Donsker theorems for occupation measures of multi-dimensional periodic diffusions

Neil Deo*

Abstract

We study the empirical process arising from a multi-dimensional diffusion process with periodic drift and diffusivity. The smoothing properties of the generator of the diffusion are exploited to prove the Donsker property for certain classes of smooth functions. We partially generalise the finding from the one-dimensional case studied in [29]: that the diffusion empirical process exhibits stronger regularity than in the classical case of i.i.d. observations. As an application, precise asymptotics are deduced for the Wasserstein-1 distance between the time-T occupation measure and the invariant measure in dimensions $d \leq 3$.

1 Introduction

Consider the multi-dimensional diffusion process $(X_t)_{t \geq 0}$ which is the solution to the stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}^d, \quad t \geq 0. \tag{1.1}$$

Here, $(W_t)_{t \geq 0}$ is a d-dimensional Brownian motion, and the *drift field* $b : \mathbb{R}^d \to \mathbb{R}^d$ and the *diffusivity* $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are assumed to be Lipschitz continuous; thus the law of X is uniquely defined on the cylindrical sigma-algebra of $[0, \infty) \times \mathbb{R}^d$. Under mild conditions on b and σ, as $t \to \infty$ the law of X_t converges to the *invariant measure* $\mu = \mu_{b,\sigma}$, defined on a suitable state space. Moreover, the *ergodic theorem* holds for suitable measurable sets $A \subset \mathbb{R}^d$:

$$\hat{\mu}_T(A) := \frac{1}{T} \int_0^T 1_A(X_t)\,\mathrm{d}t \underset{T \to \infty}{\to} \mu(A) \quad \text{almost surely.} \tag{1.2}$$

This random measure $\hat{\mu}_T$, called the *time-T occupation measure*, is the natural analogue of the empirical measure in the diffusion setting. The ergodic property does not only hold for sets, but also for sufficiently regular functions f:

$$\hat{\mu}_T(f) := \frac{1}{T} \int_0^T f(X_t)\,\mathrm{d}t \underset{T \to \infty}{\to} \int f\,\mathrm{d}\mu =: \mu(f) \quad \text{almost surely.} \tag{1.3}$$

*University of Cambridge, UK. E-mail: and30@cam.ac.uk

With the ergodic property (1.3) in hand, the next natural question is whether or not $\hat{\mu}_T(f)$ obeys a central limit theorem, and indeed this is the case (see, for example, [6]): for suitable functions f, we have that

$$\sqrt{T}\left(\hat{\mu}_T(f) - \mu(f)\right) \xrightarrow{\mathcal{L}} N(0, \Gamma(f,f)) \quad \text{as } T \to \infty, \qquad (1.4)$$

where the covariance operator Γ is determined by b and σ. By the Cramér-Wold device, an analogous result holds for any suitable finite-dimensional class of functions $\mathcal{F}$.

In this paper, we investigate whether such a central limit theorem holds uniformly over a potentially *infinite-dimensional* class $\mathcal{F}$. This question was addressed in the scalar diffusion setting ($d = 1$) in [29]: they prove that such a central limit theorem holds whenever the limit process exists as a tight Borel probability measure on $\ell^\infty(\mathcal{F})$, the set of bounded functions on $\mathcal{F}$. This contrasts with the i.i.d. sampling model, in which the existence of the limiting process as a tight Borel measure is not sufficient for a uniform central limit theorem to hold: see Proposition 4.4.7 and the preceding discussion in [17]. In the scalar diffusion model, no such gap exists. Moreover, as (1.4) dictates the covariance structure of the limit, one deduces necessary and sufficient conditions based on the class $\mathcal{F}$ and the operator Γ for a Donsker theorem to hold over $\mathcal{F}$. Using these conditions, it is shown that the diffusion empirical process is smoother than its i.i.d. counterpart, in the sense that there exist function classes $\mathcal{F}$ which are Donsker for the diffusion empirical process but not for a classical (i.i.d.) empirical process.

However, the case of scalar diffusions is considerably easier than the multidimensional case due to the existence of local time, on which the proofs of [29] and several other works rely. Diffusion local time ceases to exist when $d \geq 2$, and so alternative methods must be used to deduce concentration properties of the diffusion process X. Here, we use the approach of [23], based on an analysis of the *generator* of the diffusion process as a partial differential operator. While other works on the statistics of multi-dimensional diffusions have relied on functional inequalities to deduce the necessary concentration properties of (X_t) [8, 27], this requires the diffusion to be *reversible*, or equivalently requires that b is a gradient vector field. The PDE arguments of [23] do not require this additional structural assumption. We restrict to the *periodic case* (as in [18, 23]), in which $b = b(\cdot + k)$ and $\sigma = \sigma(\cdot + k)$ for any $k \in \mathbb{Z}^d$. This effectively confines the diffusion process X to the d-dimensional torus, in the sense that the values of $X \mod \mathbb{Z}^d$ contain all of the statistical information about b and σ. Moreover, although the process X is not recurrent on $\mathbb{R}^d$, the invariant measure μ is well-defined on the torus, and it is in this sense that the ergodic and central limit theorems (1.2)-(1.4) hold. For a discussion on the non-periodic case, see Section 2.4.

Using these techniques, we uncover a similar story to the one-dimensional case: that the empirical process for diffusions exhibits stronger regularity than in the classical i.i.d. setting. As a demonstration, Theorem 2.5 below implies that the weak convergence of $\hat{\mu}_T$ to μ in Wasserstein space occurs at a rate of $T^{-1/2}$ in dimensions $d \leq 3$; in the i.i.d. setting, the parametric rate of convergence only occurs when $d = 1$. This increased regularity is due to the smoothing properties of the inverse of the generator of the diffusion process, which appears in the covariance structure of the limit process: see (2.9) below.

2 Main results

2.1 Definitions & notation

Let $\mathbb{T}^d$ be the d-dimensional torus, isomorphic to $(0,1]^d$ if opposite points on the cube are identified. We denote by $L^2(\mathbb{T}^d)$ the usual L^2 space with respect to the Lebesgue measure $\mathrm{d}x$, equipped with its inner product $\langle f, g \rangle = \langle f, g \rangle_{L^2} := \int fg \, \mathrm{d}x$. (If ever a

region of integration is unspecified, it should be taken to be the whole of $\mathbb{T}^d$.) Let μ be a Borel probability measure on $\mathbb{T}^d$ which is absolutely continuous with respect to Lebesgue measure, and whose density is bounded and bounded away from zero. Then the inner product $\langle f, g \rangle_\mu := \int fg \, \mathrm{d}\mu$ induces an equivalent norm on $L^2(\mathbb{T}^d)$. We further define the subspaces

$$L_0^2(\mathbb{T}^d) := \left\{ f \in L^2(\mathbb{T}^d) : \int f \, \mathrm{d}x = 0 \right\}, \quad L_\mu^2(\mathbb{T}^d) := \left\{ f \in L^2(\mathbb{T}^d) : \int f \, \mathrm{d}\mu = 0 \right\}.$$

