Homogenization for deterministic maps and multiplicative noise

Georg A. Gottwald\textsuperscript{1} and Ian Melbourne\textsuperscript{2}

\textsuperscript{1}School of Mathematics and Statistics, University of Sydney, Sydney 2006 New South Wales, Australia
\textsuperscript{2}Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

A recent paper of Melbourne & Stuart (2011 A note on diffusion limits of chaotic skew product flows. Nonlinearity 24, 1361–1367 (doi:10.1088/0951-7715/24/4/018)) gives a rigorous proof of convergence of a fast–slow deterministic system to a stochastic differential equation with additive noise. In contrast to other approaches, the assumptions on the fast flow are very mild. In this paper, we extend this result from continuous time to discrete time. Moreover, we show how to deal with one-dimensional multiplicative noise. This raises the issue of how to interpret certain stochastic integrals; it is proved that the integrals are of Stratonovich type for continuous time and neither Stratonovich nor Itô for discrete time. We also provide a rigorous derivation of super-diffusive limits where the stochastic differential equation is driven by a stable Lévy process. In the case of one-dimensional multiplicative noise, the stochastic integrals are of Marcus type both in the discrete and continuous time contexts.

1. Introduction

There is considerable interest in understanding how stochastic behaviour emerges from deterministic systems, both in the mathematics and applications literature. A simple mechanism for emergent stochastic behaviour is via homogenization of multiscale systems [1].

Recently, Melbourne & Stuart [2] embarked on a programme to develop a rigorous theory of homogenization based on new ideas in the theory of dynamical systems. The aim is to avoid excessive mixing assumptions on the fast dynamics, because these are very difficult to establish even for uniformly hyperbolic (Axiom A or Anosov) flows (general references for the
ergodic theory of uniformly hyperbolic maps and flows include [3–6]). Instead, the theory relies only on relatively mild statistical properties that are known to hold very widely and are independent of mixing assumptions (see [2] or remark 2.1 below for further details).

The mechanism for emergent stochastic behaviour in deterministic systems, whereby fast chaotic dynamics induces white noise in the slow variables, is much studied in the applied literature [1,7–10]. See also the programme outlined by Mackay [11]. The aim here, continuing and extending the work in Melbourne & Stuart [2], is to obtain rigorous results for large classes of fast–slow systems under unusually mild assumptions.

In particular, Melbourne & Stuart [2] studied fast–slow ODEs of the form

\[ \begin{align*}
\dot{x}(\epsilon) &= \epsilon^{-1}f_0(y(\epsilon)) + f(x(\epsilon),y(\epsilon)), \quad x(\epsilon)(0) = \xi, \\
\dot{y}(\epsilon) &= \epsilon^{-2}g(y(\epsilon)), \quad y(\epsilon)(0) = \eta,
\end{align*} \]

(1.1)

where \( x(\epsilon) \in \mathbb{R}^d, y(\epsilon) \in \mathbb{R}^\ell, \epsilon > 0 \). It is assumed that the vector fields \( f_0 : \mathbb{R}^\ell \to \mathbb{R}^d, f : \mathbb{R}^d \times \mathbb{R}^\ell \to \mathbb{R}^d \) and \( g : \mathbb{R}^\ell \to \mathbb{R}^\ell \) satisfy certain regularity conditions and that the fast \( y \) dynamics possesses a compact attractor \( \Lambda \subset \mathbb{R}^\ell \) with ergodic invariant probability measure \( \mu \) satisfying certain mild chaoticity assumptions. Finally, it is required that \( \int_{\Lambda} f_0 \, d\mu = 0 \), so that we are in the situation of homogenization rather than averaging. The conclusion [2] is that \( x(\epsilon) \to w X \) in \( C([0, \infty), \mathbb{R}^d) \) as \( \epsilon \to 0 \), where \( X \) is the solution to a stochastic differential equation (SDE) of the form

\[ \text{d}X = \sqrt{\Sigma} \, \text{d}W + F(X) \, \text{d}t, \quad X(0) = \xi. \] (1.2)

Here, \( W \) is unit \( d \)-dimensional Brownian motion and \( F(x,y) = \int_{\Lambda} f(x,y) \, d\mu(y) \). (Throughout, we use \( \to w \) to denote weak convergence in the sense of probability measures [12].)

In this paper, we consider the twin goals of (i) allowing multiplicative noise when \( d = 1 \), and (ii) proving analogous results for discrete time to those for flows.

In §1, we will focus on the case of discrete time. First, we consider the case where there is no multiplicative noise. Consider the equation

\[ \begin{align*}
x^{(\epsilon)}(n+1) &= x^{(\epsilon)}(n) + \epsilon f_0(y(n)) + \epsilon^2 f(x^{(\epsilon)}(n), y(n), \epsilon), \quad x^{(\epsilon)}(0) = \xi,
\end{align*} \] (1.3)

where \( x^{(\epsilon)}(n) \in \mathbb{R}^d \), and the fast variables \( y(n) \) are generated by a map \( g : \mathbb{R}^\ell \to \mathbb{R}^\ell \) with compact attractor \( \Lambda \subset \mathbb{R}^\ell \) and ergodic invariant measure \( \mu \). We require that \( \Lambda \) is mildly chaotic as in Melbourne & Stuart [2]. That is, we assume a weak invariance principle (WIP) and a large deviation principle (LDP). The precise definitions are recalled in §2. A consequence is that \( n^{-1/2} \sum_{j=0}^{n-1} f_0(y(j)) \) converges in distribution as \( n \to \infty \) to a \( d \)-dimensional normal distribution with mean zero and \( d \times d \) covariance matrix \( \Sigma \).

