Bernoulli **1(1/2)**, 1995, 081–123

Time-scales for Gaussian approximation and its breakdown under a hierarchy of periodic spatial heterogeneities

RABI N. BHATTACHARYA[1*] and FRIEDRICH GÖTZE[2]

[1]*Department of Mathematics, Indiana University Bloomington, Bloomington IN 47405, USA*
[2]*Fakultät für Mathematik, Universität Bielefeld, Universitätsstrasse 25, 33501 Bielefeld, Germany*

The solution of the Itô equation $\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{b}\{\boldsymbol{X}(t)\}\mathrm{d}t + \boldsymbol{\beta}\{\boldsymbol{X}(t)/a\}\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(t)$ is analysed for $t \to \infty, a \to \infty$. In the range $1 \ll t \ll a^{2/3}$, $\boldsymbol{X}(t)$ is asymptotically Gaussian if $\boldsymbol{b}$ is periodic, $\boldsymbol{\beta}$ Lipschitzian; here the large-scale fluctuations may be ignored. In the range $t \gg a^2$, with both $\boldsymbol{b}$ and $\boldsymbol{\beta}$ periodic and divergence-free, a integral, Gaussian approximation is again valid under an appropriate hypothesis on the geometry of $\boldsymbol{\beta}$; here for some coordinates of $\boldsymbol{X}(t)$ the dispersivity, or variance per unit time, may grow at the extreme rate $O(a^2)$ while stabilizing for others. As shown by examples, Gaussian approximation generally breaks down at intermediate time-scales. These results translate into asymptotics of a class of Fokker–Planck equations which arise in the prediction of contaminant transport in an aquifer under multiple scales of spatial heterogeneity. In particular, contrary to popular belief, the growth in dispersivity is always slower than linear.

Keywords: diffusion processes; Gaussian limits; time-scales

1. Introduction

A physical law or description is in general valid only for a certain specific range of spatial and/or temporal scales. At different ranges of scales different laws are enunciated. The present study is motivated to a considerable extent by a problem in environmental engineering in which a hierarchy of spatial scales arises naturally. The paper may be viewed as an analysis by probabilistic methods of the (deterministic) Fokker–Planck equation that arises in this and similar contexts. Alternatively, it may be regarded as an asymptotic analysis of a class of stochastic differential equations with multiple spatial scales.

In Section 1.1 we provide an expository outline of the mathematical contents of the paper. Section 1.2 is devoted to the application mentioned above.

1.1. AN OUTLINE

Consider a class of stochastic differential equations on $\mathbb{R}^k, k > 1$, of the form

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{b}\{\boldsymbol{X}(t)\}\,\mathrm{d}t + \boldsymbol{\beta}\{\boldsymbol{X}(t)/a\}\,\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(t), \tag{1.1}$$

* To whom correspondence should be addressed.

1350-4864

where $\boldsymbol{b}, \boldsymbol{\beta}$ are non-constant vector fields with bounded first-order derivatives, a a large scalar parameter, $\boldsymbol{B}$ a standard k-dimensional Brownian motion independent of $\boldsymbol{X}(0)$, and $\boldsymbol{D}$ a $k \times k$ symmetric positive definite matrix whose positive definite square root is denoted $\sqrt{\boldsymbol{D}}$. One may think of (1.1) as the equation of motion of an object (e.g. a solute particle) moving in a velocity field $\boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{\beta}(\boldsymbol{x}/a)$, subject to a random perturbation. The position $\boldsymbol{X}(\cdot)$ of the object is a diffusion process whose transition probability density $p_a(t;\boldsymbol{x},\boldsymbol{y})$ satisfies Kolmogorov's backward equation

$$\frac{\partial p_a}{\partial t} = \boldsymbol{L} p_a \equiv \frac{1}{2}\sum_{j,j'} D_{jj'}\partial^2 p_a/\partial x_j \partial x_{j'} + \sum_j \{b_j(\boldsymbol{x}) + \beta_j(\boldsymbol{x}/a)\}\partial p_a/\partial x_j \tag{1.2}$$

and the forward equation

$$\frac{\partial p_a}{\partial t} = \boldsymbol{L}^* p_a \equiv \frac{1}{2}\sum_{j,j'} D_{jj'}\partial^2 p_a/\partial y_j \partial y_{j'} - \sum_j \partial/\partial y_j[\{b_j(\boldsymbol{y}) + \beta_j(\boldsymbol{y}/a)\}p_a]. \tag{1.3}$$

Imagine the object starting at a point $\boldsymbol{x}_0$ at time $t = 0$. Since the large-scale velocity $\boldsymbol{\beta}(\boldsymbol{x}/a)$ is nearly constant at positions close to $\boldsymbol{x}_0$, the object in motion will be governed for an initial period of time approximately by the equation

$$\mathrm{d}\boldsymbol{Y}(t) = \boldsymbol{b}\{\boldsymbol{Y}(t)\}\,\mathrm{d}t + \boldsymbol{\beta}\{\boldsymbol{x}_0/a\}\,\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(t), \qquad \boldsymbol{Y}(0) = \boldsymbol{x}_0, \tag{1.4}$$

where one may replace $\boldsymbol{\beta}(\boldsymbol{x}_0/a)$ by $\boldsymbol{\beta}(\boldsymbol{0})$ to get rid of a from equation (1.4). This initial period will of course depend on how large a is. In Theorem 3.1 in Section 3 it is shown, by an application of the Cameron–Martin–Girsanov theorem, that the L^1-distance between the transition probability densities of $\boldsymbol{X}$ and $\boldsymbol{Y}$ goes to zero if $t \ll a^{2/3}$, i.e. if $t/a^{2/3} \to 0$. Now if $\boldsymbol{Y}(t)$ is asymptotically Gaussian, it follows that $\boldsymbol{X}(t)$ has a valid Gaussian approximation with the same parameters in the range $1 \ll t \ll a^{2/3}$. In particular, this is the case if $\boldsymbol{b}(\cdot)$ is periodic. It is shown by examples in Section 2 that this Gaussian approximation is sometimes valid in the wider range $1 \ll t \ll a$. It is also seen from Example 2(i) and Remark 2.5 in Section 2 that the range of validity may not extend beyond $1 \ll t \ll a^{2/3}$ without the use of a centring adjustment for $\boldsymbol{X}(t)$. Whether such an adjustment is always possible is not known to us.

After an initial period as described above, the object begins to feel the effect of the fluctuations in the large-scale velocity field $\boldsymbol{\beta}(\cdot/a)$. The nature of these effects depends on $\boldsymbol{\beta}$, and from this point on, unless otherwise specified, we assume that $\boldsymbol{b}, \boldsymbol{\beta}$ are both periodic with a common period lattice. Without any essential loss of generality, we take the lattice to be $\mathbb{Z}^k$, i.e.

$$\boldsymbol{b}(\boldsymbol{x}+\boldsymbol{n}) = \boldsymbol{b}(\boldsymbol{x}), \boldsymbol{\beta}(\boldsymbol{x}+\boldsymbol{n}) = \boldsymbol{\beta}(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathbb{R}^k, \boldsymbol{n} \in \mathbb{Z}^k. \tag{1.5}$$

Also, $\boldsymbol{b}$ and $\boldsymbol{\beta}$ are divergence-free:

$$\sum_{j=1}^k \partial b_j(\boldsymbol{x})/\partial x_j \equiv 0 \equiv \sum_{j=1}^k \partial\beta_j(\boldsymbol{x})/\partial x_j. \tag{1.6}$$

Finally, assume that a is a positive integer.

Section 2 illustrates by examples the dramatic changes in behaviour that occur in the range $a \ll t \leq O(a^2)$. Generally speaking, the Gaussian approximation breaks down in the region between $O(a)$ and $O(a^2)$. If, for some special structure, $\boldsymbol{X}(t)$ becomes asymptotically Gaussian in some part of this range, as we see in a modification of Example 1 (see Remark 2.2, and the relation

(2.7) in Section 2), the scaling is most likely to be quite different from that in other ranges, indicating a 'phase change'.

It remains to analyse the behaviour of $\boldsymbol{X}(t)$ for $t \gg a^2$. Much effort (and all of Section 4) goes into this analysis. From the assumptions it follows that $\dot{\boldsymbol{X}}(\cdot) := \boldsymbol{X}(\cdot) \bmod a \equiv (X_1(\cdot) \bmod a, \ldots, X_k(\cdot) \bmod a)$ is a diffusion on the torus $\mathcal{T}_a := \{\boldsymbol{x} \bmod a : \boldsymbol{x} \in \mathbb{R}^k\}$ (see, for example, Bhattacharya and Waymire 1990, p. 518). In view of (1.6), the uniform distribution on $\mathcal{T}_a$ is the unique invariant probability for $\dot{\boldsymbol{X}}(\cdot)$. Further, for a fixed positive integer a, a central limit theorem holds (Bensoussan *et al.* 1978, Chapter 3; Bhattacharya 1985):

$$\frac{1}{\sqrt{t}}\{\boldsymbol{X}(t) - \boldsymbol{X}(0) - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})\} \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{K}), \tag{1.7}$$

where $\xrightarrow{\mathcal{L}}$ denotes convergence in law, i.e. weak convergence, $N(\boldsymbol{0}, \boldsymbol{K})$ is the normal distribution with mean vector $\boldsymbol{0}$ and dispersion matrix $\boldsymbol{K}$, and for any real- or vector-valued function f on $\mathbb{R}^k$,

$$\bar{f} := \int_{[0,1)^k} f(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}. \tag{1.8}$$

Also, the dispersion matrix $\boldsymbol{K}$ is given by

$$\boldsymbol{K} := a^{-k} \int_{[0,a)^k} \{\operatorname{grad} \boldsymbol{\gamma}(\boldsymbol{y}) - \mathbf{I}\} \boldsymbol{D} \{\operatorname{grad} \boldsymbol{\gamma}(\boldsymbol{y}) - \mathbf{I}\}^{\mathrm{T}}\, \mathrm{d}\boldsymbol{y}, \tag{1.9}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ is given by the unique mean-zero periodic (with period a) solutions γ_i of

$$\boldsymbol{L}\gamma_i(\boldsymbol{x}) = b_i(\boldsymbol{x}) - \bar{b}_i + \beta_i(\boldsymbol{x}/a) - \bar{\beta}_i \qquad 1 \le i \le k. \tag{1.10}$$

Here $\boldsymbol{L}$ is the backward operator defined in (1.2). To derive (1.7) one may use Itô's lemma (see, for example, Bhattacharya and Waymire 1990, p. 585, or Rogers and Williams 1987, pp. 60–62) to write

$$\begin{aligned}\boldsymbol{X}(t) - \boldsymbol{X}(0) - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}}) &\equiv \int_0^t [\boldsymbol{b}\{\dot{\boldsymbol{X}}(s)\} - \bar{\boldsymbol{b}} + \boldsymbol{\beta}\{\dot{\boldsymbol{X}}(s)/a\} - \bar{\boldsymbol{\beta}}]\, \mathrm{d}s + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t) \\ &= \boldsymbol{\gamma}\{\dot{\boldsymbol{X}}(t)\} - \boldsymbol{\gamma}\{\dot{\boldsymbol{X}}(0)\} - \int_0^t [\operatorname{grad} \boldsymbol{\gamma}\{\dot{\boldsymbol{X}}(s)\} - \mathbf{I}]\sqrt{\boldsymbol{D}}\, \mathrm{d}\boldsymbol{B}(s).\end{aligned} \tag{1.11}$$

The proof of (1.7) is completed by using the martingale central limit theorem (see, for example, Bhattacharya and Waymire 1990, pp. 513–515; Billingsley 1961; or Hall and Heyde 1980, pp. 58, 59). Much of this paper is concerned with the range and nature of validity of approximation (1.7) as $a \to \infty$, and we devote the next few paragraphs to this topic.

There is significant contact here with the theory of Diaconis (1988). The specific result that we use (Proposition 4.3) is an extension of a result of Fill (1991). (See also Diaconis and Stroock 1991.) Like Fill, we deal with a non-reversible Markov process $\dot{\boldsymbol{X}}(\cdot)$. But we have a continuous-parameter Markov process, and the state space is compact – not finite. In view of (1.6), the 'reversiblization' of $\boldsymbol{L}$ – in Fill's (1991) terminology – is the self-adjoint operator

$$\mathcal{D} := \frac{1}{2}\sum_{j,j'} D_{jj'}\, \partial^2/\partial x_i \partial x_j = \tfrac{1}{2}(\boldsymbol{L} + \boldsymbol{L}^*) \tag{1.12}$$

on $\mathcal{T}_a$. Using an adaptation of Fill's estimate, Trotter's product formula for semigroups, and a bound for the fundamental solution $p_a(t; \boldsymbol{x}, \boldsymbol{y})$ (see, for example, Aronson 1967), it is shown (see

(4.48)) that

$$\sup_{x} \int_{[0,a)^k} |\dot{p}_a(t;x,y) - a^{-k}|\,\mathrm{d}y \le c\exp(-2\pi^2\alpha t/a^2), \tag{1.13}$$

where $\dot{p}_a$ is the transition probability density of $\dot{X}(\cdot), a^{-k}$ its invariant density, and c does not depend on a.

Although (1.13) asserts weak dependence of $\dot{X}(\cdot)$ at time scales $t \gg a^2$ in the sense that the strong mixing rate goes to zero, much additional work is necessary to derive a meaningful Gaussian approximation. To describe this work, it is useful and natural to consider the rescaled process $Y(t) := X(a^2t)/a$, satisfying

$$Y(t) = Y(0) + \int_0^t a[b\{aY(s)\} + \beta\{Y(s)\}]\,\mathrm{d}s + \sqrt{D}\bar{B}(t), \tag{1.14}$$

where $\bar{B}(t) := B(a^2t)/a$ is again a standard Brownian motion. Since $b(a\cdot)$ and β are both periodic with period 1, $\ddot{Y}(t) := Y(t) \bmod 1$ is a diffusion on the unit torus $\mathcal{J}_1 := \{x \bmod 1 : x \in \mathbb{R}^k\}$, and it has Lebesgue measure on $\mathcal{J}_1$ as its unique invariant probability. The generator of this diffusion $\ddot{Y}(\cdot)$ is $\mathrm{A}_a := \mathcal{D} + a\{b(a\cdot) + \beta\}\cdot\nabla$. Deriving a Gaussian approximation for $X(t)$ for $t \gg a^2$ is equivalent to deriving it for $Y(t)$ for $t \gg 1$. Now for a *fixed* positive integer a, it follows using Itô's lemma that $\{Y_j(t) - Y_j(0) - at(\bar{b}_j + \bar{\beta}_j)\}/\sqrt{t}$ converges in distribution to $N(0, a^2\|g_j\|_1^2 + D_{jj})$ as $t \to \infty$ (see Bensoussan *et al.* 1978; or Bhattacharya 1982; 1985). Here $g_j(x) = \gamma_j(ax)/a^2$ is the unique mean-zero periodic solution of $\mathrm{A}_a g_j = b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j$, and $\|\cdot\|_1$ is the norm in the complex Hilbert space

$$H^1 := \left\{h \in L^2(\mathcal{J}_1, \mathrm{d}x) : h \perp 1, \int_{\mathcal{J}_1} |\nabla h(x)|^2\,\mathrm{d}x < \infty\right\},$$
$$\langle h_1, h_2\rangle_1 := \frac{1}{2}\sum_{j,j'} D_{jj'} \int_{\mathcal{J}_1} \frac{\partial h_1(x)}{\partial x_j}\frac{\partial \bar{h}_2(x)}{\partial x_{j'}}\,\mathrm{d}x, \tag{1.15}$$

$\bar{h}_2$ being the complex conjugate of h_2. It is necessary to derive the rate of growth of $a^2\|g_j\|_1^2$ as $a \to \infty$ in order to scale $Y_j(t)$ properly and to ensure that the scaled process converges to a non-singular Gaussian as both $a \to \infty, t \to \infty$. In order to derive this growth rate one may think of the approximation of the operator A_a by $\bar{\mathrm{A}}_a := \mathcal{D} + a(\bar{b} + \beta)\cdot\nabla$. Since $b(a\cdot)$ is periodic with a very small period $1/a, b(a\cdot)$ converges weakly to $\bar{b}$ in $L^2(\mathcal{J}_1, \mathrm{d}x)$, as $a \to \infty$. Lemma 4.5 estimates the error of this approximation. An approximation to g_j is then the solution h_j to $\bar{\mathrm{A}}_a h_j = \beta_j - \bar{\beta}_j \simeq b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j$. The advantage of looking at $\bar{\mathrm{A}}_a$ and h_j is that one may then use the spectral method developed in Bhattacharya *et al.* (1989). In particular, it follows that $\|h_j\|_1^2 \to \|(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N\|_1^2$ as $a \to \infty$, where f_N denotes the projection of f onto the null space N in H^1 of the operator $\mathcal{D}^{-1}(\bar{b} + \beta)\cdot\nabla$. Thus if $(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N \neq \mathbf{0}$, then the proper scaling for the jth coordinate of $Y(t)$ is $\{Y_j(t) - Y_j(0) - at(\bar{b}_j + \bar{\beta}_j)\}/a\sqrt{t}$. That this works is based on Lemma 4.6 and the weak dependence estimates in Corollary 4.4. Theorem 4.7 can then be derived by an appeal to a result of Götze and Hipp (1983).

In the case $(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N = \mathbf{0}$ and $\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j)$ belongs to the range of $\mathcal{D}^{-1}(\bar{b} + \beta(\cdot))\cdot\nabla$, $\|g_j\|_1^2$ and $\|h_j\|_1^2 \to 0$ as $a \to \infty$. Indeed, Lemma 4.8 shows that in this case $a^2\|g_j\|_1^2$ remains bounded as $a \to \infty$. The proper norming is then given by $\{Y_j(t) - Y_j(0) - at(\bar{b}_j + \bar{\beta}_j)\}/\sqrt{t}$.

Theorem 4.9, providing the appropriate result in this case, does not follow directly from Götze and Hipp (1983). It is proved by the classical method of omitting relatively small blocks in between bigger blocks which are nearly independent, along with various estimates contained in the lemmas mentioned above. Theorem 4.10 combines the two preceding theorems, assuming that for some set J_1 of coordinate indices $(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N$, $j \in J_1$, are linearly independent elements of the null space N of $\mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}) \cdot \nabla$, and $\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j)$, $j \notin J_1$, are linearly independent elements of the range of $\mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}) \cdot \nabla$. Under this assumption it is relatively simple to deduce the following results from Theorem 4.10 (see Bhattacharya and Ranga Rao 1976, pp. 22–24):

$$\sup_{C \in \mathcal{C}} |P(\{\boldsymbol{X}(t) - \boldsymbol{x} - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})/\sqrt{t} \in C\}) - N(\boldsymbol{0}, \boldsymbol{K})(C)| \to 0 \qquad \text{as } a \to \infty, t/a^2 \to \infty, \tag{1.16}$$

uniformly with respect to all initial states $\boldsymbol{x}$. Here $\mathcal{C}$ is the class of all Borel measurable convex subsets of $\mathbb{R}^k$, and $\boldsymbol{K}$ is given by (1.9). The results of Götze and Hipp (1983; 1994) and Lahiri (1993) suggest that the left-hand side is likely to be bounded by $ca/\sqrt{t}$, where c is a constant depending only on $\boldsymbol{b}$, $\boldsymbol{\beta}$ and $\boldsymbol{D}$.

1.2. AN APPLICATION

Consider a Fokker–Planck equation of the form

$$\frac{\partial c(t, \boldsymbol{y})}{\partial t} = \boldsymbol{L}^* c, \qquad t > 0, \boldsymbol{y} \in \mathbb{R}^k, \tag{1.17}$$

subject to the initial condition

$$\lim_{t \downarrow 0} c(t, \boldsymbol{y}) = c_0(\boldsymbol{y}) \qquad \boldsymbol{y} \in \mathbb{R}^k, \tag{1.18}$$

where $\boldsymbol{L}^*$ is as in (1.3), and c_0 is a non-negative continuous function with compact support. Then $c(t, \boldsymbol{y})$ has the representation

$$c(t, \boldsymbol{y}) := \int_{\mathbb{R}^k} c_0(\boldsymbol{x}) p_a(t; \boldsymbol{x}, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}. \tag{1.19}$$

In view of this, the asymptotic properties of $c(t, \boldsymbol{y})$ for $t \to \infty$, $a \to \infty$, are identical with those of the distribution of $\boldsymbol{X}(t)$ discussed in Section 1.1, provided c_0 does not depend on a. Indeed, by multiplying c_0 by an appropriate constant if necessary, $c(t, \boldsymbol{y})$ may be considered to be the density of $\boldsymbol{X}(t)$, when $\boldsymbol{X}(0)$ has density c_0. Hence the analysis of this paper may also be viewed as that of the solution of (1.18) and (1.19). This is a purely mathematical application. On the other hand, such an equation governs the concentration of a contaminant or a solute in an aquifer saturated with water in motion in a velocity field $\boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{\beta}(\boldsymbol{x}/a)$. Therefore, the application to the problem of determining the spread of the contaminant with time is also immediate. Here c_0 represents an initial concentration due to a localized injection of contaminants into the aquifer. The scope of this application is, of course, limited by the model assumptions. But the present analysis provides significant insights into the broader physical problem. We discuss this in the next paragraphs.