For $s \in \mathbb{R}$, let $H^s(\mathbb{T}^d)$ denote the usual Sobolev space of functions, defined by duality when $s < 0$. Recall that $L^2(\mathbb{T}^d) = H^0(\mathbb{T}^d)$. The Sobolev spaces comprise the case $p = q = 2$ on the scale of Besov spaces $B_{pq}^s(\mathbb{T}^d)$, $s \in \mathbb{R}, 1 \le p, q \le \infty$; see Chapter 4 of [17] for definitions. The usual (isotropic) Hölder spaces $C^s(\mathbb{T}^d)$, $s \ge 0$ are closely related to the spaces $B_{\infty\infty}^s, s \ge 0$: specifically, $C^s(\mathbb{T}^d) = B_{\infty\infty}^s(\mathbb{T}^d)$ with equivalent norms for non-integer $s \ge 0$, and $C^s(\mathbb{T}^d) \subsetneq B_{\infty\infty}^s(\mathbb{T}^d)$ when $s \in \mathbb{N} \cup \{0\}$. Finally, we introduce the L^p-Sobolev spaces: for $k \in \mathbb{N}$ and $1 \le p \le \infty$, we define

$$W^{k,p}(\mathbb{T}^d) := \left\{ f \in L^p(\mathbb{T}^d) : D^k f \text{ exists and } D^k f \in L^p\left((\mathbb{T}^d)^{\otimes k}\right) \right\};$$

here $D^k f$ is the kth-order weak derivative of f. Note that for $k \in \mathbb{N}$, $W^{k,2}(\mathbb{T}^d) = H^k(\mathbb{T}^d)$.

We employ the same notation for vector fields $f = (f_1, \ldots, f_d)$, by a slight abuse of notation: in this case, $f \in H^s(\mathbb{T}^d) = (H^s(\mathbb{T}^d))^{\otimes d}$ means that each $f_j \in H^s(\mathbb{T}^d)$, with norm $\|f\|_{H^s}^2 = \sum_{j=1}^d \|f_j\|_{H^s}^2$. When no confusion may arise, we omit the domain and simply write C^s, H^s, etc. The codomain will always be clear from context.

Let μ be a (Borel) probability measure on some metric space. We write $Z \sim \mu$ if Z is a random variable with law μ. Given a collection of random variables $(Z_T)_{T \ge 0}$, we then write $Z_T \to^{\mathcal{L}} Z$, or $Z_T \to^{\mathcal{L}} \mu$, to denote the usual notion of weak convergence as $T \to \infty$, see Chapter 11 of [11] or Section 3.7 of [17] for definitions. We also write $Z_T \to^P Z$ to mean that Z_T converges in probability to Z as $T \to \infty$ under the measure P.

Let $(Z_f)_{f \in \mathcal{F}}$ be a stochastic process indexed by a set $\mathcal{F}$. Assume that the process Z is bounded, and its finite-dimensional marginals correspond to those of a tight Borel probability measure on $\ell^\infty(\mathcal{F})$, the space of all bounded functions $\mathcal{F} \to \mathbb{R}$, equipped with the supremum norm $\|\cdot\|_{\mathcal{F}}$. Then Z has a measurable version taking values in a separable subset of $\ell^\infty(\mathcal{F})$ almost surely, which we take to be Z itself. Given bounded processes $(Z_f^n)_{f \in \mathcal{F}}$, $n \ge 1$, we say that Z^n *converges in law to* Z *in* $\ell^\infty(\mathcal{F})$, written

$$Z^n \to^{\mathcal{L}} Z \quad \text{in } \ell^\infty(\mathcal{F}), \tag{2.1}$$

if for every bounded, continuous function $H : \ell^\infty(\mathcal{F}) \to \mathbb{R}$, we have $E^* H(Z^n) \to EH(Z)$, where E^* denotes outer expectation.

We will sometimes use the symbols $\lesssim, \gtrsim, \simeq$ to denote one- or two-sided inequalities which hold up to multiplicative constants that may be either universal or 'fixed' in the relevant context (in the latter case, this will always be qualified). Finally, we write ∂_i as shorthand for the weak partial derivative $\frac{\partial}{\partial x_i}$ on $\mathbb{R}^d$ or $\mathbb{T}^d$, for $1 \le i \le d$, and ∇ for the weak gradient.

2.2 Diffusions with periodic drift; invariant measures

We consider the SDE (1.1) where the drift $b : \mathbb{R}^d \to \mathbb{R}^d$ and diffusivity $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are Lipschitz continuous and one-periodic, so that $b(x) = b(x + k)$, $\sigma(x) = \sigma(x + k)$ for all $x \in \mathbb{R}^d, k \in \mathbb{Z}^d$. We may therefore consider b, σ as maps on $\mathbb{T}^d$; in general, we identify functions on the torus with their periodic extensions to $\mathbb{R}^d$ in this way. Under this assumption, a strong pathwise solution $(X_t)_{t \ge 0}$ of the SDE (1.1) exists (e.g. Theorem

24.3 in [4]); we write $P = P_{b,\sigma} = P_{b,\sigma,x}$ for the law of X when $X_0 = x$ on the cylindrical sigma-algebra of the path space $C([0,\infty) \to \mathbb{R}^d)$. We suppress the dependence on the initial value x as our results do not depend on it; we further suppress dependence on b and σ where no confusion may arise, as we consider these to be fixed.

As discussed previously, the periodicity of the drift b and diffusivity σ permit us to consider the diffusion X to take values in $\mathbb{T}^d$, without the loss of any statistical information about the parameters b, σ. (Note that the periodicity of b means that X is not even recurrent as a process on $\mathbb{R}^d$.) Once viewed as taking values in $\mathbb{T}^d$, the process X possesses an invariant measure $\mu = \mu_{b,\sigma}$ under mild conditions on b and σ (see Proposition 2.1 below), and the ergodic property (1.3) holds for all $f \in C(\mathbb{T}^d)$.