Define \( \hat{x}^{(\epsilon)}(t) = x^{(\epsilon)}(te^{-2}) \) for \( t = 0, \epsilon^2, 2\epsilon^2, \ldots \) and linearly interpolate to obtain \( \hat{x}^{(\epsilon)} \in C([0, \infty), \mathbb{R}^d) \).

Our first main result is a direct analogue of the continuous time result of Melbourne & Stuart [2, theorem 1.1].

**Theorem 1.1.** Consider equation (1.3). Assume that \( f_0 \) and \( f \) are locally Lipschitz in \( x \) and \( y \), and that \( \lim_{t \to 0} f(x, y, \epsilon) = f(x, y, 0) \) uniformly on compact subsets of \( \mathbb{R}^d \times \Lambda \). Assume that \( \Lambda \) satisfies the WIP and LDP as described below. Suppose that \( \int_{\Lambda} f_0 \, d\mu = 0 \) and set \( F(x) = \int_{\Lambda} f(x, y, 0) \, d\mu(y) \).

Suppose that solutions \( X \) to the SDE (1.2) exist on \([0, \infty)\) with probability one. Then, \( \hat{x}^{(\epsilon)} \to w X \) in \( C([0, \infty), \mathbb{R}^d) \) as \( \epsilon \to 0 \).

**Remark 1.2.** (a) In Melbourne & Stuart [2], it was assumed that \( f \) (and hence \( F \)) is globally Lipschitz and so the condition on global existence for \( X \) is automatic. In the course of this paper, such global conditions become rather excessive, so we relax them from the outset (see §3a).

(b) In the generality of this paper, including theorem 1.1, it is not possible to write down a formula for the covariance matrix \( \Sigma \). However, in many situations, including the Young tower
situation \([13,14]\) discussed in \(\S 2\), it is possible to prove convergence of second moments \([15]\)
leading to the expression
\[
\Sigma = \lim_{n \to \infty} n^{-1} \sum_{j=0}^{n-1} \left( \sum_{j=0}^{n-1} f_0(y(j)) \right)^T d\mu. \tag{1.4}
\]

Under the additional assumption of summable decay of correlations (which is valid for Young
towers that satisfy the WIP and are mixing), we obtain the well-known Green–Kubo formula
\[
\Sigma = \int_A f_0 f_1^T d\mu + \sum_{n=1}^{\infty} \int_A f_0(y(n)) f_0(y(0)) d\mu + \sum_{n=1}^{\infty} \int_A f_0(y(0)) f_0(y(n))^T d\mu. \tag{1.5}
\]
In particular, we note that (1.4) is valid for general uniformly hyperbolic maps, and (1.5) is valid
under the additional assumption that the map is mixing.

(We have written formula (1.5), so that it applies equally to invertible and non-invertible maps.
Of course in the case of invertible maps \(\Sigma = \sum_{n=1}^{\infty} \int_A f_0(y(n)) f_0(y(0))^T d\mu\).

Next, we turn to the case of multiplicative noise in the discrete time setting with \(d = 1\). Consider
the equation
\[
x^{(e)}(n + 1) = x^{(e)}(n) + \epsilon h(x^{(e)}(n)) f_0(y(n)) + \epsilon^2 f(x^{(e)}(n), y(n), \epsilon), \tag{1.6}
\]
where \(x^{(e)}(n) \in \mathbb{R}\) and the fast variables \(y(n)\) are generated as described earlier.

As \(\epsilon \to 0\), we expect that \(\hat{x}^{(e)}(t) = x^{(e)}(t \epsilon^{-2})\) converges weakly to solutions \(X\) of an SDE of the form
\[
dX = h(X) \, dW + F(X) \, dt,
\]
but there is the issue of how to interpret the stochastic integral \(\int h(X) \, dW\). In the continuous
time setting, one expects the limiting SDE to be Stratonovich \([16,17]\), and this is indeed the case
(see theorem 3.3). However, the discrete time case is very different, as shown by Givon &
Kupferman \([18]\). They considered the special case where \(h(x) = x\) and \(f(x, y, \epsilon) = \lambda x \) (\(\lambda\) constant),
and showed that, in general, the stochastic integral is neither Itô nor Stratonovich except in the
special case where \(y(n)\) is an iid sequence—in that case the integral is Itô. Their proof exploited
the linearity in a crucial way. Here, we extend their results, relaxing linearity and allowing \(f\) to
depend on \(y\).

**Theorem 1.3.** Let \(d = 1\) and consider equation (1.6). Assume that \(f_0\) and \(f\) are locally Lipschitz in \(x\) and
\(y\), and that \(\lim_{\epsilon \to 0} f(x, y, \epsilon) = f(x, y, 0)\) uniformly on compact subsets of \(\mathbb{R} \times A\). Assume that \(h : \mathbb{R} \to \mathbb{R}\)
is \(C^1\) and non-vanishing. Assume that \(A\) satisfies the WIP and LDP as described below. Suppose that
\(\int_A f_0 \, d\mu = 0\) and set \(F(x) = \int_A f(x, y, 0) \, d\mu(y)\).

