned} g_N(x_1, \dots, x_N) &:= g(x_1, \dots, x_N) \\ &\qquad + \sum_{\ell=1}^{N} (-1)^\ell \sum_{\underline{j} \in E^\ell_<} g(x_1, \dots, x^*_{j_1}, \dots, x^*_{j_\ell}, \dots, x_N) \end{aligned}$$

with $x^*_1 := \cdots := x^*_N := \infty$. One easily checks that $g_N(\infty, x_2, \dots, x_N) = 0$. Using this, one finds that g_N belongs to $C_{k,s}((\mathbb{R}^d)^N)$ and $\operatorname{supp} g_N \subseteq K^N$. Hence,

$$g(\underline{x}) = g_N(\underline{x}) + \sum_{n=1}^{N-1} \sum_{\underline{j} \in E^n_<} h_n(\underline{x}_{\underline{j}})$$

with $h_n \in \mathcal{G}^0_{n,K}$ for $n = 1, 2, \dots, N-1$ and $h_n = 0$ on $(K^c)^n$. Now one may apply the same decomposition to each of the functions h_n (instead of g), and one successively arrives at (3.5) with $g_0 = 0$. The linearity and continuity of the maps $g \mapsto g_\ell$ is obvious from this. To prove uniqueness of the representation (3.5) assume that $g \equiv 0$. Then, by letting all N variables tend to inifinity, we conclude that $g_0 = 0$. After that, letting all but one variable tend to infinity, we find that $g_1 \equiv 0$, and so on.

b) The proof of b) is the same.

c) The separability of $\mathcal{G}^{\infty}_{N,K}$ with respect to ρ_N follows from assertion b) and the separability of $\mathbb{R}$ and of the spaces $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$, $\ell = 1, \dots, N$. Since $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$ is dense in the subspace of $C_{k,s}((\mathbb{R}^d)^\ell)$ consisting of functions g with $\operatorname{supp} g \subseteq K^\ell$ for $\ell = 1, \dots, N$, we also conclude from a) and b) that $\mathcal{G}^{\infty}_{N,K}$ is dense in $\mathcal{G}^{0}_{N,K}$. $\square$

As before, we consider $C_k(\mathcal{M})$ to be endowed with the supremum norm. Given a compact $K \subset \mathbb{R}^d$, let $C_K(\mathcal{M})$ denote the subspace of $C_k(\mathcal{M})$ consisting of functions f with $\operatorname{arg\,dom} f \subseteq K$. Recall that the operators $T = T^N$ were introduced in Section 2.

LEMMA 3.8. *Fix $N \in \mathbb{N}$ and a compact set $K \subset \mathbb{R}^d$ arbitrarily.*
a) The linear operator F^N defined by

$$F^N f := f \circ \varepsilon^N$$

is continuous and surjective both as map from $C_K(\mathcal{M})$ into $\mathcal{G}^{0}_{N,K}$ and as map from $\mathcal{D}_K(\mathcal{M})$ into $\mathcal{G}^{\infty}_{N,K}$.
b) The linear operator G^N defined by

$$G^N g(\mu) := \langle \mu^{\otimes N}, T^{-1} g \rangle, \qquad \mu \in \mathcal{M},$$

is continuous and injective both as map from $\mathcal{G}^{0}_{N,K}$ into $C_K(\mathcal{M})$ and as map from $\mathcal{G}^{\infty}_{N,K}$ into $\mathcal{D}_K(\mathcal{M})$.
c) $F^N \circ G^N$ is the identity operator on $\mathcal{G}^{0}_{N,K}$.

PROOF. a) Clearly $f \circ \varepsilon^N$ belongs to $\mathcal{G}^{0}_{N,K}$ for each $f \in C_K(\mathcal{M})$ and the map $F^N \colon C_K(\mathcal{M}) \to \mathcal{G}^{0}_{N,K}$ is continuous. Using the definition and properties of the derivatives $D^m f$, one verifies that $f \circ \varepsilon^N$ is smooth for $f \in \mathcal{D}_K(\mathcal{M})$ and that the map $F^N \colon \mathcal{D}_K(\mathcal{M}) \to \mathcal{G}^{\infty}_{N,K}$ is continuous.

b) According to Proposition 2.2, the linear operator T^{-1} is well-defined and continuous both as operator on $\mathcal{G}^{0}_{N,K}$ and on $\mathcal{G}^{\infty}_{N,K}$. Moreover, by Lemma 3.7, the function $f(\mu) := \langle \mu^{\otimes N}, h \rangle$ is the sum of a constant and of functions $f_\ell(\mu) = \langle \mu^{\otimes \ell}, g_\ell \rangle$ with $g_\ell \in C_{k,s}((\mathbb{R}^d)^\ell)$ and $\operatorname{supp} g_\ell \subseteq K^\ell$ (resp. $g_\ell \in \mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$) for $h \in C_K(\mathcal{M})$ (resp. $h \in \mathcal{D}_K(\mathcal{M})$), $\ell = 1, \dots, N$. But these functions clearly belong to $C_K(\mathcal{M})$ (resp. $\mathcal{D}_K(\mathcal{M})$). Moreover, the maps $h \mapsto f_\ell$, $\ell = 1, \dots, N$, are linear and continuous. This proves the linearity and continuity of G^N.

c) Assertion c) is obvious from the definitions of T, F^N, and G^N. It also yields the surjectivity of F^N and the injectivity of G^N. $\square$

We now introduce *Bernstein operators* $B^N \colon C_k(\mathcal{M}) \to C_k(\mathcal{M})$, $N \in \mathbb{N}$, by

$$B^N f(\mu) := \langle \mu^{\otimes N}, f \circ \varepsilon^N \rangle, \qquad \mu \in \mathcal{M}.$$

More explicitly, this may be written as

$$B^N f(\mu) = \int \cdots \int f\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}\right) \mu(dx_1) \dots \mu(dx_N). \tag{3.6}$$

We will call $B^N f$ *Bernstein polynomial approximation of f of order N*. Combining assertion a) of Lemma 3.8 with Lemma 3.7, one finds that $B^N f$ has indeed a polynomial structure:

$$B^N f(\mu) = \sum_{\ell=0}^{N} \langle \mu^{\otimes \ell}, g_{N,l} \rangle, \qquad \mu \in \mathcal{M}.$$

Thereby, if $f \in C_K(\mathcal{M})$ then $g_{N,\ell} \in C_{k,s}((\mathbb{R}^d)^\ell)$ and $\operatorname{supp} g_{N,\ell} \subseteq K^\ell$ for $\ell = 1, \dots, N$. If $f \in \mathcal{D}_K(\mathcal{M})$, then $g_{N,\ell} \in \mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$. By convention, $\langle \mu^{\otimes 0}, g_{N,0} \rangle := g_{N,0}$ is a constant. Moreover, those lemmas tell us that the maps $f \mapsto g_{N,\ell}$ are linear and continuous for $\ell = 0, 1, \dots, N$. From this we conclude that the Bernstein operators B^N are linear and continuous both as operators on $C_K(\mathcal{M})$ and as operators on $\mathcal{D}_K(\mathcal{M})$ for each compact K.

THEOREM 3.9. *(Weierstrass' polynomial approximation)*
Let K be an arbitrary compact in $\mathbb{R}^d$.
a) $B^N f \to f$ in $C_K(\mathcal{M})$ for each $f \in C_K(\mathcal{M})$.
b) $B^N f \to f$ in $\mathcal{D}_K(\mathcal{M})$ for each $f \in \mathcal{D}_K(\mathcal{M})$.

PROOF. a) The uniform convergence $B^N f \to f$ is a simple variation of the law of large numbers. To see this, fix $f \in C_K(\mathcal{M})$ and $\delta > 0$ arbitrarily. Since f is uniformly continuous (Lemma 3.4), we find $r \in \mathbb{N}$ and $\phi_1, \dots, \phi_r \in C_b(\mathbb{R}^d)$ such that

$$|f(\mu) - f(\nu)| < \delta/2 \qquad \text{for } (\mu, \nu) \in \mathcal{U}, \tag{3.7}$$

where

$$\mathcal{U} := \{(\mu, \nu) \in \mathcal{M} \times \mathcal{M} \colon |\langle \mu, \phi_k \rangle - \langle \nu, \phi_k \rangle| < 1 \text{ for } k = 1, \dots, r\}.$$

Abbreviating as before $\varepsilon(\underline{x}) = N^{-1} \sum_{i=1}^N \delta_{x_i}$ and denoting the supremum norm of f by $\|f\|_0$, we obtain

$$\begin{aligned}
&|B^N f(\mu) - f(\mu)| \\
&\le \left(\int_{(\varepsilon(\underline{x}),\mu) \in \mathcal{U}} \cdots \int + \int_{(\varepsilon(\underline{x}),\mu) \notin \mathcal{U}} \cdots \int \right) |f(\varepsilon(\underline{x})) - f(\mu)| \, \mu(dx_1) \dots \mu(dx_N) \\
&\le \frac{\delta}{2} + 2\|f\|_0 \, \mu^{\otimes N}\big((\varepsilon(\underline{x}), \mu) \notin \mathcal{U}\big) \\
&\le \frac{\delta}{2} + 2\|f\|_0 \sum_{k=1}^{r} \mu^{\otimes N} \left(\left| \frac{1}{N} \sum_{i=1}^{N} \big(\phi_k(x_i) - \langle \mu, \phi_k \rangle\big) \right| \ge 1 \right) \\
&\le \frac{\delta}{2} + \frac{2}{N} \|f\|_0 \sum_{k=1}^{r} \left\langle \mu, (\phi_k - \langle \mu, \phi \rangle)^2 \right\rangle.
\end{aligned}$$

In addition to (3.7), we have also used Chebyshev's inequality. For large N and all $\mu \in \mathcal{M}$, the expression on the right is smaller than δ, and we are done.

b) Let us now assume that $f \in \mathcal{D}_K(\mathcal{M})$. We already know that then $B^N f \in \mathcal{D}_K(\mathcal{M})$ for all N. We next want to find appropriate expressions for the derivatives of $B^N f$ in terms of the derivatives of f. Recall that the function

$$g_N(x_1, \dots, x_N) := f\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right), \qquad x_1, \dots, x_N \in \mathbb{R}^d, \tag{3.8}$$

belongs to $\mathcal{G}^\infty_{N,K}$ for each N. We introduce difference operators $\Delta^{(m)}$, $m = 0, 1, 2, \dots$, which act on functions $g \in \mathcal{G}^\infty_{N,K}$ according to the following rules:

$$\begin{aligned}
\Delta^{(0)} g &:= g, \\
\Delta^{(m)} g(x_1, \dots, x_N) &:= (N-m+1)\big[\Delta^{(m-1)} g(x_1, \dots, x_m, \dots, x_N) \\
&\qquad - \Delta^{(m-1)} g(x_1, \dots, x_m^*, \dots, x_N)\big] \qquad \text{for } m = 1, 2, \dots, N, \\
\Delta^{(m)} g &:= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for } m > N.
\end{aligned}$$