However, rather than viewing $\mu_{b,\sigma}$ as an object determining ergodic averages, we consider it as the solution of a certain partial differential equation arising from the *infinitesimal generator* of the diffusion process X, defined as the second order partial differential operator $L : H^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ given by

$$L = L_{b,\sigma} := \sum_{i,j=1}^d a_{ij}(\cdot)\partial_i\partial_j + \sum_{i=1}^d b_i(\cdot)\partial_i = \sum_{i=1}^d \partial_i\left(\sum_{j=1}^d a_{ij}(\cdot)\partial_j\right) + \sum_{i=1}^d \tilde{b}_i(\cdot)\partial_i, \qquad (2.2)$$

where $a := \frac{1}{2}\sigma\sigma^\top$ and $\tilde{b} : \mathbb{T}^d \to \mathbb{R}^d$ is defined by $\tilde{b}_i = b_i - \sum_j \partial_j a_{ij}$. Integrating by parts, one computes the L^2-adjoint operator of L to be

$$L^* = L_{b,\sigma}^* = \sum_{i=1}^d \partial_i\left(\sum_{j=1}^d a_{ij}(\cdot)\partial_j\right) - \sum_{i=1}^d \tilde{b}_i(\cdot)\partial_i - \sum_{i=1}^d \partial_i\tilde{b}_i. \qquad (2.3)$$

A standard argument (as in Section 2.2 of [23], for example) establishes that the invariant measure μ must then weakly solve

$$L^*\mu = 0 \quad \text{in } \mathbb{T}^d. \qquad (2.4)$$

As $\sigma : \mathbb{T}^d \to \mathbb{R}^{d \times d}$ is assumed to be Lipschitz, it is uniformly bounded. We further assume that σ is such that the induced $a = \frac{1}{2}\sigma\sigma^\top$ is *strictly elliptic*: that is, there exist constants $0 < \lambda, \Lambda < \infty$ such that

$$\lambda|\xi|^2 \le \sum_{i,j=1}^d a_{ij}(x)\xi_i\xi_j \le \Lambda|\xi|^2, \quad \forall x \in \mathbb{T}^d, \xi \in \mathbb{R}^d. \qquad (2.5)$$

Under this assumption, the second-order differential operators L and L^* are strictly elliptic and one may utilise the standard theory for such operators ([5, 16]).

For sufficiently regular b, σ, the adjoint equation (2.4) characterises the invariant measure, as shown by the following result.

Proposition 2.1. *Assume* $b \in C^1(\mathbb{T}^d), \sigma \in C^2(\mathbb{T}^d)$, *and assume further that* $a = \frac{1}{2}\sigma\sigma^\top$ *satisfies (2.5). Then a unique periodic solution* $\mu = \mu_{b,\sigma}$ *to (2.4) satisfying* $\int \mathrm{d}\mu = 1$ *exists. Moreover,* μ *is Lipschitz and bounded away from zero on* $\mathbb{T}^d$.

The proof is identical to that of [23, Proposition 1], which covers the special case $\sigma \equiv \mathrm{id}$. The proof extends naturally to the strictly elliptic case considered here using standard elliptic PDE theory; in particular, the relevant heat kernel estimates from [24] hold for such strictly elliptic operators. Furthermore, standard arguments tell us that there is a well-defined *solution operator* $L^{-1} : L^2(\mathbb{T}^d) \to H^2(\mathbb{T}^d)$ such that

$$L(L^{-1}[f]) = f \quad \forall f \in L_\mu^2. \qquad (2.6)$$

Moreover, L^{-1} is a bounded linear operator from L^2 to H^2, and more generally can be thought of as a '2-smoothing' operator between Besov spaces: see Lemma A.1 below.

2.3 Central limit theorems for $\hat{\mu}_T$

Our central limit theorem concerns the empirical process

$$\mathbb{G}_T(f) := \sqrt{T}\left(\hat{\mu}_T(f) - \mu(f)\right), \quad f \in \mathcal{F}, \tag{2.7}$$

where $\mathcal{F}$ is some class of test functions. We wish to establish the weak convergence of $\mathbb{G}_T$ to some limit process $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$, in the sense of (2.1). This is accomplished by checking the following two conditions (see [17, Theorem 3.7.23] or [30, Theorems 1.5.4 and 1.5.7]):

1. The finite-dimensional marginals of $\mathbb{G}_T$ converge weakly to those of $\mathbb{G}$.
2. There exists a pseudo-metric ρ on $\mathcal{F}$ such that $(\mathcal{F}, \rho)$ is totally bounded and for all $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{T \to \infty} \Pr\left\{ \sup_{\rho(f,g) \leq \delta} |\mathbb{G}_T(f) - \mathbb{G}_T(g)| > \varepsilon \right\} = 0. \tag{2.8}$$

We refer to (2.8) as the *asymptotic equicontinuity condition*; it ensures that the sequence $\mathbb{G}_T$ is asymptotically tight, and consequently that the limit process $\mathbb{G}$ exists as a tight Borel probability measure on $\ell^\infty(\mathcal{F})$.

The first condition determines the form of the limiting process $\mathbb{G}$ (when it exists): an application of the martingale central limit theorem (see e.g. [13, Theorem 7.1.4]) dictates the covariance structure of $\mathbb{G}$ to be given by (2.9) below. It then remains to check that the asymptotic equicontinuity condition holds. Via chaining arguments, this can be reduced to suitably controlling the complexity of the class $\mathcal{F}$ as measured by metric entropy with respect to the pseudo-distance ρ, for a 'good' choice of ρ. Here, one may leverage the smoothing property of the solution operator L^{-1} to define a rather weak pseudo-distance ρ (see (3.3) below), with respect to which the process $\mathbb{G}_T$ is sub-Gaussian uniformly in T. These considerations lead to the following result.

Theorem 2.2. *Let $d \geq 2$. Assume that there exists $\beta > 0$ such that $b \in C^{1+\beta}(\mathbb{T}^d)$, and that $\sigma \in C^2(\mathbb{T}^d)$ satisfies the strict ellipticity condition (2.5) for some $\lambda > 0$. Let $1 \leq p, q \leq \infty$, $s > \left(\max\left\{\frac{d}{2}, \frac{d}{p}\right\} - 1\right)$, and let $\mathcal{F}$ be a bounded subset of B_{pq}^s. Let $\mathbb{G}_T$ be the empirical process defined in (2.7). Then as $T \to \infty$,*

$$\mathbb{G}_T \to^{\mathcal{L}} \mathbb{G} \quad \text{in } \ell^\infty(\mathcal{F}),$$

where $\mathbb{G} = \mathbb{G}_{b,\sigma}$ is centred Gaussian process on $\mathcal{F}$ with covariance kernel

$$E\mathbb{G}(f)\mathbb{G}(g) = \langle \sigma^\top \nabla L^{-1}[\overline{f}], \sigma^\top \nabla L^{-1}[\overline{g}] \rangle_\mu, \tag{2.9}$$

whose law defines a tight Borel probability measure on $\ell^\infty(\mathcal{F})$. Here, $\mu = \mu_{b,\sigma}$ is the invariant measure of the diffusion process with drift field b and diffusivity σ, $L^{-1} = L_{b,\sigma}^{-1}$ is the inverse operator of the generator of the diffusion as defined in (2.6), and for $f \in \mathcal{F}$, $\overline{f} := f - \int f \, d\mu$.