Consider the Stratonovich SDE
\[
dX = \sigma h(X) \circ dW + \left( F(X) - \frac{1}{2} h(X) h'(X) \int_A f_0^2 \, d\mu \right) dt, \quad X(0) = \xi, \tag{1.7}
\]
where \(W\) is unit one-dimensional Brownian motion. Suppose that solutions \(X\) to the SDE exist on \([0, \infty)\)
with probability one. Then, \(\hat{x}^{(e)} \to _w X\) in \(C([0, \infty), \mathbb{R})\) as \(\epsilon \to 0\).

**Remark 1.4.** Note as in Givon & Kupferman \([18]\) that the correction term in this SDE is Itô
if and only if \(\sigma^2 = \int_A f_0^2 \, d\mu\). This holds in the independent case but is generally false. (For
example, the Green–Kubo formula (1.5) specialized to the case \(d = 1\) yields \(\sigma^2 = \int_A f_0^2 \, d\mu +
2 \sum_{n=1}^{\infty} \int_A f_0(y(n)) f_0(y(0)) \, d\mu\).)

See Givon & Kupferman \([18, \text{section 6}]\) and also references \([19,20]\) for discussions concerning
corrections that are neither Itô nor Stratonovich.

The assumption that \(h\) is non-vanishing means that we can write \(h = 1/r'\), where \(r\) is a
monotone differentiable function. By a change in variables, \(Z = r(X)\), it is then possible to reduce
to the situation where there is no multiplicative noise, see §§3b and 4b. In the process, the drift term \( F(X) \) in the SDE is transformed into \( \tilde{F}(Z) \), where
\[
\tilde{F} = \left( \frac{F}{h} \right) \circ r^{-1},
\]
(see proposition 3.4).

These considerations lead to an extension of theorem 1.3, where \( h \) is not required to be non-vanishing. Instead, we require only that there is a monotone differentiable function \( r \) such that \( r' = 1/h \) and such that \( \xi = X(0) \) lies in the domain of \( r \). (It suffices that \( h(\xi) \neq 0 \), in which case we can choose \( r(x) = \int_0^x 1/h(y) \, dy \) for \( x \) near \( \xi \).) Define \( \tilde{F} = (F/h) \circ r^{-1} \) whenever this expression makes sense.

**Proposition 1.5.** Assume the same set-up as in theorem 1.3 except that \( h \) is not assumed to be non-vanishing. Consider the SDE
\[
dZ = \sigma \, dW + \tilde{F}(Z) \, dt, \quad Z(0) = r(\xi).
\]
Suppose that solutions \( Z \) to the SDE exist on \([0, \infty)\) with probability one. Then, solutions \( X \) to the SDE (1.7) exist on \([0, \infty)\) with probability one, and \( \tilde{x}^{(\varepsilon)} \to_w X \) in \( C([0, \infty), \mathbb{R}) \) as \( \varepsilon \to 0 \).

For example, suppose that \( h(x) = x \) and \( f(x, y) = xq(x, y) \), \( F(x) = xQ(x) \). If \( \xi > 0 \), then we choose \( r(x) = \log x \) yielding \( \tilde{F}(z) = Q(e^z) \) which is well-behaved in many situations (e.g. \( Q(x) \) constant, \( Q(x) = -x^p \), \( p \geq 0 \)). The case \( \xi < 0 \) is similar with \( r(x) = \log(-x) \).

**Remark 1.6** (higher-dimensional multiplicative noise). A more general setting that includes both situations described earlier is where the slow equations have the form
\[
x^{(\varepsilon)}(n+1) = x^{(\varepsilon)}(n) + \varepsilon h(x^{(\varepsilon)}(n))f_0(y(n)) + \varepsilon^2 f(x^{(\varepsilon)}(n), y(n), \varepsilon), \quad x^{(\varepsilon)}(0) = \xi,
\]
where \( x^{(\varepsilon)}(n) \in \mathbb{R}^d \), \( f_0 : \Lambda \to \mathbb{R}^d \), \( f : \mathbb{R}^d \times \Lambda \times \mathbb{R} \to \mathbb{R}^d \) and \( h : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^d) \). Theorem 1.1 deals with the case \( h(x) = I_d \), \( d \) general, and theorem 1.3 covers the case \( d = 1 \), \( h \) general. It is well known that multiplicative noise presents more serious problems in higher dimensions, and it will be necessary to make assumptions beyond the WIP and LDP in general. This is the subject of future work. However, the methods in this paper generalize immediately to the higher-dimensional situation whenever \( h \) has the particular form \( h = (Dr)^{-1} \) for some \( C^2 \) diffeomorphism \( r : \mathbb{R}^d \to \mathbb{R}^d \).