Here $x_1^* := \dots = x_N^* := \infty$. As a function of its first m variables $x_1, \dots, x_m$, $\Delta^{(m)} g(x_1, \dots, x_N)$ is symmetric and has compact support contained in K^m, $m = 1, 2, \dots, N$. Moreover, $\Delta^{(m)} g(x_1, \dots, x_N)$ will not change if we vary one of the variables $x_{m+1}, \dots, x_N$ outside of K. Using this, (3.6), and (3.8), we find that

$$\begin{aligned}
&\partial^\alpha D^m B^N f(\mu)(x_1, \dots, x_m) \\
&\qquad = \int \cdots \int \partial^\alpha \Delta^{(m)} g_N(x_1, \dots, x_N)\, \mu(dx_{m+1}) \dots \mu(dx_N)
\end{aligned} \tag{3.9}$$

for $m = 0, 1, \dots, N$ and all $m \times d$ multi-indices α. Using (3.8), the definition of the operators $\Delta^{(m)}$, and (3.3), we obtain

$$\begin{aligned}
\Delta^{(m)} g_N(x_1, \dots, x_N) &= \frac{N(N-1)\dots(N-m+1)}{N^m} \\
\times \int_0^1 \cdots \int_0^1 D^m f &\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} + \sum_{k=1}^m \frac{\theta_k}{N}(\delta_\infty - \delta_{x_k})\right)(x_1, \dots, x_m)\, d\theta_1 \dots d\theta_m
\end{aligned}$$

successively for $m = 1, 2, \dots, N$. From this and the uniform continuity of the derivatives $\partial^\alpha D^m f$ (cf. the proof of Lemma 3.4), we derive that

$$\partial^\alpha \Delta^{(m)} g_N(x_1, \dots, x_N) = \partial^\alpha D^m f\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)(x_1, \dots, x_m) + \overline{o}_u(1) \tag{3.10}$$

for $m = 0, 1, \dots, N$ and all $m \times d$ multi-indices α. Here $\overline{o}_u(1)$ denotes a function which, for fixed m and α, tends to zero as $N \to \infty$ uniformly in $x_1, \dots, x_N$.

Note that in the above integral expression for $\Delta^{(m)} g_N$ the differentiation with respect to the variables $x_1, \ldots, x_m$ in the argument

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}+\sum_{k=1}^{m}\frac{\theta_k}{N}(\delta_\infty-\delta_{x_k})$$

leads to a factor N^{-1}, so that these derivatives may be neglected asymptotically. Therefore in (3.10) the differential operator ∂^α does not act on the variables in $N^{-1}\sum\delta_{x_i}$. Substituting (3.10) in (3.9), we finally arrive at the representation

$$\begin{aligned}(3.11)\quad &\partial^\alpha D^m B^N f(\mu)(x_1,\ldots,x_m)\\ &=\int\cdots\int\partial^\alpha D^m f\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}\right)(x_1,\ldots,x_N)\,\mu(dx_{m+1})\ldots\mu(dx_N)+\overline{\overline{o}}_u(1)\end{aligned}$$

for $m=0,1,\ldots,N$ and all $m\times d$ multi-indices α. Thereby, for fixed m and α, the function $\overline{\overline{o}}_u(1)$ tends to zero as $N\to\infty$ uniformly in $\mu\in\mathcal{M}$ and $x_1,\ldots,x_m\in\mathbb{R}^d$.

In order to show that $B^N f\to f$ in $\mathcal{D}_K(\mathcal{M})$, it only remains to check that $\partial^\alpha D^m B^N f\to\partial^\alpha D^m f$ uniformly as $N\to\infty$ for all m and α. Since $\partial^\alpha D^m f$ is uniformly continuous, this now follows from (3.11) in the same way as the uniform convergence $B^N f\to f$ was derived from (3.6) in the proof of part a). □

We are now ready to state our separability result.

Lemma 3.10. *The following is valid for each compact subset K of $\mathbb{R}^d$.*
a) The Fréchet space $\mathcal{D}_K(\mathcal{M})$ is separable.
b) $\mathcal{D}_K(\mathcal{M})$ is dense in $C_K(\mathcal{M})$ with respect to the uniform topology.

Proof. a) Fix $f\in\mathcal{D}_K(\mathcal{M})$ arbitrarily and consider the polynomial approximations $f_N:=B^N f$ which have the form

$$f_N(\mu)=\langle\mu^{\otimes N},g_N\rangle,\tag{3.12}$$

with $g_N\in\mathcal{G}^\infty_{N,K}$. According to Lemma 3.7 c), the spaces $\mathcal{G}^\infty_{N,K}$ are separable with respect to the metric ρ_N. For each N, let $\mathcal{S}_{N,K}$ be a countable dense subset of $\mathcal{G}^\infty_{N,K}$. Choose $\tilde g_N\in\mathcal{S}_{N,K}$ so that

$$\rho_N(\tilde g_N,g_N)<e^{-N}\tag{3.13}$$

and consider the polynomials

$$\tilde f_N(\mu):=\langle\mu^{\otimes N},\tilde g_N\rangle.\tag{3.14}$$

Since these polynomials are taken from a countable collection of functions in $\mathcal{D}_K(\mathcal{M})$ which does not depend on f, the separability of $\mathcal{D}_K(\mathcal{M})$ will be shown as soon as we will have proved that $\tilde{f}_N \to f$ in $\mathcal{D}_K(\mathcal{M})$. According to Theorem 3.9 b), $f_N \to f$ in $\mathcal{D}_K(\mathcal{M})$. Therefore it will be enough to check that $\partial^\alpha D^m \tilde{f}_N - \partial^\alpha D^m f_N \to 0$ uniformly as $N \to \infty$ for all m and α. We know from the proof of part b) of Theorem 3.9 that

$$\partial^\alpha D^m f_N(\mu)(x_1, \dots, x_m) = \int \cdots \int \partial^\alpha \Delta^{(m)} g_N(x_1, \dots, x_N)\, \mu(dx_{m+1}) \dots \mu(dx_N).$$

The same is true for f_N and g_N replaced by $\tilde{f}_N$ and $\tilde{g}_N$, respectively. But from the definition of $\Delta^{(m)}$ and (3.13) we conclude that

$$\left|\partial^\alpha \Delta^{(m)} \tilde{g}_N(x_1, \dots, x_N) - \partial^\alpha \Delta^{(m)} g_N(x_1, \dots, x_N)\right| \leq N^m 2^{N+|\alpha|} e^{-N}$$

for all m, all $m \times d$ multi-indices α, $N > |\alpha|$, and $x_1, \dots, x_N \in \mathbb{R}^d$. Since the expression on the right tends to zero as $N \to \infty$, this finally yields the uniform convergence $\partial^\alpha D^m \tilde{f}_N - \partial^\alpha D^m f_N \to 0$.

b) If $f \in C_K(\mathcal{M})$, then $g_N \in \mathcal{G}^0_{N,K}$. According to part c) of Lemma 3.7, $\mathcal{G}^\infty_{N,K}$ is dense in $\mathcal{G}^0_{N,K}$ in the uniform topology. Hence, for each N we may choose $\tilde{g}_N \in \mathcal{G}^\infty_{N,K}$ so that $\|\tilde{g}_N - g_N\|_{N,0} \to 0$. Then the functions $\tilde{f}_N$ defined by (3.14) belong to $\mathcal{D}_K(\mathcal{M})$. Comparing (3.12) with (3.14), we obtain $\tilde{f}_N - f_N \to 0$ uniformly. Since $f_N \to f$ uniformly by Theorem 3.9 a), we conclude that $\tilde{f}_N \to f$ uniformly. In other words, f may be approached uniformly by functions from $\mathcal{D}_K(\mathcal{M})$, and we are done. □

We next need the following fact.

LEMMA 3.11. *The space $\mathcal{M}(\mathcal{M})$ is a topological subspace of $\mathcal{D}'(\mathcal{M})$.*

PROOF. Clearly $\mathcal{M}(\mathcal{M})$ may be considered as a subset of $\mathcal{D}'(\mathcal{M})$, and the topology of weak convergence is at least as strong as the subspace topology induced by $\mathcal{D}'(\mathcal{M})$. To prove that it is not strictly stronger, let us fix a probability measure $Q_0 \in \mathcal{M}(\mathcal{M})$ and a uniformly continuous function $f \in C_b(\mathcal{M})$ arbitrarily. We will show that there exist functions $f_1, f_2 \in C_k(\mathcal{M})$ such that

(3.15)
$$\{Q \in \mathcal{M}(\mathcal{M}) \colon |\langle Q - Q_0, f\rangle| < 1\} \subseteq \bigcup_{i=1}^{2} \{Q \in \mathcal{M}(\mathcal{M}) \colon |\langle Q - Q_0, f_i\rangle| < 1\}.$$

The desired assertion will then follow from this and the observation that, according to Lemma 3.10 b), $\mathcal{D}(\mathcal{M})$ is dense in $C_k(\mathcal{M})$ in the uniform topology, so that in (3.15) f_1 and f_2 may be replaced by functions from $\mathcal{D}(\mathcal{M})$.

Let (ψ_n) be a sequence in $C_k(\mathbb{R}^d)$ such that $0 \leq \psi_n \leq 1$ and $\psi_n \uparrow 1$ pointwise. Given $n \in \mathbb{N}$ and $\mu \in \mathcal{M}$, define $F_n(\mu) \in \mathcal{M}$ by

$$\langle F_n(\mu), \phi\rangle := \langle \mu, \psi_n \phi\rangle + \langle \mu, 1 - \psi_n\rangle \phi(0), \qquad \phi \in C_b(\mathbb{R}^d).$$

One readily checks that $F_n\colon \mathcal{M} \to \mathcal{M}$ is continuous, $F_n(\mu) \to \mu$ for each $\mu \in \mathcal{M}$, and $f \circ F_n$ belongs to $C_k(\mathcal{M})$ with $\operatorname{arg\,dom} f \circ F_n \subseteq \operatorname{supp} \psi_n$ for each n. Given $n \in \mathbb{N}$, abbreviate

$$\Psi_n(\mu) := \langle \mu, 1 - \psi_n \rangle, \qquad \mu \in \mathcal{M}.$$

Note that $\Psi_n \in C_k(\mathcal{M})$ for each n and $\Psi_n \to 0$ boundedly and pointwise.