Remark 2.3. The theorem also holds (with an identical proof) in the case $d = 1$ whenever $s > 0$. Note that the empirical process is more regular than in the i.i.d. case, which requires $s > d/2$ in any dimension for the limit to even be tight on $\ell^\infty(\mathcal{F})$.

The next proposition establishes that the assumption $s > \frac{d}{2} - 1$ is necessary and sufficient for the limit process $\mathbb{G}$ to exist as a tight Borel measure on $\ell^\infty(\mathcal{F})$.

Proposition 2.4. *Let $d \geq 2$, and let $b, \sigma, \mu = \mu_{b,\sigma}, L^{-1} = L_{b,\sigma}^{-1}$ be as in Theorem 2.2. Let $1 \leq p, q \leq \infty$, $s \geq 0$, and let $\mathcal{F}$ be the unit ball of $B_{pq}^s(\mathbb{T}^d)$. Then the centred Gaussian process $\mathbb{G}$ defined by (2.9) is a tight Borel measure on $\ell^\infty(\mathcal{F})$ if and only if $s > \frac{d}{2} - 1$.*

Thus in the case $p \in [2, \infty]$, the condition $s > \frac{d}{2} - 1$ in Theorem 2.2 is sharp. However, in the case $1 \leq p < 2$, when $s \in (\frac{d}{2} - 1, \frac{d}{p} - 1]$, while the limit process $\mathbb{G}$ exists as a tight Borel measure on $\ell^\infty(\mathcal{F})$, we do not know whether the empirical process $\mathbb{G}_T$ converges to it. For the analogous parameter configuration in the i.i.d. setting, $s \in (d/2, d/p)$, the limit exists as a tight Borel measure, but the empirical process does not in fact converge to this limit ([17, Proposition 4.4.7]). In the scalar diffusion setting, it was established in [29] that no such phenomenon takes place, and the empirical process converges to the limit experiment whenever it exists. What occurs in the multidimensional case is currently unclear; see Section 2.4 for a discussion of this matter.

Theorem 2.2 is related to the Bernstein-von-Mises results in [23] (Theorems 5 and 6) in the case $\sigma \equiv \mathrm{id}$ and $d \leq 3$. In place of the occupation measure $\hat{\mu}_T$, their results are for a sequence $\mu_{\hat{b}_T}$, where $\hat{b}_T$ is a suitable estimator of the drift field b (the posterior mean), and $\mu_{\hat{b}_T}$ denotes the invariant measure associated to the drift field $\hat{b}_T$. In the frequentist literature, a uniform central limit theorem similar to Theorem 2.2 was deduced in the scalar setting for a suitable kernel estimator $\hat{b}_T$ in [1], building on the results of [29].

As an application of Theorem 2.2, we derive convergence rates for the occupation measure $\hat{\mu}_T$ in the Wasserstein distance in dimensions $d \leq 3$. Recall that the Wasserstein-1 distance between Borel measures ν, ρ on $\mathbb{T}^d$ is defined as

$$W_1(\nu, \rho) = \inf_{\gamma \in \mathcal{C}(\nu, \rho)} \int_{\mathbb{T}^d \times \mathbb{T}^d} \|x - y\| \, \mathrm{d}\gamma(x, y), \qquad (2.10)$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{T}^d$ and $\mathcal{C}(\nu, \rho)$ is the set of all *couplings* of ν and ρ: Borel measures on $\mathbb{T}^d \times \mathbb{T}^d$ whose first and second marginals (on $\mathbb{T}^d$) are ν and ρ respectively. By the Kantorovich-Rubinstein duality formula, one may rewrite the Wasserstein-1 distance as

$$W_1(\hat{\mu}_T, \mu) = \sup_{f \in \mathrm{Lip}_1(\mathbb{T}^d)} \left[\int f \, \mathrm{d}\hat{\mu}_T - \int f \, \mathrm{d}\mu \right], \qquad (2.11)$$

where $\mathrm{Lip}_1(\mathbb{T}^d)$ is the class of all functions $f : \mathbb{T}^d \to \mathbb{R}$ that are 1-Lipschitz.

Theorem 2.5. *Let $d \leq 3$. Let b, σ be as in Theorem 2.2; let μ be the invariant measure of the associated diffusion process and let $\hat{\mu}_T$ be the time-T occupation measure as defined in (1.2). Then*

$$\sqrt{T} \, W_1(\hat{\mu}_T, \mu) \to^{\mathcal{L}} \|\mathbb{G}\|_{\mathcal{F}} \quad \text{as } T \to \infty,$$

where $\mathbb{G}$ is the centred Gaussian process on $\mathcal{F} := \mathrm{Lip}_1(\mathbb{T}^d) \cap L^2_\mu(\mathbb{T}^d)$ with covariance kernel given by (2.9), and $\|\mathbb{G}\|_{\mathcal{F}}$ is finite almost surely.

Proof. By linearity, one may replace $\mathrm{Lip}_1(\mathbb{T}^d)$ in (2.11) with the class $\mathcal{F}$ defined in the statement; observe that $\mathcal{F}$ is a bounded subset of $B^1_{\infty\infty}(\mathbb{T}^d)$. Theorem 2.2 now applies with $s = 1$ and $d \leq 3$, and this gives the result after noting that $W_1(\hat{\mu}_T, \mu)$ is measurable (as the supremum in (2.11) is attained over a countable set) and $\|\cdot\|_{\mathcal{F}} : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ is continuous.

Lastly, by considering the covariance structure of $\mathbb{G}$, one shows that $\|\mathbb{G}\|_{\mathcal{F}}$ is sub-Gaussian with respect to a multiple of the distance ρ_L defined in (3.3) below. Analogous arguments to those of Section 3.2 establish that $\|\mathbb{G}\|_{\mathcal{F}}$ is almost surely finite. □

Remark 2.6 (Concentration, finite sample bounds)**.** Using Exercise 2.3.1 of [17], one can show that $\|\mathbb{G}\|_{\mathcal{F}}$ concentrates about its finite expectation with sub-Gaussian tails. As was noted in Lemma 1 of [23], a similar analysis of $\|\mathbb{G}_T\|_{\mathcal{F}}$ can be used to derive non-asymptotic sub-Gaussian concentration bounds for $W_1(\hat{\mu}_T, \mu)$. Similar finite sample results were proved in [7], using transport-entropy inequalities.

In particular, Theorem 2.5 tells us that in dimensions $d \leq 3$,

$$W_1(\hat{\mu}_T, \mu) = O_P(T^{-1/2}).$$

When $d = 2, 3$, the empirical measure in the diffusion model converges faster than for a smooth density in the i.i.d. sampling model, in which the empirical measure converges in W_1-distance at a rate of $n^{-1/2}(\log n)^{1/2}$ for $d = 2$ and $n^{-1/3}$ for $d = 3$, where n is the sample size ([2, 10]; see also [15]). The $d = 1$ case of the classical i.i.d. setting, in which W_1 convergence does occur at the parametric rate, was comprehensively analysed in [9].