In order to clarify the general notion of a hierarchy of Gaussian approximations and their associated temporal and spatial scales, we begin with Einstein's (1905–8) diffusion equation and

Taylor's (1953) study of solute dispersion in a long capillary. Einstein's equation

$$\frac{\partial c}{\partial t} = \tfrac{1}{2} D_0 \Delta c - \boldsymbol{v} \cdot \nabla c \qquad t > 0, \boldsymbol{y} \in \mathbb{R}^3, \tag{1.20}$$

governs the concentration $c(t, \boldsymbol{y})$ of a solute (in dilute concentration) in a homogeneous isotropic fluid with a constant, or zero, velocity $\boldsymbol{v}$. Here D_0 is the molecular diffusion coefficient and Δ is the Laplacean. This equation is derived from the fact that the total displacement of a solute molecule over a period of time t may be viewed as the sum of a large number of nearly independent displacements suffered by it because of its interaction with surrounding liquid molecules, provided t is large compared to the time-scale of molecular interactions. By the central limit theorem, this displacement is Gaussian, and its probability density satisfies (1.20). Integrating over the initial concentration of solute molecules one arrives at (1.20) at the hydrodynamical scales of space and time. Taylor's equation for the concentration of a solute immersed in a liquid flowing through a long circular capillary of cross-section radius a under a constant pressure applied at one end is given by

$$\begin{aligned} \frac{\partial c}{\partial t} &= \tfrac{1}{2} D_0 \Delta c - \frac{\partial}{\partial y_1}\left\{ u_0 \left(1 - \frac{y_2^2 + y_3^2}{a^2}\right) c \right\} \qquad t > 0, y_2^2 + y_3^2 < a^2, \\ \frac{\partial c}{\partial \nu} &= 0 \qquad y_2^2 + y_3^2 = a^2. \end{aligned} \tag{1.21}$$

Here y_1 is the (horizontal) coordinate along the capillary length, while y_2 and y_3 are the coordinates of the (vertical) cross-section; u_0 is the maximum velocity attained at the centre of the cross-section. The second equation in (1.21) is the no-flux boundary condition in the direction of the normal ν to the boundary. The first equation is locally (i.e. within a region of diameter negligible compared to a) the same as (1.20) with $\boldsymbol{v} = (v_1, 0, 0), v_1 = u_0\{1 - (y_2^2 + y_3^2)/a^2\}$. At times t large compared to the time for a solute molecule to reach the capillary boundary from the centre, the horizontal displacement becomes Gaussian, its dispersivity, i.e. variance or dispersion per unit time, being $D_0 + a^2 u_0^2 / 96 D_0$. A probabilistic derivation of Taylor's theory, as completed by Aris (1956), is given in Bhattacharya and Gupta (1984), based on the fact that the first coordinate of the Markov process $\boldsymbol{X}(t) = (X_1(t), X_2(t), X_3(t))$ whose transition probability density satisfies (1.21) has the representation

$$X_1(t) = X_1(0) + \int_0^t u_0 \left(1 - \frac{X_2^2(s) + X_3^2(s)}{a^2}\right) \mathrm{d}s + \sqrt{D_0} B_1(t). \tag{1.22}$$

Because $(X_2(s), X_3(s))$ is an ergodic Markov process on the disc $\{ y_2^2 + y_3^2 \le a^2\}$, the time integral in (1.22) becomes asymptotically Gaussian.

For solute motion in an aquifer, laboratory-scale experiments indicate that the solute concentration satisfies an equation such as (1.20), but with a D_0 much larger than the molecular diffusion coefficient (Fried and Combarnous 1971). Measurements at a larger spatial and temporal scale – the so-called Darcy scale – show a larger dispersivity than that at the laboratory scale (Fried and Combarnous 1971). At scales much larger than the Darcy scale yet larger dispersivities are observed (Molinary *et al.* 1977; Lallemond-Barres and Peaudecerf 1978; Anderson 1979). The usual way of measuring dispersivity and its growth is to inject a tracer substance into the aquifer at a point and monitor the concentration distribution at regular intervals. At each of these times a Gaussian is fitted to the observed distribution. The dispersivity in each direction is then plotted against time or against the spatial distance between the point of injection and the fitted mean. Some authors have devised concentration prediction formulae which are Gaussian at all times, with steadily increasing dispersivity (see, for example, Dagan 1984).

Neither physical measurements nor theoretical considerations support the view that there can be valid Gaussian approximations accompanied by a steady increase in dispersivity continuously in time. Going back to the Taylor example, the Gaussian approximation for $X_1(t)$ at the larger scale is due to the ergodicity of the horizontal velocity along the particle path, causing an increase in dispersivity. This dispersivity is not going to increase any further with time.

In our analysis Gaussian approximation holds at three scales. The first of these is at a spatial scale at which fluctuations in $\boldsymbol{b}(\cdot)$ are negligible (see (3.1)). At this scale the dispersivity is given by the matrix $\boldsymbol{D}$. The corresponding time-scale is determined by the time the particle can traverse only this small spatial region, and it is $o(\|\boldsymbol{D}\|^{1/2})$. The second scale is analysed in Section 3. Here the spatial scale is much larger than 1 but much smaller compared to a, and the time-scale is $o(a^{2/3})$ or possibly $o(a)$. The enhanced dispersivity is given by (3.15), and more explicitly by Examples 1(i), 2(i) in Section 2. The third spatial scale is larger than $O(a)$, and the time-scale larger than $O(a^2)$. This is analysed in Section 4. The enhanced dispersivities are indicated in Theorems 4.7 and 4.9, and in Examples 1(iv), 2(ii) in Section 2. In order to introduce even higher scales, one may simply rescale time and space and write an equation such as (3.1) at the last scale considered, with $\boldsymbol{D}$ replaced by the dispersivity at the third scale considered above. Such a notion of hierarchical Gaussian approximations at widely separated scales in an aquifer was introduced in Bhattacharya and Gupta (1983), and is rigorously analysed here for the case of periodic velocity fields. The 'divergence-free' condition (1.6) refers to the incompressibility of the fluid in a saturated aquifer. In between two successive scales in the hierarchy the Gaussian approximation generally breaks down, as shown by Example 1 (ii), (iii). The examples in Section 2 model the so-called layered media (Gupta and Bhattacharya 1986).

It would be an important and challenging task to extend the theory presented in this paper to velocity fields other than periodic ones. As a first step, a central limit theorem (CLT) is needed for a fixed a, or without the term $\beta(\boldsymbol{x}/a)$, for such velocity fields. One class of velocity fields which has received much attention is the class of ergodic random velocity fields. In the case where the diffusion matrix $\{(a_{ij}(\boldsymbol{x})) : \boldsymbol{x} \in \mathbb{R}^k\}$ is an ergodic random field and the generator of the diffusion is self-adjoint and expressible in the divergence from $\boldsymbol{L} = \frac{1}{2}\sum_{i,j}(\partial/\partial x_i)\,a_{ij}(\boldsymbol{x})\,\partial/\partial x_j$, Papanicolaou and Varadhan (1979) and Kozlov (1980) have derived a CLT for $\boldsymbol{X}(t)$. Results from Gelhar and Axness (1983) and Winter *et al.* (1984) provide strong indications that such a CLT may be valid also in the non-self-adjoint case of a constant diffusion matrix and an arbitrary Lipschitzian ergodic random velocity field. Finally, one may consider (deterministic) almost periodic velocity fields. Kozlov (1979) provides a CLT in the self-adjoint case with an almost periodic diffusion matrix $(a_{ij}(\boldsymbol{x}))$ and Δ in divergence form as above. Bhattacharya and Ramasubramanian (1988) prove a CLT in the non-self-adjoint almost periodic case assuming a 'resonance' condition on rationally independent frequencies of the almost periodic drift and diffusion coefficients.

The present study differs from earlier ones in that it looks at not one, but a hierarchy of Gaussian approximations based on multiple scales of spatial heterogeneities, and derives the time-scales for these approximations. One must introduce such a hierarchical structure of spatial heterogeneities in order to justify Gaussian approximations with enhanced dispersivities at such widely separated scales as the laboratory scale, the Darcy scale and the field scales. See also Sposito *et al.* (1986) on the existence of multiple scales in aquifers. Although the mathematical task of deriving such results for almost periodic $\boldsymbol{b}, \boldsymbol{\beta}$, or when $\boldsymbol{b}, \boldsymbol{\beta}$ are ergodic random fields seems formidable, the results in the periodic case provide valuable insights for more general situations.

We conclude with the remark that the growth in dispersivity with time (at those times where a Gaussian approximation is valid) in the periodic case is *sublinear* (see also Remark 2.6 in Section 2). This contrasts with the folklore of linear growth (see, for example, Molinary *et al.* 1977).

2. Examples

Example 1

In both sets of examples in this section the time t and the spatial scale parameter a go to infinity simultaneously. As the relative speeds with which t and a go to infinity change, the asymptotic behaviour of the diffusion undergoes dramatic phase changes. Consider a two-dimensional diffusion $\boldsymbol{X}(t) = (X_1(t), X_2(t))$ governed by the following stochastic differential equation with two scales of heterogeneity:

$$\begin{aligned} \mathrm{d}X_1(t) &= \{c_0 + c_1 \sin(2\pi X_2(t)) + c_2 \cos(2\pi X_2(t)/a)\}\,\mathrm{d}t + \mathrm{d}B_1(t), \\ \mathrm{d}X_2(t) &= \mathrm{d}B_2(t), \end{aligned} \tag{2.1}$$

where c_1, c_2 are non-zero constants, a is a 'large' positive integer, $\boldsymbol{B}(t) = (B_1(t), B_2(t)), t \geq 0$, a two-dimensional standard Brownian motion, and $\boldsymbol{X}(0) = (X_1(0), X_2(0))$ is independent of $\{\boldsymbol{B}(t) : t \geq 0\}$. The integral form of (2.1) is

$$\begin{aligned} X_1(t) &= X_1(0) + c_0 t + \int_0^t \{c_1 \sin(2\pi X_2(s)) + c_2 \cos(2\pi X_2(s)/a)\}\,\mathrm{d}s + B_1(t), \\ X_2(t) &= X_2(0) + B_2(t). \end{aligned} \tag{2.2}$$

We are interested in the asymptotic behaviour of $\boldsymbol{X}(t)$ for 'large' t and a. The following results are derived under 'Proofs' below by direct computation. For simplicity, let the initial $\boldsymbol{X}(0)$ be constant, $\boldsymbol{X}(0) = \boldsymbol{x} = (x_1, x_2)$.

Case (i): If $1 \ll t \ll a$, i.e. $t \to \infty, t/a \to 0$, then

$$\frac{X_1(t) - x_1 - c_0 t - c_2 t \cos(2\pi x_2/a)}{\sqrt{t}} \xrightarrow{\mathcal{L}} N\left(0, \frac{c_1^2}{2\pi^2} + 1\right) \tag{2.3}$$

uniformly for all $\boldsymbol{x} = (x_1, x_2)$.

Case (ii): If $a^{4/3} \ll t \ll a^2$, i.e. $t/a^{4/3} \to \infty, t/a^2 \to 0$, then

$$\frac{X_1(t) - x_1 - t(c_0 + c_2)}{t^2/a^2} \xrightarrow{\mathcal{L}} \mathcal{L}\left(-2c_2\pi^2 \int_0^1 B_2^2(s)\,\mathrm{d}s\right) \tag{2.4}$$

uniformly for all $\boldsymbol{x} = (x_1, x_2)$, with x_2 in a compact set. Here $\mathcal{L}(Z)$ denotes the law, or distribution, of Z.

Case (iii): If $t/a^2 \to \theta \in (0, \infty)$, then

$$\frac{X_1(t) - x_1 - c_0 t}{t} \xrightarrow{\mathcal{L}} \mathcal{L}\left(\frac{c_2}{\theta}\int_0^\theta \cos(2\pi B_2(s))\, \mathrm{d}s\right) \tag{2.5}$$

uniformly in $\boldsymbol{x}$ such that x_2 lies in a compact set.

Case (iv): If $t \gg a^2$, i.e. $t/a^2 \to \infty$, then

$$\frac{X_1(t) - x_1 - c_0 t}{a\sqrt{t}} \xrightarrow{\mathcal{L}} N\left(0, \frac{c_2^2}{2\pi^2}\right) \tag{2.6}$$

uniformly in $\boldsymbol{x}$.

Remark 2.1

In the range $1 \ll t \ll a$, the second term in the drift velocity, namely $\cos(2\pi x_2/a)$, is nearly constant; hence it does not contribute to the growth of $\boldsymbol{X}(t)$. In general (see Sections 3 and 4), one expects asymptotic normality at the two ends of the spectrum: $1 \ll t \ll a$, and $t \gg a^2$, but not in between.

Remark 2.2

If $\cos(2\pi x_2/a)$ is replaced by $\sin(2\pi x_2/a)$ in (2.1), then in cases (i) and (iii) cos is simply replaced by sin, while case (iv) remains unchanged. But case (ii) changes as follows:

Case (ii)': If $a \ll t \ll a^2$, then

$$\frac{X_1(t) - x_1 - c_0 t}{t^{3/2}/a} \xrightarrow{\mathcal{L}} N\left(0, \frac{4\pi^2 c_2^2}{3}\right) \tag{2.7}$$

uniformly for all $\boldsymbol{x}$.

Although a normal approximation holds here, the norming by $t^{3/2}/a$ is essentially different from those in cases (i) and (iii) since $t^{3/2}/a$ is of larger order than $t^{1/2}$ and of smaller order than t in this range. Thus here also a 'phase transition' occurs from one scale to another. As is implicit in the proofs below, the different natures of the limit laws in (2.4) and (2.7) arise because cos is an even function while sin is odd. Also, the appearance of the Gaussian law in (2.7) is due to the fact that $X_2(s)$ is Gaussian (namely, standard Brownian motion) at all scales. This is a very special situation.

Example 2

Here $\mathrm{d}X_1(t) = \{c_0 + c_1 \sin(2\pi X_2(t)) + c_2 \sin(2\pi X_2(t)/a)\}\,\mathrm{d}t + \mathrm{d}B_1(t)$. But $X_2(t)$ is a Brownian motion with a drift: $\mathrm{d}X_2(t) = \delta\,\mathrm{d}t + \mathrm{d}B_2(t), \delta \neq 0$.

Case (i): If $1 \ll t \ll a$, then

$$\frac{X_1(t) - x_1 - c_0 t - c_2\int_0^t \sin\left(2\pi\left(\frac{x_2 + s\delta}{a}\right)\right)\mathrm{d}s}{\sqrt{t}} \xrightarrow{\mathcal{L}} N\left(0, \frac{c_1^2}{2(\delta^2 + \pi^2)} + 1\right). \tag{2.8}$$

Case (ii): If $t \gg a^2$, then

$$\frac{X_1(t) - x_1 - c_0 t}{\sqrt{t}} \xrightarrow{\mathcal{L}} N\left(0, \frac{c_1^2}{2(\delta^2 + \pi^2)} + \frac{c_2^2}{2\delta^2} + 1\right). \tag{2.9}$$

Remark 2.3

In the range $1 \ll t \ll a$, the second term in the drift velocity is nearly constant and does not contribute to the dispersivity. In the range $t \gg a^2$, however, both terms contribute to dispersivity, unlike Example 1(iv). Also, the asymptotic variance per unit time, or dispersivity, is $O(1)$ in *both* ranges here.

Proofs

Example 1

Case (i): Since $t \to \infty$ and $t/a \to 0$,

$$\begin{aligned} &\mathrm{E}\left|\frac{1}{\sqrt{t}}\int_0^t c_2 \cos(2\pi X_2(s)/a)\,\mathrm{d}s - c_2\sqrt{t}\cos(2\pi x_2/a)\right| \\ &\le \frac{1}{\sqrt{t}}\int_0^t \frac{2\pi|c_2|}{a}\mathrm{E}\,|X_2(s) - x_2|\,\mathrm{d}s \le \frac{2\pi|c_2|}{a\sqrt{t}}\int_0^t \sqrt{s}\,\mathrm{d}s \\ &= \frac{4\pi|c_2|}{3}(t/a) \to 0. \end{aligned} \tag{2.10}$$

Therefore, the left-hand side of (2.3) has the same asymptotic distribution as that of

$$\frac{1}{\sqrt{t}}\int_0^t c_1 \sin(2\pi X_2(s))\,\mathrm{d}s + \frac{1}{\sqrt{t}}B_1(t). \tag{2.11}$$

The first term may be expressed by Itô's lemma (see, for example, Friedman 1975, p. 90; or Bhattacharya and Waymire 1990), p. 585) as

$$\frac{1}{\sqrt{t}}\left\{f(X_2(t)) - f(X_2(0)) - \int_0^t f'(X_2(s))\,\mathrm{d}B_2(s)\right\}, \tag{2.12}$$

where f satisfies

$$\tfrac{1}{2}f''(y) = c_1 \sin(2\pi y) \qquad \text{or } f''(y) = 2c_1\sin(2\pi y)$$

so that one may take, by integration,

$$f'(y) = -\frac{c_1}{\pi}\cos(2\pi y), \qquad f(y) = -\frac{c_1}{2\pi^2}\sin(2\pi y) \tag{2.13}$$

Using (2.13), (2.12) becomes

$$\frac{1}{\sqrt{t}}\left\{-\frac{c_1}{2\pi^2}(\sin(2\pi\{B_2(t) + x_2\}) - \sin(2\pi x_2))\right\} + \frac{1}{\sqrt{t}}\int_0^t \frac{c_1}{\pi}\cos(2\pi\{B_2(s) + x_2\})\,\mathrm{d}B_2(s). \tag{2.14}$$

The first term goes to zero almost surely, as $t \to \infty$, while the average quadratic variation of the stochastic integral is

$$\frac{c_1^2}{\pi^2 t}\int_0^t \cos^2(2\pi\{B_2(s)+x_2\})\,\mathrm{d}s. \tag{2.15}$$

The last expression converges in probability to

$$\frac{c_1^2}{\pi^2}\int_{[0,1)} \cos^2(2\pi y)\,\mathrm{d}y = \frac{c_1^2}{2\pi^2}, \tag{2.16}$$

in view of the fact that $\cos^2(2\pi X_2(s)) = \cos^2(2\pi\ddot{X}_2(s))$ (where $\ddot{y} = y \bmod 1$), and the diffusion $\ddot{X}_2(s)$ (Brownian motion on the unit circle) is a Markov process which has the uniform distribution as its unique invariant probability, and which satisfies Doeblin's condition. It then follows from the martingale central limit theorem, applied to the martingale $\{(c_1/\pi)\int_0^t \cos(2\pi X_2(s))\,\mathrm{d}B_2(s) : t \geq 0\}$, that the second term in (2.14) converges in distribution to $N(0, c_1^2/2\pi^2)$. Since $B_1(t)/\sqrt{t}$ is $N(0,1)$ and is independent of $\{B_2(s) : s \geq 0\}$, (2.3) follows.

Case (ii): Since $t^2/a^2 \gg t^{1/2}$ in this case, it follows from the calculations in case (i) that

$$\frac{1}{t^2/a^2}\int_0^t c_1 \sin(2\pi X_2(s))\,\mathrm{d}s \to 0 \text{ in probability.} \tag{2.17}$$

Hence it is enough to consider

$$\frac{1}{t^2/a^2}\int_0^t c_2\cos(2\pi X_2(s)/a)\,\mathrm{d}s \equiv \frac{c_2}{t^2/a^2}\int_0^t \cos(2\pi\{B_2(s)/a + x_2/a\})\,\mathrm{d}s. \tag{2.18}$$

Now $\{B_2(s)/a : s \geq 0\} \overset{\mathrm{d}}{=} \{B_2(s/a^2) : s \geq 0\}$, where $\overset{\mathrm{d}}{=}$ denotes equality in distribution. Therefore, (2.18) is equal in distribution to

$$\begin{aligned}\frac{c_2}{t^2/a^2}\int_0^t \cos(2\pi\{B_2(s/a^2) + x_2/a\})\,\mathrm{d}s &= \frac{c_2 a^2}{t^2/a^2}\int_0^{t/a^2} \cos\left(2\pi\left\{B_2(s') + \frac{x_2}{a}\right\}\right)\mathrm{d}s' \\ &\simeq \frac{c_2}{(t/a^2)^2}\int_0^{t/a^2} \cos[2\pi\{B_2(s')\}]\,\mathrm{d}s',\end{aligned} \tag{2.19}$$

where $\simeq$ indicates here, and elsewhere, that the difference between the two sides of the relation goes to zero in probability or pointwise. Now by a Taylor expansion,

$$\begin{aligned}\frac{1}{t^2/a^2}\int_0^t c_2\cos(2\pi B_2(s)/a)\,\mathrm{d}s &\overset{\mathrm{d}}{=} \frac{c_2}{t^2/a^2}\int_0^t \cos(2\pi B_2(s/a^2))\,\mathrm{d}s \\ &= \frac{c_2 a^2}{t^2/a^2}\int_0^{t/a^2} \cos(2\pi B_2(s'))\,\mathrm{d}s' \\ &= \frac{c_2 a^2}{t^2/a^2}\int_0^{t/a^2} (1 - 2\pi^2 B_2^2(s') + O(B_2^4(s')))\,\mathrm{d}s' \\ &= \frac{c_2}{t/a^2} - \frac{2c_2\pi^2}{(t/a^2)^2}\int_0^{t/a^2} B_2^2(s')\,\mathrm{d}s' + O_p\left(\frac{1}{(t/a^2)^2}\int_0^{t/a^2} s'^2\,\mathrm{d}s'\right).\end{aligned} \tag{2.20}$$

Now

$$\frac{1}{(t/a^2)^2}\int_0^{t/a^2}(s')^2\,\mathrm{d}s' = O(t/a^2) \to 0,$$

so that we need to consider, apart from the non-random first term in (2.20), only the second term. But, again by Brownian scaling, for all positive A one has

$$\frac{1}{A^2}\int_0^A B_2^2(s')\,\mathrm{d}s' \stackrel{\mathrm{d}}{=} \int_0^1 B_2^2(s)\,\mathrm{d}s. \tag{2.21}$$

Note that $\{(1/\sqrt{A})\,B_2(s') : s' \geq 0\} \stackrel{\mathrm{d}}{=} \{B_2(s'/A) : s' \geq 0\}$. Finally, $B_1(t)/(t^2/a^2) = o_p(1)$.