Finally, we consider the case where the fast dynamics is not sufficiently chaotic to support the central limit theorem. In this case, we can still hope to prove homogenization theorems, but the limiting SDE is driven by a stable Lévy process. Again, we obtain results for ODEs and maps, with additive noise in general dimensions and multiplicative noise for \( d = 1 \). The interpretation of the stochastic integral in the limiting SDE is the same for both continuous and discrete time, and is of Marcus type (see §5 for further details.)

The remainder of this paper is organized as follows. In §2, we recall the setting of Melbourne & Stuart [2] as regards the chaoticity assumptions on the fast dynamics, though now in the context of discrete time. In §3, we consider fast–slow ODEs, relaxing the uniformity of the Lipschitz conditions in Melbourne & Stuart [2] and permitting multiplicative noise for \( d = 1 \) (with a restricted extension to higher dimensions). In §3, we consider fast–slow maps and prove the results stated in this introduction. In §5, we state and prove results where the limiting SDE is driven by a stable Lévy process. In §6, we present some numerical results. Conclusions are given in §7.

# 2. Assumptions on the fast dynamics

Melbourne & Stuart [2] made mild assumptions on the fast dynamics that are satisfied by large classes of dynamical systems. The formulation there is for continuous time. Here, we discuss discrete time (both contexts are required in this paper).


Let $g: \Lambda \to \Lambda$, where $\Lambda$ is a compact subset of $\mathbb{R}^\ell$, and $\mu$ is an ergodic invariant measure supported on $\Lambda$. Given $y_0 = \eta \in \Lambda$, we define the fast variables $y(n)$, $n \geq 0$, by setting $y(n+1) = g(y(n))$.

Let $f_0: \Lambda \to \mathbb{R}^d$ be a Lipschitz observable of mean zero. Define $W_n(t) = n^{-1/2} \sum_{j=0}^{n-1} f_0(y(j))$ for $t = 0, 1/n, 2/n, \ldots$ and linearly interpolate to obtain a continuous function $W_n: [0, \infty) \to \mathbb{R}^d$. We assume the weak invariance principle (WIP), namely that $W_n \to \sqrt{\Sigma} W$ in $C([0, \infty), \mathbb{R}^d)$, where $W$ is unit $d$-dimensional Brownian motion and $\Sigma$ is a $d \times d$ covariance matrix.

Next, let $f: \mathbb{R}^d \times \Lambda \to \mathbb{R}^d$ be globally Lipschitz and define $F: \mathbb{R}^d \to \mathbb{R}^d$, $F(x) = \int_{\Lambda} f(x, y) \, d\mu(y)$. We assume the following large deviation principle (LDP):

$$
\mu \left( \left\{ \frac{1}{N} \sum_{j=0}^{N-1} f(x, y(j)) - F(x) \right\} > a \right) \leq b(a, N),
$$

where $b(a, N)$ is independent of $x$ and $b(a, N) \to 0$ as $N \to \infty$ for all $a > 0$.

**Remark 2.1.** As discussed in Melbourne & Stuart [2, remark 1.3(a)], the WIP and LDP hold for a large class of maps and flows. These include, but go far beyond, Axiom A diffeomorphisms and flows, Hénon-like attractors and Lorenz attractors. Young [13,14] introduced a class of non-uniformly hyperbolic maps with exponential and polynomial decay of correlations. For maps, the WIP holds when the correlations are summable, and the LDP holds in all cases. For flows, it suffices that there is a Poincaré map with these properties and then the WIP and LDP lift to these flows (irrespective of the mixing properties of the flow). Precise statements about the validity of the WIP and LDP can be found in references [21–23].

These mild chaoticity assumptions are used throughout this paper in exactly the same way as in Melbourne & Stuart [2]. For example, in the situation of (1.1), the WIP suffices when the slow coupling term $f(x^{(e)}, y^{(e)})$ is independent of $y^{(e)}$. (The argument can be found in Melbourne & Stuart [2] and is also present in the proof of theorem 1.1 in §4a.) The LDP is required to control the $y^{(e)}$ dependence in $f(x^{(e)}, y^{(e)})$, see lemma A.2.

### 3. Extensions of the results for flows

In this section, we extend the results in Melbourne & Stuart [2] by (i) relaxing the global Lipschitz conditions and (ii) allowing multiplicative noise.

#### (a) Relaxing the global Lipschitz condition on $f$

Consider the fast–slow system (1.1). We suppose throughout that the fast equations possess a ‘mildly chaotic’ compact attractor $\Lambda \subset \mathbb{R}^\ell$ satisfying the WIP and LDP. It is natural to assume that $f_0: \Lambda \to \mathbb{R}^d$, $f: \mathbb{R}^d \times \Lambda \to \mathbb{R}^d$ and $g: \Lambda \to \Lambda$ are locally Lipschitz to ensure existence and uniqueness of solutions to the various initial value problems arising above. Boundedness and uniformity of Lipschitz constants on $\Lambda$ then follows from compactness. However, in Melbourne & Stuart [2], it is further assumed (mainly for simplicity) that $f$ is bounded with a uniform Lipschitz constant on the whole of $\mathbb{R}^d \times \Lambda$.

Here, we show that the result of Melbourne & Stuart [2] holds without the global Lipschitz condition provided solutions to the limiting SDE exist for all time with probability one. The formulation of Melbourne & Stuart [2, theorem 1.1] is unchanged if in addition solutions to the fast–slow system exist for all time for $\mu$-almost every $\eta$. Otherwise, we require the following modification. Throughout, we regard $\xi$ as fixed.