Since f is uniformly continuous and because of the structure of the functions F_n, one finds $\delta > 0$ such that

$$\{\mu \in \mathcal{M}\colon\ |f(\mu) - f \circ F_n(\mu)| > 1/8\} \subseteq \{\mu \in \mathcal{M}\colon\ \Psi_n(\mu) > \delta\}$$

for all n. Hence,

$$\begin{aligned} |\langle Q - Q_0, f\rangle| &\le\ |\langle Q - Q_0, f \circ F_n\rangle| + |\langle Q - Q_0, f - f \circ F_n\rangle| \\ &\le\ |\langle Q - Q_0, f \circ F_n\rangle| \\ &\qquad + \frac{1}{4} + 2\|f\|_0 \left[Q(\Psi_n > \delta) + Q_0(\Psi_n > \delta)\right] \end{aligned}$$

for each n and all $Q \in \mathcal{M}(\mathcal{M})$. Here $\|f\|_0$ denotes the supremum norm of f. By Chebyshev's inequality,

$$\begin{aligned} Q(\Psi_n > \delta) + Q_0(\Psi_n > \delta) &\le \delta^{-1}\langle Q, \Psi_n\rangle + \delta^{-1}\langle Q_0, \Psi_n\rangle \\ &\le \delta^{-1}\,|\langle Q - Q_0, \Psi_n\rangle| + 2\delta^{-1}\langle Q_0, \Psi_n\rangle. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} |\langle Q - Q_0, f\rangle| \le\ & |\langle Q - Q_0, f \circ F_n\rangle| \\ &+ \frac{1}{4} + 2\|f\|_0\delta^{-1}\left[|\langle Q - Q_0, \Psi_n\rangle| + 2\langle Q_0, \Psi_n\rangle\right]. \end{aligned}$$

Since $\Psi_n \to 0$ boundedly and pointwise, we may fix n so large that

$$4\|f\|_0\delta^{-1}\langle Q_0, \Psi_n\rangle < 1/4.$$

For such n and all $Q \in \mathcal{M}(\mathcal{M})$,

$$|\langle Q - Q_0, f\rangle| \le \frac{1}{2} + \frac{1}{4}\,|\langle Q - Q_0, f_1\rangle| + \frac{1}{4}\,|\langle Q - Q_0, f_2\rangle|,$$

where $f_1 := 4f \circ F_n$ and $f_2 := 8\|f\|_0\delta^{-1}\Psi_n$ and both functions belong to $C_k(\mathcal{M})$. This implies the desired inclusion (3.15). $\square$

Let us now introduce the *de Finetti operators* Φ^N on $\mathcal{D}'(\mathcal{M})$ as the adjoints of the Bernstein operators B^N on $\mathcal{D}(\mathcal{M})$. Given $N \in \mathbb{N}$ and $\Lambda \in \mathcal{D}'(\mathcal{M})$, we will call $\Phi^N\Lambda$ *de Finetti approximation of* Λ *of order* N. Note that Φ^N maps $\mathcal{M}(\mathcal{M})$ into $\mathcal{M}(\mathcal{M}^N)$ and

$$\Phi^N Q = \widehat{\varepsilon}^N\left(\int \mu^{\otimes N}\, Q(d\mu)\right) \qquad \text{for } Q \in \mathcal{M}(\mathcal{M}).$$

The map $\widehat{\varepsilon}^N \colon \mathcal{M}_s((\mathbb{R}^d)^N) \to \mathcal{M}(\mathcal{M}^N)$ was defined in Section 2, and $\mathcal{M}(\mathcal{M}^N)$ is considered as closed subspace of $\mathcal{M}(\mathcal{M})$. Clearly Φ^N is also continuous as a map from $\mathcal{M}(\mathcal{M})$ into $\mathcal{M}(\mathcal{M}^N)$. The following corollary is an immediate consequence of Theorem 3.9.

COROLLARY 3.12. *(De Finetti approximation)*
a) For each $Q \in \mathcal{M}(\mathcal{M})$, $\Phi^N Q \to Q$ in $\mathcal{M}(\mathcal{M})$.
b) For each $\Lambda \in \mathcal{D}'(\mathcal{M})$, $\Phi^N \Lambda \to \Lambda$ in $\mathcal{D}'(\mathcal{M})$.

3.3. PRELIMINARIES ON DISTRIBUTION-VALUED FUNCTIONS.

DEFINITION 3.13. A function $\theta(\cdot)\colon [0,T] \to \mathcal{D}'(\mathcal{M})$ will be called *absolutely continuous* if for every compact K in $\mathbb{R}^d$ there exist a neighborhood U_K of 0 in $\mathcal{D}_K(\mathcal{M})$ and an absolutely continuous function $H_K\colon [0,T] \to \mathbb{R}$ such that

$$|\langle \theta(s), f\rangle - \langle \theta(t), f\rangle| \le |H_K(s) - H_K(t)| \tag{3.16}$$

for all $s, t \in [0,T]$ and $f \in U_K$.

Note that this is a straightforward adaption to our situation of the definition of absolute continuity of $\mathcal{D}'(\mathbb{R}^d)$-valued functions given in Dawson and Gärtner [2], Section 4.1.

LEMMA 3.14. *Let $\theta(\cdot)\colon [0,T] \to \mathcal{D}'(\mathcal{M})$ be absolutely continuous. Then the real-valued function $\langle \theta(\cdot), f\rangle$ is absolutely continuous for each $f \in \mathcal{D}(\mathcal{M})$. Moreover, there exists a Borel measurable function $\dot\theta(\cdot)\colon [0,T] \to \mathcal{D}'(\mathcal{M})$ such that*

$$\dot\theta(t) = \lim_{h\to 0} h^{-1}[\theta(t+h) - \theta(t)] \qquad \text{in } \mathcal{D}'(\mathcal{M})$$

for Lebesgue almost all $t \in [0,T]$.

PROOF. This is mainly a repetition of the proof of Lemma 4.2 in Dawson and Gärtner [2], where $\mathcal{D}'(\mathbb{R}^d)$-valued functions have been considered. The proof essentially relies on an application of the Banach-Alaoglu Theorem and the separability of the Fréchet spaces $\mathcal{D}_K(\mathcal{M})$ proved in Lemma 3.10. The existence of a Borel measurable version of $\dot\theta(\cdot)$ is a consequence of the Borel measurability of the real-valued functions $\langle \dot\theta(\cdot), f\rangle$, $f \in \mathcal{D}(\mathcal{M})$, and the separability of $\mathcal{D}'_K(\mathcal{M})$ with respect to the weak* topology for each compact $K \subset \mathbb{R}^d$. Note that the separability of $\mathcal{D}'_K(\mathcal{M})$ is also caused by the separability of $\mathcal{D}_K(\mathcal{M})$. $\square$

LEMMA 3.15. *(Integration by parts)*
Assume that $\theta(\cdot)\colon [0,T] \to \mathcal{D}'(\mathcal{M})$ is absolutely continuous and $f(\cdot)\colon [0,T] \to \mathcal{D}(\mathcal{M})$ is continuously differentiable. Then

$$\langle \theta(T), f(T)\rangle - \langle \theta(0), f(0)\rangle = \int_0^T \langle \dot\theta(u), f(u)\rangle\, du + \int_0^T \langle \theta(u), \dot f(u)\rangle\, du. \tag{3.17}$$

Here $\dot\theta(u)$ and $\dot f(u)$ denote the derivatives of $\theta(u)$ and $f(u)$ in $\mathcal{D}'(\mathcal{M})$ and $\mathcal{D}(\mathcal{M})$, respectively.

PROOF. Let U_K and H_K be as in Definition 3.13. The sets $\{f(u)\colon u \in [0,T]\}$ and $\{\dot f(u)\colon u \in [0,T]\}$ are compact and, in particular, bounded in $\mathcal{D}(\mathcal{M})$. Hence, both images are contained in $\mathcal{D}_K(\mathcal{M})$ for some compact K. Moreover, there exists a positive constant c such that these sets are contained in cU_K. From this and (3.16) we conclude that

$$|\langle\dot\theta(u), f(v)\rangle| \le c|\dot H_K(u)| \tag{3.18}$$

and

$$|\langle\dot\theta(u), \dot f(v)\rangle| \le c|\dot H_K(u)| \tag{3.19}$$

for Lebesgue almost all $u \in [0,T]$ and all $v \in [0,T]$. Note that the bound on the right is Lebesgue integrable on $[0,T]$.

Let us next check that the integrals on the right of (3.17) are well-defined. The function $\langle\dot\theta(u), f(v)\rangle$ is Borel measurable in u and continuous in v and therefore jointly measurable. In particular, the first integrand on the right of (3.17) is Borel measurable. Because of (3.18), it is also Lebesgue integrable. The second integrand is continuous.