Case (iii): Here, once again from the calculations in case (i), the term in (2.2) involving $\sin(2\pi X_2(s))$ may be neglected, since $t \gg t^{1/2}$. The other term may be expressed as

$$\begin{aligned}
\frac{1}{t}\int_0^t c_2\cos(2\pi X_2(s)/a)\,\mathrm{d}s &\simeq \frac{1}{t}\int_0^t c_2\cos(2\pi B_2(s)/a)\,\mathrm{d}s \\
&\stackrel{\mathrm{d}}{=} \frac{1}{t}\int_0^t c_2\cos(2\pi B_2(s/a^2))\,\mathrm{d}s = \frac{a^2}{t}\int_0^{t/a^2} c_2\cos(2\pi B_2(s'))\,\mathrm{d}s' \\
&\to \frac{1}{\theta}\int_0^\theta c_2\cos(2\pi B_2(s'))\,\mathrm{d}s'.
\end{aligned} \tag{2.22}$$

Since $B_1(t)/t = o_p(1)$, (2.5) is proved.

Case (iv): As above, because of the large scaling factor $a\sqrt{t}$, the only term that needs to be considered is

$$\frac{1}{a\sqrt{t}}\int_0^t c_2\cos(2\pi X_2(s)/a)\,\mathrm{d}s = \frac{1}{a\sqrt{t}}\int_0^t \frac{1}{2}f''(X_2(s))\,\mathrm{d}s, \tag{2.23}$$

where

$$\tfrac{1}{2}f''(y) = c_2\cos(2\pi y/a), \tag{2.24}$$

so that one may take

$$f'(y) = \frac{c_2 a}{\pi}\sin(2\pi y/a), \qquad f(y) = -\frac{c_2 a^2}{2\pi^2}\cos(2\pi y/a). \tag{2.25}$$

By Itô's lemma, (2.23) becomes

$$\begin{aligned}
&\frac{1}{a\sqrt{t}}\left\{f(X_2(t)) - f(X_2(0)) - \int_0^t f'(X_2(s))\,\mathrm{d}B_2(s)\right\} \\
&\quad = \frac{1}{a\sqrt{t}}\left(-\frac{c_2 a^2}{2\pi^2}\right)\{\cos(2\pi X_2(t)/a) - \cos(2\pi x_2/a)\} \\
&\qquad - \frac{1}{a\sqrt{t}}\int_0^t \frac{c_2 a}{\pi}\sin(2\pi X_2(s)/a)\,\mathrm{d}B_2(s) \\
&\quad \simeq \left(-\frac{c_2}{\pi}\right)\frac{1}{\sqrt{t}}\int_0^t \sin(2\pi X_2(s)/a)\,\mathrm{d}B_2(s).
\end{aligned} \tag{2.26}$$

The average quadratic variation of the second integral is

$$\frac{1}{t}\int_0^t \sin^2(2\pi(B_2(s)+x_2)/a)\,\mathrm{d}s \stackrel{\mathrm{d}}{=} \frac{1}{t}\int_0^t \sin^2\left(2\pi\left(B_2\left(\frac{s}{a^2}\right)+\frac{x_2}{a}\right)\mathrm{d}s\right.$$

$$= \frac{1}{t/a^2}\int_0^{t/a^2} \sin^2\left(2\pi(B_2(s')+\frac{x_2}{a}\right)\mathrm{d}s'. \tag{2.27}$$

Since the Brownian motion $(B_2(s')+x_2/a) \bmod 1$ on the unit circle is ergodic and satisfies Doeblin's condition, and since $t/a^2 \to \infty$, the last expression in (2.27) converges in probability (uniformly with respect to x_2) to the phase average

$$\int_{[0,1)} \sin^2(2\pi y)\,\mathrm{d}y = \tfrac{1}{2}. \tag{2.28}$$

The proof of (2.6) is now complete by (2.26)–(2.28) and the martingale central limit theorem. □

Remark 2.4

One may avoid Itô's lemma in these and several other cases in this paper by using the fact that the normalized integral $t^{-1/2}\int_0^t g(U_s)\,\mathrm{d}s$ of a function g on the state space of an ergodic Markov process U_s, with invariant probability π and infinitesimal generator A on $L^2(\pi)$, is asymptotically normal $N(0,\sigma^2)$, provided g is in the range of A. Here $\sigma^2 = -2\langle f, g\rangle$, where f satisfies $\mathrm{A}f = g$ (Bhattacharya and Waymire 1990, p. 513; or Bhattacharya 1982).

Example 2

Case (i): First note that

$$\mathrm{E}\left|\frac{1}{\sqrt{t}}c_2\int_0^t \{\sin(2\pi X_2(s)/a) - \sin(2\pi(x_2+s\delta)/a)\}\,\mathrm{d}s\right|$$

$$\leq \frac{|c_2|}{\sqrt{t}}\int_0^t \mathrm{E}\,|B_2(s)/a|\,\mathrm{d}s \leq \frac{c_2}{a\sqrt{t}}\int_0^t s^{1/2}\,\mathrm{d}s = O\left(\frac{t}{a}\right) \to 0. \tag{2.29}$$

Thus

$$\frac{1}{\sqrt{t}}\left\{X_1(t) - x_1 - c_0 t - c_2\int_0^t \sin(2\pi(x_2+s\delta)/a)\right\}\mathrm{d}s$$

$$\simeq \frac{1}{\sqrt{t}}c_1\int_0^t \sin(2\pi X_2(s))\,\mathrm{d}s + \frac{1}{\sqrt{t}}B_1(t)$$

$$= \frac{1}{\sqrt{t}}c_1\int_0^t \sin(2\pi \ddot{X}_2(s))\,\mathrm{d}s + \frac{1}{\sqrt{t}}B_1(t). \tag{2.30}$$

Here $\ddot{X}_2(s) \equiv (B_2(s)+x_2+s\delta) \bmod 1$ is a diffusion on the circle whose generator is $\frac{1}{2}\mathrm{d}^2/\mathrm{d}x^2$. Therefore, by Remark 2.4, the last normalized integral is asymptotically $N(0,\sigma^2)$, where, writing $g(x) = \sin(2\pi x)$, one has

$$\sigma^2 = -2\langle f, g\rangle, \tag{2.31}$$

f being a solution of the equation $Af = g$, i.e.

$$\tfrac{1}{2}f''(x) + \delta f'(x) = \sin(2\pi x). \tag{2.32}$$

By integration one obtains

$$f'(x) = \frac{\delta}{\delta^2+\pi^2}\left(\sin(2\pi x) - \frac{\pi\cos(2\pi x)}{\delta}\right),$$
$$f(x) = -\frac{\delta}{\delta^2+\pi^2}\left(\frac{\cos(2\pi x)}{2\pi} + \frac{\sin(2\pi x)}{2\delta}\right). \tag{2.33}$$

The invariant measure on $[0,1)$ is again the uniform distribution, so that a direct integration yields

$$\sigma^2 = -2\int_{[0,1)} f(x)g(x)\,\mathrm{d}x = \frac{2\delta}{\delta^2+\pi^2}\left(\frac{1}{4\delta}\right) = \frac{1}{2(\delta^2+\pi^2)}. \tag{2.34}$$

From (2.30) assertion (2.8) follows.

Remark 2.5
The centring by $\int_0^t c_2 \sin(2\pi(x_2 + s\delta)/a)\,\mathrm{d}s$ takes into account the non-zero mean of $X_2(s)$. If we centred by $c_2 t \sin(2\pi x_2/a)$, as in Example 1(i), then the normal convergence would hold in the narrower region $1 \ll t \ll a^{2/3}$.

Case (ii): It turns out in this case that no term on the right-hand side of the expression

$$\frac{X_1(t) - x_1 - c_0 t}{\sqrt{t}} = \frac{1}{\sqrt{t}}\int_0^t \{c_1 \sin(2\pi X_2(s)) + c_2 \sin(2\pi X_2(s)/a)\}\,\mathrm{d}s + \frac{B_1(t)}{\sqrt{t}} \tag{2.35}$$

can be neglected. To see clearly the role of time-scale we use Itô's lemma here to write the normalized Riemann integral above as

$$\frac{f(X_2(t)) - f(X_2(0))}{\sqrt{t}} - \frac{1}{\sqrt{t}}\int_0^t f'(X_2(s))\,\mathrm{d}B_2(s), \tag{2.36}$$

where f is a solution of

$$\tfrac{1}{2}f''(x) + \delta f'(x) = c_1 \sin(2\pi x) + c_2 \sin(2\pi x/a). \tag{2.37}$$

On integration, as in case (i), we get

$$f'(x) = \frac{c_1\delta}{\delta^2+\pi^2}\left\{\sin(2\pi x) - \frac{\pi}{\delta}\cos(2\pi x)\right\}$$
$$+ \frac{c_2 a\delta}{\pi^2 + a^2\delta^2}\left\{a\sin(2\pi x/a) - \frac{\pi}{\delta}\cos(2\pi x/a)\right\} = I_1(x) + I_2(x), \tag{2.38}$$

say, and

$$f(x) = -\frac{c_1\delta}{\delta^2+\pi^2}\left\{\frac{\cos(2\pi x)}{2\pi} + \frac{\sin(2\pi x)}{2\delta}\right\} - \frac{c_2 a\delta}{\pi^2+a^2\delta^2}\left\{\frac{a^2}{2\pi}\cos(2\pi x/a) + \frac{a}{2\delta}\sin(2\pi x/a)\right\}. \tag{2.39}$$

Since $t \gg a^2$, the first term in (2.36) goes to zero in probability (uniformly in $x = (x_1, x_2) = X(0)$). The quadratic variation of the stochastic integral is given by

$$\int_0^t I_1^2\{X_2(s)\}\,ds + \int_0^t I_2^2\{X_2(s)\}\,ds + 2\int_0^t I_1\{X_2(s)\}I_2\{X_2(s)\}\,ds. \tag{2.40}$$

Under case (i) it was shown that time average of the first term in (2.40) converges in probability to $c_1^2/2(\delta^2 + \pi^2)$.

Next,

$$\frac{1}{t}\int_0^t I_2^2\{X_2(s)\}\,ds = \frac{1}{t}\int_0^t \left\{\frac{c_2 a^2 \delta}{\pi^2 + a^2\delta^2}\sin(2\pi X_2(s)/a) - \frac{c_2 \pi \delta a}{\pi^2 + a^2\delta^2}\cos(2\pi X_2(s)/a)\right\}^2 ds. \tag{2.41}$$

The second integral is of the order $O(1/a)$ and goes to zero. Hence

$$\begin{aligned}\frac{1}{t}\int_0^t I_2^2(X_2(s))\,ds &\simeq \left(\frac{c_2^2}{\delta^2}\right)\frac{1}{t}\int_0^t \sin^2(2\pi X_2(s)/a)\,ds \\ &\stackrel{d}{=} \left(\frac{c_2^2}{\delta^2}\right)\frac{1}{t}\int_0^t \sin^2\left(2\pi\left(B_2(s/a^2) + \frac{x_2 + s\delta}{a}\right)\right) ds \\ &= \left(\frac{c_2^2}{\delta^2}\right)\frac{a^2}{t}\int_0^{t/a^2} \sin^2\left(2\pi\left(B_2(s') + \frac{x_2}{a} + as'\delta\right)\right) ds'.\end{aligned} \tag{2.42}$$

Since the Brownian motion $(B_2(s') + y) \bmod 1$ on the unit circle has a transition probability density which approaches equilibrium (i.e. the uniform) density exponentially fast in L^1-norm, uniformly with respect to the initial state, it follows that the time average in (2.42) (recall $t/a^2 \to \infty$) approaches the equilibrium average, namely,

$$\int_{[0,1)} \sin^2(2\pi y)\,dy = \tfrac{1}{2} \tag{2.43}$$

in probability. Therefore,

$$\frac{1}{t}\int_0^t I_2^2\{X_2(s)\}\,ds \to \frac{c_2^2}{2\delta^2} \text{ in probability.} \tag{2.44}$$

It remains to consider the product term in (2.40). Clearly,

$$\begin{aligned}\frac{2}{t}\int_0^t I_1\{X_2(s)\}I_2\{X_2(s)\}\,ds &\simeq \frac{2}{t}\int_0^t I_1\{X_2(s)\}\frac{c_2}{\delta}\sin(2\pi X_2(s)/a)\,ds \\ &= \frac{2c_1c_2}{\delta^2 + \pi^2}\left(\frac{1}{t}\right)\int_0^t \left\{\sin(2\pi X_2(s)) - \frac{\pi}{\delta}\cos(2\pi X_2(s))\right\}\sin(2\pi X_2(s)/a)\,ds.\end{aligned} \tag{2.45}$$

Now the stochastic process $\{X_2(s)/a \equiv (B_2(s) + x_2 + s\delta)/a : s \geq 0\}$ has the same distribution as $\{B_2(s/a^2) + (x_2 + s\delta)/a : s \geq 0\}$. Therefore, by the time change $s' = s/a^2$, the average time integral in (2.45) (as a stochastic process for $t \geq 0$) has the same distribution as

$$\frac{a^2}{t}\int_0^{t/a^2}\left\{\sin\left(2\pi(aB_2(s') + x_2 + a^2 s'\delta)\right) - \frac{\pi}{\delta}\cos\left(2\pi(aB_2(s') + x_2 + a^2 s'\delta)\right)\right\}$$
$$\cdot \sin\left(2\pi\left(B_2(s') + \frac{x_2}{a} + as'\delta\right)\right) \mathrm{d}s'. \tag{2.46}$$

In the integrand one may now replace everywhere $B_2(s')$ by $\ddot{B}_2(s') \equiv B_2(s') \bmod 1$ – the Brownian motion on the unit circle. Note that in (2.45) we could only replace $B_2(s)$ by $\dot{B}_2(s) \equiv B_2(s) \bmod a$, but as a increases toward infinity the approach of $\dot{B}_2(s)$ to equilibrium on the circle $\mathcal{J}_a \equiv \{x \bmod a : x \in \mathbb{R}^1\}$ is quite slow. Now if h is a bounded (uniformly in a) measurable function on the unit circle, then, in view of the exponentially fast approach of $\ddot{B}_2(s)$ to equilibrium *uniformly* with respect to the initial state, one has

$$\sup_a \mathrm{E}\left|\frac{1}{T}\int_0^T h\{\ddot{B}_2(s)\}\,\mathrm{d}s - \int_{[0,1)} h(y)\,\mathrm{d}y\right| \to 0 \qquad \text{as } T \to \infty. \tag{2.47}$$

Therefore, (2.46) goes to zero, and we have proved that the normalized stochastic integral in (2.36) converges in distribution to $N(0, (c_1^2/2(\delta^2 + \pi^2)) + (c_2^2/2\delta^2))$. Hence the right-hand side of (2.35) converges to the normal law appearing in (2.9).

Remark 2.6

In all cases above where normal approximation holds, the dispersivity, i.e. the asymptotic variance of $X_1(t)$ per unit of time, is $o(t)$. In other words, the growth in dispersivity is *sublinear* in all cases. To emphasize this point we list the dispersivities here. In Example 1, the dispersivities are $O(1)$ in case (i), and $O(a^2) = o(t)$ in case (iv); in case (ii)′ it is $O(t^2/a^2) = o(t)$. In Example 2, in both cases the dispersivity is $O(1)$.

3. The general case $1 \ll t \ll a$

Consider the Itô equation

$$\mathrm{d}\boldsymbol{X}(t) = \{\boldsymbol{b}(\boldsymbol{X}(t)) + \boldsymbol{\beta}(\boldsymbol{X}(t)/a)\}\,\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(t), \tag{3.1}$$

where (i) $\boldsymbol{b}, \boldsymbol{\beta}$ are Lipschitzian vector fields (i.e. functions on $\mathbb{R}^k$ into $\mathbb{R}^k$); (ii) a is a scalar; (iii) $\boldsymbol{D}$ is a $k \times k$ positive definite matrix with the positive definite square root $\sqrt{\boldsymbol{D}}$; and (iv) $\boldsymbol{B}(t)$ is a standard k-dimensional Brownian motion independent of $\boldsymbol{X}(0)$. We wish to compare the asymptotics of $\boldsymbol{X}(t)$ with those of $\boldsymbol{Y}(t)$ governed by

$$\mathrm{d}\boldsymbol{Y}(t) = [\boldsymbol{b}\{\boldsymbol{X}(t)\} + \boldsymbol{\beta}(\boldsymbol{0})]\,\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(t) \tag{3.2}$$

when a and t are large, but t is small compared to a. Let us, for simplicity, consider the same non-random initial points for $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$, and express (3.1) and (3.2) in integral form:

$$\boldsymbol{X}(t) = \boldsymbol{x}_0 + \int_0^t [\boldsymbol{b}\{\boldsymbol{X}(s)\} + \boldsymbol{\beta}\{\boldsymbol{X}(s)/a\}]\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t), \tag{3.3}$$

$$\boldsymbol{Y}(t) = \boldsymbol{x}_0 + \int_0^t [\boldsymbol{b}\{\boldsymbol{Y}(s)\} + \boldsymbol{\beta}(\boldsymbol{0})]\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t). \tag{3.4}$$

By the Cameron–Martin–Girsanov theorem (see, for example, Friedman 1975, p. 169; Gihman and Skorohod 1979, pp. 279, 280), if f is a real-valued bounded measurable function on $\mathbb{R}^k$, then

$$\mathrm{E}\{f(\boldsymbol{X}(t))\} = \mathrm{E}\{f(\boldsymbol{Y}(t))\}\exp\{Z(t,a)\} \tag{3.5}$$

where

$$Z(t,a) := \int_0^t (\sqrt{\boldsymbol{D}})^{-1}[\boldsymbol{\beta}\{\boldsymbol{Y}(s)/a\} - \boldsymbol{\beta}(\boldsymbol{0})]\,\mathrm{d}\boldsymbol{B}(s) - \frac{1}{2}\int_0^t [\boldsymbol{\beta}\{\boldsymbol{Y}(s)/a\} - \boldsymbol{\beta}(\boldsymbol{0})]^{\mathrm{T}}\boldsymbol{D}^{-1}[\boldsymbol{\beta}\{\boldsymbol{Y}(s)/a\} - \boldsymbol{\beta}(\boldsymbol{0})]\,\mathrm{d}s. \tag{3.6}$$

Therefore,

$$|\mathrm{E}\{f(\boldsymbol{X}(t))\} - \mathrm{E}\{f(\boldsymbol{Y}(t))\}| = |\mathrm{E}\{f(\boldsymbol{Y}(t))\}[\exp\{Z(t,a)\} - 1]|. \tag{3.7}$$

It is easy to see that

$$\mathrm{E}\{Z^2(t,a)\} \le 2\left[\frac{c_1^2}{\alpha}\int_0^t \mathrm{E}\{|\boldsymbol{Y}(s)|^2/a^2\}\,\mathrm{d}s + \frac{c_1^4}{4\alpha^2}\left(\int_0^t \mathrm{E}\{|\boldsymbol{Y}(s)|^2/a^2\}\,\mathrm{d}s\right)^2\right], \tag{3.8}$$

where c_1 is the Lipschitzian constant for $\boldsymbol{\beta}(\cdot)$ and α is the smallest eigenvalue of $\boldsymbol{D}$. Further, $\mathrm{E}|\boldsymbol{Y}(s)|^2 \le 3(|\boldsymbol{x}_0|^2 + c_2 s^2 + \Gamma s)$, where $c_2 = (\|\boldsymbol{b}\|_\infty + |\boldsymbol{\beta}(\boldsymbol{0})|)^2$ and Γ is the largest eigenvalue of $\boldsymbol{D}$. It now follows that the right-hand side of (3.8) $\to 0$ if

$$t/a^{2/3} \to 0. \tag{3.9}$$

Therefore, $\exp\{Z(t,a)\} - 1 \to 0$ in probability as $t/a^{2/3} \to 0$. Also, writing

$$\mathrm{d}Z(t,a) = \boldsymbol{V}(t)\cdot\mathrm{d}\boldsymbol{B}(t) - \tfrac{1}{2}|\boldsymbol{V}(t)|^2\,\mathrm{d}t,$$

one has

$$\begin{aligned}
\mathrm{E}[\exp\{Z(t,a)\}]^2 &= \mathrm{E}[\exp\{2Z(t,a)\}] = \mathrm{E}\left[\exp\left\{2\int_0^t \boldsymbol{V}(s)\cdot\mathrm{d}\boldsymbol{B}(s) - \int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\right] \\
&= \mathrm{E}\left[\exp\left\{2\int_0^t \boldsymbol{V}(s)\cdot\mathrm{d}\boldsymbol{B}(s) - 4\int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\exp\left\{3\int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\right] \\
&\le \left(\mathrm{E}\left[\exp\left\{4\int_0^t \boldsymbol{V}(s)\cdot\mathrm{d}\boldsymbol{B}(s) - 8\int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\right]\right)^{1/2}\left(\mathrm{E}\left[\exp\left\{6\int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\right]\right)^{1/2} \\
&= \left(\mathrm{E}\left[\exp\left\{6\int_0^t |\boldsymbol{V}(s)|^2\,\mathrm{d}s\right\}\right]\right)^{1/2} \le \left(\mathrm{E}\left[\exp\left\{(6c_1^2/\alpha a^2)\int_0^t |\boldsymbol{Y}(s)|^2\,\mathrm{d}s\right\}\right]\right)^{1/2} \\
&\le \left(\mathrm{E}\left[\exp\left\{(c/a^2)\int_0^t (|\boldsymbol{x}_0|^2 + s^2 + |\boldsymbol{B}(s)|^2)\,\mathrm{d}s\right\}\right]\right)^{1/2}.
\end{aligned}$$

Taking $t = \varphi(a) = O(a^{2/3})$ as $a \to \infty$, we then obtain on any given compact set of initial values $\boldsymbol{x}_0$,

$$\mathrm{E}(\exp\{Z(\varphi(a),a)\})^2 \le \left[c''\mathrm{E}\exp\left\{(c/a^2)\int_0^{\varphi(a)} |\boldsymbol{B}(s)|^2\,\mathrm{d}s\right\}\right]^{1/2}.$$

By Brownian scaling, the expectation on the right equals

$$\mathrm{E}\left[\exp\left\{(c/a^2)\int_0^{\varphi(a)}\varphi(a)\,|\,\boldsymbol{B}(s/\varphi(a))\,|^2\,\mathrm{d}s\right\}\right]$$
$$= \mathrm{E}\left[\exp\left\{(c/a^2)\varphi^2(a)\int_0^1|\,\boldsymbol{B}(s')\,|^2\,\mathrm{d}s'\right\}\right]$$
$$\le \mathrm{E}\left\{\exp\left(c'''\int_0^1|\,\boldsymbol{B}(s')\,|^2\,\mathrm{d}s'\right)\right\} < \infty,$$

where c''' does not depend on a. This establishes the uniform integrability of $(\exp\{Z(\varphi(a), a)\} : a \geq 1)$. Hence (3.7) $\to 0$ if (3.9) holds. Indeed, the argument shows that this latter convergence is uniform over the class of all measurable f satisfying $\|f\|_\infty \le 1$. In other words, one has the L^1-convergence

$$\int_{\mathbb{R}^k}|\,p_a(t;\boldsymbol{x}_0,\boldsymbol{y}) - p(t;\boldsymbol{x}_0,\boldsymbol{y})\,|\,\mathrm{d}\boldsymbol{y} \to 0, \tag{3.10}$$

if (3.9) holds. Here p_a and p are the transition probability densities of the diffusions $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$, respectively.