2.4 Discussion

A shortcoming of Theorem 2.2 is that it fails to address the case $1 \leq p < 2$, $s \in (\frac{d}{2} - 1, \frac{d}{p} - 1]$. For these parameters, the limit process $\mathbb{G}$ defined by (2.9) exists as a tight Borel measure on $\ell^\infty(\mathcal{F})$ by Proposition 2.4. However, the relevance of the limit to the empirical process $\mathbb{G}_T$ is no longer clear, since the Itô-Krylov formula (3.1) does not necessarily hold. This formula from [20] appears to be the most general such result for weakly differentiable functions, holding for functions in $W^{2,r}(\mathbb{T}^d)$ for any $r > d$. In fact, stochastic chain rules exist for only once weakly differentiable functions, as in [3, 14, 26]; however, in these formulae the quadratic covariation term is replaced by a suitable analogue, and so they do not yield an occupation time formula as in (3.1) below. Moreover, such formulae require delicate initial conditions, often depending on the function f under consideration, and so are unsuitable for developing uniform central limit theorems.

One may ask whether the assumption of periodicity of the drift field b and the diffusivity σ may be removed. In principle, this should be possible by making use of the results of [25] concerning the Poisson equation on $\mathbb{R}^d$, as in [1]. However, it is not obvious how precisely to adapt our key regularity estimate Lemma A.1 below; in particular, the choice of which Besov spaces to use must be informed by metric entropy considerations if there is to be any hope of proving the asymptotic equicontinuity condition (3.2) via chaining methods. One candidate is the family of weighted Besov spaces studied in [19] and successfully deployed in [22] in the classical i.i.d. empirical process setting. However, this generalisation requires many complicated technicalities to be resolved, and is left as a matter for future research.

3 Proof of Theorem 2.2

For this section, we fix the dimension $d \geq 2$, as well as a drift field b and diffusivity σ satisfying the conditions of Theorem 2.2. Fix $1 \leq p, q \leq \infty$, $s > (\max\{d/2, d/p\} - 1)$ and let $\mathcal{F} \subset B_{pq}^s(\mathbb{T}^d)$ be bounded. For $f \in \mathcal{F}$, we write $\overline{f} = f - \int f \, d\mu$. Let $\mathbb{G}_T$ be as in (2.7).

3.1 Convergence of finite-dimensional marginals

Our first task is to establish the covariance structure of the candidate limit process $\mathbb{G}$. By linearity of the process $f \mapsto \mathbb{G}_T(f)$ and the Cramér-Wold device, it suffices to consider a single $f \in \mathcal{F} \cap L_\mu^2$. The following argument was first used in [6]; we require a slightly augmented version for the function classes we consider.

As $f \in B_{pq}^s$, by Lemma A.1 we have that $L^{-1}[f] \in B_{pq}^{2+s}$. Using standard Besov embeddings (see, for example, [17, Proposition 4.3.10]) and the Sobolev-Gagliardo-Nirenberg inequality, for $p \geq 2$ we have the sequence of continuous embeddings $B_{pq}^{s+2} \subset B_{22}^{s+2-\delta} = W^{s+2-\delta,2} \subset W^{2,r}$ for some small $\delta > 0$ and some $r > d$. If $1 \leq p < 2$, there is an extra initial step: we use the embedding $B_{pq}^{s+2} \subset B_{2q}^{s-\frac{d}{p}+\frac{d}{2}+2}$ and then the fact that $s > \frac{d}{p} - 1$. Thus for $f \in \mathcal{F}$, $L^{-1}[f] \in W^{2,r}(\mathbb{T}^d)$ and so we may apply the Itô-Krylov formula

[20, Theorem 2.10.1] to obtain

$$\mathbb{G}_T(f) = \frac{1}{\sqrt{T}} \left(L^{-1}[f](X_T) - L^{-1}[f](x) \right) - \frac{1}{\sqrt{T}} \int_0^T \nabla L^{-1}[f](X_s)^\top \sigma(X_s) \, \mathrm{d}W_s. \qquad (3.1)$$

Observe that $\sup_{f \in \mathcal{F}} \|L^{-1}[f]\|_\infty < \infty$ so the first term is $O_P(T^{-1/2})$, and is thus asymptotically negligible. Now let T_n be any sequence increasing to infinity. Define processes

$$M_t^n := \frac{1}{\sqrt{T_n}} \int_0^{T_n t} \nabla L^{-1}[f](X_s)^\top \sigma(X_s) \cdot \mathrm{d}W_s, \quad t \geq 0.$$

As $L^{-1}[f] \in C^1$ by a further Sobolev embedding, each M^n is in fact a continuous martingale with quadratic variation given by

$$\langle M^n \rangle_t = \frac{1}{T_n} \int_0^{T_n t} \left\| \sigma(X_s)^\top \nabla L^{-1}[f](X_s) \right\|_2^2 \, \mathrm{d}s \xrightarrow{P} t \| \sigma^\top \nabla L^{-1}[f] \|_\mu^2$$

as $n \to \infty$, using the ergodic average property (1.3), which is proved for $f \in \mathcal{F}$ as in [23, Lemma 6]. Then by the martingale central limit theorem ([13, Theorem 7.1.4]), as $T \to \infty$ we have that

$$\frac{1}{\sqrt{T}} \int_0^T \nabla L^{-1}[f](X_s)^\top \sigma(X_s) \cdot \mathrm{d}W_s \to^{\mathcal{L}} N \left(0, \| \sigma^\top \nabla L^{-1}[f] \|_\mu^2 \right),$$

and so by (3.1) and linearity, we have for any $k \in \mathbb{N}, f_1, \ldots, f_k \in \mathcal{F}$, that as $T \to \infty$, $(\mathbb{G}_T(f_1), \ldots, \mathbb{G}_T(f_m)) \to^{\mathcal{L}} N_m(0, C)$, where the covariance matrix C is given by $C_{ij} = \left\langle \sigma^\top \nabla L^{-1}[\bar{f}_i], \sigma^\top \nabla L^{-1}[\bar{f}_j] \right\rangle_\mu$. Thus the finite-dimensional marginals of $\mathbb{G}_T$ converge to those of $\mathbb{G}$ as defined by (2.9).