Let $[0, \tau_e]$ be the maximal interval of existence for a solution $x^{(e)}$ and define

$$
x^{(e)}_w(t) = \begin{cases} 
x^{(e)}(t), & 0 \leq t \leq \frac{\tau_e}{2} \\
x^{(e)}\left(\frac{\tau_e}{2}\right), & t \geq \frac{\tau_e}{2}.
\end{cases}
$$
(If $x(\epsilon)$ exists on $[0, \infty)$, then set $x_{\epsilon}^{(e)} \equiv x^{(e)}$.) We say that $x^{(e)}$ converges weakly to $X$ in $C([0, \infty), \mathbb{R}^d)$ if $\tau_\epsilon \to \infty$ in probability and $x_{\epsilon}^{(e)}$ converges weakly to $X$ in $C([0, \infty), \mathbb{R}^d)$.

**Theorem 3.1.** Assume that the fast equation (with $\epsilon = 1$) has a mildly chaotic compact invariant set $\Lambda$ with invariant ergodic probability measure $\mu$. Suppose that $f_0 : \Lambda \to \mathbb{R}^d$ and $f : \mathbb{R}^d \times \Lambda \to \mathbb{R}^d$ are locally Lipschitz, and that $\int_A f_0 \, d\mu = 0$. Define $F(x) = \int_A f(x, y) \, d\mu(y)$ and let $\xi \in \mathbb{R}^d$.

Let $x^{(e)}, y^{(e)}$ denote the solutions to the fast–slow system (1.1). Assume that solutions $X$ to the SDE (1.2) exist on $[0, \infty)$ with probability one. Then, $x^{(e)} \to_w X$ in $C([0, \infty), \mathbb{R}^d)$ as $\epsilon \to 0$.

In the remainder of this section, we prove theorem 3.1 by reducing it to the situation in Melbourne & Stuart [2]. The ideas are standard, but care has to be taken, because we are talking about weak convergence of solutions, and solutions to the fast-slow equation can blow up arbitrarily quickly.

Let $R > 0$ and define $f_R : \mathbb{R}^d \times \Lambda \to \mathbb{R}^d$ to be a globally Lipschitz function that agrees with $f$ on $B_R(\xi) = \{ x \in \mathbb{R}^d : |x - \xi| \leq R \}$ (where $|x - \xi|$ is Euclidean distance). Let $x^{(e),R}, y^{(e)}$ denote solutions to (1.1) with $f$ replaced by $f_R$. These solutions exist for all time; in particular, $x^{(e),R} \in C([0, \infty), \mathbb{R}^d)$. Similarly, let $X^R$ denote the solution to the SDE (1.2) with $F$ replaced by $F_R$, where $F_R(x) = \int_A f_R(x, y) \, d\mu(y)$. (Note that $\Sigma$ depends only on $f_0$ and $g$ and hence is independent of $R$.) By Melbourne & Stuart [2], $x^{(e),R} \to_w X^R$ in $C([0, \infty), \mathbb{R}^d)$ as $\epsilon \to 0$.

Next, we define stopping times $\tau_R, \tau_{\epsilon,R} \in (0, \infty) \cup \{ \infty \}$, where $\tau_R \geq 0$ is least such that $X^R \in \partial B_R(\xi) = \{ x \in \mathbb{R}^d : |x - \xi| = R \}$ and $\tau_{\epsilon,R} \geq 0$ is least such that $x^{(e),R} \in \partial B_R(\xi)$. By construction, $X^R \equiv X$ on $[0, \tau_R]$ and $x^{(e),R} \equiv x^{(e)}$ on $[0, \tau_{\epsilon,R}]$.

**Proposition 3.2.** $\tau_\epsilon \to \infty$, $\tau_R \to \infty$ and $\tau_{\epsilon,R} \to \infty$ in probability as $R \to \infty$, $\epsilon \to 0$.

**Proof.** By assumption $\tau_\epsilon \to \infty$ almost surely, and hence in probability, as $R \to \infty$. In other words, for any $T > 0$, $\delta > 0$, there exists $R$ such that $P(\tau_R > T) > 1 - \delta$.

Next, for fixed $R > 0$, given $v \in C([0, \infty], \mathbb{R})$ we set $\psi(v) = \sup \{ t \geq 0 : \eta(t) \in B_R(v(0)) \} \in (0, \infty) \cup \{ \infty \}$. This defines a continuous map $\psi$. Moreover, $\tau_{\epsilon,R} = \psi(x^{(e),R})$ and $\tau_R = \psi(X^R)$, so it follows from the continuous mapping theorem that $\tau_{\epsilon,R} \to_d \tau_R$ as $\epsilon \to 0$.

In particular, with $T$ and $R$ as in the first paragraph, we have that there exists $\epsilon_0 > 0$ such that $\mu(\tau_{\epsilon,R} > T) > P(\tau_R > T) - \delta > 1 - 2\delta$ for all $\epsilon \in (0, \epsilon_0)$. Altogether, we have shown that for any $T > 0$, $\delta > 0$, there exists $R$ and $\epsilon_0$ such that $\mu(\tau_{\epsilon,R} > T) > 1 - 2\delta$ for all $\epsilon \in (0, \epsilon_0)$, so $\tau_{\epsilon,R} \to \infty$ in probability.