Next, assume (v) that $\boldsymbol{b}$ is periodic. Without essential loss of generality, we will assume that the period lattice is the standard lattice $\mathbb{Z}^k$, i.e.

$$\boldsymbol{b}(\boldsymbol{x}+\boldsymbol{n}) = \boldsymbol{b}(\boldsymbol{x}) \qquad \forall \boldsymbol{x}\in\mathbb{R}^k, \forall \boldsymbol{n}\in\mathbb{Z}^k. \tag{3.11}$$

Write

$$\ddot{y} := \boldsymbol{y} \bmod 1 \equiv (y_1 \bmod 1, \ldots, y_k \bmod 1) \qquad (\boldsymbol{y} = (y_1, \ldots, y_k)). \tag{3.12}$$

Then

$$\boldsymbol{Y}(t) = \boldsymbol{x}_0 + \int_0^t \boldsymbol{b}\{\ddot{\boldsymbol{Y}}(s)\}\,\mathrm{d}s + t\boldsymbol{\beta}(\boldsymbol{0}) + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t), \tag{3.13}$$

where $\ddot{\boldsymbol{Y}}(s)$ is a diffusion on the torus $\mathcal{J}_1 := \{\boldsymbol{y} \bmod 1 : \boldsymbol{y}\in\mathbb{R}^k\}$ (see, for example, Bhattacharya and Waymire 1990, p. 518). Let $\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ denote the unique invariant probability of this diffusion. It now follows from Bensoussan *et al.* (1978, Chapter 3) and Bhattacharya (1985), that

$$\frac{\boldsymbol{Y}(t) - \boldsymbol{x}_0 - t\boldsymbol{\beta}(\boldsymbol{0}) - t\bar{\boldsymbol{b}}}{\sqrt{t}} \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{C}) \qquad \text{as } t\to\infty, \tag{3.14}$$

where $\bar{\boldsymbol{b}}$ and $\boldsymbol{C} = (C_{ij})$ are given by

$$\bar{\boldsymbol{b}} = \int_{[0,1)^k}\boldsymbol{b}(\boldsymbol{x})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = (\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_k),$$
$$\boldsymbol{C} = \int_{[0,1)^k}\{\mathrm{grad}\,\boldsymbol{\psi}(\boldsymbol{x}) - \mathbf{I}\}\boldsymbol{D}\{\mathrm{grad}\,\boldsymbol{\psi}(\boldsymbol{x}) - \mathbf{I}\}^{\mathrm{T}}\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{3.15}$$

Here $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)$, grad $\boldsymbol{\psi}$ is the $k\times k$ matrix whose ith row is grad ψ_i, ψ_i being the mean-zero (under π) periodic solution of

$$\frac{1}{2}\sum_{r,r'} D_{rr'}\frac{\partial^2\psi_i(\boldsymbol{x})}{\partial x_r \partial x_{r'}} + \sum_{r=1}^{k}\{b_r(\boldsymbol{x}) + \beta_r(\boldsymbol{0})\}\frac{\partial\psi_i(x)}{\partial x_r} = b_i(\boldsymbol{x}) - \bar{b}_i, \qquad 1 \le i \le k. \tag{3.16}$$

From (3.10) and (3.14) the following result is obtained.

Theorem 3.1 (a). If $\boldsymbol{b}, \boldsymbol{\beta}$ are non-constant Lipschitzian vector fields, then the L^1-distance between the transition probability densities $p_a(t; \boldsymbol{x}, \boldsymbol{y})$ and $p(t; \boldsymbol{x}, \boldsymbol{y})$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$ goes to zero if $t/a^{2/3} \to 0$. (b). If, in addition, $\boldsymbol{b}$ is periodic with period 1, i.e. with period lattice $\mathbb{Z}^k$, then

$$\frac{\boldsymbol{X}(t) - \boldsymbol{x}_0 - t\boldsymbol{\beta}(\boldsymbol{0}) - t\bar{\boldsymbol{b}}}{\sqrt{t}} \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{C}) \qquad \text{as } t \to \infty, a \to \infty, t/a^{2/3} \to 0, \tag{3.17}$$

the convergence being uniform for any compact set of initial points $\boldsymbol{x}_0$.

Remark 3.1
Note that $\boldsymbol{\beta}$ is not required to be periodic in Theorem 3.1, and a need not be an integer. Also, (3.10) does not require $\boldsymbol{b}$ to be periodic either. Thus if $\boldsymbol{b}$, $\boldsymbol{\beta}$ are Lipschitzian then the asymptotics of $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are the same for $1 \ll t \ll a^{2/3}$. Indeed, as long as (3.14) holds for some $\bar{\boldsymbol{b}}$ and $\boldsymbol{C}$, (3.17) also holds. For example, there are central limit theorems for diffusions with almost periodic coefficients, and for diffusions whose coefficients are ergodic random fields (Papanicolaou and Varadhan 1979; Kozlov 1980; Bhattacharya and Ramasubramanian 1988).

Remark 3.2
The range of validity of (3.17) cannot in general be extended beyond $1 \ll t \ll a^{2/3}$. In Example 2(i), in Section 2, the range is extended to $1 \ll t \ll a$ by resorting to a more delicate centring. As observed in Remark 2.5, with the present centring the convergence to a Gaussian would fail in Example 2 if t is not $o(a^{2/3})$. Whether one may find an appropriate centring, in the general situation studied in this section, in order to have a valid Gaussian approximation over the range $1 \ll t \ll a$ is not clear to us. Note that if $\mathrm{E}|\boldsymbol{Y}(s)|^2 = O(s)$, instead of $O(s^2)$, then (3.8) $\to 0$ as $t/a \to 0$. Thus for the special case $\bar{\boldsymbol{b}} = \boldsymbol{0}, \boldsymbol{\beta}(\boldsymbol{0}) = \boldsymbol{0}$, one can replace $t = o(t^{2/3})$ by $t = o(a)$ in Theorem 3.1.

4. The general case $t \gg a^2$

Consider the Itô equation (3.1). Throughout this section we will assume the following:

Assumptions
1. $\boldsymbol{b}, \boldsymbol{\beta}$ are continuously differentiable periodic vector fields on $\mathbb{R}^k$ having the common period lattice $\mathbb{Z}^k$, i.e.

$$\boldsymbol{b}(\boldsymbol{x} + \boldsymbol{n}) = \boldsymbol{b}(\boldsymbol{x}), \qquad \boldsymbol{\beta}(\boldsymbol{x} + \boldsymbol{n}) = \boldsymbol{\beta}(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathbb{R}^k, \boldsymbol{n} \in \mathbb{Z}^k. \tag{4.1}$$

2. The vector fields $\boldsymbol{b}, \boldsymbol{\beta}$ are divergence-free:

$$\operatorname{div} \boldsymbol{b}(\boldsymbol{x}) \equiv 0 \equiv \operatorname{div} \boldsymbol{\beta}(\boldsymbol{x}). \tag{4.2}$$

3. $\boldsymbol{B}(t)$ is a k-dimensional standard Brownian motion and $\boldsymbol{D}$ is a positive definite $k \times k$ matrix whose positive definite square root is denoted by $\sqrt{\boldsymbol{D}}$.
4. a is a positive integer.

We will use the notation

$$\dot{\boldsymbol{x}} := \boldsymbol{x} \bmod a \equiv (x_1 \bmod a, x_2 \bmod a, \ldots, x_k \bmod a). \tag{4.3}$$

For the moment, let us consider a fixed positive integer a. Then $\dot{\boldsymbol{X}}(t)$ is an ergodic Markov process having an exponentially decaying ϕ-mixing rate. In view of (4.2), its unique invariant probability is the uniform distribution on $\mathcal{J}_a$. Note that for measure-theoretic purposes, i.e. for the integration of periodic functions, we may identify $\mathcal{J}_a$ with $[0, a)^k$. It is known that $\boldsymbol{X}(t)$ is asymptotically, as $t \to \infty$, Gaussian (Bensoussan *et al.* 1978; Bhattacharya 1985). To express the parameters of this limiting Gaussian in terms of $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}$, denote by $\boldsymbol{L}$ the infinitesimal generator of $\dot{\boldsymbol{X}}(t)$ on $L^2([0,a)^k, a^{-k}\,\mathrm{d}\boldsymbol{x})$:

$$\boldsymbol{L}h(\boldsymbol{x}) := \frac{1}{2}\sum_{i,j=1}^{k} D_{ij}\frac{\partial^2 h(\boldsymbol{x})}{\partial x_i \partial x_j} + \sum_{i=1}^{k}\{b_i(\boldsymbol{x}) + \beta_i(\boldsymbol{x}/a)\}\frac{\partial h(\boldsymbol{x})}{\partial x_i}; \tag{4.4}$$

$$\bar{f} := \int_{[0,1)^k} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{4.5}$$

For each j there is a unique mean-zero periodic solution γ_j of the equation

$$\boldsymbol{L}\gamma_j(\boldsymbol{x}) = b_j(\boldsymbol{x}) + \beta_j(\boldsymbol{x}/a) - \bar{b}_j - \bar{\beta}_j. \tag{4.6}$$

This follows from the fact that the range of Δ is $1^\perp$ – the set of all mean-zero elements of $L^2([0,a)^k, a^{-k}\,\mathrm{d}\boldsymbol{x})$ (see Bhattacharya 1982; or Bensoussan *et al.* 1978, Chapter 3). Write $\boldsymbol{\gamma} := (\gamma_1, \gamma_2, \ldots, \gamma_k)$, $\operatorname{grad}\boldsymbol{\gamma} = (\operatorname{grad}\gamma_1, \ldots, \operatorname{grad}\gamma_k)$, and $\operatorname{grad}\boldsymbol{\gamma}_j = (\partial\gamma_j/\partial x_1, \ldots, \partial\gamma_j/\partial x_k)$. Then, by Itô's lemma,

$$\begin{aligned}
\boldsymbol{X}(t) - \boldsymbol{X}(0) - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}}) &= \int_0^t [\boldsymbol{b}\{\boldsymbol{X}(s)\} + \boldsymbol{\beta}\{\boldsymbol{X}(s)/a\} - \bar{\boldsymbol{b}} - \bar{\boldsymbol{\beta}}]\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t) \\
&= \int_0^t \Delta\boldsymbol{\gamma}\{\boldsymbol{X}(s)\}\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t) \\
&= \boldsymbol{\gamma}\{\boldsymbol{X}(t)\} - \boldsymbol{\gamma}\{\boldsymbol{X}(0)\} - \int_0^t \operatorname{grad}\boldsymbol{\gamma}\{\boldsymbol{X}(s)\}\sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(s) + \sqrt{\boldsymbol{D}}\boldsymbol{B}(t) \\
&= \boldsymbol{\gamma}(\dot{\boldsymbol{X}}(t)) - \boldsymbol{\gamma}(\dot{\boldsymbol{X}}(0)) - \int_0^t [\operatorname{grad}\boldsymbol{\gamma}\{\dot{\boldsymbol{X}}(s)\} - \mathbf{I}]\sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(s).
\end{aligned} \tag{4.7}$$

Dividing by $\sqrt{t}$, it now follows from the martingale central limit theorem that

$$\frac{1}{\sqrt{t}}\{\boldsymbol{X}_t - \boldsymbol{X}_0 - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{K}), \tag{4.8}$$

where $N(\mathbf{0}, \boldsymbol{K})$ denotes the normal distribution with mean vector $\mathbf{0}$ and dispersion matrix $\boldsymbol{K}$ given by

(see, for example, Bhattacharya 1985)

$$K := a^{-k}\int_{[0,a)^k}\{\text{grad}\,\boldsymbol{\gamma}(\boldsymbol{x})-\mathbf{I}\}\boldsymbol{D}\{\text{grad}\,\boldsymbol{\gamma}(\boldsymbol{x})-\mathbf{I}\}^{\mathrm{T}}\,\mathrm{d}x$$

$$= a^{-k}\int_{[0,a)^k}\{\text{grad}\,\boldsymbol{\gamma}(\boldsymbol{x})\}\{\text{grad}\,\boldsymbol{\gamma}(\boldsymbol{x})\}^{\mathrm{T}}\,\mathrm{d}s + \boldsymbol{D}. \tag{4.9}$$

Our goal in this article is to analyse the asymptotics of the distribution of $\boldsymbol{X}(t)$ as both t and a go to ∞. Note that $\boldsymbol{K}$ depends on a, and $\mathcal{J}_a$ grows as $a \to \infty$. It is, therefore, convenient to introduce the scaled process

$$\boldsymbol{Y}(t) := \boldsymbol{X}(a^2 t)/a, \tag{4.10}$$

which is governed by the stochastic differential equation

$$\mathrm{d}\boldsymbol{Y}(t) = a[\boldsymbol{b}\{a\boldsymbol{Y}(t)\} + \boldsymbol{\beta}\{\boldsymbol{Y}(t)\}]\,\mathrm{d}t + \sqrt{\boldsymbol{D}}\,\mathrm{d}\bar{\boldsymbol{B}}(t), \tag{4.11}$$

where $\bar{\boldsymbol{B}}(t) := \boldsymbol{B}(a^2 t)/a$ is again a standard Brownian motion on $\mathbb{R}^k$. Since $\boldsymbol{x} \to \boldsymbol{b}(a\boldsymbol{x}) + \boldsymbol{\beta}(\boldsymbol{x})$ is periodic with period 1 in each coordinate (i.e. the period lattice is $\mathbb{Z}^k$), the process

$$\ddot{Y}(t) := \boldsymbol{Y}(t) \bmod 1 = \dot{\boldsymbol{X}}(a^2 t)/a \tag{4.12}$$

is a diffusion on the torus $\mathcal{J}_1$. Also,

$$\frac{\boldsymbol{X}(a^2 t) - \boldsymbol{X}(0) - a^2 t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})}{a\sqrt{t}} = \frac{\boldsymbol{Y}(t) - \boldsymbol{Y}(0) - at(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})}{\sqrt{t}}. \tag{4.13}$$

The Markov process $\ddot{Y}(t)$ on $\mathcal{J}_1$ has the uniform distribution (on $\mathcal{J}_1$) as the unique invariant probability. Let A_a denote its infinitesimal generator,

$$\mathrm{A}_a h(\boldsymbol{x}) := \frac{1}{2}\sum_{i,j=1}^{k} D_{ij}\frac{\partial^2 h(\boldsymbol{x})}{\partial x_i\,\partial x_j} + a\sum_{j=1}^{k}\{b_j(a\boldsymbol{x}) + \beta_j(\boldsymbol{x})\}\frac{\partial h(\boldsymbol{x})}{\partial x_j} \tag{4.14}$$

for all periodic $h : h(\boldsymbol{x}+\boldsymbol{n}) = h(\boldsymbol{x})\,\forall \boldsymbol{x} \in \mathbb{R}^k, \boldsymbol{n} \in \mathbb{Z}^k$. For each j, let g_j be the unique mean-zero periodic solution of

$$\mathrm{A}_a g_j(\boldsymbol{x}) = b_j(a\boldsymbol{x}) + \beta_j(\boldsymbol{x}) - \bar{b}_j - \bar{\beta}_j. \tag{4.15}$$

By Itô's lemma,

$$\boldsymbol{Y}(t) - \boldsymbol{Y}(0) - at(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}}) \equiv a\int_0^t [\boldsymbol{b}\{a\boldsymbol{Y}(s)\} + \boldsymbol{\beta}\{\boldsymbol{Y}(s)\} - \bar{\boldsymbol{b}} - \bar{\boldsymbol{\beta}}]\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\bar{\boldsymbol{B}}(t)$$

$$= a\int_0^t \mathrm{A}_a g\{\boldsymbol{Y}(s)\}\,\mathrm{d}s + \sqrt{\boldsymbol{D}}\bar{\boldsymbol{B}}(t) = g\{\boldsymbol{Y}(t)\} - g\{\boldsymbol{Y}(0)\} - \int_0^t a\,\text{grad}\,g\{\boldsymbol{Y}(s)\}\,\mathrm{d}\bar{\boldsymbol{B}}(s) + \sqrt{\boldsymbol{D}}\bar{\boldsymbol{B}}(t)$$

$$= g\{\ddot{Y}(t)\} - g\{\ddot{Y}(0)\} - \int_0^t [a\,\text{grad}\,g\{\ddot{Y}(s)\} - \boldsymbol{I}]\sqrt{\boldsymbol{D}}\,\mathrm{d}\bar{\boldsymbol{B}}(s). \tag{4.16}$$

Note that

$$g_j(\boldsymbol{x}) = \gamma_j(a\boldsymbol{x})/a^2 \qquad 1 \le j \le k. \tag{4.17}$$

To proceed further, we introduce the complex Hilbert spaces H^0, H^1 as in Bhattacharya *et al.* (1989):

$$\begin{aligned}
H^0 &:= \left\{h : h \text{ periodic}, \int_{[0,1)^k} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0, \int_{[0,1)^k} |h(\boldsymbol{x})|^2\,\mathrm{d}\boldsymbol{x} < \infty\right\},\\
\langle h_1, h_2\rangle_0 &:= \int_{[0,1)^k} h_1(\boldsymbol{x}) h_2^-(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x};\\
H^1 &:= \left\{h : h \in H^0, \int_{[0,1)^k} |\nabla h(\boldsymbol{x})|^2\,\mathrm{d}\boldsymbol{x} < \infty\right\},\\
\langle h_1, h_2\rangle_1 &:= \frac{1}{2}\int_{[0,1)^k} \sum_{j,j'} D_{jj'} \frac{\partial h_1(\boldsymbol{x})}{\partial x_j} \frac{\partial h_2^-(\boldsymbol{x})}{\partial x_{j'}}\,\mathrm{d}\boldsymbol{x}\\
&= -\int_{[0,1)^k} \frac{1}{2}\sum_{j,j'} D_{jj'} \frac{\partial^2 h_1(\boldsymbol{x})}{\partial x_j\,\partial x_{j'}} h_2^-(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= -\langle \mathcal{D} h_1, h_2\rangle_0,
\end{aligned} \tag{4.18}$$

where f^- denotes the complex conjugate of f, and $\mathcal{D}$ is the second-order elliptic operator

$$\mathcal{D}h(\boldsymbol{x}) := \frac{1}{2}\sum_{j,j'=1}^{k} D_{jj'} \frac{\partial^2 h(\boldsymbol{x})}{\partial x_j\,\partial x_{j'}}. \tag{4.19}$$

One may express the operator A_a on H^1 as

$$\mathrm{A}_a = \mathcal{D}[\mathrm{I} + a\mathrm{S}_a], \tag{4.20}$$

where I is the identity operator and S_a is the operator on H^1 defined by

$$\mathrm{S}_a h(\boldsymbol{x}) := \mathcal{D}^{-1}[\boldsymbol{b}(a\boldsymbol{x}) + \boldsymbol{\beta}(\boldsymbol{x})] \cdot \nabla h(\boldsymbol{x}), \tag{4.21}$$

∇ denoting the gradient $(\partial/\partial x_1, \ldots, \partial/\partial x_k)$. Note that S_a is skew symmetric and compact (see Bhattacharya *et al.* 1989). Below $\mathbb{Z}^{++}$ denotes the set of all positive integers.