Clearly the function $u \mapsto \langle\theta(u), f(T)\rangle$ is absolutely continuous with derivative $\langle\dot\theta(u), f(T)\rangle$, and $u \mapsto \langle\theta(0), f(u)\rangle$ is continuously differentiable with derivative $\langle\theta(0), \dot f(u)\rangle$. Therefore

$$\begin{aligned}
&\langle\theta(T), f(T)\rangle - \langle\theta(0), f(0)\rangle\\
&\qquad = \langle\theta(T)-\theta(0), f(T)\rangle + \langle\theta(0), f(T)-f(0)\rangle\\
&\qquad = \int_0^T \langle\dot\theta(u), f(T)\rangle\,du + \int_0^T \langle\theta(0), \dot f(u)\rangle\,du\\
&\qquad = \int_0^T \langle\dot\theta(u), f(u)\rangle\,du + \int_0^T \langle\dot\theta(u), f(T)-f(u)\rangle\,du\\
&\qquad\quad + \int_0^T \langle\theta(u), \dot f(u)\rangle\,du - \int_0^T \langle\theta(u)-\theta(0), \dot f(u)\rangle\,du
\end{aligned}$$

Now, for each $u \in [0,T]$, $v \mapsto \langle\dot\theta(u), f(v)\rangle$ is continuously differentiable with derivative $\langle\dot\theta(u), \dot f(v)\rangle$ and $v \mapsto \langle\theta(v), \dot f(u)\rangle$ is absolutely continuous with derivative $\langle\dot\theta(v), \dot f(u)\rangle$. Thus,

$$\int_0^T \langle\dot\theta(u), f(T)-f(u)\rangle\,du = \int_0^T\int_u^T \langle\dot\theta(u), \dot f(v)\rangle\,dv\,du \tag{3.20}$$

and

$$\int_0^T \langle\theta(u)-\theta(0), \dot f(u)\rangle\,du = \int_0^T\int_0^u \langle\dot\theta(v), \dot f(u)\rangle\,dv\,du. \tag{3.21}$$

The integrand $\langle\dot\theta(u), \dot f(v)\rangle$ is Borel measurable in u and continuous in v. Because of (3.19), it is jointly Lebesgue integrable. Hence, we may apply Fubini's

Theorem to see that the integrals (3.20) and (3.21) coincide, and we arrive at (3.17). □

LEMMA 3.16. *Suppose that* $\theta(\cdot)\colon [0,T] \to \mathcal{D}'(\mathcal{M})$ *is absolutely continuous. Then the de Finetti approximations* $\Phi^N\theta(\cdot)$*,* $N \in \mathbb{N}$*, are also absolutely continuous.*

PROOF. Since $\langle \Phi^N\theta(t), f\rangle = \langle \theta(t), B^N f\rangle$ for all $t \in [0,T]$ and $f \in \mathcal{D}(\mathcal{M})$, our assertion is immediate from the definition of absolute continuity and the continuity of the Bernstein operators B^N on $\mathcal{D}_K(\mathcal{M})$. □

According to Proposition 2.1, the map $\widehat{\varepsilon}^N$ induces a one-to-one correspondence between $\mathcal{M}_s((\mathbb{R}^d)^N)$-valued and $\mathcal{M}(\mathcal{M}^N)$-valued paths. Given a measure $\mu \in \mathcal{M}_s((\mathbb{R}^d)^N)$, we will call the measures $\mu^{(\ell)} \in \mathcal{M}_s((\mathbb{R}^d)^\ell)$ defined by $\mu^{(\ell)}(A) := \mu(A \times (\mathbb{R}^d)^{N-\ell})$, $\ell = 1,\dots,N-1$, partial marginals of μ. For convenience we set $\mu^{(N)} := \mu$ and define $\langle \mu^{(0)}, g\rangle := g$ to be a constant.

LEMMA 3.17. *Given a measure-valued path* $\mu(\cdot)\colon [0,T] \to \mathcal{M}_s((\mathbb{R}^d)^N)$*, consider the path* $Q(\cdot)\colon [0,T] \to \mathcal{M}(\mathcal{M}^N)$ *defined by* $Q(t) := \widehat{\varepsilon}^N(\mu(t))$*,* $t \in [0,T]$*. Then* $Q(\cdot)$ *is absolutely continuous as* $\mathcal{D}'(\mathcal{M})$*-valued function if and only if* $\mu(\cdot)$ *and all its partial marginals are absolutely continuous as* $\mathcal{D}'((\mathbb{R}^d)^\ell)$*-valued functions,* $\ell = 1,\dots,N$*.*

PROOF. 1^0 Suppose that the measure-valued paths $\mu^{(\ell)}(\cdot)$, $\ell = 1,\dots,N$, are absolutely continuous. This means that, for each compact K in $\mathbb{R}^d$, there exist open neighborhoods $U_K^{(\ell)}$ of 0 in $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$ and absolutely continuous functions $H_K^{(\ell)}\colon [0,T] \to \mathbb{R}$ such that

$$|\langle \mu^{(\ell)}(s), g_\ell\rangle - \langle \mu^{(\ell)}(t), g_\ell\rangle| \le |H_K^{(\ell)}(s) - H_K^{(\ell)}(t)| \tag{3.22}$$

for $s,t \in [0,T]$, $g_\ell \in U_K^{(\ell)}$, and all $\ell = 1,\dots,N$. We want to show that this implies the absolute continuity of $Q(\cdot)$.

For $t \in [0,T]$ and $f \in \mathcal{D}_K(\mathcal{M})$, we have

$$\langle Q(t), f\rangle = \langle \mu(t), F^N f\rangle, \tag{3.23}$$

where the map $F^N\colon \mathcal{D}_K(\mathcal{M}) \to \mathcal{G}^\infty_{N,K}$ is continuous by Lemma 3.8 a). Combining this with statement b) of Lemma 3.7, we find that (3.23) may be rewritten in the form

$$\langle Q(t), f\rangle = \sum_{\ell=0}^{N} \langle \mu^{(\ell)}(t), g_\ell\rangle. \tag{3.24}$$

Thereby g_ℓ belongs to $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$ and the maps $f \mapsto g_\ell$ from $\mathcal{D}_K(\mathcal{M})$ into $\mathcal{D}_{K^\ell,s}((\mathbb{R}^d)^\ell)$ are continuous for $\ell = 1,\dots,N$. The intersection U_K of the preimages of the sets $U_K^{(\ell)}$ with respect to these maps is therefore an open neighborhood of 0 in $\mathcal{D}_K(\mathcal{M})$. Combining (3.24) with (3.22), we obtain

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

for $s, t \in [0,T]$ and $f \in U_K$, where

$$H_K := \sum_{\ell=1}^{N} H_K^{(\ell)}$$

is absolutely continuous, and we are done.

2^0 Now suppose that $Q(\cdot)$ is absolutely continuous. We want to show that then the paths $\mu^{(\ell)}(\cdot)$, $\ell = 1, \dots, N$, are also absolutely continuous. To this end, we introduce symmetrization operators $S^{(\ell)}\colon \mathcal{D}_{K^\ell}((\mathbb{R}^d)^\ell) \to \mathcal{G}_{N,K}^\infty$, $\ell = 1, \dots, N$, by

$$S^{(\ell)}\phi(x_1, \dots, x_N) := \frac{1}{N(N-1)\dots(N-\ell+1)} \sum_{\pi} \phi(x_{\pi(1)}, \dots, x_{\pi(\ell)}),$$

where the sum runs over all injective maps $\pi\colon \{1, \dots, \ell\} \to \{1, \dots, N\}$. Using (3.23) and Lemma 3.8 c), we find that

$$\langle \mu^{(\ell)}(t), \phi \rangle = \langle Q(t), G^N \circ S^{(\ell)}\phi \rangle \tag{3.25}$$

for $t \in [0,T]$ and $\phi \in \mathcal{D}_{K^\ell}((\mathbb{R}^d)^\ell)$. Since both G^N and $S^{(\ell)}$ are continuous, $G^N \circ S^{(\ell)}\colon \mathcal{D}_{K^\ell}((\mathbb{R}^d)^\ell) \to \mathcal{D}_K(\mathcal{M})$ is also continuous. But this together with (3.25) implies the absolute continuity of $\mu^{(\ell)}(\cdot)$. $\square$

4. Identification of the rate function.

4.1. Empirical processes of N-tuples. Let $\xi_{ij}(t)$, $i = 1, \dots, M$, $j = 1, \dots, N$, be MN independent copies of our diffusion process in $\mathbb{R}^d$ with generator L_t. We begin by fixing N and viewing $(\xi_{i1}(t), \dots, \xi_{iN}(t))$, $i = 1, \dots, M$, as a system of M independent diffusions in $(\mathbb{R}^d)^N$ described with generator

$$L_t^N := \sum_{j=1}^{N} L_{t,j},$$

where $L_{t,j}$ is the operator L_t applied to the j-th coordinate. The probability laws of these processes on $C([0,T]; (\mathbb{R}^d)^N)$ will be denoted by $P_{\underline{x},s}^N$, $(\underline{x}, s) \in (\mathbb{R}^d)^N \times [0,T]$, and we will write $P_{\underline{x}}^N$ instead of $P_{\underline{x},0}^N$. We consider the associated empirical processes $X^{MN}(\cdot)$ defined by

$$X^{MN}(t) := \frac{1}{M} \sum_{i=1}^{M} \delta_{(\xi_{i1}(t), \dots, \xi_{iN}(t))}.$$

The probability law on $C([0,T]; \mathcal{M}((\mathbb{R}^d)^N))$ of the process $X^{MN}(\cdot)$ starting at $\nu \in \mathcal{M}^M((\mathbb{R}^d)^N)$ will be denoted by $\mathcal{Q}_\nu^{MN}$.

It follows from Dawson and Gärtner [2], Theorem 4.5, that the family $\{\mathcal{Q}_\nu^{MN};\ \nu \in \mathcal{M}^M((\mathbb{R}^d)^N)\}$ satisfies the large deviation principle as $M \to \infty$ with scale M and rate function $I^N\colon C([0,T];\mathcal{M}((\mathbb{R}^d)^N)) \to [0,\infty]$ given by

$$I^N(\mu(\cdot)) := \frac{1}{2}\int_0^T \|\dot{\mu}(t) - (L_t^N)^*\mu(t)\|^2_{\mu(t),t}\,dt \tag{4.1}$$

if $\mu(\cdot)$ is absolutely continuous as a $\mathcal{D}'((\mathbb{R}^d)^N)$-valued function and $I^N(\mu(\cdot)) = \infty$ otherwise. Here

$$\|\vartheta\|^2_{\mu,t} := \sup_{\phi\in\mathcal{D}((\mathbb{R}^d)^N)} \frac{|\langle\vartheta,\phi\rangle|^2}{\langle\mu, |\boldsymbol{\nabla}^N\phi|_t^2\rangle},$$

where

$$|\boldsymbol{\nabla}^N\phi|_t^2 := \sum_{j=1}^N |\boldsymbol{\nabla}_j\phi|_t^2,$$

with $|\boldsymbol{\nabla}_j\phi|_t^2$ being the application of (1.4) to the j-th coordinate of ϕ.

Given $\mu \in \mathcal{M}((\mathbb{R}^d)^N)$, let us denote by $\mu_s \in \mathcal{M}_s((\mathbb{R}^d)^N)$ its symmetrization:

$$\mu_s := \frac{1}{N!}\sum_\sigma \mu\circ\sigma^{-1},$$

where the sum runs over all permutations $\sigma\colon (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ of the N coordinates. We next prove the following symmetry property of the rate function I^N.

Lemma 4.1. *For each $\mu(\cdot) \in C([0,T];\mathcal{M}((\mathbb{R}^d)^N))$, we have*

$$I^N(\mu_s(\cdot)) = \min\left\{I^N(\nu(\cdot))\colon\ \nu(0) = \mu(0),\ \nu_s(\cdot) = \mu_s(\cdot)\right\}.$$

The peculiarity here is that the minimizer is not symmetric if $\mu(0)$ is not symmetric.