Because $\tau_\epsilon > \tau_{\epsilon,R}$ for all $R$, it follows immediately that $\tau_\epsilon \to \infty$ in probability.

Now fix $T > 0$. By proposition 3.2, for any $\delta > 0$, there exists $\epsilon_0$ and $R$ such that $\mu(\tau_{\epsilon,R} > 2T) > 1 - \delta$ for all $\epsilon \in (0, \epsilon_0)$. We choose $R$ so that in addition $\mu(\tau_R > 2T) > 1 - \delta$.

For this choice of $R$, let $X_T = X^R$ and $x_T^{(e)} = x^{(e),R}$. Then, we have defined families of random elements $X_T$ and $x_T^{(e)} \in C([0, \infty), \mathbb{R}^d)$ such that $x_T^{(e)} \to_w X_T$. Moreover, neglecting a set of measure $\delta$, we have $\tau_T \geq \tau_{\epsilon,R} > 2T$ so $\tau_T/2 > T$ and hence $x_T^{(e)} \equiv x_T^{(e)} \equiv x_T^{(e)}$ on $[0, T]$. Finally, neglecting a set of measure $\delta$, we have $X_T \equiv X$. Hence $x_T^{(e)}$ converges weakly to $X$ in $C([0, T), \mathbb{R}^d)$. Because $T$ is arbitrary, $x_T^{(e)} \to_w X$ in $C([0, \infty), \mathbb{R}^d)$.

(b) Flows with multiplicative noise

Next, we consider the case of multiplicative noise when $d = 1$. Consider the fast–slow system

\[
\begin{align*}
x^{(e)}(t) &= \epsilon^{-1} h(x^{(e)}(t)) y_0(y^{(e)}(t)) + f(x^{(e)}(t), y^{(e)}(t)), \\
y^{(e)}(t) &= \epsilon^{-2} g(y^{(e)}(t)), \quad y^{(e)}(0) = \xi,
\end{align*}
\]

with $x^{(e)} \in \mathbb{R}$, $y^{(e)} \in \mathbb{R}$.

**Theorem 3.3.** (a) Assume that $g, f_0, f, F, W$ and $\Sigma$ are as in theorem 3.1 (but with the restriction that $d = 1$ and $\Sigma$ is denoted by $\sigma^2 > 0$). Suppose that $h : \mathbb{R} \to \mathbb{R}$ is $C^1$ and non-vanishing.
Let $\xi \in \mathbb{R}$ and consider the Stratonovich SDE
\[
dX = \sigma h(X) \circ dW + F(X) \, dt, \quad X(0) = \xi. \tag{3.2}
\]

Let $x^{(\epsilon)}, y^{(\epsilon)}$ denote the solutions to the fast–slow system (3.1). Assume that solutions $X$ to this SDE exist on $[0, \infty)$ with probability one. Then, $x^{(\epsilon)} \to_w X$ in $C([0,\infty), \mathbb{R})$ as $\epsilon \to 0$.

(b) More generally, assume the above set-up but without the assumption that $h$ is non-vanishing. Suppose that we can write $h = 1/r'$ on an interval containing $\xi$. Write $\tilde{F} = (F/h) \circ r^{-1}$ where defined. Suppose that solutions $Z$ to the SDE (1.8) exist on $[0, \infty)$ with probability one.

Then, solutions $X$ to the SDE (3.2) exist on $[0, \infty)$ with probability one, and $x^{(\epsilon)} \to_w X$ in $C([0,\infty), \mathbb{R})$ as $\epsilon \to 0$.

Theorem 3.3 is proved by reducing it to theorem 3.1. We require a preliminary elementary result.

**Proposition 3.4.** Let $r : \mathbb{R} \to \mathbb{R}$ be a $C^2$ diffeomorphism. Suppose that $W$ is a one-dimensional unit Brownian motion and $\sigma^2 > 0$. Consider the Stratonovich SDE
\[
dX = \sigma (r'(X))^{-1} \circ dW + F(X) \, dt, \quad X(0) = \xi. \tag{3.2}
\]

Then, $X$ is a solution to this SDE if and only if $Z = r(X)$ satisfies the SDE
\[
dZ = \sigma \, dW + \tilde{F}(Z) \, dt, \quad Z(0) = r(\xi),
\]
where $\tilde{F} = (r'F) \circ r^{-1}$.

**Proof.** Suppose that $X$ satisfies the first SDE. Because the Stratonovich integral satisfies the usual chain rule, $Z = r(X)$ satisfies
\[
dZ = r'(X) \circ dX = \sigma \, dW + r'(X)F(X) \, dt = \sigma \, dW + \tilde{F}(Z) \, dt.
\]
The converse direction is identical. ■

**Proof of theorem 3.3.** Write $h = 1/r'$ and let $Z = r(X)$ where $X$ satisfies the SDE (3.2). The assumptions on $h$ guarantee that $r$ is a $C^2$ diffeomorphism. By proposition 3.4, $Z$ satisfies the SDE (1.8).