Lemma 4.1 There exists a positive number c_1 such that

$$\sup_{a \in \mathbb{Z}^{++}} \|g_j\|_1^2 \leq c_1 \qquad 1 \leq j \leq k. \tag{4.22}$$

Also,

$$\|g_j\|_0^2 \leq \frac{1}{2\pi^2\alpha} \|g_j\|_1^2 \qquad 1 \leq j \leq k, \tag{4.23}$$

where α is the smallest eigenvalue of the matrix $\boldsymbol{D} = (D_{jj'})$.

Proof
First note that

$$\langle \mathrm{S}_a f, f\rangle_1 = 0 \qquad \forall f \in H^1, \tag{4.24}$$

since S_a is skew symmetric. Therefore,

$$\|(I + aS_a)f\|_1^2 = \|f\|_1^2 + a^2\|S_a f\|_1^2 \geq \|f\|_1^2. \tag{4.25}$$

Now one may express (4.15) as

$$(I + aS_a)g_j = \mathcal{D}^{-1}[b_j(a\boldsymbol{x}) - \bar{b}_j + \beta_j(\boldsymbol{x}) - \bar{\beta}_j]. \tag{4.26}$$

By (4.25), (4.26) and (4.18),

$$\begin{aligned}\|g_j\|_1^2 &\leq \|(I + aS_a)g_j\|_1^2 = \|\mathcal{D}^{-1}[b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j]\|_1^2 \\ &= \langle b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j, -\mathcal{D}^{-1}[b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j]\rangle_0. \end{aligned} \tag{4.27}$$

On H^0 the largest eigenvalue of $-\mathcal{D}^{-1}$ is at most $(2\pi^2\alpha)^{-1}$. Hence

$$\begin{aligned}\|g_j\|_1^2 &\leq \frac{1}{2\pi^2\alpha}\|b_j(a\cdot) - \bar{b}_j + \beta_j - \bar{\beta}_j\|_0^2 \\ &\leq \frac{1}{\pi^2\alpha}[\|b_j(a\cdot) - \bar{b}_j\|_0^2 + \|\beta_j - \bar{\beta}_j\|_0^2] \\ &= \frac{1}{\pi^2\alpha}[\|b_j(\cdot) - \bar{b}_j\|_0^2 + \|\beta_j - \bar{\beta}_j\|_0^2] = c_{1j}, \end{aligned} \tag{4.28}$$

say. Let $c_1 = \max\{c_{1j} : 1 \leq j \leq k\}$. To prove (4.23), note that

$$\begin{aligned}\|g_j\|_1^2 &= \langle -\mathcal{D}g_j, g_j\rangle_0 = \sum_{\boldsymbol{n}\neq\boldsymbol{0}}\sum_{j,j'} 2\pi^2 D_{jj'} n_j n_{j'} |\hat{g}_j(\boldsymbol{n})|^2 \\ &\geq 2\pi^2\alpha \sum_{\boldsymbol{n}\neq\boldsymbol{0}} |\boldsymbol{n}^2| |\hat{g}_j(\boldsymbol{n})|^2 \geq 2\pi^2\alpha\|g_j\|_0^2. \end{aligned} \tag{4.29}$$

□

One consequence of (4.22) is that the term $(g(\ddot{\boldsymbol{Y}}(t)) - g(\ddot{\boldsymbol{Y}}(0)))/\sqrt{t}$ in (4.16) may be neglected. To see this assume, for the sake of simplicity, that $\ddot{\boldsymbol{Y}}(0)$ has the uniform distribution. Then $\ddot{\boldsymbol{Y}}(t)$, $t \geq 0$, is stationary and

$$\begin{aligned} E\left|\frac{1}{\sqrt{t}}[g\{\ddot{\boldsymbol{Y}}(t)\} - g\{\ddot{\boldsymbol{Y}}(0)\}]\right|^2 &= \sum_{j=1}^{k}\frac{1}{t}E|g_j\{\ddot{\boldsymbol{Y}}(t)\} - g_j\{\ddot{\boldsymbol{Y}}(0)\}|^2 \\ &\leq \sum_{j=1}^{k}\frac{4}{t}\|g_j\|_0^2 \leq \frac{4kc_1}{2\pi^2\alpha t} = \frac{2kc_1}{\pi^2\alpha t}. \end{aligned} \tag{4.30}$$

In order to prove the asymptotic normality of $\boldsymbol{Y}(t)$ we will need an estimate of the rate of convergence of the distribution of $\ddot{\boldsymbol{Y}}(t)$, starting from an arbitrary state, to equilibrium. Let $\ddot{p}_a(t;\boldsymbol{x},\boldsymbol{y})$ denote the transition probability density of $\ddot{\boldsymbol{Y}}$. The following lemma follows from Proposition 4.3 below, which is an extension of a result from Fill (1991) to continuous-parameter Markov processes.

Lemma 4.2 One has

$$\int_{[0,1)^k} |\ddot{p}_a(t;\boldsymbol{x},\boldsymbol{y}) - 1|\, \mathrm{d}\boldsymbol{y} \le c' \exp\{-2\pi^2 \alpha t\} \qquad \forall t > 0, \boldsymbol{x} \in [0,1)^k, \tag{4.31}$$

where α is the smallest eigenvalue of $\boldsymbol{D} = (D_{ij})$, and c' is a constant which does not depend on a.

Remark 4.1
Since $\dot{X}(t) = a\ddot{Y}(t/a^2)$, (4.31) is equivalent to the inequality

$$\int_{[0,a)^k} |\dot{p}_a(t;\boldsymbol{x},\boldsymbol{y}) - a^{-k}|\, \mathrm{d}\boldsymbol{y} \le c' \exp(-2\pi^2 \alpha t/a^2) \qquad \forall t > 0, \boldsymbol{x} \in [0,a)^k,$$

where $\dot{p}_a$ is the transition probability density of $\dot{\boldsymbol{X}}$.

To state the proposition from which Lemma 4.2 follows, consider a continuous-parameter Markov process $\{U(t) : t \ge 0\}$ on a metric space M, which has a transition probability density $u(t;x,y)$ with respect to some σ-finite measure ν on (the Borel σ-field of) M. Suppose the process has an invariant probability with density π (with respect to ν) such that the Markov process with this initial (invariant) distribution is ergodic. Define the transition probability density $v(t;x,y)$ for the 'time-reversed' process by

$$v(t;x,y) := \frac{\pi(y)}{\pi(x)} u(t;y,x). \tag{4.32}$$

It is simple to check that v, like u, also has π as an invariant probability density. Let T_t, $\tilde{T}_t$ denote the transition operators with kernels u, v, respectively, and let A, Ã denote the infinitesimal generators of these semigroups of operators on $L^2(M, \pi\, \mathrm{d}\nu)$. Define B to be the closed operator determined by

$$\mathrm{B}f = \tfrac{1}{2}(\mathrm{A} + \tilde{\mathrm{A}})f \qquad \forall f \in D_{\mathrm{A}} \cap D_{\tilde{\mathrm{A}}}, \tag{4.33}$$

where, for an operator C on L^2, $\mathrm{D_C}$ denotes the domain of C. We assume $\mathrm{D_A} \cap \mathrm{D_{\tilde{A}}}$ is dense in L^2. Note that B is self-adjoint. Denote by $R_t, t > 0$, the semigroup of operators generated by B.

Proposition 4.3 Suppose B has a discrete spectrum $\lambda_0 = 0 > -\lambda_1 \ge -\lambda_2 \ge -\lambda_3 \ge \ldots$, 0 being a simple eigenvalue and each eigenvalue having finite multiplicity. Then for every initial distribution with density η, the density η_t of $U(t)$ satisfies the inequality

$$\int |\eta_t(y) - \pi(y)|\, \nu(\mathrm{d}y) \le \mathrm{e}^{-\lambda_1 t} \left(\int \frac{(\eta(y) - \pi(y))^2}{\pi(y)} \nu(\mathrm{d}y) \right)^{1/2}. \tag{4.34}$$

Proof
Let $1, \phi_1, \phi_2, \phi_3, \ldots$ be a complete orthonormal sequence corresponding to the eigenvalues $0 > -\lambda_1 \ge -\lambda_2 \ge -\lambda_3 \ge \ldots$. If $\psi \in L^2(M, \pi\, \mathrm{d}\nu)$ and $\int \psi\pi\, \mathrm{d}\nu = 0$, then $R_t\psi$ has the eigenfunction expansion

$$R_t\psi = \sum_{n \ge 1} \mathrm{e}^{-\lambda_n t} \langle \psi, \phi_n \rangle \phi_n, \tag{4.35}$$

where $\langle , \rangle$ denotes the inner product on $L^2(M, \pi \, \mathrm{d}\nu)$. Also, write $\| \cdot \|$ for the L^2-norm. Now let

$$\psi_0(x) = \frac{\eta(x)}{\pi(x)} - 1, \tag{4.36}$$

assuming $\psi_0 \in L^2$. By Trotter's product formula (see Ethier and Kurz 1986, p. 33) one has, for every $\varphi \in L^2(M, \pi \, \mathrm{d}\nu)$,

$$R_t \varphi = \lim_{j \to \infty} (T_{t/2j} \tilde{T}_{t/2j})^j \varphi = \lim_{j \to \infty} S_j^j \varphi, \tag{4.37}$$

where $S_j := T_{t/2j} \tilde{T}_{t/2j}$. Fix $t > 0$, and write

$$\eta^{(m)}(y) := \eta_{mt/2j}(y), \psi_m(y) := \frac{\eta^{(m)}(y)}{\pi(y)} - 1 \qquad 0 \le m \le 2j. \tag{4.38}$$

Then $\eta^{(0)} = \eta, \eta^{(2j)} = \eta_t$, and

$$\begin{aligned} \eta^{(m)}(y) &= \int \eta^{(m-1)}(x) u(t/2j; x, y) \nu(\mathrm{d}x) \\ &= \int \eta^{(m-1)}(x) v(t/2j; y, x) \frac{\pi(y)}{\pi(x)} \nu(\mathrm{d}x), \\ \psi_m(y) &= \int \left(\frac{\eta^{(m-1)}(x)}{\pi(x)} - 1 \right) v(t/2j; y, x) \nu(\mathrm{d}x) = \tilde{T}_{t/2j} \psi^{(m-1)}(y). \end{aligned} \tag{4.39}$$

Thus if a_j denotes the norm of S_j on $1^{\perp}$, one has

$$\begin{aligned} \| \psi^{(m)} \|^2 &= \| \tilde{T}_{t/2j} \psi^{(m-1)} \|^2 = \langle S_j \psi^{(m-1)}, \psi^{(m-1)} \rangle \\ &\le a_j \| \psi^{(m-1)} \|^2 \qquad m = 1, 2, \ldots, 2j. \end{aligned} \tag{4.40}$$

Iterating, one gets

$$\int \frac{\{\eta_t(y) - \pi(y)\}^2}{\pi(y)} \nu(\mathrm{d}y) \equiv \| \psi^{(2j)} \|^2 \le a_j^{2j} \int \frac{\{\eta(y) - \pi(y)\}^2}{\pi(y)} \nu(\mathrm{d}y). \tag{4.41}$$

Now a_j^j and $\exp\{-\lambda_1 t\}$ are the norms of S_j^j and R_t, respectively, on $1^{\perp}$. Since the latter are self-adjoint, it is not difficult to check using (4.37) that $a_j^j \to \exp\{-\lambda_1 t\}$ as $j \to \infty$. For this use the fact that for a self-adjoint operator C on a Hilbert space H having a non-negative spectrum, $\| C \| = \sup\{\langle C\varphi, \varphi \rangle : \varphi \in H, \| \varphi \| = 1\}$. Relation (4.41) now implies

$$\int \frac{(\eta_t(y) - \pi(y))^2}{\pi(y)} \nu(\mathrm{d}y) \le \mathrm{e}^{-2\lambda_1 t} \int \frac{(\eta(y) - \pi(y))^2}{\pi(y)} \nu(\mathrm{d}y). \tag{4.42}$$

Now use Cauchy–Schwarz to get

$$\begin{aligned} \int | \eta_t(x) - \pi(x) | \nu(\mathrm{d}x) &\equiv \int \frac{| \eta_t(x) - \pi(x) |}{\sqrt{\pi(x)}} \sqrt{\pi(x)} \nu(\mathrm{d}x) \\ &\le \left(\int \frac{(\eta_t(x) - \pi(x))^2}{\pi(x)} \nu(\mathrm{d}x) \right)^{1/2}. \end{aligned} \tag{4.43}$$

□

Proof of Lemma 4.2

To apply Proposition 4.3 to our problem of interest, let $p_a(t;\boldsymbol{x},\boldsymbol{y}), \dot{p}_a(t;\boldsymbol{x},\boldsymbol{y})$ and $\ddot{p}_a(t;\boldsymbol{x},\boldsymbol{y})$ denote the transition probability densities of $\boldsymbol{X}(t), \dot{\boldsymbol{X}}(t)$ and $\ddot{\boldsymbol{Y}}(t)$, respectively. Since $\boldsymbol{X}(t)$ (being governed by the Itô equation (3.1) and having the infinitesimal generator $\boldsymbol{L}$ in (4.4)) has a drift which is uniformly (with respect to a) bounded, as are its first-order derivatives, it follows from the Cameron–Martin–Girsanov theorem, or from standard estimates in partial differential equations (see, for example, Aronson 1967) that there exist positive constants c, c' independent of a, such that

$$p_a(1;\boldsymbol{x},\boldsymbol{y}) \leq c' \exp(-c\,|\boldsymbol{x}-\boldsymbol{y}|^2) \qquad \forall \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^k. \tag{4.44}$$

It then follows that

$$\sup_{\substack{a\in\mathbb{Z}^{++}\\ \boldsymbol{x},\boldsymbol{y}\in[0,a)^k}} a^k \dot{p}_a(1;\boldsymbol{x},\boldsymbol{y}) \equiv \sup a^k \sum_{\boldsymbol{n}\in\mathbb{Z}^k} p_a(1;\boldsymbol{x},\boldsymbol{y}+a\boldsymbol{n}) = c'' < \infty,$$

$$\sup_{\substack{a\in\mathbb{Z}^{++}\\ \boldsymbol{x},\boldsymbol{y}\in[0,1)^k}} \ddot{p}_a(1/a^2;\boldsymbol{x},\boldsymbol{y}) \leq c''. \tag{4.45}$$

Now take $M = \mathcal{J}_1 = [0,1)^k$, $\ddot{\boldsymbol{Y}}(t) = U(t)$ in Proposition 4.3. Then $u(t;\boldsymbol{x},\boldsymbol{y}) = \ddot{p}_a(t;\boldsymbol{x},\boldsymbol{y}) = v(t;\boldsymbol{y},\boldsymbol{x})$, $\pi(\boldsymbol{y}) = 1$, and $\mathrm{A} = \mathrm{A}_a = \mathcal{D} + (\boldsymbol{b}(a\,\cdot) + \boldsymbol{\beta}(\cdot))\cdot\nabla$, $\tilde{\mathrm{A}} = \mathcal{D} - a\ \ (\boldsymbol{b}(a\,\cdot) + \boldsymbol{\beta}(\cdot))\cdot\nabla$ (in view of (4.2)), so that

$$\tfrac{1}{2}(\mathrm{A}+\tilde{\mathrm{A}}) = \mathcal{D} \equiv \frac{1}{2}\sum_{j,j'} D_{jj'} \frac{\partial^2}{\partial x_j\,\partial x_{j'}}, \tag{4.46}$$

$$\begin{aligned} -\lambda_1 &:= \sup_{\|f\|_0=1,\, f\perp 1} \langle f, \mathcal{D}f\rangle_0 \\ &= \sup_{\|f\|_0=1,\, f\perp 1} \left\{(-2\pi^2)\sum_{\boldsymbol{n}\neq\boldsymbol{0}} |\hat{f}(\boldsymbol{n})|^2 \sum_{j,j'} D_{jj'} n_j n_{j'}\right\} \\ &\leq -2\pi^2\alpha, \end{aligned} \tag{4.47}$$

α being the smallest eigenvalue of the matrix $\boldsymbol{D}$. Now take $t - 1/a^2$ for t in Proposition 4.3, and let $\eta(\boldsymbol{y})$ be $\ddot{p}_a(1/a^2;\boldsymbol{x},\boldsymbol{y})$. Then

$$\begin{aligned} &\int \frac{\{\eta(\boldsymbol{y}) - \pi(\boldsymbol{y})\}^2}{\pi(\boldsymbol{y})}\,\nu(\mathrm{d}\boldsymbol{y}) \\ &= \int_{[0,a)^k} \{\ddot{p}_a(1/a^2;\boldsymbol{x},\boldsymbol{y}) - 1\}^2\,\mathrm{d}\boldsymbol{y} \\ &\leq (c''-1)^2 = c^2, \end{aligned}$$

say, by (4.45). Therefore, (4.34) provides the bound

$$\begin{aligned} \sup_{\boldsymbol{x}\in[0,1)^k} &= \int_{[0,1)^k} |\ddot{p}_a(t;\boldsymbol{x},\boldsymbol{y}) - 1|\,\mathrm{d}\boldsymbol{y} \\ &\leq c\exp\{-2\pi^2\alpha(t-1/a^2)\} \\ &\leq c'\exp(-2\pi^2\alpha t), \qquad t > 1/a^2, \end{aligned} \tag{4.48}$$

where we have taken $c' = c\exp(2\pi^2\alpha)$. To extend inequality (4.48) from $t \geq 1/a^2$ to $t > 0$ take $c' = \max\{c, 2\}\exp(2\pi^2\alpha)$. Finally, since $\ddot{Y}(t) = \dot{X}(a^2t)/a$, (4.31) follows from (4.48). □

The following corollary to Lemma 4.2 is fairly straightforward. To state it, let E_x denote expectation when $\ddot{Y}(0) = \boldsymbol{x}$, and E that under the uniform initial distribution. Also, let $\mathrm{var}_x, \mathrm{cov}_x$ denote variance and covariance under E_x, while for those under E the subscript x will be dropped. Let $\ddot{T}_t$ denote the transition operator

$$(\ddot{T}_t f)(x) := \mathrm{E}_x[f\{\ddot{Y}(t)\}]. \tag{4.49}$$

We will denote the 'supremum norm' as usual by $\|\cdot\|_\infty$, and $\bar{f}$ denotes expectation on $[0, 1)^k$ under Lebesgue measure.

Corollary 4.4 There exist positive constants $c_i (i = 1, 2, 3, 4, 5, 6)$ independent of a such that

$$\|\ddot{T}_t f - \bar{f}\|_\infty \leq c_1 \|f\|_\infty \exp(-2\pi^2\alpha t); \tag{4.50}$$

$$|\mathrm{cov}_x(f\{\ddot{Y}(s)\}, g\{\ddot{Y}(t)\})| \leq c_2 \|f\|_\infty \|g\|_\infty \exp\{-2\pi^2\alpha(t-s)\} \qquad \text{for } t \geq s; \tag{4.51}$$

$$|\mathrm{cov}_x(f\{\ddot{Y}(s)\}, g\{\ddot{Y}(t)\}) - \mathrm{cov}(f\{\ddot{Y}(s)\}, g\{\ddot{Y}(t)\})| \leq c_3 \|f\|_\infty \|g\|_\infty \exp(-2\pi^2\alpha t) \qquad \text{for } t \geq s; \tag{4.52}$$

$$\left\| \frac{1}{\sqrt{t}} \left(\int_0^t \ddot{T}_s f(\cdot)\, \mathrm{d}s - t\bar{f} \right) \right\|_\infty \leq c_4 \|f\|_\infty / \sqrt{t}; \tag{4.53}$$

$$\frac{1}{t} \left\| \mathrm{cov}_x\left(\int_0^t f\{\ddot{Y}(s)\}\, \mathrm{d}s, \int_0^t g\{\ddot{Y}(s)\}\, \mathrm{d}s \right) - \mathrm{cov}\left(\int_0^t f\{\ddot{Y}(s)\}\, \mathrm{d}s, \int_0^t g\{\ddot{Y}(s)\}\, \mathrm{d}s \right) \right\|_\infty$$
$$\leq c_5 \|f\|_\infty \|g\|_\infty / t; \tag{4.54}$$

$$\left| \frac{1}{t} \mathrm{cov}\left(\int_0^t f\{\ddot{Y}(s)\}\, \mathrm{d}s, \int_0^t g\{\ddot{Y}(s)\}\, \mathrm{d}s \right) - \int_0^\infty \overline{(f - \bar{f})(\ddot{T}_u g - \bar{g})}\, \mathrm{d}u \right.$$
$$\left. - \int_0^\infty \overline{(g - \bar{g})(\ddot{T}_u f - \bar{f})}\, \mathrm{d}u \right| \leq c_6 \|f\|_\infty \|g\|_\infty / t. \tag{4.55}$$

Proof
Inequality (4.50) is an immediate consequence of Lemma 4.2.
To derive (4.51), write

$$|\mathrm{cov}_x(f\{\ddot{Y}(s)\}, g\{\ddot{Y}(t)\})| = |\mathrm{E}_x([f\{\ddot{Y}(s)\} - \ddot{T}_s f(\boldsymbol{x})][\ddot{T}_{t-s} g\{\ddot{Y}(s)\} - \bar{g}])|. \tag{4.56}$$

Now apply (4.50) to the second factor on the right-hand side.