3.2 Asymptotic equicontinuity

It now remains to prove that the asymptotic equicontinuity condition (2.8) holds for $\mathbb{G}_T$, for some pseudo-distance ρ with respect to which $\mathcal{F}$ is totally bounded. By Markov's inequality, it is sufficient to prove that

$$\lim_{\delta \to 0} \limsup_{T \to \infty} E \left[\sup_{f,g \in \mathcal{F}: \rho(f,g) \leq \delta} \left| \mathbb{G}_T(f) - \mathbb{G}_T(g) \right| \right] = 0. \qquad (3.2)$$

We first address the choice of pseudo-distance. Recall the decomposition of $\mathbb{G}_T(f)$ in (3.1):

$$\mathbb{G}_T(f) = \frac{1}{\sqrt{T}} \underbrace{\left(L^{-1}[f](X_T) - L^{-1}[f](x) \right)}_{=:Y_T(f)} - \frac{1}{\sqrt{T}} \underbrace{\int_0^T \nabla L^{-1}[f](X_s)^\top \sigma(X_s) \, \mathrm{d}W_s}_{=:Z_T(f)}$$

where for fixed $f \in \mathcal{F}$, $(Z_t(f))_{t \geq 0}$ is a continuous local martingale. As noted previously, $T^{-1/2} Y_T(f)$ is asymptotically negligible and it suffices to prove the asymptotic equicontinuity condition (3.2) with $\frac{1}{\sqrt{T}} Z_T$ in place of $\mathbb{G}_T$.

The quadratic variation of $Z(f)$ can be bounded above:

$$\langle Z(f) \rangle_T = \int_0^T \| \sigma(X_s)^\top \nabla L^{-1}[f](X_s) \|^2 \, \mathrm{d}s \leq \Lambda \int_0^T \| \nabla L^{-1}[f](X_s) \|^2 \, \mathrm{d}s,$$

where $\Lambda < \infty$ is as in (2.5). Thus from the argument of Lemma 1 in [23], the process $\left(T^{-1/2} Z_T(f) \right)_{f \in \mathcal{F}}$ is sub-Gaussian with respect to the pseudo-distance ρ_L given by

$$\rho_L^2(f,g) = \Lambda \sum_{i=1}^d \left\| \partial_i L^{-1}[f-g] \right\|_\infty^2. \qquad (3.3)$$

The next lemma, similar to Lemma 3 of [23], establishes a useful upper bound for ρ_L.

Lemma 3.1. *Let $b, \sigma \in C^{1+\beta}(\mathbb{T}^d)$ for some $\beta > 0$. Then for any non-integer γ such that $0 < \gamma \leq \beta$,*

$$\rho_L(f, g) \lesssim \|f - g\|_{B_{\infty\infty}^{-1+\gamma}},$$

where the constant depends on $b, \sigma, d, \beta, \gamma$.

Thus $T^{-1/2}Z_T$ is also sub-Gaussian with respect to (a multiple of) the $B_{\infty\infty}^{-1+\gamma}$-norm, for arbitrarily small $\gamma > 0$; we now confirm (3.2) with $T^{-1/2}Z_T$ in place of $\mathbb{G}_T$, with respect to the $B_{\infty\infty}^{-1+\gamma}$-norm. Note that $\mathcal{F}$ is totally bounded with respect to this norm so long as $\gamma \leq s + 1$, which we assume. For any pseudo-distance ρ, let $N(\mathcal{F}, \rho, \varepsilon)$ denote the covering number of $\mathcal{F}$ by ρ-balls of radius ε. By Dudley's chaining bound for sub-Gaussian processes (see e.g. [17, Theorem 2.3.7]), if

$$\int_0^\infty \sqrt{\log N(\mathcal{F}, \|\cdot\|_{B_{\infty\infty}^{-1+\gamma}}, \varepsilon)}\, \mathrm{d}\varepsilon < \infty, \tag{3.4}$$

then

$$E\left[\sup_{f,g\in\mathcal{F}:\|f-g\|_{B_{\infty\infty}^{-1+\gamma}}\leq\delta} \frac{1}{\sqrt{T}}\big|Z_T(f) - Z_T(g)\big|\right] \lesssim \int_0^\delta \sqrt{\log 2N(\mathcal{F}, \|\cdot\|_{B_{\infty\infty}^{-1+\gamma}}, \varepsilon)}\, \mathrm{d}\varepsilon,$$

where the constant is universal. In view of the finiteness of the entropy integral in (3.4), the right-hand side converges to zero as $\delta \to 0$. Since the constants in the previous display do not depend on T, it follows that (3.4) implies (3.2) for $T^{-1/2}Z_T$, and thus for $\mathbb{G}_T$. To prove (3.4), we make use of the following covering number bounds.

Lemma 3.2. *Let $1 \leq p, p' \leq \infty$ and let $s, t \in \mathbb{R}$ be such that $s - t > \max\left(\frac{d}{p} - \frac{d}{p'}, 0\right)$. Let $\mathcal{B} = \mathcal{B}(s, p, M)$ be the norm-ball of radius M in $B_{p\infty}^s(\mathbb{T}^d)$. Then $\mathcal{B}$ satisfies*

$$\log N(\mathcal{B}, \|\cdot\|_{B_{p'1}^t}, \varepsilon) \simeq \left(\frac{M}{\varepsilon}\right)^{\frac{d}{s-t}} \quad \forall \varepsilon > 0,$$

where the constants depend on p, p', s, t, d.

Remark 3.3. Using the embeddings $B_{pq}^s \subset B_{p\infty}^s$ and $B_{p'1}^t \subset B_{p'q'}^t$ for any $1 \leq q, q' \leq \infty$, the lemma also holds with any choices of third Besov index.

We now bound the entropy integral: letting D be the $\|\cdot\|_{B_{\infty\infty}^{-1+\gamma}}$-diameter of $\mathcal{F}$ (which is certainly finite), for $\varepsilon > D$ the integrand in (3.4) is zero and so the range of integration is effectively $[0, D]$. Since $\mathcal{F} \subset B_{pq}^s$ is bounded, we may then apply Lemma 3.2 with $t = -1 + \gamma, p' = \infty$ (here we require the assumption $s > d/p - 1$) to obtain that

$$\int_0^D \sqrt{\log N(\mathcal{F}, \|\cdot\|_{B_{\infty\infty}^{-1+\gamma}}, \varepsilon)}\, \mathrm{d}\varepsilon \lesssim \int_0^D \varepsilon^{-\frac{d}{2(1+s-\gamma)}}\, \mathrm{d}\varepsilon;$$

this integral is finite if and only if $\frac{d}{2(1+s-\gamma)} < 1$, or equivalently $d < 2(s + 1 - \gamma)$. Since we may choose $\gamma > 0$ arbitrarily small, the entropy integral (3.4) can be made finite whenever $d < 2(s + 1)$, or equivalently (recalling that $d \geq 2$), whenever $s > \frac{d}{2} - 1$.

4 Additional proofs

Proof of Proposition 2.4. Without loss of generality, we may assume that $\mathcal{F} = \mathcal{F} \cap L_\mu^2$.