Proof. 1^0 We know from Dawson and Gärtner [2], Lemma 4.6, that

$$I^N(\mu(\cdot)) = \min\left\{\widehat{I}^N(Q)\colon\ Q\circ\pi_\cdot^{-1} = \mu(\cdot)\right\}, \tag{4.2}$$

$\mu(\cdot) \in C([0,T];\mathcal{M}((\mathbb{R}^d)^N))$, where

$$\widehat{I}^N(Q) := \sup_F\left[\langle Q, F\rangle - \langle Q\circ\pi_0^{-1}, \log E_\cdot^N e^F\rangle\right],$$

$Q \in \mathcal{M}(C([0,T];(\mathbb{R}^d)^N))$. Here the supremum is taken over all bounded continuous functions F on $C([0,T];(\mathbb{R}^d)^N)$, $\pi_t\colon C([0,T];(\mathbb{R}^d)^N) \to (\mathbb{R}^d)^N$, $t\in[0,T]$, denote the canonical projections, and $E_{\underline{x}}^N$ stands for expectation with respect to $P_{\underline{x}}^N$.

As a Legendre transform, $\widehat{I}^N$ is convex. Since the family of diffusion laws $P^N_{\cdot}$ is permutation invariant, $\widehat{I}^N$ is also permutation invariant. This implies that I^N is convex and permutation invariant. Moreover, if $\mu(\cdot)$ is symmetric, then the measure Q for which the minimum in (4.2) is attained, may also be chosen symmetric. For, if Q minimizes the expression on the right side of (4.2), then, by convexity and permutation invariance, its symmetrization is also a minimizer.

2^0 Given a path $\mu(\cdot)$ with $I^N(\mu(\cdot)) < \infty$ and a minimizer Q for (4.2), we next show that Q is absolutely continuous with respect to $P^N_{\mu(0)} = \int \mu(0)(d\underline{x})\, P^N_{\underline{x}}$ and

$$I^N(\mu(\cdot)) = \left\langle Q, \log \frac{dQ}{dP^N_{\mu(0)}} \right\rangle. \tag{4.3}$$

To prove this, we use the following estimate:

$$\begin{aligned}
\widehat{I}^N(Q) &= \sup_F \left[\langle Q, F\rangle - \langle \mu(0), \log E^N_{\cdot} e^F\rangle\right] \\
&\geq \sup_F \left[\langle Q, F\rangle - \log E^N_{\mu(0)} e^F\right] \\
&\geq \sup_f \left[\langle Q, F_f\rangle - \log E^N_{\mu(0)} e^{F_f}\rangle\right] \\
&= \sup_f \langle Q, F_f\rangle \\
&= \sup_f \Bigg[\langle \mu(T), f(T)\rangle - \langle \mu(0), f(0)\rangle \\
&\qquad - \int_0^T ds \left\langle \mu(s), \left(\frac{\partial}{\partial s} + L^N_s\right) f(s) + \frac{1}{2}|\nabla^N f(s)|^2_s \right\rangle\Bigg] \\
&= I^N(\mu(\cdot)).
\end{aligned}$$

The last three suprema are taken over all functions $f\colon (\mathbb{R}^d)^N \times [0,T] \to \mathbb{R}$ with compact support which are twice continuously differentiable with respect to the spatial variables and continuously differentiable in time. In that expressions, $F_f := F_{f,T}$, where

$$\begin{aligned}
F_{f,t}(\underline{x}(\cdot)) :=& f(\underline{x}(t), t) - f(\underline{x}(0), 0) \\
&- \int_0^t ds \left[\left(\frac{\partial}{\partial s} + L^N_s\right) f(\underline{x}(s), s) + \frac{1}{2}|\nabla^N f|^2_s(\underline{x}(s), s)\right],
\end{aligned}$$

$t \in [0,T]$. We have first used Jensen's inequality, then the fact that $e^{F_{f,t}}$ is an exponential $P^N_{\mu(0)}$-martingale for each f, and finally a variational representation of $I^N(\mu(\cdot))$ from Dawson and Gärtner [2], Lemma 4.8. Since, by assumption, $\widehat{I}^N(Q) = I^N(\mu(\cdot))$, all expressions in our estimate are in fact equal to $I^N(\mu(\cdot))$. The second variational expression is Sanov's rate function for empirical measures of i.i.d. random variables with law $P^N_{\mu(0)}$. Since it is finite, we conclude that Q

is absolutely continuous with respect to $P^N_{\mu(0)}$ and the considered expression equals $\langle Q, \log dQ/dP^N_{\mu(0)}\rangle$, see e.g. Deuschel and Stroock [5], Lemma 3.2.13. In this way we arrive at (4.3).

The above estimates also show the following. For each $Q \in \mathcal{M}(C([0,T]; (\mathbb{R}^d)^N)$ and $\mu(\cdot) := Q \circ \pi_\cdot^{-1}$ such that Q is absolutely continuous with respect to $P^N_{\mu(0)}$, we have

$$I^N(\mu(\cdot)) \leq \left\langle Q, \log \frac{dQ}{dP^N_{\mu(0)}} \right\rangle. \tag{4.4}$$

3^0 As a consequence of the convexity and permutation invariance of I^N, we obtain

$$I^N(\mu_s(\cdot)) \leq I^N(\mu(\cdot)) \tag{4.5}$$

for all $\mu(\cdot)$. Now fix a path $\mu(\cdot)$ with $I^N(\mu_s(\cdot)) < \infty$ arbitrarily. We want to construct a path $\nu(\cdot) \in C([0,T]; \mathcal{M}((\mathbb{R}^d)^N))$ so that $\nu(0) = \mu(0)$, $\nu_s(\cdot) = \mu_s(\cdot)$, and

$$I^N(\nu(\cdot)) \leq I^N(\mu_s(\cdot)). \tag{4.6}$$

A combination of (4.5) and (4.6) then yields the assertion of our lemma.

We know from 2^0 that there exists a symmetric measure $Q_s \in \mathcal{M}(C([0,T]; (\mathbb{R}^d)^N))$ such that $Q_s \circ \pi_\cdot^{-1} = \mu_s(\cdot)$, Q_s is absolutely continuous with respect to $P^N_{\mu_s(0)}$, and

$$I^N(\mu_s(\cdot)) = \left\langle Q_s, \log \frac{dQ_s}{dP^N_{\mu_s(0)}} \right\rangle. \tag{4.7}$$

Note that $\mu(0)$ and all its permutations are absolutely continuous with respect to $\mu_s(0)$. As a consequence, $P^N_{\mu(0)}$ is absolutely continuous with respect to $P^N_{\mu_s(0)}$, and

$$\frac{dP^N_{\mu(0)}}{dP^N_{\mu_s(0)}}(\underline{x}(\cdot)) = \frac{d\mu(0)}{d\mu_s(0)}(\underline{x}(0)), \qquad P^N_{\mu_s(0)}\text{-a.s.} \tag{4.8}$$

for any version of $d\mu(0)/d\mu_s(0)$, and analogous formulas are valid for the permutations of $P^N_{\mu(0)}$. We define a measure $Q \in \mathcal{M}(C([0,T]; (\mathbb{R}^d)^N))$ by fixing a version of $d\mu(0)/d\mu_s(0)$ and setting

$$\frac{dQ}{dQ_s}(\underline{x}(\cdot)) = \frac{d\mu(0)}{d\mu_s(0)}(\underline{x}(0)), \qquad Q_s\text{-a.s.} \tag{4.9}$$

The symmetrization of Q coincides with Q_s. Set

$$\nu(\cdot) := Q \circ \pi_\cdot^{-1}.$$

Then $\nu(0) = \mu(0)$ and $\nu_s(\cdot) = \mu_s(\cdot)$. It remains to verify (4.6). Using (4.8) and (4.9) together with the fact that Q_s is absolutely continuous with respect to $P^N_{\mu_s(0)}$, we find that Q is absolutely continuous with respect to $P^N_{\mu(0)}$ with

$$\frac{dQ}{dP^N_{\mu(0)}} = \frac{dQ_s}{dP^N_{\mu_s(0)}}, \qquad P^N_{\mu(0)}\text{-a.s.},$$

for any version of $dQ_s/dP^N_{\mu_s(0)}$. Moreover, the density on the right is permutation invariant $P^N_{\mu(0)}$-a.s. Together with (4.4) from step 2^0, this yields

$$I^N(\nu(\cdot)) \le \left\langle Q, \log\frac{dQ}{dP^N_{\mu(0)}}\right\rangle = \left\langle Q, \log\frac{dQ_s}{dP^N_{\mu_s(0)}}\right\rangle = \left\langle Q_s, \log\frac{dQ_s}{dP^N_{\mu_s(0)}}\right\rangle.$$

Combining this with (4.7), we finally arrive at assertion (4.6). □

4.2. Level II large deviations as $M \to \infty$. The level II empirical process $\Xi^{MN}(\cdot)$ will now be considered as a process in $\mathcal{M}(\mathcal{M}^N_I)$. Its laws on $C([0,T];\mathcal{M}(\mathcal{M}^N_I))$ will be denoted by $\tilde{\mathcal{P}}^{MN}_\nu$, $\nu \in \mathcal{M}^{MN}_{II}$. The original laws $\mathcal{P}^{MN}_\nu$ are then obtained as measure images with respect to the continuous embedding $C([0,T];\mathcal{M}(\mathcal{M}^N_I)) \hookrightarrow \mathcal{C}_{II} = C([0,T];\mathcal{M}_{II})$. Using the notation of Section 2, we find that

$$\Xi^{MN}(t) = \widehat{\varepsilon}^N(X^{MN}(t)), \qquad t \in [0,T], \tag{4.10}$$

where $\widehat{\varepsilon}^N$ is considered as continuous map from $\mathcal{M}((\mathbb{R}^d)^N)$ into $\mathcal{M}(\mathcal{M}^N_I)$. By an application of the contraction principle to the large deviation principle of Section 4.1, we will conclude from this that, for fixed N, the family $\{\tilde{\mathcal{P}}^{MN}_\nu; \nu \in \mathcal{M}^{MN}_{II}\}$ satisfies the large deviation principle as $M \to \infty$. The aim of this subsection is to derive an integral representation for the associated rate function. To this end we need two auxiliary lemmas.