In order to obtain (4.52), proceed from the expression within modulus bars on the right-hand side of (4.56) to get

$$|\mathrm{cov}_x(f\{\ddot{Y}(s)\}, g\{\ddot{Y}(t)\}) - \mathrm{E}_x([f\{\ddot{Y}(s)\} - \bar{f}][\ddot{T}_{t-s} g\{\ddot{Y}(s)\} - \bar{g}])|$$
$$\leq c_3' \|f\|_\infty \|g\|_\infty \exp(-2\pi^2\alpha s) \exp\{-2\pi^2\alpha(t-s)\},$$

the error being caused by replacing $\ddot{T}_s f(\boldsymbol{x})$ by $\bar{f}$. Now the E_x term above may be expressed as $\ddot{T}_s h(\boldsymbol{x})$

where $h(y) := \{f(y) - \bar{f}\}\{\ddot{T}_{t-s}g(y) - \bar{g}\}$ so that, using (4.50),

$$\|\ddot{T}_s h(\cdot) - \bar{h}\|_\infty \le c_1 \|h\|_\infty \exp\{-2\pi^2\alpha s\} \le 2c_1^2 \|f\|_\infty \|g\|_\infty \exp\{-2\pi^2\alpha t\}.$$

Since $\bar{h}$ is precisely the covariance under equilibrium, (4.52) is obtained.

Inequality (4.53) is immediate from (4.50).

To derive (4.54) write

$$\begin{aligned} &\left| \operatorname{cov}_x\left(\int_0^t f\{\ddot{Y}(s)\}\,\mathrm{d}s, \int_0^t g\{\ddot{Y}(s)\}\,\mathrm{d}s \right) - \operatorname{cov}\left(\int_0^t f\{\ddot{Y}(s)\}\,\mathrm{d}s, \int_0^t g\{\ddot{Y}(s)\}\,\mathrm{d}s \right) \right| \\ &\quad \le \left| \int_0^t \int_0^s \{\operatorname{cov}_x(f\{\ddot{Y}(s')\}, g\{\ddot{Y}(s)\}) - \operatorname{cov}(f\{\ddot{Y}(s')\}, g\{\ddot{Y}(s)\})\}\,\mathrm{d}s'\,\mathrm{d}s \right| \\ &\quad + \left| \int_0^t \int_0^s \{\operatorname{cov}_x(g\{\ddot{Y}(s')\}, f\{\ddot{Y}(s)\}) - \operatorname{cov}(g\{\ddot{Y}(s')\}, f\{\ddot{Y}(s)\})\}\,\mathrm{d}s'\,\mathrm{d}s \right|. \end{aligned} \tag{4.57}$$

In view of (4.52), the right-hand side of (4.57) is bounded by

$$\begin{aligned} &2c_3 \|f\|_\infty \|g\|_\infty \int_0^t \int_0^s \exp(-2\pi^2\alpha s)\,\mathrm{d}s'\,\mathrm{d}s \\ &= 2c_3 \|f\|_\infty \|g\|_\infty \int_0^t s \exp(-2\pi^2\alpha s)\,\mathrm{d}s \\ &\le c_3' \|f\|_\infty \|g\|_\infty, \end{aligned} \tag{4.58}$$

from which (4.54) follows.

Finally, the 'cov' term in (4.55) may be expressed as

$$\frac{1}{t}\left\{ \int_0^t \int_0^s \overline{\ddot{T}_s\{(f-\bar{f})(\ddot{T}_u g - \bar{g})\}}\,\mathrm{d}u\,\mathrm{d}s + \int_0^t \int_0^s \overline{\ddot{T}_s\{(g-\bar{g})(\ddot{T}_u f - \bar{f})\}}\,\mathrm{d}u\,\mathrm{d}s \right\}. \tag{4.59}$$

Now, because of stationarity, $\overline{\ddot{T}_s h} = \bar{h}$, so that

$$\begin{aligned} &\left| \frac{1}{t}\int_0^t \int_0^s \overline{\ddot{T}_s\{(f-\bar{f})(\ddot{T}_u g - \bar{g})\}}\,\mathrm{d}u\,\mathrm{d}s - \int_0^\infty \overline{(f-\bar{f})(\ddot{T}_u g - \bar{g})}\,\mathrm{d}u \right| \\ &= \left| \frac{1}{t}\int_0^t \int_s^\infty \overline{(f-\bar{f})(\ddot{T}_u g - \bar{g})}\,\mathrm{d}u\,\mathrm{d}s \right| \le c_1 \|f\|_\infty \|g\|_\infty \frac{1}{t}\int_0^t \int_s^\infty \exp(-2\pi^2\alpha u)\,\mathrm{d}u\,\mathrm{d}s \\ &= \frac{c_1}{2\pi^2\alpha} \|f\|_\infty \|g\|_\infty \frac{1}{t}\int_0^t \exp(-2\pi^2\alpha s)\,\mathrm{d}s \le c_6' \|f\|_\infty \|g\|_\infty / t. \end{aligned}$$

□

Corollary 4.4 paves the way for deriving central limit theorems for $X(t)$ at time-scales $t \gg a^2$, by showing that $\ddot{Y}(t)$ is weakly dependent with an exponentially decaying strong mixing rate. We still need to find appropriate scalings under which the dispersion of the scaled $X(t)$, or $Y(t)$, stabilizes away from zero and infinity. As we shall see, the growth in the asymptotic variance of a component of $X(t)$, or $Y(t)$, depends crucially on properties of the individual functions β_j, $1 \le j \le k$. In the first case that we consider the asymptotic variance of $X_j(t)$ is $O(a^2)$, while in the second case it is $O(1)$. In order to derive verifiable criteria for such growths we introduce the operator $\bar{\mathrm{A}}_a$ defined on H^1 by

$$\begin{aligned}\bar{\mathrm{A}}_a f(\boldsymbol{x}) &:= \mathcal{D}f(\boldsymbol{x}) + a(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\boldsymbol{x}))\cdot\nabla f(\boldsymbol{x})\\ &= \mathcal{D}[\mathrm{I} + a\bar{\mathrm{S}}]f(\boldsymbol{x}),\end{aligned} \tag{4.60}$$

where $\bar{\mathrm{S}}$ is the skew-symmetric compact operator (on H^1) defined as

$$\bar{\mathrm{S}}f(\boldsymbol{x}) := \mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\boldsymbol{x}))\cdot\nabla f(\boldsymbol{x}). \tag{4.61}$$

Now let h_j be the solution (in H^1) of

$$\bar{\mathrm{A}}_a h_j(\boldsymbol{x}) = \beta_j(\boldsymbol{x}) - \bar{\beta}_j \qquad 1 \le j \le k. \tag{4.62}$$

Let N denote the null space of $\bar{\mathrm{S}}$ or, equivalently, of $(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\cdot))\cdot\nabla$ in H^1. The projection of an element f of H^1 on N will be denoted by f_N.

Lemma 4.5 For all $f \in H^1$

$$\begin{aligned}|\langle\mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j), f\rangle_1| = |\langle b_j(a\,\cdot) - \bar{b}_j, f\rangle_0| &\le \frac{\|b_j\|_0}{a}\left(\sum_{|\boldsymbol{n}|\ge a}|\boldsymbol{n}|^2|\hat{f}(\boldsymbol{n})|^2\right)^{1/2}\\ &\le \frac{\|b_j\|_0}{(2\pi^2\alpha)^{1/2}a}\|f\|_1,\end{aligned} \tag{4.63}$$

where α is smallest eigenvalue of $\boldsymbol{D} = (D_{jj'})$. This implies that if $\mathcal{F} \subset H^1$ is relatively compact then

$$\sup_{f\in\mathcal{F}}|\langle b_j(a\,\cdot) - \bar{b}_j, f\rangle_0| = o\left(\frac{1}{a}\right) \qquad \text{as } a \to \infty. \tag{4.64}$$

If $\mathcal{F}$ in (4.64) is only relatively compact in H^0 then the right-hand side of (4.64) is $o(1)$.

Proof
Let $f \in H^1$. One may suppose $(\partial^r/\partial x_i^r)b_j(\boldsymbol{x}) \in L^2([0,1)^k, \mathrm{d}\boldsymbol{x})$ for $1 \le i \le k$ and for all $r \le k_0 := [k/2] + 1$. Then

$$\langle\mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j), f\rangle_1 = \langle b_j(a\,\cdot) - \bar{b}_j, f\rangle_0 = \sum_{\boldsymbol{n}\ne 0}(b_j(a\,\cdot) - \bar{b}_j)\hat{}(\boldsymbol{n})\,\hat{f}^{-}(\boldsymbol{n}) \tag{4.65}$$

where

$$\hat{f}(\boldsymbol{n}) := \int_{[0,1)^k} f(\boldsymbol{x})\,\mathrm{e}^{-2\pi i\boldsymbol{n}\cdot\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} \qquad \boldsymbol{n} \in \mathbb{Z}^k.$$

By the assumption on $b_j(\cdot)$,

$$\begin{aligned}\sum_{\boldsymbol{n}\ne 0}|\hat{b}_j(\boldsymbol{n})| &= \sum_{\boldsymbol{n}\ne 0}\frac{|\boldsymbol{n}|^{k_0}|\hat{b}_j(\boldsymbol{n})|}{|\boldsymbol{n}|^{k_0}}\\ &\le \left(\sum_{\boldsymbol{n}\ne 0}|\boldsymbol{n}|^{2k_0}|\hat{b}_j(\boldsymbol{n})|^2\right)^{1/2}\left(\sum_{\boldsymbol{n}\ne 0}\frac{1}{|\boldsymbol{n}|^{2k_0}}\right)^{1/2}\\ &< \infty.\end{aligned} \tag{4.66}$$

Therefore, the Fourier series

$$\sum_{n\neq 0} \hat{b}_j(\boldsymbol{n})\, e^{2\pi i \boldsymbol{n}\cdot \boldsymbol{x}} \to b_j(\boldsymbol{x}) - \bar{b}_j \qquad \text{uniformly in } \boldsymbol{x}. \tag{4.67}$$

In particular,

$$\begin{aligned} b_j(a\boldsymbol{x}) - \bar{b}_j &= \sum_{n\neq 0} \hat{b}_j(\boldsymbol{n})\, e^{2\pi i \boldsymbol{n}\cdot a\boldsymbol{x}} = \sum_{n\neq 0} \hat{b}_j(\boldsymbol{n})\, e^{2\pi i a\boldsymbol{n}\cdot \boldsymbol{x}} \\ &= \sum_{m=an:\, n\neq 0} \hat{b}_j(\boldsymbol{m}/a)\, e^{2\pi i \boldsymbol{m}\cdot \boldsymbol{x}}. \end{aligned} \tag{4.68}$$

Hence, writing $\boldsymbol{n}$ n.d.a to indicate some coordinate of $\boldsymbol{n}$ is not divisible by a,

$$(b_j(a\,\cdot) - \bar{b}_j)^{\hat{}}(\boldsymbol{n}) = \begin{cases} 0 & \text{if } |\boldsymbol{n}| < a \text{ or } \boldsymbol{n} \text{ n.d.}a, \\ \hat{b}_j(\boldsymbol{n}/a) & \text{otherwise.} \end{cases} \tag{4.69}$$

Using (4.69) in (4.65) we get

$$\begin{aligned} |\langle b_j(a\,\cdot) - \bar{b}_j, f\rangle_0| &\le \sum_{|n|\ge a} |\hat{b}_j(\boldsymbol{n}/a)\hat{f}(\boldsymbol{n})| \\ &\le \left(\sum_{|n|\ge a} |\hat{b}_j(\boldsymbol{n}/a)|^2\right)^{1/2} \left(\sum_{|n|\ge a} |\hat{f}(\boldsymbol{n})|^2\right)^{1/2} \\ &\le \|b_j\|_0 \left(\sum_{|n|\ge a} \frac{1}{a^2} |\boldsymbol{n}|^2 |\hat{f}(\boldsymbol{n})|^2\right)^{1/2} \le \frac{\|b_j\|_0}{a} \left(\sum_{|n|\ge a} |\boldsymbol{n}|^2 |\hat{f}(\boldsymbol{n})|^2\right)^{1/2} \\ &\le \|b_j\|_0 \|f\|_1 / \{(2\pi^2\alpha)^{1/2} a\}, \end{aligned}$$

since

$$2\pi^2\alpha \sum_{n} |\boldsymbol{n}|^2 |\hat{f}(\boldsymbol{n})|^2 \le \sum_{n} \sum_{j,j'} 2\pi^2 D_{jj'} n_j n_{j'} |\hat{f}(\boldsymbol{n})|^2 = \|f\|_1^2. \qquad \square$$

It follows from (4.16) and Lemma 4.1 that, at least under the uniform initial distribution,

$$\frac{1}{a\sqrt{t}}\{Y_j(t) - Y_j(0) - at(\bar{b}_j + \bar{\beta}_j)\} \simeq \frac{1}{\sqrt{t}} \int_0^t \operatorname{grad} g_j\{\ddot{Y}(s)\} \cdot \sqrt{\boldsymbol{D}}\, d\boldsymbol{B}(s). \tag{4.70}$$

The stochastic integral on the right is a martingale whose squared variation, averaged over time, is

$$\frac{1}{t} \int_0^t \operatorname{grad} g_j\{\ddot{Y}(s)\} \cdot \boldsymbol{D} \operatorname{grad} g_j\{\ddot{Y}(s)\}\, ds. \tag{4.71}$$

The expected value of (4.71), under equilibrium, is (see (4.18))

$$\int_{[0,1)^k} \operatorname{grad} g_j(\boldsymbol{y}) \cdot \boldsymbol{D} \operatorname{grad} g_j(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} = 2\,||g_j||_1^2. \tag{4.72}$$

It is, therefore, important to know how $||g_j||_1^2$ behaves as $a \to \infty$.

Lemma 4.6 As $a \to \infty$, $||g_j||_1^2 \to ||(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N||_1^2$, where f_N denotes the projection of an element f of H_1 onto the null space N of the operator $(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\cdot)) \cdot \nabla$ or, equivalently, of $\bar{\mathrm{S}} = \mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\cdot)) \cdot \nabla$.

Proof

Recall that $\bar{\mathrm{S}}$ is a compact skew symmetric operator on H^1 and has, therefore, a spectral representation (see Bhattacharya *et al.* 1989)

$$\bar{\mathrm{S}} f = \sum_{r=1}^{\infty} \mathrm{i}\lambda_r \langle f, \phi_r \rangle_1 \phi_r \tag{4.73}$$

where $\lambda_r, r \geq 1$, is a sequence of non-zero real numbers converging to zero, and $\phi_r, r \geq 1$, is a sequence of orthonormal eigenfunctions of $\bar{\mathrm{S}}$:

$$\bar{\mathrm{S}}\phi_r = \mathrm{i}\lambda_r \phi_r \qquad r \geq 1. \tag{4.74}$$

The sequence $\{\phi_r : r \geq 1\}$ is complete in $N^{\perp}$. Note that an arbitrary $f \in H^1$ has the representation

$$f = f_N + \sum_{r=1}^{\infty} \langle f, \phi_r \rangle_1 \phi_r, \tag{4.75}$$

from which (4.73) follows, using (4.74).

We now show that the solution h_j to the equation $\bar{\mathrm{A}}_a h_j = \beta_j(\cdot) - \bar{\beta}_j$ converges in H^1 to $(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j \cdot))_N$ (see (4.18)–(4.21), (4.60)–(4.62) to recall notation). One has by definition,

$$(\mathrm{I} + a\bar{\mathrm{S}})\, h_j = \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j), \tag{4.76}$$

so that, expanding h_j and $\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j)$, as in (4.75), we get

$$h_j = (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N + \sum_{r=1}^{\infty} \frac{1}{1 + \mathrm{i}a\lambda_r} \langle \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j), \phi_r \rangle_1 \phi_r.$$

Therefore,

$$||h_j - (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N||_1^2 = \sum_{r=1}^{\infty} \frac{1}{1 + a^2\lambda_r^2} \langle \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j), \phi_r \rangle_1^2$$
$$\to 0 \qquad \text{as } a \to \infty. \tag{4.77}$$

Next we compare g_j and h_j. Writing

$$(\mathrm{I} + a\mathrm{S}_a)\, g_j = \mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j + \beta_j(\cdot) - \bar{\beta}_j), \tag{4.78}$$

and comparing it with (4.76), one gets

$$(\mathrm{I} + a\mathrm{S}_a)(g_j - h_j) = \mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j) - a\mathcal{D}^{-1}(\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j. \tag{4.79}$$

Expanding $g_j - h_j$ in eigenfunctions (see (4.75)), (4.79) becomes

$$(g_j - h_j)_N = (\mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j))_N - a(\mathcal{D}^{-1}(\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j)_N, \tag{4.80}$$

and

$$\langle g_j - h_j, \phi_r \rangle_1 = \frac{1}{1 + \mathrm{i}a\lambda_r} \langle \mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j), \phi_r \rangle_1 - \frac{a}{1 + \mathrm{i}a\lambda_r} \langle \mathcal{D}^{-1}(\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j), \phi_r \rangle_1 \qquad r = 1, 2, \ldots. \tag{4.81}$$

The first term on the right is bounded by $(1 + a^2\lambda_r^2)^{-1/2} c/a$, by Lemma 4.5. For the second term, write

$$\langle \mathcal{D}^{-1}(\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j, \phi_r \rangle_1 = \langle (\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j, \phi_r \rangle_0 = \sum_{s=1}^{k} \left\langle \frac{\partial}{\partial x_s} \{(b_s(a\,\cdot) - \bar{b}_s) h_j\}, \phi_r \right\rangle_0 = -\sum_{s=1}^{k} \left\langle b_s(a\,\cdot) - \bar{b}_s, h_j^- \frac{\partial \phi_r}{\partial x_s} \right\rangle_0. \tag{4.82}$$

Since $\{h_j^-(\partial\phi_r/\partial x_s) : a = 1, 2, \ldots\}$ is relatively compact in H^0, by Lemma 4.1, it follows from Lemma 4.5 that the quantity in (4.82) is $o(1)$ as $a \to \infty$. Hence

$$\langle g_j - h_j, \phi_r \rangle_1 \to 0 \qquad \text{as } a \to \infty, r = 1, 2, \ldots. \tag{4.83}$$

Since $\{g_j : a = 1, 2, \ldots\}$ is H^1-bounded by Lemma 4.1, and $\{h_j : a = 1, 2, \ldots\}$ is also H^1-bounded by (4.77), $\{g_j - h_j : a = 1, 2, \ldots\}$ is relatively weakly compact in H^1. Therefore, (4.83) implies that $(g_j - h_j)_{N^\perp}$ converges to 0 weakly in H^1. Since S_a is skew symmetric one may write

$$\begin{aligned}
\|g_j\|_1^2 &= \langle g_j, (\mathrm{I} + a\mathrm{S}_a) g_j \rangle_1 = \langle g_j, \mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j + \beta_j(\cdot) - \bar{\beta}_j) \rangle_1 \\
&\simeq \langle g_j, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 \\
&= \langle g_j - h_j, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 + \langle h_j, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 \\
&\simeq \langle g_j - h_j, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 + \|(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N\|_1^2 \\
&\simeq \langle (g_j - h_j)_N, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 + \|(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N\|_1^2,
\end{aligned} \tag{4.84}$$

the symbol $\simeq$ indicating that the difference between the two sides goes to zero. The first approximation in (4.84) uses Lemma 4.5, the second uses (4.77), while the last approximation follows from the fact that $(g_j - h_j)_{N^\perp}$ converges to zero weakly in H^1. Next, using (4.80),

$$\begin{aligned}
&\langle (g_j - h_j)_N, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) \rangle_1 \\
&\quad = \langle \mathcal{D}^{-1}(b_j(a\,\cdot) - \bar{b}_j), (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \rangle_1 \\
&\qquad - a \langle \mathcal{D}^{-1}(\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j, (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \rangle_1 \\
&\quad \simeq -a \langle (\boldsymbol{b}(a\,\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla h_j, (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \rangle_0 \\
&\quad = -a \sum_{s=1}^{k} \left\langle \frac{\partial}{\partial x_s} \{(b_s(a\,\cdot) - \bar{b}_s) h_j\}, (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \right\rangle_0 \\
&\quad = a \sum_{s=1}^{k} \left\langle b_s(a\,\cdot) - \bar{b}_s, h_j^- \frac{\partial}{\partial x_s} (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \right\rangle_0.
\end{aligned} \tag{4.85}$$

Since $h_j^- \to (\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N^-$ in H^1 as $a \to \infty$, and $(\partial/\partial x_s)(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N \in H^1$, it follows that $\{h_j^-(\partial/\partial x_s)(\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j))_N : a = 1, 2, \ldots\}$ is relatively compact in H^1. Indeed, this follows from the general inequality

$$\| uv \|_1^2 \leq \frac{\Gamma}{\alpha}(\| u \|_0 \| v \|_1 + \| u \|_1 \| v \|_0)^2, \qquad u, v \in H^1. \tag{4.86}$$

Here Γ is the largest eigenvalue of the matrix $\boldsymbol{D}$, and α is the smallest. Thus invoking Lemma 4.5 again we see that the last expression in (4.85) is $o(1)$. The conclusion of the lemma now follows from (4.84). □

We are now ready to prove one of the main results of this paper.