Next, let $z^{(\epsilon)} = r \circ x^{(\epsilon)}$. The $\tilde{z}^{(\epsilon)}$ equation in (3.1) becomes
\[
z^{(\epsilon)}(t) = \epsilon^{-1} f_0(y^{(\epsilon)}(t)) + \tilde{f}(z^{(\epsilon)}(t), y^{(\epsilon)}(t)), \tag{3.3}
\]
where
\[
\tilde{f}(z, y) = \frac{f(r^{-1}z, y)}{h(r^{-1}z)}.
\]

Because $\tilde{F}(z) = \int_A \tilde{f}(z, y) \, d\mu(y)$ and $\tilde{f}$ is locally Lipschitz, we are now in the situation of theorem 3.1, and it follows that solutions of (3.3) converge weakly to solutions of (1.8). That is, $r \circ x^{(\epsilon)} \to_w r \circ X$. Applying the $C^1$ map $r^{-1}$, it follows from the continuous mapping theorem that $x^{(\epsilon)} \to_w X$ as required. ■

**Remark 3.5.** As mentioned in §1, this result has a restricted extension to higher dimensions. Consider the fast–slow equations (3.1) with $x^{(\epsilon)} \in \mathbb{R}^d$ and suppose that $h : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^d)$ can be written as $h = (Dr)^{-1}$ for some $C^2$ diffeomorphism $r : \mathbb{R}^d \to \mathbb{R}^d$. Then, the substitution $z^{(\epsilon)} = r(x^{(\epsilon)})$ yields the equation (3.3) with $\tilde{f}(z, y) = h(r^{-1}z)^{-1}f(r^{-1}z, y)$. Again, it follows from theorem 3.1 that $z^{(\epsilon)} \to_w Z$, where $dZ = \sqrt{\Sigma} \, dW + \tilde{F}(Z) \, dt$, $\tilde{F}(z) = h(r^{-1}z)^{-1}F(r^{-1}z)$. The change in variables formulae for Stratonovich SDEs shows that $X = r^{-1}(Z)$ satisfies $dX = \sqrt{\Sigma} h(X) \circ dW + F(X) \, dt$. Again, $x^{(\epsilon)} \to_w X$ by the continuous mapping theorem.
4. Proofs of the results for maps

Here we prove the discrete time homogenization results stated in §1. By the argument in the proof of theorem 3.1, we may suppose from the outset that all relevant Lipschitz constants are uniform on the whole of $\mathbb{R}^d$ and that $\lim_{\epsilon \to 0} f(x, y, \epsilon) = f(x, y, 0)$ uniformly on the whole of $\mathbb{R}^d \times \Lambda$.

Throughout, it is more convenient to work with the piecewise constant function $\tilde{x}^{(\epsilon)}(t) = x^{(\epsilon)}([t\epsilon^{-2}])$ rather than with the linearly interpolated function $\hat{x}^{(\epsilon)}(t)$. (Here, $[t\epsilon^{-2}]$ denotes the integer part of $t\epsilon^{-2}$.) Because the process $\tilde{x}^{(\epsilon)}(t) = x^{(\epsilon)}([t\epsilon^{-2}])$ is not continuous, we can no longer work within $C([0, \infty), \mathbb{R}^d)$. Instead, we prove weak convergence in the space $D([0, \infty), \mathbb{R}^d)$ of càdlàg functions (right-continuous functions with left-hand limits, see Billingsley [12, ch. 3]) with the supremum norm.

It is clear that $\sup_{t \in [0, T]} |\tilde{x}^{(\epsilon)}(t) - \tilde{x}(t)| \to 0$ as $\epsilon \to 0$. Hence, weak convergence of $\tilde{x}^{(\epsilon)}$ in $D([0, \infty), \mathbb{R}^d)$ is equivalent to weak convergence of $\tilde{x}^{(\epsilon)}$ in $D([0, \infty), \mathbb{R}^d)$. This, in turn, is equivalent to weak convergence of $\tilde{x}^{(\epsilon)}$ in $C([0, \infty), \mathbb{R}^d)$ (see the last line of p. 124 of Billingsley [12]).

(a) Proof of theorem 1.1

Write $d(\epsilon) = \sup_{x \in \mathbb{R}^d, y \in A} |f(x, y, \epsilon) - f(x, y, 0)|$, so $\lim_{\epsilon \to 0} d(\epsilon) = 0$. First note that

$$x^{(\epsilon)}(n) = \xi + \epsilon \sum_{j=0}^{n-1} f_0(y(j)) + \epsilon^2 \sum_{j=0}^{n-1} f(x^{(\epsilon)}(j), y(j), \epsilon).$$

Hence

$$\tilde{x}^{(\epsilon)}(t) = \xi + W^{(\epsilon)}(t) + \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} f(x^{(\epsilon)}(j), y(j), 0) + K_1^{(\epsilon)}(t)$$

$$= \xi + W^{(\epsilon)}(t) + K_1^{(\epsilon)}(t) + K_2^{(\epsilon)}(t) + \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} F(\tilde{x}^{(\epsilon)}(j)),$$

where $W^{(\epsilon)}(t) = \epsilon \sum_{j=0}^{[t\epsilon^{-2}]-1} f_0(y(j))$, $|K_1^{(\epsilon)}(t)| \leq Td(\epsilon)$, and

$$K_2^{(\epsilon)}(t) = \epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} (f(x^{(\epsilon)}(j), y(j), 0) - F(x^{(\epsilon)}(j))).$$