First recall that the operator $\mathcal{L}^N_t\colon \mathcal{D}(\mathcal{M}_I) \to C_k(\mathcal{M}_I)$ is defined by

$$\mathcal{L}^N_t f(\mu) := \frac{1}{2N}\langle \mu, \Sigma_t D^2 f(\mu)\rangle + \langle \mu, L_t Df(\mu)\rangle,$$

where $\Sigma_t\colon \mathcal{D}((\mathbb{R}^d)^2) \to C_k(\mathbb{R}^d)$ is given by

$$\Sigma_t\phi(x) := \sum_{\alpha,\beta=1}^d a^{\alpha,\beta}(x,t)\frac{\partial^2\phi}{\partial x^\alpha\partial y^\beta}(x,x).$$

Lemma 4.2. *For each N, $t \in [0,T]$, and $f \in \mathcal{D}(\mathcal{M}_I)$, we have*

$$L^N_t(f\circ\varepsilon^N) = (\mathcal{L}^N_t f)\circ\varepsilon^N$$

and

$$|\boldsymbol{\nabla}^N(f\circ\varepsilon^N)|^2_t = \frac{1}{N}\left\langle \varepsilon^N, |\boldsymbol{\nabla} Df(\varepsilon^N)|^2_t\right\rangle.$$

The first identity shows that $\mathcal{L}_t^N$ is the 'restriction' of the Feller generator of the level I empirical process $\Xi^N(\cdot)$ to $\mathcal{D}(\mathcal{M}_I)$. To see this, note that $\Xi^N(t) = \varepsilon^N((\xi_1(t), \ldots, \xi_N(t)))$ and the diffusion $(\xi_1(\cdot), \ldots, \xi_N(\cdot))$ is governed by L_t^N.

Proof of Lemma 4.2. We know from Lemma 3.8 a) that $f \circ \varepsilon^N$ belongs to $\mathcal{G}_N^\infty$. Therefore our expressions are well-defined. A simple calculation shows that

$$\frac{\partial}{\partial x_j^\alpha}(f \circ \varepsilon^N)(\underline{x}) = \frac{1}{N} \frac{\partial}{\partial x^\alpha} Df(\varepsilon^N(\underline{x}))(x_j)$$

and

$$\begin{aligned}\frac{\partial^2}{\partial x_j^\alpha \partial x_j^\beta}(f \circ \varepsilon^N)(\underline{x}) =& \frac{1}{N} \frac{\partial^2}{\partial x^\alpha \partial x^\beta} Df(\varepsilon^N(\underline{x}))(x_j) \\ &+ \frac{1}{N^2} \frac{\partial^2}{\partial x^\alpha \partial y^\beta} D^2 f(\varepsilon^N(\underline{x}))(x_j, x_j),\end{aligned}$$

where the partial derivatives $\partial/\partial x^\alpha$ and $\partial/\partial y^\beta$ act on the spatial variables x and y in $Df(\mu)(x)$ and $D^2 f(\mu)(x, y)$ only. Substituting the above expressions in the definitions of L_t^N and $\boldsymbol{\nabla}^N$, we obtain the desired identities. $\square$

The functions of the form $f \circ \varepsilon^N$, $f \in \mathcal{D}(\mathcal{M}_I)$, are smooth but, in general, do not have compact support. They belong to the space $\mathcal{G}_N^\infty$ introduced in Section 2. This forces us to modify the representation (4.1) of the rate function I^N.

Lemma 4.3. *Fix $\mu(\cdot) \in C([0, T]; \mathcal{M}_s((\mathbb{R}^d)^N))$ arbitrarily. If the path $\mu(\cdot)$ and all of its marginals are absolutely continuous, then*

$$I^N(\mu(\cdot)) = \frac{1}{2} \int_0^T \sup_{\phi \in \mathcal{G}_N^\infty} \frac{\left|\langle \dot{\mu}(t), \phi\rangle - \langle \mu(t), L_t^N \phi\rangle\right|^2}{\left\langle \mu(t), |\boldsymbol{\nabla}^N \phi|_t^2 \right\rangle}\, dt. \tag{4.11}$$

Otherwise $I^N(\mu(\cdot)) = \infty$.

Even if $\phi \in \mathcal{G}_N^\infty$ does not have compact support, $\langle \dot{\mu}(t), \phi\rangle$ is well-defined. For, according to Lemma 3.7 b), each $\phi \in \mathcal{G}_N^\infty$ has a unique representation of the form

$$\phi(\underline{x}) = \phi_0 + \sum_{\ell=1}^N \sum_{\underline{j} \in E_<^\ell} \phi_\ell(\underline{x}_{\underline{j}})$$

with $\phi_\ell \in \mathcal{D}_s((\mathbb{R}^d)^\ell)$ for $\ell = 1, \ldots, N$. Let $\mu_{\underline{j}}^\ell(\cdot)$ denote the marginals of $\mu(\cdot)$ with respect to the variables $\underline{x}_{\underline{j}}$, $\underline{j} \in E_<^\ell$. If these paths are absolutely continuous, then we may define

$$\langle \dot{\mu}(t), \phi\rangle := \sum_{\ell=1}^N \sum_{\underline{j} \in E_<^\ell} \langle \dot{\mu}_{\underline{j}}^\ell(t), \phi_\ell\rangle \tag{4.12}$$

which makes sense for all $\phi \in \mathcal{G}_N^\infty$ and Lebesgue-a.a. $t \in [0, T]$.

PROOF OF LEMMA 4.3. Let $\mu(\cdot)$ be an arbitrary symmetric path.

If $I^N(\mu(\cdot)) < \infty$, then $\mu(\cdot)$ and all of its marginals are absolutely continuous. To see this, observe that the empirical processes associated with the marginals of $(\xi_{i1}(\cdot), \dots, \xi_{iN}(\cdot))$, $i = 1, \dots, M$, satisfy the large deviation principle as $M \to \infty$. Since the corresponding rate functions may be identified by an application of the contraction principle to $X^{MN}(\cdot)$, their values at the marginal paths $\mu^{\ell}_{\underline{j}}(\cdot)$ do not exceed $I^N(\mu(\cdot))$ and are therefore finite. This implies the absolute continuity of $\mu^{\ell}_{\underline{j}}(\cdot)$.

Now suppose that $\mu(\cdot)$ and all of its marginals are absolutely continuous. Given an arbitrary time interval $[s, t] \subseteq [0, T]$, we consider large deviations for the process $X^{MN}(\cdot)$ on $[s, t]$ as $M \to \infty$. The associated rate function $I^N_{s,t}$ has the form

$$I^N_{s,t}(\mu(\cdot)) = \frac{1}{2} \int_s^t \left\| \dot{\mu}(u) - (L^N_u)^* \mu(u) \right\|^2_{\mu(u),u} \, du. \tag{4.13}$$

For each $\phi \in \mathcal{G}^\infty_N$,

$$\exp\left\{ \phi(\underline{x}(u)) - \phi(\underline{x}(s)) - \int_s^u dv \, \left[L^N_v \phi(\underline{x}(v)) + \frac{1}{2} |\nabla^N \phi|^2_v(\underline{x}(v)) \right] \right\},$$

$u \in [s, t]$, is a bounded $P^N_{\mu(s),s}$-martingale, where $P^N_{\mu(s),s} = \int \mu(s)(d\underline{x}) \, P^N_{\underline{x},s}$. From this we conclude that

$$I^N_{s,t}(\mu(\cdot)) \geq \langle \mu(t), \phi \rangle - \langle \mu(s), \phi \rangle - \int_s^t du \, \left\langle \mu(u), L^N_u \phi + \frac{1}{2} |\nabla^N \phi|^2_u \right\rangle,$$

cf. step 2^0 in the proof of Lemma 4.1 or [2], Lemma 4.9. Using definition (4.12), we also get

$$\langle \mu(t), \phi \rangle - \langle \mu(s), \phi \rangle = \int_s^t du \, \langle \dot{\mu}(u), \phi \rangle.$$

Hence,

$$I^N_{s,t}(\mu(\cdot)) \geq \int_s^t du \, \left[\langle \dot{\mu}(u), \phi \rangle - \langle \mu(u), L^N_u \phi + \frac{1}{2} |\nabla^N \phi|^2_u \rangle \right].$$

Comparing this with (4.13), we conclude that

$$\frac{1}{2} \left\| \dot{\mu}(u) - (L^N_u)^* \mu(u) \right\|^2_{\mu(u),u} \geq \langle \dot{\mu}(u), \phi \rangle - \langle \mu(u), L^N_u \phi + \frac{1}{2} |\nabla^N \phi|^2_u \rangle \tag{4.14}$$

for Lebesgue-a.a. $u \in [0, T]$ and all $\phi \in \mathcal{G}^\infty_N$. Since the spaces $\mathcal{G}^\infty_{N,K}$ are separable (Lemma 3.7 c)), the corresponding Lebesgue null sets are contained in a universal set of Lebesgue measure zero. Hence, on the right of (4.14) we may take the supremum over all $\phi \in \mathcal{G}^\infty_N$ to obtain

$$\left\| \dot{\mu}(u) - (L^N_u)^* \mu(u) \right\|^2_{\mu(u),u} \geq \sup_{\phi \in \mathcal{G}^\infty_N} \frac{\left| \langle \dot{\mu}(u), \phi \rangle - \langle \mu(u), L^N_u \phi \rangle \right|^2}{\left\langle \mu(u), |\nabla^N \phi|^2_u \right\rangle}$$

for Lebesgue-a.a. $u \in [0,T]$. Since $\mu(\cdot)$ is symmetric and $\mathcal{D}_s((\mathbb{R}^d)^N) \subset \mathcal{G}_N^\infty$, the expressions on both sides are in fact equal a.e., and we are done. $\square$

We are now ready to identify the rate function of the level II process $\Xi^{MN}(\cdot)$ as $M \to \infty$.

LEMMA 4.4. *For each N, the family $\{\tilde{\mathcal{P}}_\nu^{MN};\, \nu \in \mathcal{M}_{II}^{MN}\}$ satisfies the large deviation principle as $M \to \infty$ with scale M and a rate function $\tilde{S}^N$ which admits the representation*

$$\tilde{S}^N(Q(\cdot)) = \frac{N}{2}\int_0^T \left\| \dot{Q}(t) - (\mathcal{L}_t^N)^* Q(t) \right\|_{Q(t),t}^2 dt$$

if $Q(\cdot) \in C([0,T]; \mathcal{M}(\mathcal{M}_I^N))$ is absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued function and $\tilde{S}^N(Q(\cdot)) = \infty$ otherwise.