Theorem 4.7 In addition to Assumptions 1–4 stated at the beginning of this section, assume that for some $p, 1 \leq p \leq k$, the functions

$$\{(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N : 1 \leq j \leq p\} \text{ are linearly independent in } H^1, \tag{4.87}$$

where f_N denotes the projection of f on the null space N of $\mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}) \cdot \nabla$. Then, if $t \gg a^2$, i.e.

$$\frac{t}{a^2} \to \infty \qquad \text{as } a \to \infty, \tag{4.88}$$

one has, uniformly with respect to all initial states $\boldsymbol{x} = \boldsymbol{X}(0)$, the following convergence for the vector of the first p coordinates of the solution of the Itô equation (3.1):

$$\left\{\frac{1}{a\sqrt{t}}(X_j(t) - x_j - t(\bar{b}_j + \bar{\beta}_j)) : 1 \leq j \leq p\right\} \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{\Sigma}_1), \tag{4.89}$$

where $\boldsymbol{\Sigma}_1 \equiv (\sigma_{ij})$ is given by

$$\sigma_{ij} = \langle(\mathcal{D}^{-1}(\beta_i - \bar{\beta}_i))_N, (\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N\rangle_1 + \langle(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N, (\mathcal{D}^{-1}(\beta_i - \bar{\beta}_i))_N\rangle_1 \qquad 1 \leq i, j \leq p. \tag{4.90}$$

Proof
We need to prove that an arbitrary linear combination with coefficients ξ_j of the random variables on the left-hand side of (4.89) converges in law to a normal distribution with mean zero and variance $\sum_{j,j'=1}^{p}$. To avoid messier notation, we will prove this convergence for the special case $\xi_j = 1, \xi_i = 0$ for $1 \leq i \neq j \leq p$. The general case is entirely analogous. Thus we wish to prove

$$\frac{1}{a^2\sqrt{t}}(X_j(a^2 t) - x_j - a^2 t(\bar{b}_j + \bar{\beta}_j)) \xrightarrow{\mathcal{L}} N(0, 2\| (\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N \|_1^2) \tag{4.91}$$

as

$$t \to \infty, a \to \infty. \tag{4.92}$$

The left-hand side of (4.91) equals (with $\boldsymbol{Y}(0) = \boldsymbol{x}/a$)

$$\begin{aligned}
&\frac{1}{a\sqrt{t}}\{Y_j(t) - Y_j(0) - at(\bar{b}_j + \bar{\beta}_j)\} \\
&= \frac{1}{\sqrt{t}}\int_0^t [b_j\{a\ddot{\boldsymbol{Y}}(s)\} + \beta_j\{\ddot{\boldsymbol{Y}}(s)\} - \bar{b}_j - \bar{\beta}_j]\,\mathrm{d}s + \frac{(\sqrt{\boldsymbol{D}}\boldsymbol{B})_j(a^2t)}{a^2\sqrt{t}} \\
&\simeq \frac{1}{\sqrt{t}}\int_0^t [b_j\{a\ddot{\boldsymbol{Y}}(s)\} + \beta_j\{\ddot{\boldsymbol{Y}}(s)\} - \bar{b}_j - \bar{\beta}_j]\,\mathrm{d}s \\
&= \frac{1}{\sqrt{t}}\int_0^t f\{\ddot{\boldsymbol{Y}}(s)\}\,\mathrm{d}s, \qquad (4.93)
\end{aligned}$$

say. The approximation '$\simeq$' (indicating that the difference goes to zero in probability) follows from the fact that $\boldsymbol{B}(a^2t)/a^2\sqrt{t} \stackrel{\mathcal{L}}{=} \boldsymbol{B}(1)/a$. Now, by Corollary 4.4 (inequality (4.53)), the last expression in (4.93) differs by a negligible quantity $O(1/\sqrt{t})$ from

$$\begin{aligned}
&\frac{1}{\sqrt{t}}\int_0^t (b_j\{a\ddot{\boldsymbol{Y}}(s)\} + \beta_j\{\ddot{\boldsymbol{Y}}(s)\} - \mathrm{E}_{\boldsymbol{x}/a}[b_j\{a\ddot{\boldsymbol{Y}}(s)\} + \beta_j\{\ddot{\boldsymbol{Y}}(s)\}])\,\mathrm{d}s \\
&= \frac{1}{\sqrt{t}}\int_0^t (f\{\ddot{\boldsymbol{Y}}(s)\} - \mathrm{E}_{\boldsymbol{x}/a}[f\{\ddot{\boldsymbol{Y}}(s)\}])\,\mathrm{d}s, \qquad (4.94)
\end{aligned}$$

say. Write

$$V_{r,a} := \int_{r-1}^{r} (f\{\ddot{\boldsymbol{Y}}(s)\} - \mathrm{E}_{\boldsymbol{x}/a}[f\{\ddot{\boldsymbol{Y}}(s)\}])\,\mathrm{d}s \qquad r = 1, 2, \ldots. \qquad (4.95)$$

We need to prove that, for every sequence of positive integers $\{\varphi(a) : a = 1, 2, \ldots\}$ going to infinity,

$$\frac{1}{\sqrt{\varphi(a)}}\sum_{r=1}^{\varphi(a)} V_{r,a} \xrightarrow{\mathcal{L}} N(0, 2\|(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N\|_1^2). \qquad (4.96)$$

Now the variance under equilibrium of the last expression in (4.94) is

$$\frac{2}{t}\int_0^t\int_0^s \overline{(f\ddot{T}_u f)}\,\mathrm{d}u \simeq 2\int_0^\infty \overline{f\ddot{T}_u f}\,\mathrm{d}u = -2\langle g_j, \mathrm{A}_a g_j\rangle_0 = 2\|g_j\|_1^2 \qquad (4.97)$$

by Corollary 4.4 (inequality (4.55)) and Bhattacharya (1982) (see Remark 2.4); the error of approximation '$\simeq$' is $O(1/t)$ uniformly in a, as $t \to \infty$. By inequality (4.54) of Corollary 4.4, the variance of (4.94) is estimated by $2\|g_j\|_1^2$, also with an error $O(1/t)$. By Lemma 4.6, $2\|g_j\|_1^2 \to 2\|(\mathcal{D}^{-1}(\beta_j - \bar{\beta}_j))_N\|_1^2 > 0$. The proof of (4.96) may now be completed by the classical method, using characteristic functions and forming approximately independent blocks omitting smaller blocks in between (see the proof of Theorem 4.9 below). A better proof follows from Theorem (2.10) of Götze and Hipp (1983) (see also Lahiri 1993), which provides refinements of the CLT. □

Remark 4.2

The requirement (4.87) holds if $\{\beta_j - \bar{\beta}_j : 1 \le j \le p\}$ are linearly independent elements of the null space of $(\bar{\boldsymbol{b}} + \boldsymbol{\beta})\cdot\nabla$ (see Bhattacharya *et al.* 1989, Lemma 3.1).

Remark 4.3

All the hypotheses of Theorem 4.7 are satisfied by Example 1 in Section 2, for $p = 1$. Here $k = 2$, $\boldsymbol{b}(\boldsymbol{x}) = (c_0 + c_1 \sin(2\pi x_2), 0)$, $\boldsymbol{\beta}(\boldsymbol{x}) = (c_2 \cos(2\pi x_2), 0)$, $\boldsymbol{D} = \mathbf{I}$. Note that $\bar{\boldsymbol{b}} = (c_0, 0)$, $\bar{\boldsymbol{\beta}}_1 = (0, 0)$, $\beta_1(\boldsymbol{x}) - \bar{\beta}_1 = c_2 \cos(2\pi x_2)$ is annihilated by $(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\boldsymbol{x})) \cdot \nabla \equiv \{c_0 + c_2 \cos(2\pi x_2)\}\partial/\partial x_1$. The asymptotic variance, given by (4.89) and (4.90), is

$$
\begin{aligned}
&-2\langle \cos(2\pi x_2), \mathcal{D}^{-1}\{c_2 \cos(2\pi x_2)\}\rangle_0 = -4c_2^2 \left\langle \cos(2\pi x_2), -\frac{\cos(2\pi x_2)}{4\pi^2} \right\rangle_0 \\
&\quad = \frac{c_2^2}{\pi^2} \langle \cos(2\pi x_2), \cos(2\pi x_2)\rangle_0 = \frac{c_2^2}{2\pi^2},
\end{aligned}
\tag{4.98}
$$

which accords with (2.6).

To prove the next main result we will need the following lemma.

Lemma 4.8 If $\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j)$ belongs to the range of $\bar{\mathrm{S}} = \mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}(\cdot)) \cdot \nabla$, then

$$
\sup\{a^2 \| g_j \|_1^2 : a = 1, 2, \ldots\} < \infty.
\tag{4.99}
$$

Proof

Let $p \in H^1$ be such that $\mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j) = \bar{\mathrm{S}}p$. Then, by (4.63),

$$
\begin{aligned}
\| g_j \|_1^2 &= \langle g_j, (\mathrm{I} + a\mathrm{S}_a) g_j \rangle_1 \\
&= \langle g_j, \mathcal{D}^{-1}(b_j(a\cdot) - \bar{b}_j)\rangle_1 + \langle g_j, \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j)\rangle_1 \\
&\le c \| g_j \|_1 \| b_j \|_0 / a + \langle g_j, \bar{\mathrm{S}}p\rangle_1.
\end{aligned}
\tag{4.100}
$$

Now write

$$
\begin{aligned}
\langle g_j, \bar{\mathrm{S}}p\rangle &= \langle g_j, \mathrm{S}_a p\rangle_1 - \langle g_j, (\mathrm{S}_a - \bar{\mathrm{S}})p\rangle_1 \\
&= -\langle \mathrm{S}_a g_j, p\rangle_1 - \langle g_j, \mathcal{D}^{-1}(\boldsymbol{b}(a\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla p\rangle_1 \\
&= -\frac{1}{a}\langle -g_j + \mathcal{D}^{-1}(b_j(a\cdot) - \bar{b}_j) + \mathcal{D}^{-1}(\beta_j(\cdot) - \bar{\beta}_j), p\rangle_1 - \langle g_j, \mathcal{D}^{-1}(\boldsymbol{b}(a\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla p\rangle_1 \\
&= \frac{1}{a}\langle g_j, p\rangle_1 - \frac{1}{a}\langle \mathcal{D}^{-1}(b_j(a\cdot) - \bar{b}_j), p\rangle_1 - \frac{1}{a}\langle \bar{\mathrm{S}}p, p\rangle_1 - \langle g_j, \mathcal{D}^{-1}(\boldsymbol{b}(a\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla p\rangle_1 \\
&\le c' \| g_j \|_1 / a + o\left(\frac{1}{a^2}\right) - o + c'' \| g_j \|_1 / a.
\end{aligned}
\tag{4.101}
$$

For the last inequality we have used Lemma 4.5, inequality (4.86) and the fact that $\partial p/\partial x_r \in H^1$, $r = 1, 2, \ldots, k$, to get

$$
\begin{aligned}
|\langle g_j, \mathcal{D}^{-1}(\boldsymbol{b}(a\cdot) - \bar{\boldsymbol{b}}) \cdot \nabla p\rangle_1 &= \left| \left\langle g_j, \sum_{r=1}^{k} (b_r(a\cdot) - \bar{b}_r) \frac{\partial p}{\partial x_r} \right\rangle_0 \right| \\
&\le \sum_{r=1}^{k} \left| \left\langle g_j \frac{\partial p^-}{\partial x_r}, b_r(a\cdot) - \bar{b}_r \right\rangle_0 \right| \le c'' \| g_j \|_1 / a.
\end{aligned}
\tag{4.102}
$$

Combining (4.100) and (4.101) we get

$$\| ag_j \|_1^2 \leq c''' \| ag_j \|_1 + o(1),$$

from which (4.99) follows. □

Theorem 4.9 below is stated somewhat differently from Theorem 4.7, in view of the fact that we are unable to assert convergence of $a^2 \| g_j \|_1^2$ as $a \to \infty$. That is, we do not have an analogue of Lemma 4.6. The result is, therefore, best stated as a normal approximation theorem rather than a convergence theorem. For this introduce the following distance ρ on the space $\mathcal{P}(\mathbb{R}^k)$ of all probability measures on (the Borel σ-field of) $\mathbb{R}^k$. Let $\mathcal{C}$ denote the class of all Borel measurable convex subsets of $\mathbb{R}^k$. Define

$$\rho(Q_1, Q_2) := \sup \{ | Q_1(C) - Q_2(C) | : C \in \mathcal{C} \}, \qquad \text{for } Q_1, Q_2 \in \mathcal{P}(\mathbb{R}^k). \tag{4.103}$$

For the statement of the theorem we will write $Q(a, t)$ for the distribution of the random vector

$$\left\{ \frac{1}{\sqrt{t}} \{ X_j(t) - X_j(0) - t(\bar{b}_j + \bar{\beta}_j) \} : p + 1 \leq j \leq k \right\}.$$

Theorem 4.9 Suppose in the hypothesis of Theorem 4.7 we replace (4.87) by the requirement that for some p, $0 \leq p \leq k - 1$,

$$\{ \mathcal{D}^{-1}(\beta_j - \bar{\beta}_j) : p + 1 \leq j \leq k \} \text{ are linearly independent elements of the range of } \mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}) \cdot \nabla. \tag{4.104}$$

Then, uniformly with respect to all initial states $\boldsymbol{X}(0) = \boldsymbol{x}$,

$$\rho(Q(a, t), N(0, \boldsymbol{\Sigma}_2)) \to 0, \tag{4.105}$$

provided $t \gg a^2$; here $\boldsymbol{\Sigma}_2 = (\eta_{ij})$ is a $(k - p) \times (k - p)$ matrix whose elements are given by

$$\eta_{ij} := \langle ag_i, ag_j \rangle_1 + \langle ag_j, ag_i \rangle_1 + D_{ij}, \qquad p + 1 \leq i, j \leq k, \tag{4.106}$$

and whose eigenvalues are bounded away from zero and infinity.

Proof

As in the case of the proof of Theorem 4.7, we will prove the appropriate convergence for one coordinate of $\boldsymbol{X}$, say X_j, instead of proving it for an arbitrary linear combination of X_js, since the proof for the latter case is entirely analogous to that for X_j.

Define the random variables

$$\begin{aligned} V_r \equiv V_{r,a} &:= X_j(r) - X_j(r-1) - (\bar{b}_j + \bar{\beta}_j) - \mathrm{E}(X_j(r) - X_j(r-1) - \bar{b}_j - \bar{\beta}_j \,|\, \boldsymbol{X}(0) = \boldsymbol{x}) \\ &= \int_{r-1}^{r} \{ b_j(\dot{\boldsymbol{X}}(s)) + \beta_j(\dot{\boldsymbol{X}}(s)/a) - \dot{T}_s(b_j + \beta_j(\cdot/a))(\boldsymbol{x}) \} \, \mathrm{d}s + (\sqrt{\boldsymbol{D}})_j (\boldsymbol{B}(r) - \boldsymbol{B}(r-1)) \\ &= \int_{r-1}^{r} \{ f(\dot{\boldsymbol{X}}(s)) - \dot{T}_s f(\boldsymbol{x}) \} \, \mathrm{d}s + (\sqrt{\boldsymbol{D}})_j (\boldsymbol{B}(r) - \boldsymbol{B}(r-1)), \end{aligned} \tag{4.107}$$

say. Here $\dot{T}_s$ is the transition operator for $\dot{\boldsymbol{X}}(\cdot)$,

$$(\dot{T}_s f)(\boldsymbol{x}) := \mathrm{E}(f(\dot{\boldsymbol{X}}(s)) \,|\, \dot{\boldsymbol{X}}(0) = \boldsymbol{x}), \tag{4.108}$$

and $(\sqrt{\boldsymbol{D}})_j$ denotes the jth row of $\sqrt{\boldsymbol{D}}$. We wish to prove asymptotic normality (in the sense of (4.105)) of

$$\frac{1}{\sqrt{\varphi(a)}}\left(\int_0^{\varphi(a)}[b_j\{\dot{\boldsymbol{X}}(s)\}+\beta_j\{\dot{\boldsymbol{X}}(s)/a\}-\bar{b}_j-\bar{\beta}_j]\,\mathrm{d}s+(\sqrt{\boldsymbol{D}})_j\boldsymbol{B}\{\varphi(a)\}\right)$$

$$=\frac{1}{\sqrt{\varphi(a)}}\sum_{r=1}^{\varphi(a)}\{X_j(r)-X_j(r-1)-\bar{b}_j-\bar{\beta}_j\} \tag{4.109}$$

as $a\to\infty$, for a sequence of integers $\varphi(a)$ such that

$$\delta(a):=\frac{\varphi(a)}{a^2}\to\infty. \tag{4.110}$$

First note that, in view of Remark 4.1, Corollary 4.4 is easily modified for functions of $\dot{\boldsymbol{X}}(s)$. If we continue to denote by E_x, var_x, cov_x the expectation, variance and covariance when the initial value is $\dot{\boldsymbol{X}}(0)=\boldsymbol{x}$ (recall that the equilibrium average $\bar{f}$ is with respect to the uniform distribution on $[0,a)^k$), and replace $\ddot{T}_t$ by $\dot{T}_t$, then Corollary 4.4 may be stated for functions of $\dot{\boldsymbol{X}}$, instead of $\ddot{\boldsymbol{Y}}$, by simply replacing the time variables $t, t-s$, on the right-hand sides of the inequalities (4.50)–(4.55) by t/a^2, $(t-s)/a^2$, respectively. In particular, it follows from the modified version of (4.53) that the asymptotic normality of (4.109) is equivalent to that of

$$\frac{1}{\sqrt{\varphi(a)}}\sum_{r=1}^{\varphi(a)}V_r, \tag{4.111}$$

i.e. the difference between (4.109) and (4.111) goes to zero if (4.110) holds. Here and elsewhere in this proof, unless stated otherwise, we consider the initial state fixed at $\boldsymbol{x}$.

Next, by Itô's lemma one obtains (4.7), so that the quantity (4.109) may be expressed as

$$\frac{X_j(1)-X_j(0)-\bar{b}_j-\bar{\beta}_j}{\sqrt{\varphi(a)}}+\frac{\gamma_j(\dot{\boldsymbol{X}}(\varphi(a)))-\gamma_j(\dot{\boldsymbol{X}}(1))}{\sqrt{\varphi(a)}}-\frac{1}{\sqrt{\varphi(a)}}\int_1^{\varphi(a)}[\operatorname{grad}\gamma_j\{\dot{\boldsymbol{X}}(s)\}-\mathbf{I}_j]\sqrt{\boldsymbol{D}}\,\mathrm{d}\boldsymbol{B}(s), \tag{4.112}$$

where $\mathbf{I}_j$ is the jth row of the identity matrix $\mathbf{I}$. The expected square of the first term goes to zero uniformly with respect to the initial value $\boldsymbol{X}(0)=\boldsymbol{x}$, as $a\to\infty$. The expected squared value of the second term is bounded by

$$\frac{2}{\varphi(a)}\mathrm{E}_x(\gamma_j^2[\dot{\boldsymbol{X}}\{\varphi(a)\}]+\gamma_j^2\{\dot{\boldsymbol{X}}(1)\}). \tag{4.113}$$

Now, by (4.45), (4.17) and Lemma 4.8,

$$\frac{1}{\varphi(a)}\mathrm{E}_x[\gamma_j^2\{\dot{\boldsymbol{X}}(1)\}]\le c''\frac{a^{-k}}{\varphi(a)}\int_{[0,a)^k}\gamma_j^2(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

$$=\frac{c''a^4}{\varphi(a)}\int_{[0,1)^k}g_j^2(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}=c''\frac{a^4\|g_j\|_0^2}{\varphi(a)}\le\frac{c'''a^4\|g_j\|_1^2}{\varphi(a)}$$

$$=\frac{c_3'''a^2\|g_j\|_1^2}{\delta(a)}\to 0\qquad\text{as } a\to\infty. \tag{4.114}$$

By the Markov property, $\mathrm{E}_x(\gamma_j^2[\dot{X}\{\varphi(a)\}]) \le c_3'''a^2\|g_j\|_1^2/\delta(a)$. Thus the expected square of the sum of the first two terms in (4.112) goes to zero uniformly with respect to x, as $a \to \infty$. The integral term has the expected squared value

$$\begin{aligned}
&\frac{1}{\varphi(a)}\int_1^{\varphi(a)} \dot{T}_{s-1}\dot{T}_1\left[\operatorname{grad}\gamma_j(\cdot)\cdot \boldsymbol{D}\operatorname{grad}\gamma_j(\cdot) - 2\sum_i D_{ji}\frac{\partial\gamma_j(\cdot)}{\partial x_i} + D_{jj}\right](x)\,\mathrm{d}s \\
&\simeq \frac{1}{\varphi(a)}\int_1^{\varphi(a)} \overline{\dot{T}_1[\;]}\,\mathrm{d}s \\
&= \frac{1}{\varphi(a)}\int_1^{\varphi(a)} (2a^2\|g_j\|_1^2 + D_{jj})\,\mathrm{d}s \simeq 2a^2\|g_j\|_1^2 + D_{jj}. \qquad (4.115)
\end{aligned}$$

The average under the integral sign for the first approximation ($\simeq$) in (4.115) is with respect to the uniform distribution on $[0,a)^k$. By (4.45), (4.17), and Lemma 4.8,

$$|\dot{T}_1[\operatorname{grad}\gamma_j(\cdot)\cdot \boldsymbol{D}\operatorname{grad}\gamma_j(\cdot)](y)| \le c''a^{-k}\int_{[0,a)^k}\operatorname{grad}\gamma_j(z)\cdot \boldsymbol{D}\operatorname{grad}\gamma_j(z)\,\mathrm{d}z = c''a^2\|g_j\|_1^2 \le c_1',$$

$$\begin{aligned}
&\left|\dot{T}_1\left\{2\sum_i D_{ji}\frac{\partial\gamma_j}{\partial x_i}(\cdot)\right\}(y)\right| \le 2c''a^{-k}\sum_i|D_{ji}|\int_{[0,a)^k}\left|\frac{\partial\gamma_j}{\partial z_i}\right|(z)\,\mathrm{d}z \\
&\le c_1''\int_{[0,1)^k} a\,|\operatorname{grad} g_j(z)|\,\mathrm{d}z \le c_1'''a\|g_j\|_1 \le c_2', \qquad (4.116)
\end{aligned}$$

where c_1', c_2' do not depend on a or y. Thus, from the modified version of inequality (3.50) in Corollary 4.4 mentioned earlier, the quantity $\overline{\dot{T}_1[\;]}$ appearing in the integrand in the second line of (4.115) differs from the integrand on the first line by no more than $c_2\exp\{-2\pi^2\alpha(s-1)/a^2\}$ for some constant c_2 which does not depend on a or x. Thus the error in the first approximation is no more than

$$\begin{aligned}
&\frac{1}{\varphi(a)}\int_1^{\varphi(a)} c_2\exp\{-2\pi^2\alpha(s-1)/a^2\}\,\mathrm{d}s \le \frac{c_2}{\delta(a)}\int_0^{\delta(a)}(1/2\pi^2\alpha)\,\mathrm{e}^{-u}\,\mathrm{d}u \\
&\le \frac{c_2/(2\pi^2\alpha)}{\delta(a)} \to 0 \qquad \text{as } a\to\infty. \qquad (4.117)
\end{aligned}$$

The validity of the second approximation in (4.115) is obvious. Hence,

$$\operatorname{var}_x\left(\frac{1}{\sqrt{\varphi(a)}}\sum_{r=1}^{\varphi(a)} V_r\right) \simeq 2a^2\|g_j\|_1^2 + D_{jj}, \qquad (4.118)$$

the difference going to zero uniformly with respect to x.