We first consider the case $s > \frac{d}{2} - 1$. Recall the *intrinsic pseudo-distance* $\rho_\mathbb{G}$ induced by the Gaussian process $\mathbb{G}$, defined by $\rho_\mathbb{G}^2(f, g) = E(\mathbb{G}(f) - \mathbb{G}(g))^2$. Then $\mathbb{G}$ is sub-Gaussian with respect to $\rho_\mathbb{G}$ (see, for example, the discussion after Definition 2.3.5 in [17]). By using the equivalence of μ to Lebesgue density, the strict ellipticity of σ,

and Lemma A.1, we have that for $f, g \in L^2_\mu$, $\rho_{\mathbb{G}}(f, g) \leq C(b, \sigma, d) \|f - g\|^2_{H^{-1}}$. Thus $\mathbb{G}$ is sub-Gaussian with respect to the H^{-1}-norm; to prove tightness of $\mathbb{G}$ as a measure on $\ell^\infty(\mathcal{F})$, it suffices to check condition (3.4) with $\|\cdot\|_{H^{-1}}$ in place of $\|\cdot\|_{B^{-1+\gamma}_{\infty\infty}}$. For any values of p, q we may apply the covering number bound Lemma 3.2 to obtain that

$$\int_0^\infty \sqrt{\log N(\mathcal{F}, \|\cdot\|_{H^{-1}}, \varepsilon)} \, \mathrm{d}\varepsilon \lesssim \int_0^D \varepsilon^{-\frac{d}{2(1+s)}} \, \mathrm{d}\varepsilon,$$

where $D > 0$ is the H^{-1}-diameter of $\mathcal{F}$. This integral is finite since $s > \frac{d}{2} - 1$, confirming condition (3.4). Thus $\mathbb{G}$ is a tight Borel measure on $\ell^\infty(\mathcal{F})$.

Next, we consider the case $s < \frac{d}{2} - 1$; then for some $\delta > 0$, we have $s = \frac{d}{2} - 1 - \delta$. Clearly it suffices to consider the case where $p = \infty$ and $\delta > 0$ arbitrarily small. By Sudakov's Theorem [17, Corollary 2.4.13], if

$$\liminf_{\varepsilon \to 0} \varepsilon \sqrt{\log N(\mathcal{F}, \rho_{\mathbb{G}}, \varepsilon)} = \infty, \tag{4.1}$$

then $\sup_{f \in \mathcal{F}} |\mathbb{G}(f)| = \infty$ a.s., and so $\mathbb{G}$ is not sample bounded; in particular, it is not tight.

Using the strict ellipticity assumption on $\sigma\sigma^\top$ (2.5), we have that for $f, g \in \mathcal{F}$,

$$\rho^2_{\mathbb{G}}(f, g) = \left\| \sigma^\top \nabla L^{-1}[f - g] \right\|^2_\mu \gtrsim \left\| \nabla L^{-1}[f - g] \right\|^2_\mu. \tag{4.2}$$

Now, $L^{-1}[f - g] \in L^2_\mu$, and so the Poincaré inequality ([16, Chapter 7]) with reference density μ yields that $\|L^{-1}[f - g]\|_\mu \lesssim \|\nabla L^{-1}[f - g]\|_\mu$. Hence, again using the equivalence of $\|\cdot\|_{L^2}$ and $\|\cdot\|_\mu$,

$$\|L^{-1}[f - g]\|^2_{H^1} = \|L^{-1}[f - g]\|^2_{L^2} + \|\nabla L^{-1}[f - g]\|^2_{L^2} \lesssim \|\nabla L^{-1}[f - g]\|^2_{L^2}.$$

In combination with (4.2) and the fact that $L : H^1 \to H^{-1}$ is a continuous linear operator (see [5], Part II, Lemma 3.9), we deduce that $\rho_{\mathbb{G}}(f, g) \gtrsim \|f - g\|_{H^{-1}}$. It therefore suffices to prove (4.1) with $\|\cdot\|_{H^{-1}}$ in place of $\rho_{\mathbb{G}}$. Recall that $\mathcal{F}$ is the unit ball of B^s_{pq} (intersected with L^2_μ); we may assume that $s > 0$ and then apply Lemma 3.2 to obtain that for $\varepsilon > 0$ sufficiently small, $\log N(\mathcal{F}, \|\cdot\|_{H^{-1}}, \varepsilon) \gtrsim \varepsilon^{-\frac{d}{s+1}}$, and so using that $s = d/2 - 1 - \delta$ for some $\delta > 0$,

$$\varepsilon \sqrt{\log N(\mathcal{F}, \|\cdot\|_{H^{-1}}, \varepsilon)} \gtrsim \varepsilon^{1 - \frac{d}{2(s+1)}} = \varepsilon^{-\frac{2\delta}{d - 2\delta}}.$$

This quantity diverges as $\varepsilon \to 0$, establishing (4.1) when $s < \frac{d}{2} - 1$.

The case $s = \frac{d}{2} - 1$ follows a slightly more refined version of the argument for $s < \frac{d}{2} - 1$, as outlined in Remark 12 of [21]; we do not give details here. $\qquad\square$

Proof of Lemma 3.1. For any $\gamma > 0$ such that $\gamma \notin \mathbb{N}$ and any function h, standard Besov embeddings give that $\sum_{i=1}^d \|\partial_i g\|^2_\infty \leq \|h\|^2_{C^1} \leq \|h\|^2_{C^{1+\gamma}} \lesssim \|h\|^2_{B^{1+\gamma}_{\infty\infty}}$, for a constant depending on d, γ only. Thus $\rho_L(f, g) \lesssim \|L^{-1}[f - g]\|^2_{B^{1+\gamma}_{\infty\infty}}$. The result then follows from an application of Lemma A.1. $\qquad\square$

Proof of Lemma 3.2. By a scaling argument, we need only consider the case $M = 1$. Recall the *entropy numbers* $e_k, k \geq 1$ of $\mathcal{B}$: e_k is defined to be the smallest radius r such that $\mathcal{B}$ can be covered by at most 2^k ρ-balls of radius r.

Theorem 2 in [12, Section 3.3] establishes that for $\mathcal{B}$ as in the statement of the lemma, the entropy numbers e_k satisfy $e_k \simeq k^{-\frac{s-t}{d}}$, with constants as stated. The result now follows from the fundamental relationship between entropy numbers and covering numbers (see Exercise 2.5.5 and the preceding discussion in [28]):

$$e_k \simeq k^\alpha \Leftrightarrow \log N(\mathcal{B}, \rho, \varepsilon) \simeq \varepsilon^{1/\alpha}. \qquad\square$$

A   A PDE regularity estimate

For a more detailed exposition of the PDE theory of the generator L as defined in (2.2) and its adjoint L^*, we refer to [23, Section 6] and [5, Chapter II.3].

Consider the Poisson equation

$$Lu = f \quad \text{in } \mathbb{T}^d \tag{A.1}$$

where $f \in L^2(\mathbb{T}^d)$ is given. A key property of the solution operator L^{-1} defined by (2.6) is that it is '2-smoothing': see, for example, [5, Chapter 3, Theorem 3]. We require a quantitative version of this property, similar to equation (74) in [23] or Lemma 8 in [18]. We reformulate these results slightly for more general Besov spaces.