PROOF. Fix N arbitrarily. We want to apply the contraction principle with respect to the map (4.10). To this end, we consider an arbitrary sequence of initial measures $\nu_M \in \mathcal{M}_{II}^{MN}$ such that $\nu_M \to \nu_0$ in $\mathcal{M}(\mathcal{M}_I^N)$. Then we find measures $\mu_M \in \mathcal{M}^M((\mathbb{R}^d)^N)$ such that $\widehat{\varepsilon}^N(\mu_M) = \nu_M$ for all M. In general, the measures μ_M are not symmetric and do not converge in $\mathcal{M}((\mathbb{R}^d)^N)$. But, because of Proposition 2.1, their symmetrizations converge to the symmetric preimage of ν_0 with respect to $\widehat{\varepsilon}^N$. Therefore the sequence (μ_M) is tight, and each limit point μ_0 is mapped by $\widehat{\varepsilon}^N$ to ν_0.

Now fix $Q(\cdot) \in C([0,T]; \mathcal{M}(\mathcal{M}_I^N))$ with $Q(0) = \nu_0$ arbitrarily. Because of Proposition 2.1, there exists a unique path $\mu_s(\cdot) \in C([0,T]; \mathcal{M}_s((\mathbb{R}^d)^N))$ such that

$$Q(t) = \widehat{\varepsilon}^N(\mu_s(t)) \qquad \text{for all } t \in [0,T]. \tag{4.15}$$

For each subsequence (μ_{M_n}) with $\mu_{M_n} \to \mu_0$ for some $\mu_0 \in \mathcal{M}((\mathbb{R}^d)^N)$, we may apply the contraction principle with respect to the continuous map (4.10) to find that the sequence $(\tilde{\mathcal{P}}_{\nu_{M_n}}^{M_n N})$ satisfies the large deviation principle as $n \to \infty$. The value of the corresponding rate function at $Q(\cdot)$ is

$$\min\left\{ I^N(\tilde{\mu}(\cdot))\colon \tilde{\mu}(0) = \mu_0,\, \tilde{\mu}_s(\cdot) = \mu_s(\cdot) \right\}.$$

According to Lemma 4.1, this minimum is independent of the particular limit point μ_0 (having symmetrization $\mu_s(0)$) and coincides with $I^N(\mu_s(\cdot))$. This shows that $\{\tilde{\mathcal{P}}_\nu^{MN};\, \nu \in \mathcal{M}_{II}^{MN}\}$ indeed satisfies the large deviation principle as $M \to \infty$ with scale M and rate function $\tilde{S}^N$ given by

$$\tilde{S}^N(Q(\cdot)) = I^N(\mu_s(\cdot)), \tag{4.16}$$

where $\mu_s(\cdot)$ is defined via (4.15).

We know from Lemma 3.17 that $Q(\cdot)$ is absolutely continuous if and only if $\mu_s(\cdot)$ and all of its marginals are absolutely continuous. Hence, supposing that $Q(\cdot)$ is absolutely continuous, we obtain

$$\langle \dot{\mu}_s(t), f \circ \varepsilon^N \rangle = \langle \dot{Q}(t), f \rangle \qquad \text{Lebesgue-a.e.}$$

for each $f \in \mathcal{D}(\mathcal{M}_I)$, where the expression on the left may be defined by (4.12), cf. also Lemma 3.14. Moreover, an application of the formulas in Lemma 4.2 shows that

$$\frac{\left|\langle \dot{\mu}_s(t), f \circ \varepsilon^N \rangle - \langle \mu_s(t), L_t^N (f \circ \varepsilon^N)\rangle\right|^2}{\left\langle \mu_s(t), |\boldsymbol{\nabla}^N (f \circ \varepsilon^N)|_t^2 \right\rangle} = N \frac{\left|\langle \dot{Q}(t), f\rangle - \langle Q(t), \mathcal{L}_t^N f\rangle\right|^2}{\left\langle Q(t), \langle \mu, |\boldsymbol{\nabla} Df(\mu)|_t^2 \rangle \right\rangle}$$

for Lebesgue-a.a. t and all $f \in \mathcal{D}(\mathcal{M}_I)$. But, according to Lemma 3.8 a), the transformation $f \mapsto f \circ \varepsilon^N$ maps $\mathcal{D}(\mathcal{M}_I)$ onto $\mathcal{G}_N^\infty$. Hence, using Lemma 4.3, we find that

$$I^N(\mu_s(\cdot)) = \frac{N}{2} \int_0^T \left\| \dot{Q}(t) - (\mathcal{L}_t^N)^* Q(t) \right\|_{Q(t),t}^2 dt.$$

Together with (4.16) this yields the desired integral representation of $\tilde{S}^N(Q(\cdot))$ provided that $Q(\cdot)$ is absolutely continuous. If $Q(\cdot)$ is not absolutely continuous, then $\mu_s(\cdot)$ is also not absolutely continuous, and therefore $\tilde{S}^N(Q(\cdot)) = I^N(\mu_s(\cdot)) = \infty$. $\square$

4.3. Completion of the proofs. The laws of the level I empirical processes $\Xi^N(\cdot)$ satisfy the large deviation principle with scale N. We may therefore apply assertion c) of Theorem 2.9 in [4] with respect to the canonical projection $\mathcal{M}(\mathcal{C}_I) \to \mathcal{C}_{II}$ to conclude that the laws $\mathcal{P}_\nu^{MN}$ of the level II processes $\Xi^{MN}(\cdot)$ satisfy the large deviation principle both for fixed N as $M \to \infty$ with scale M and for $M, N \to \infty$ with scale MN. That theorem also tells us that the associated rate functions S^N and S are related to each other by

$$\operatorname*{epi\,lim}_{N\to\infty} \frac{1}{N} S^N = S.$$

In particular, this proves Theorem 1.1.

As a straightforward consequence of Lemma 4.4,

$$\frac{1}{N} S^N(Q(\cdot)) = \frac{1}{2} \int_0^T \left\| \dot{Q}(t) - (\mathcal{L}_t^N)^* Q(t) \right\|_{Q(t),t}^2 dt \tag{4.17}$$

if $Q(\cdot)$ belongs to $C([0,T]; \mathcal{M}(\mathcal{M}_I^N))$ and is absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued function. Otherwise $S^N(Q(\cdot)) = \infty$. The *final step in the proof of Theorem 1.2* is to show by using (4.17) that the above epilimit is given by the appropriate integral representation. This is achieved by the following two lemmas.

Let $S^\infty(Q(\cdot))$ denote the desired representation of our rate function which is equal to the expression on the right of (1.8) if $Q(\cdot)$ is absolutely continuous and $+\infty$ otherwise. Recall that $\mathcal{C}_{II} = C([0,T]; \mathcal{M}_{II})$.

LEMMA 4.5. *Given $Q^N(\cdot) \in C([0,T];\mathcal{M}(\mathcal{M}_I^N))$ and $Q(\cdot) \in \mathcal{C}_{II}$, suppose that $Q^N(\cdot) \to Q(\cdot)$ in $\mathcal{C}_{II}$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} S^N(Q^N(\cdot)) \geq S^\infty(Q(\cdot)).$$

LEMMA 4.6. *Given $Q(\cdot) \in \mathcal{C}_{II}$, let $Q^N(t) := \Phi^N Q(t)$, $t \in [0,T]$, be the corresponding $\mathcal{M}(\mathcal{M}_I^N)$-valued de Finetti approximations. Then $Q^N(\cdot)$ belongs to $C([0,T];\mathcal{M}(\mathcal{M}_I^N))$, $Q^N(\cdot) \to Q(\cdot)$ in $\mathcal{C}_{II}$, and*

$$\limsup_{N\to\infty} \frac{1}{N} S^N(Q^N(\cdot)) \leq S^\infty(Q(\cdot)). \tag{4.18}$$

PROOF OF LEMMA 4.5. Fix $Q^N(\cdot) \in C([0,T];\mathcal{M}(\mathcal{M}_I^N))$ and $Q(\cdot) \in \mathcal{C}_{II}$ with $Q^N(\cdot) \to Q(\cdot)$ in $\mathcal{C}_{II}$ arbitrarily. Denote by $\mathcal{C}^1 := C^1([0,T];\mathcal{D}(\mathcal{M}_I))$ the space of continuously differentiable maps $[0,T] \to \mathcal{D}(\mathcal{M}_I)$. Consider the functionals

$$\begin{aligned} J^N(f) := \big\langle Q^N(T), f(T)\big\rangle - \big\langle Q^N(0), f(0)\big\rangle - \int_0^T dt\, \left\langle Q^N(t), \left(\frac{\partial}{\partial t} + \mathcal{L}_t^N\right) f(t)\right\rangle \\ - \frac{1}{2}\int_0^T dt\, \left\langle Q^N(t), \langle \mu, |\nabla D f(\mu,t)|_t^2\rangle\right\rangle, \end{aligned}$$

$f \in \mathcal{C}^1$. Define $J(f)$ in the same way but with $Q^N(t)$ and $\mathcal{L}_t^N$ replaced by $Q(t)$ and $\mathcal{L}_t$, respectively. Arguing as in the proof of Lemma 4.8 in Dawson and Gärtner [2] but using our Lemmas 3.14 and 3.15 in place of the Lemmas 4.2 and 4.3 of [2] and taking into account Lemma 3.10, we obtain

$$\frac{1}{N} S^N(Q^N(\cdot)) = \sup_{f\in\mathcal{C}^1} J^N(f)$$

and

$$S^\infty(Q(\cdot)) = \sup_{f\in\mathcal{C}^1} J(f). \tag{4.19}$$

Hence,

$$\liminf_{N\to\infty} \frac{1}{N} S^N(Q^N(\cdot)) \geq \lim_{N\to\infty} J^N(f) = J(f) \tag{4.20}$$

for all $f \in \mathcal{C}^1$. Here we have used that $Q^N(\cdot) \to Q(\cdot)$ in $\mathcal{C}_{II}$ and $\mathcal{L}_t^N f(t) \to \mathcal{L}_t f(t)$ in $C_b(\mathcal{M}_I)$ uniformly in $t \in [0,T]$. Combining (4.20) with (4.19), we arrive at the desired assertion. □

Proof of Lemma 4.6. 1^0 By definition, the measures $Q^N(t) = \Phi^N Q(t)$ are concentrated on $\mathcal{M}(\mathcal{M}_I^N)$. Since the operators Φ^N are continuous, the paths $Q^N(\cdot)$ are continuous in $\mathcal{D}'(\mathcal{M}_I)$ and, by Lemma 3.11, also in $\mathcal{M}_{II}$. To show that $Q^N(\cdot) \to Q(\cdot)$ in $\mathcal{C}_{II}$, it suffices to check that $Q^N(t_N) \to Q(t_0)$ in $\mathcal{M}_{II}$ whenever $t_N \to t_0$. Since the de Finetti operators Φ^N are adjoint to the Bernstein operators B^N and because of the Weierstrass approximation theorem (Theorem 3.9), we get

$$\langle Q^N(t_N), f\rangle = \langle Q(t_N), B^N f\rangle \to \langle Q(t_0), f\rangle$$

for each $f \in \mathcal{D}(\mathcal{M}_I)$. Together with Lemma 3.11 this yields the convergence $Q^N(t_N) \to Q(t_0)$ in $\mathcal{M}_{II}$.