We will now prove the asymptotic normality of (4.111) by representing it approximately as the sum of a number of nearly independent block sums. For this write

$$\begin{aligned}
\eta(a) &:= [\delta^{1/4}(a)a^2], \\
\psi(a) &:= [\delta^{7/8}(a)a^2], \\
m(a) &:= \left[\frac{\varphi(a)}{\eta(a)+\psi(a)}\right] \sim \delta^{1/8}(a), \qquad (4.119)
\end{aligned}$$

where $[z]$ denotes the integer part of z, and '$\sim$' means that the ratio of its two sides goes to 1 (as $a \to \infty$). Define the block sums

$$Z_1 := \sum_{r=1}^{\psi(a)} V_r,\ Z_2 := \sum_{r=1}^{\psi(a)} V_{r+\psi(a)+\eta(a)}, \ldots, Z_{m(a)} := \sum_{r=1}^{\psi(a)} V_{r+(m(a)-1)(\psi(a)+\eta(a))}. \tag{4.120}$$

The omitted block sums are

$$\mathcal{E}_1 := \sum_{r=1}^{\eta(a)} V_{r+\psi(a)},\ \mathcal{E}_2 := \sum_{r=1}^{\eta(a)} V_{r+2\psi(a)+\eta(a)}, \ldots, \mathcal{E}_{m(a)} := \sum_{r=1}^{\eta(a)} V_{r+m(a)\psi(a)+(m(a)-1)\eta(a)}. \tag{4.121}$$

Then

$$\frac{1}{\sqrt{\varphi(a)}} \sum_{r=1}^{\varphi(a)} V_r \simeq \frac{1}{\sqrt{\varphi(a)}} \left(\sum_{r=1}^{m(a)} Z_r + \sum_{r=1}^{m(a)} \mathcal{E}_r \right), \tag{4.122}$$

and

$$\mathrm{E}_x \left(\frac{1}{\sqrt{\varphi(a)}} \sum_{r=1}^{m(a)} \mathcal{E}_r \right)^2 = \frac{1}{\varphi(a)} \left\{ \sum_{r=1}^{m(a)} \mathrm{var}_x \mathcal{E}_r + 2 \sum_{r=1}^{m(a)-1} \sum_{r'=1}^{m(a)-r} \mathrm{cov}_x(\mathcal{E}_r, \mathcal{E}_{r+r'}) \right\}. \tag{4.123}$$

Applying (4.118), with $\eta(a)$ in place of $\varphi(a)$, one has

$$\begin{aligned} \frac{1}{\varphi(a)} \sum_{r=1}^{m(a)} \mathrm{var}_x \mathcal{E}_r &\le c_3 \frac{m(a)}{\varphi(a)} (2a^2 \|g_j\|_1^2 + D_{jj}) \eta(a) \\ &\to 0 \end{aligned} \tag{4.124}$$

since $m(a)\eta(a)/\varphi(a) \to 0$ (and $a^2 \|g_j\|_1^2$ is bounded). Also,

$$\begin{aligned} (1/\varphi(a)) &\sum_{r=1}^{m(a)-1} \sum_{r'=1}^{m(a)-r} |\mathrm{cov}_x(\mathcal{E}_r, \mathcal{E}_{r+r'})| \\ &\le \{1/\varphi(a)\} \sum_{r=1}^{m(a)-1} \sum_{r'=1}^{m(a)-r} (\mathrm{var}_x \mathcal{E}_r)^{1/2} (\mathrm{var}_x \mathcal{E}_{r+r'})^{1/2} \\ &\le \{1/\varphi(a)\} m^2(a) (2a^2 \|g_j\|_1^2 + D_{jj}) \eta(a) \to 0. \end{aligned} \tag{4.125}$$

Also, by (4.118) (with $\psi(a)$ in place of $\varphi(a)$),

$$\frac{1}{\varphi(a)} \sum_{r=1}^{m(a)} \mathrm{var}\, Z_r - (2a^2 \|g_j\|_1^2 + D_{jj}) \simeq -\left(1 - \frac{\psi(a)\, m(a)}{\varphi(a)}\right) (2a^2 \|g_j\|_1^2 + D_{jj}) \to 0. \tag{4.126}$$

Finally, writing $f(\boldsymbol{x}) := \mathrm{E}_x \exp\{\mathrm{i}\xi Z_1/\sqrt{\varphi(a)}\}$, and using the analogue of (4.50),

$$\left|\mathrm{E}_x\left[\exp\left\{\mathrm{i}\xi\sum_{r=1}^{m(a)} Z_r/\sqrt{\varphi(a)}\right\}\right] - \mathrm{E}_x\left[\exp\left\{\mathrm{i}\xi\sum_{r=1}^{m(a)-1} Z_r/\sqrt{\varphi(a)}\right\}\right]\cdot \mathrm{E}[\exp\{\mathrm{i}\xi Z_{m(a)}/\sqrt{\varphi(a)}\}]\right|$$
$$= \left|\mathrm{E}_x\left[\exp\left\{\mathrm{i}\xi\sum_{r=1}^{m(a)-1} Z_r/\sqrt{\varphi(a)}\right\}\{\dot{T}_{\eta(a)} f(\dot{\boldsymbol{X}}[\{m(a)-1\}\{\psi(a)+\eta(a)\}-\eta(a)])-\bar{f}\}\right]\right|$$
$$\le c_4' \exp\{-2\pi^2\alpha\eta(a)/a^2\}. \tag{4.127}$$

In this manner one gets

$$\left|\mathrm{E}_x\left\{\exp\left(\mathrm{i}\xi\sum_{r=1}^{m(a)} Z_r/\sqrt{\varphi(a)} - \prod_{r=1}^{m(a)} \mathrm{E}[\exp\{\mathrm{i}\xi Z_r/\sqrt{\varphi(a)}\}]\right)\right\}\right|$$
$$\le c_4' m(a)\exp\{2\pi^2\alpha\eta(a)/a^2\} \to 0 \qquad \text{as } a\to\infty. \tag{4.128}$$

Combining (4.122), (4.123)–(4.126), (4.128), it follows that the asymptotic distribution of $(1/\sqrt{\varphi(a)})\sum_{r=1}^{\varphi(a)} V_r$, under the initial state $\boldsymbol{X}(0)=\boldsymbol{x}$, is the same as the sum of $m(a)$ i.i.d. random variables each having the same distribution as that of $Z_1/\sqrt{\varphi(a)}$ under equilibrium.

Remark 4.4

Condition (4.104) is equivalent to $\{\beta_j - \bar{\beta}_j : p+1 \le j \le k\}$ being linearly independent elements of the range of $(\bar{\boldsymbol{b}}+\boldsymbol{\beta})\cdot\nabla$.

Remark 4.5

It may appear, going through the proof of Theorem 4.9, that no use is made of the fact that the smallest eigenvalue of $\boldsymbol{\Sigma}_2$ is bounded by that of $\mathbf{D}$ from below. Since Lemma 4.8 does not assert convergence of the matrix with elements $\langle ag_i, ag_j\rangle_1 + \langle ag_j, ag_i\rangle_1$ to a non-singular matrix, the presence of $\boldsymbol{D}$ in (4.106) turns out in fact to be crucial for the validity of assertion (4.105) of the theorem. For example, if $\boldsymbol{\Sigma}_2$ converged to the matrix with all zero elements, (4.105) would not hold and one would not be able to derive a theorem such as Theorem 4.10 below.

Remark 4.6

In Example 2, Section 2, we have $k=2, \boldsymbol{b}(\boldsymbol{x}) = (c_0 + c_1\sin(2\pi x_2), \delta), \boldsymbol{\beta}(\boldsymbol{x}) = (c_2\sin(2\pi x_2), 0)$, $\boldsymbol{D}=\mathbf{I}, \bar{\boldsymbol{b}} = (c_0, \delta), \bar{\boldsymbol{\beta}} = (0,0); \beta_1(\boldsymbol{x}) - \bar{\beta}_1 \equiv c_2\sin(2\pi x_2)$ belongs to the range of $(\bar{\boldsymbol{b}}+\boldsymbol{\beta}(\cdot))\cdot\nabla \equiv \{c_0 + c_2\sin(2\pi x_2)\}\partial/\partial x_1 + \delta(\partial/\partial x_2)$. The asymptotic variance, given by (4.106), is $2a^2\|g_1\|_1^2 + 1$, where g_1 is the mean-zero periodic solution of (see (4.15))

$$\frac{1}{2}\left(\frac{\partial^2 g_1(\boldsymbol{x})}{\partial x_1^2} + \frac{\partial^2 g_1(\boldsymbol{x})}{\partial x_2^2}\right) + a(c_0 + c_1\sin(2\pi a x_2) + c_2\sin(2\pi x_2))\frac{\partial g_1(\boldsymbol{x})}{\partial x_1}$$
$$+ a\delta\frac{\partial g_1(\boldsymbol{x})}{\partial x_2} = c_1\sin(2\pi a x_2) + c_2\sin(2\pi x_2). \tag{4.129}$$

From this it follows that g_1 is a function of x_2 alone. Write $g_1(\boldsymbol{x}) = f(x_2)$. Then (4.129) becomes

$$\tfrac{1}{2}f''(x_2) + a\delta f'(x_2) = c_1 \sin(2\pi a x_2) + c_2 \sin(2\pi x_2),$$

or

$$f'(x_2) = \frac{c_1\delta}{a(\delta^2+\pi^2)}\left\{\sin(2\pi a x_2) - \frac{\pi\cos(2\pi a x_2)}{\delta}\right\} + \frac{c_2\delta}{\pi^2+a^2\delta^2}\left\{a\sin(2\pi x_2) - \frac{\pi}{\delta}\cos(2\pi x_2)\right\}.$$

Hence, for $a > 1$,

$$\begin{aligned}\|g_1\|_1^2 &= \frac{1}{2}\int_{[0,1)} (f'(x_2))^2\, dx_2 \\ &= \frac{1}{2}\left\{\frac{c_1^2\delta^2}{2a^2(\delta^2+\pi^2)^2} + \frac{c_1^2\pi^2}{2a^2(\delta^2+\pi^2)^2} + \frac{c_2^2\delta^2 a^2}{2(\pi^2+a^2\delta^2)^2} + \frac{c_2^2\pi^2}{2(\pi^2+a^2\delta^2)^2}\right\} \\ &= \frac{c_1^2}{4a^2(\delta^2+\pi^2)} + \frac{c_2^2}{4(\pi^2+a^2\delta^2)}.\end{aligned}$$

Therefore,

$$\begin{aligned}a^2\|g_1\|_1^2 &= \frac{c_1^2}{4(\delta^2+\pi^2)} + \frac{c_2^2 a^2}{4(\pi^2+a^2\delta^2)} \\ &\to \frac{c_1^2}{4(\delta^2+\pi^2)} + \frac{c_2^2}{4\delta^2}.\end{aligned}$$

Hence the asymptotic variance is as given by (2.9).

Finally, we may combine Theorems 4.7 and 4.9 to express the asymptotic normality of $\boldsymbol{X}(t)$ either in the form (4.105), or in the form of convergence in distribution to $N(\mathbf{0}, \mathbf{I})$ of an appropriate linear transformation. Here and in the statement below $\mathbf{I}$ is the $k \times k$ identity matrix.

Theorem 4.10 Under the hypotheses of Theorems 4.7 and 4.9,

$$\frac{1}{\sqrt{t}}\begin{pmatrix}\frac{1}{a}\boldsymbol{\Sigma}_1^{-1/2} & 0 \\ 0 & \boldsymbol{\Sigma}_2^{-1/2}\end{pmatrix}(\boldsymbol{X}(t) - \boldsymbol{x} - t(\bar{\boldsymbol{b}} + \bar{\boldsymbol{\beta}})) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}), \tag{4.130}$$

provided $t \gg a^2$, i.e. if

$$a \to \infty, \quad \frac{t}{a^2} \to \infty. \tag{4.131}$$

Proof

The method of proof of Theorem 4.9 works for an arbitrary linear combination of the coordinates of $\boldsymbol{X}(t)$. Also, since the null space and range of $\mathcal{D}^{-1}(\bar{\boldsymbol{b}} + \boldsymbol{\beta}) \cdot \nabla$ are orthogonal in H^1, it is simple to check that the asymptotic covariance between $t^{-1/2}(X_i(t) - x_i - t(\bar{b}_i + \bar{\beta}_i))$ and $t^{-1/2}(X_j(t) - x_j - t(\bar{b}_j + \bar{\beta}_j))$ is $o(a)$ if $1 \le i \le p, p+1 \le j \le k$. □

Remark 4.7
In Examples 1 and 2 (in Section 2) b_2, β_2 are constants, so that neither (4.87) nor (4.104) holds. Here $X_2(t)$ is Gaussian (i.e. a Brownian motion) at all times, and the method of proof of Theorem 4.9 still works for such cases. In general, one needs simply to augment the matrix $\mathbf{\Sigma}_2$ in order to include coordinates $X_r(t)$ for which β_r are constants.

Acknowledgements

Supported by the Alexander von Humboldt-Stiftung, and the Sonderforschungsbereich 343, 'Diskrete Strukturen in der Mathematik', Universität Bielefeld; and by NSF Grant DMS 9206937.

The authors gratefully acknowledge the careful reading of the manuscript by three anonymous referees and thank them for their helpful suggestions.

References

Anderson, M.P. (1979) Using models to simulate the movement of contaminants through ground water flow systems. *CRC Critical Reviews in Environmental Control*, **9**, 97–156.

Aris, R. (1956) On the dispersion of a solute in a fluid flowing through a tube. *Proc. Roy. Soc. London Ser. A*, **219**, 186–203.

Aronson, D.G. (1967) Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, **73**, 890–896.

Bensoussan, A., Lions, J.L. and Papanicolaou, G.C. (1978) *Asymptotic Analysis for Periodic Structures*. Amsterdam: North-Holland.

Bhattacharya, R.N. (1982) On the functional central limit theorem and the law of iterated logarithm for Markov processes. *Z. Wahrscheinlichkeitstheorie und Verw. Geb.*, **60**, 185–201.

Bhattacharya, R.N. (1985) A central limit theorem for diffusions with periodic coefficients. *Ann. Probab.*, **13**, 385–396.

Bhattacharya, R.N. and Gupta, V.K. (1983) A theoretical explanation of solute dispersion in saturated porous media at the Darey scale. *Water Resources Res.*, **19**, 938–944.

Bhattacharya, R.N. and Gupta, V.K. (1984) On the Taylor–Aris theory of solute dispersion in periodic porous media. *SIAM J. Appl. Math.*, **44**, 33–39.

Bhattacharya, R.N. and Ramasubramanian, S. (1988) On the central limit theorem for diffusions with almost periodic coefficients. *Sankhyā Ser. A*, **50**, 9–25.

Bhattacharya, R.N. and Ranga Rao (1976) *Normal Approximation and Asymptotic Expansions*. New York: Wiley.

Bhattacharya, R.N. and Waymire, E.C. (1990) *Stochastic Processes with Applications*. New York: Wiley.

Bhattacharya, R.N., Gupta, V.K., and Walker, H.F. (1989) Asymptotics of solute dispersion in periodic porous media. *SIAM J. Appl. Math.*, **49**, 86–98.

Billingsley, P. (1961) The Lindeberg–Levy theorem for martingales. *Proc. Amer. Math. Soc.*, **12**, 788–792.

Dagan, G. (1984) Solute transport in heterogeneous porous formations. *J. Fluid Mech.*, **145**, 151–177.

Diaconis, P. (1988) *Group Representations in Probability and Statistics*, IMS Lecture Notes. Hayward, CA: IMS.

Diaconis, P. and Stroock, D.W. (1991) Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, **1**, 36–61.

Einstein, A. (1905–8) *Investigations on the Theory of the Brownian Movement*. New York: Dover.

Ethier, S.N. and Kurz, T.G. (1986) *Markov Processes: Characterization and Convergence*. New York: Wiley.

Fill, J.A. (1991) Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, **1**, 62–87.

Fried, J.A. and Combarnous, M.A. (1971) Dispersion in porous media. *Adv. Hydrosci.*, **7**, 169–282.

Friedman, A. (1975) *Stochastic Differential Equations and Applications*, vol. 1. New York: Academic Press.

Gelhar, L.W. and Axness, C.L. (1983) Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resources Res.*, **19**, 161–180.

Gihman, I.I. and Skorohod, A.V. (1979) *Theory of Stochastic Processes*, vol. 3. New York: Springer-Verlag.

Götze, F. and Hipp, C. (1983) Asymptotic expansions of sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie und Verw. Geb.*, **64**, 211–239.

Götze, F. and Hipp, C. (1994) Asymptotic distributions of statistics in time series. *Ann. Statist*, **22**, 2062–2088.

Gupta, V.K. and Bhattacharya, R.N. (1986) Solute dispersion in multidimensional periodic saturated porous media. *Water Resources Res.*, **22**, 156–164.

Guven, O. and Molz, F.J. (1986) Deterministic and stochastic analysis of dispersion in an unbounded stratified porous medium. *Water Resources Res.*, **22**, 1565–1574.

Hall, P.G. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Applications*. New York: Academic Press.

Kozlov, S.M. (1979) Averaging operators with almost periodic, rapidly oscillating coefficients. *Math. USSR-Sb.*, **35**, 481–498.

Kozlov, S.M. (1980) Averaging of random operators. *Math. USSR-Sb.*, **37**, 167–180.

Lahiri, S.N. (1993) Refinements in asymptotic expansions for sums of weakly dependent random vectors. *Ann. Probab.*, **21**, 791–799.

Lallemond-Barres, A. and Peaudecerf, P. (1978) Recherche des relations entre la valeur de la disperivité microscopique d'un milieu aquifère, des autres caractéristiques et les conditions de mesure: étude bibliographique. *Bull. BRGM* Ser. 2, Sect. III, **4**, 277–284.

Molinary, J., Peaudecerf, P., Gaillard, B. and Launay, M. (1977) Essais conjoints en laboratoire et sur le terrain en vue, d'une approche simplifiée de la prevision des propagations des substances, miscibles dans les acquifers réels. *Proc. Symp. on Hydrodynamic Diffusion and Dispersion in Porous Media, IAHR, Pavia, Italy*, 89–102.

Papanicolaou, G.C. and Varadhan, S.R.S. (1979) Boundary problems with rapidly oscillating coefficients. *Colloq. Math. Soc. János Bolyai*, **27**, 835–875.

Rogers, L.C.G. and Williams, D. (1987) *Diffusions, Markov Processes, and Martingales*, vol. 2. New York: Wiley.

Sposito, G., Jury, W.A. and Gupta, V.K. (1986) Fundamental problems in the stochastic convection–dispersion model of solute transport in aquifers and fields. *Water Resources Res.*, **22**, 77–99.

Taylor, G.I. (1953) Dispersion of soluble matter in solvent flowing slowly through a tube. *Proc. Roy. Soc. London Ser. A*, **219**, 186–203.

Winter, C.L., Newman, C.M. and Neuman, S.P. (1984) A perturbation expansion for diffusion in a random velocity field. *SIAM J. Appl. Math.*, **44**, 425–442.

Received May 1994 and revised September 1994