Lemma A.1. *Let $t \geq 2$. Assume that $b \in C^{t-2}(\mathbb{T}^d)$, and that $\sigma \in C^{t-2}(\mathbb{T}^d)$ is such that the strict ellipticity condition (2.5) is satisfied with constant $\lambda > 0$. Then for any $f \in L^2_\mu(\mathbb{T}^d)$, there exists a unique solution $L^{-1}[f] \in L^2_0(\mathbb{T}^d)$ of (A.1) such that $LL^{-1}[f] = f$ almost everywhere. Moreover, for any $s \leq t$ and any $1 \leq p, q \leq \infty$,*

$$\left\| L^{-1}[f] \right\|_{B^s_{pq}} \lesssim \|f\|_{B^{s-2}_{pq}},$$

where the constant depends on s, t, p, q, d, λ and an upper bound for $\|b\|_{B^t_{\infty\infty}}$.

The lemma is proved exactly as in Section 6.0.1 of [23], with the generalisation of the Besov indices requiring only cosmetic changes, once one obtains the estimate

$$\|u\|_{B^s_{pq}} \lesssim \|Au\|_{B^{s-2}_{pq}} \quad \forall u \in B^s_{pq} \cap L^2_0, \tag{A.2}$$

where A is the second-order differential operator given by $A = \sum_{i,j=1}^d a_{ij}(\cdot)\partial_i\partial_j$. This replaces an analogous estimate in [23] for the case $A = \Delta$, the Laplacian. One proves the estimate (A.2) as in that reference, substituting the Fourier coefficient identity $\langle \Delta u, e_k \rangle = -(2\pi)^2 |k|^2 \langle u, e_k \rangle$ with the inequality $|\langle Au, e_k \rangle| \geq (2\pi)^2 \lambda |k|^2 |\langle u, e_k \rangle|$, which holds for $k \in \mathbb{Z}^d \setminus \{0\}$. Plugging this into a Littlewood-Paley definition of the B^s_{pq}-norm and using a Fourier multiplier argument then yields (A.2), and thus Lemma A.1.

References

[1] Cathrine Aeckerle-Willems and Claudia Strauch, *Sup-norm adaptive drift estimation for multivariate nonreversible diffusions*, The Annals of Statistics **50** (2022), no. 6, 3484–3509. MR4524505

[2] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings*, Combinatorica **4** (1984), no. 4, 259–264. MR0779885

[3] Randolf Altmeyer, *Fourier Methods for Estimating Integrated Volatility and Occupation Time Functionals*, Ph.D. thesis, Humboldt University of Berlin, 2017.

[4] Richard F. Bass, *Stochastic processes / Richard F. Bass.*, Cambridge Series on Statistical and Probabilistic Mathematics, no. 33, University Press, Cambridge, 2011. MR2856623

[5] Lipman Bers, Fritz John, and Martin Schechter, *Partial differential equations*, Lectures in Applied Mathematics; v. 3, Interscience, New York; London, 1964. MR0163043

[6] R. N. Bhattacharya, *On the functional central limit theorem and the law of the iterated logarithm for Markov processes*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **60** (1982), no. 2, 185–201. MR0663900

[7] Emmanuel Boissard, *Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance*, Electronic Journal of Probability **16** (2011), 2296–2333. MR2861675

[8] Arnak S. Dalalyan and Markus Reiss, *Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case*, Probability Theory and Related Fields **137** (2007), no. 1-2, 25–47. MR2278451

[9] Eustasio del Barrio, Evarist Gine, and Carlos Matran, *Central limit theorems for the Wasserstein distance between the empirical and the true distributions*, The Annals of Probability **27** (1999), no. 2, 1009–1071. MR1698999

[10] R. M. Dudley, *The speed of mean Glivenko-Cantelli convergence*, Annals of Mathematical Statistics **40** (1968), 40–50. MR236977

[11] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, October 2002. MR1932358

[12] D. E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers, Differential Operators*, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1996. MR1410258

[13] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York; Chichester, 1986. MR0838085

[14] Hans Föllmer and Philip Protter, *On Itô s formula for multidimensional Brownian motion*, Probability Theory and Related Fields **116** (2000), no. 1, 1–20. MR1736587

[15] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields **162** (2015), no. 3-4, 707–738. MR3383341

[16] David Gilbarg and Neil S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer Science & Business Media, January 2001. MR1814364

[17] Evarist Giné and Richard Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series on Statistical and Probabilistic Mathematics, no. 40, University Press, Cambridge, 2016. MR3588285

[18] Matteo Giordano and Kolyan Ray, *Nonparametric Bayesian inference for reversible multidimensional diffusions*, The Annals of Statistics **50** (2022), no. 5, 2872–2898. MR4500628

[19] Dorothee D. Haroske and Hans Triebel, *Wavelet bases and entropy numbers in weighted function spaces*, Mathematische Nachrichten **278** (2005), no. 1-2, 108–132. MR2111803

[20] N. V. Krylov, *Controlled Diffusion Processes*, Applications of Mathematics, vol. 14, Springer-Verlag, New York-Berlin, 1980. MR601776

[21] Richard Nickl, *Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation*, Journal of the European Mathematical Society **22** (2020), no. 8, 2697–2750. MR4118619

[22] Richard Nickl and Benedikt M. Pötscher, *Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type*, Journal of Theoretical Probability **20** (2007), no. 2, 177–199. MR2324525

[23] Richard Nickl and Kolyan Ray, *Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions*, Annals of Statistics **48** (2020), no. 3, 1383–1408. MR4124327

[24] J. R. Norris, *Long-time behaviour of heat flow: Global estimates and exact asymptotics*, Archive for Rational Mechanics and Analysis **140** (1997), no. 2, 161–195. MR1482931

[25] E. Pardoux and Yu. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Annals of Probability **29** (2001), no. 3, 1061–1085. MR1872736

[26] F Russo and P Vallois, *Ito formula for cl-functions of semimartingales*, Probability Theory and Related Fields **104** (1996), no. 1, 15. MR1367665

[27] Claudia Strauch, *Adaptive invariant density estimation for ergodic diffusions over anisotropic classes*, The Annals of Statistics **46** (2018), no. 6B, 3451–3480. MR3852658

[28] Michel Talagrand, *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*, Springer Science & Business Media, February 2014. MR3184689

[29] Aad van der Vaart and Harry van Zanten, *Donsker theorems for diffusions: Necessary and sufficient conditions*, The Annals of Probability **33** (2005), no. 4, 1422–1451. MR2150194

[30] Aad van der Vaart and Jon Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Science & Business Media, March 1996. MR1385671