2^0 As key for the proof of assertion (4.18), we next show that

$$B^N \mathcal{L}_t^N = \mathcal{L}_t B^N \qquad \text{on } \mathcal{D}(\mathcal{M}_I). \tag{4.21}$$

Remembering the definition of B^N, we obtain

$$L_t D B^N f(\mu)(x) = \sum_{j=1}^N \left\langle \mu^{\otimes(N-1)}, L_{t,j}(f \circ \varepsilon^N)(\dots, x, \dots)\right\rangle.$$

The operator $L_{t,j}$ under the sum is nothing but the operator L_t acting on the j-th coordinate (which is x), whereas integration with respect to $\mu^{\otimes(N-1)}$ is taken over the remaining $N-1$ coordinates. Hence, using the definitions of $\mathcal{L}_t$ and L_t^N and taking into account the first part of Lemma 4.2, we find that

$$\begin{aligned}
\mathcal{L}_t B^N f(\mu) &= \langle \mu, L_t D B^N f(\mu)\rangle \\
&= \langle \mu^{\otimes N}, L_t^N (f \circ \varepsilon^N)\rangle \\
&= \langle \mu^{\otimes N}, (\mathcal{L}_t^N f) \circ \varepsilon^N\rangle \\
&= B^N \mathcal{L}_t^N f(\mu).
\end{aligned}$$

This proves (4.21).

3^0 We are now going to prove (4.18). Let us assume without loss of generality that $S^\infty(Q(\cdot)) < \infty$. Then $Q(\cdot)$ is absolutely continuous and, by Lemma 3.16, the de Finetti approximations $Q^N(\cdot)$ are also absolutely continuous as $\mathcal{D}'(\mathcal{M}_I)$-valued paths. As a consequence of (4.21) and the duality between Φ^N and B^N, we find that

$$\left\langle \dot{Q}^N(t) - (\mathcal{L}_t^N)^* Q^N(t), f\right\rangle = \left\langle \dot{Q}(t) - \mathcal{L}_t^* Q(t), B^N f\right\rangle \tag{4.22}$$

for all $f \in \mathcal{D}(\mathcal{M}_I)$ and Lebesgue-a.a. $t \in [0,T]$. Here we have also used Lemma 3.10 a). We next claim that

(4.23)
$$\left\langle Q^N(t), \langle \mu, |\nabla D f(\mu)|_t^2\rangle\right\rangle \geq \left\langle Q(t), \langle \mu, |\nabla D B^N f(\mu)|_t^2\rangle\right\rangle$$

for all $f \in \mathcal{D}(\mathcal{M}_I)$ and $t \in [0,T]$. For,

$$\nabla DB^N f(\mu)(x) = \sum_{j=1}^{N} \left\langle \mu^{\otimes(N-1)}, \nabla_j (f \circ \varepsilon^N)(\dots, x, \dots) \right\rangle,$$

where ∇_j is the Riemannian gradient ∇ acting on the j-th variable (which is x) and integration is again taken over the remaining $N-1$ variables. From this we conclude that

$$|\nabla DB^N f(\mu)(x)|_t^2 \leq N \sum_{j=1}^{N} \left\langle \mu^{\otimes(N-1)}, |\nabla_j (f \circ \varepsilon^N)|_t^2(\dots, x, \dots) \right\rangle.$$

Together with the second formula in Lemma 4.2 this yields

$$\begin{aligned} \left\langle \mu, |\nabla DB^N f(\mu)|_t^2 \right\rangle &\leq N \left\langle \mu^{\otimes N}, |\nabla^N (f \circ \varepsilon^N)|_t^2 \right\rangle \\ &= \left\langle \mu^{\otimes N}, \langle \varepsilon^N, |\nabla Df(\varepsilon^N)|_t^2 \rangle \right\rangle. \end{aligned}$$

The expression on the right is the image of the function $\langle \mu, |\nabla Df(\mu)|_t^2 \rangle$ with respect to B^N. Therefore integration of both sides by $Q(t)(d\mu)$ leads to assertion (4.23).

Using (4.22) and (4.23), we find that

$$\begin{aligned} \left\| \dot{Q}^N(t) - (\mathcal{L}_t^N)^* Q^N(t) \right\|_{Q^N(t),t}^2 &= \sup_{f \in \mathcal{D}(\mathcal{M}_I)} \frac{\left| \left\langle \dot{Q}^N(t) - (\mathcal{L}_t^N)^* Q^N(t), f \right\rangle \right|^2}{\left\langle Q^N(t), \langle \mu, |\nabla Df(\mu)|_t^2 \rangle \right\rangle} \\ &\leq \sup_{f \in \mathcal{D}(\mathcal{M}_I)} \frac{\left| \left\langle \dot{Q}(t) - \mathcal{L}_t^* Q(t), B^N f \right\rangle \right|^2}{\left\langle Q(t), \langle \mu, |\nabla DB^N f(\mu)|_t^2 \rangle \right\rangle} \\ &\leq \left\| \dot{Q}(t) - \mathcal{L}_t^* Q(t) \right\|_{Q(t),t}^2. \end{aligned}$$

for Lebesgue-a.a. $t \in [0,T]$. Because of (4.17) and the corresponding integral representation for $S^\infty(Q(\cdot))$, this implies (4.18). $\square$

The proof of Theorem 1.2 is now complete. It only remains to derive Corollary 1.3. It is obvious from Theorem 1.2 that a path $\nu(\cdot) \in \mathcal{C}_{II}$ is a minimizer of the rate function S if and only if $\nu(\cdot)$ is absolutely continuous as a $\mathcal{D}'(\mathcal{M}_I)$-valued function and satisfies

$$\dot{\nu}(t) = \mathcal{L}_t^* \nu(t) \qquad \text{in } \mathcal{D}'(\mathcal{M}_I) \tag{4.24}$$

for Lebesgue-a.a. $t \in [0,T]$. The law of large numbers stated in Corollary 1.3 will therefore be an immediate consequence of the large deviation result presented in Theorem 1.1 provided that the solution $\nu(\cdot\,;\nu_0)$ to equation (4.24) with initial datum ν_0 is unique for each $\nu_0 \in \mathcal{M}_{II}$. Once uniqueness is established, a straightforward computation shows that $\nu(t;\nu_0)$ is given by formula (1.5). The continuous dependence of $\nu(\cdot\,;\nu_0)$ on ν_0 then follows from the continuity of $\mu_0 \mapsto \mu(\cdot\,;\mu_0)$ considered as map from $\mathcal{M}_I$ into $C([0,T];\mathcal{M}_I)$. The latter is a consequence of the Feller continuity of the semigroup associated with the diffusion operator L_t. To finish the proof of Corollary 1.3 it remains to verify uniqueness.

LEMMA 4.7. *For each $\nu_0 \in \mathcal{M}_{II}$ the absolutely continuous solution $\nu(\cdot)$ of equation (4.24) with initial datum $\nu(0) = \nu_0$ is unique.*

PROOF. Let $\nu(\cdot)$ be an arbitrary absolutely continuous solution of (4.24). Then the de Finetti approximations $\nu^N(\cdot) = \Phi^N \nu(\cdot)$ are also absolutely continuous and solve

$$\dot{\nu}^N(t) = (\mathcal{L}_t^N)^* \nu^N(t) \qquad \text{in } \mathcal{D}'(\mathcal{M}_I). \tag{4.25}$$

This follows from Lemma 3.16 and our key identity (4.21). According to Proposition 2.1, we find a unique path $\mu^N(\cdot) \in C([0,T];\mathcal{M}_s((\mathbb{R}^d)^N))$ such that $\nu^N(\cdot) = \widehat{\varepsilon}^N \mu^N(\cdot)$. By Lemma 3.17, the paths $\mu^N(\cdot)$ (and their partial marginals) are absolutely continuous. Therefore, taking into account the first formula in Lemma 4.2 and Lemma 3.8 a), we deduce from (4.25) that $\mu^N(\cdot)$ satisfies

$$\dot{\mu}^N(t) = (L_t^N)^* \mu^N(t) \qquad \text{in } \mathcal{D}'((\mathbb{R}^d)^N).$$

But the initial value problem for this Fokker-Planck equation is known to be unique, see Gärtner [6], Appendix B. This implies the uniqueness of the initial value problem (4.25) for each N. Since $\nu^N(t) \to \nu(t)$ for each $t \in [0,T]$ by Corollary 3.12 a), we conclude from this that the initial value problem for (4.24) is also unique, and we are done. $\square$

REFERENCES

[1] H. BAUER (1981). *Probability Theory and Elements of Measure Theory.* Academic Press, London.
[2] D. A. DAWSON AND J. GÄRTNER (1987). Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **20** 247–308.
[3] D. A. DAWSON AND J. GÄRTNER (1988). Long time behaviour of interacting diffusions. In *Stochastic calculus in application. Proc. Cambridge Symp., 1987,* J. R. Norris, ed., Longman 29–54.
[4] D. A. DAWSON AND J. GÄRTNER (1994). Multilevel large deviations and interacting diffusions. *Probab. Theory Relat. Fields* **98** 423–487.
[5] J.-D. DEUSCHEL AND D. W. STROOCK (1989). *Large Deviations.* Academic Press, San Diego.
[6] J. GÄRTNER (1988). On the McKean-Vlasov limit for interacting diffusions. *Math. Nachr.* **137** 197–248.
[7] W. RUDIN (1991). *Functional Analysis.* McGraw-Hill, New York.

FIELDS INSTITUTE
222 COLLEGE STREET
TORONTO, ONTARIO
CANADA M5T 3J1

TECHNISCHE UNIVERSITÄT BERLIN
FACHBEREICH MATHEMATIK
STR. DES 17. JUNI 136
D-10623 BERLIN, GERMANY