For $t$, an integer multiple of $\epsilon^2$, the term $\epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} F(\tilde{x}^{(\epsilon)}(j))$ is the Riemann sum of a piecewise constant function and is precisely $\int_0^t F(\tilde{x}^{(\epsilon)}(s)) \, ds$. For general $t$,

$$\epsilon^2 \sum_{j=0}^{[t\epsilon^{-2}]-1} F(\tilde{x}^{(\epsilon)}(j)) = \int_0^t F(\tilde{x}^{(\epsilon)}(s)) \, ds + K_3^{(\epsilon)}(t),$$

where $K_3^{(\epsilon)}(t) \leq \epsilon^2 |f|_\infty$. Altogether,

$$\tilde{x}^{(\epsilon)}(t) = \xi + W^{(\epsilon)}(t) + K^{(\epsilon)}(t) + \int_0^t F(\tilde{x}^{(\epsilon)}(s)) \, ds,$$

where $K^{(\epsilon)} = K_1^{(\epsilon)} + K_2^{(\epsilon)} + K_3^{(\epsilon)}$.

By exactly the same argument as in Melbourne & Stuart [2], $K_2^{(\epsilon)} \to 0$ in $L^1(D([0, T], \mathbb{R}^d); \mu)$. (For convenience, the proof is reproduced in appendix A.) It follows that $W^{(\epsilon)} + K^{(\epsilon)} \to w \sqrt{\Sigma W}$ in $D([0, T], \mathbb{R}^d)$. Now consider the continuous map $G : D([0, T], \mathbb{R}^d) \to D([0, T], \mathbb{R}^d)$ given by $G(u) = u$, where $u(t) = \xi + u(t) + \int_0^t F(v(s)) \, ds$. Then, $\tilde{x}^{(\epsilon)} = G(W^{(\epsilon)} + K^{(\epsilon)})$, so $\tilde{x}^{(\epsilon)} \to w G(\sqrt{\Sigma W}) = X$. This completes the proof of theorem 1.1.
(b) Proof of theorem 1.3 and proposition 1.5

Again, we reduce to the case where \( \tilde{F} \) is globally Lipschitz and uniformly continuous at \( \epsilon = 0 \). Also, we may suppose that \( r' \) is uniformly continuous.

Write \( Z = r(X) \), where \( X \) satisfies the SDE (1.7). By proposition 3.4, \( Z \) satisfies

\[
dZ = \sigma \, dW + \tilde{F}(Z) \, dt,
\]

where

\[
\tilde{F}(z) = r'(r^{-1}z)F(r^{-1}z) - \frac{1}{2} h'(r^{-1}z) \int_A f_0^2 \, d\mu
\]

\[
= r'(r^{-1}z)F(r^{-1}z) + \frac{1}{2} r''(r^{-1}z)[r'(r^{-1}z)]^{-2} \int_A f_0^2 \, d\mu.
\]

Define \( z^{(e)}(n) = r(x^{(e)}(n)) \). Using Taylor’s theorem to expand the \( C^2 \) map \( r \), we obtain

\[
z^{(e)}(n + 1) - z^{(e)}(n) = r(x^{(e)}(n + 1)) - r(x^{(e)}(n))
\]

\[
= r'(x^{(e)}(n))(x^{(e)}(n + 1) - x^{(e)}(n))
\]

\[
+ \frac{1}{2} r''(x^{(e)}(n))(x^{(e)}(n + 1) - x^{(e)}(n))^2 + o((x^{(e)}(n + 1) - x^{(e)}(n))^2),
\]

where the last term is uniformly \( o(e^2 \) because \( r' \) is uniformly continuous.

Substituting for \( x^{(e)}(n + 1) = x^{(e)}(n) \) using equation (1.6), equation (4.2) becomes

\[
z^{(e)}(n + 1) - z^{(e)}(n) = \epsilon f_0(y(n)) + \epsilon^2 (r'(x^{(e)}(n))f(x^{(e)}(n), y(n), 0)
\]

\[
+ \frac{1}{2} r''(x^{(e)}(n))[r'(x^{(e)}(n))]^{-2} f_0(y(n))^2 + o(1))
\]

\[
= \epsilon f_0(y(n)) + \epsilon^2 \tilde{f}(z^{(e)}(n), y(n), \epsilon),
\]

where

\[
\tilde{f}(z, y, \epsilon) = r'(r^{-1}z)f(r^{-1}z, y, 0) + \frac{1}{2} r''(r^{-1}z)[r'(r^{-1}z)]^{-2} f_0(y)^2 + o(1),
\]

uniformly in \( z, y \) as \( \epsilon \to 0 \). Because \( \tilde{F}(z) = \int_A \tilde{f}(z, y, 0) \, d\mu(y) \), it follows from theorem 1.1 that \( \tilde{z}^{(e)}(t) = z^{(e)}([t \epsilon^2]) \) converges weakly to solutions \( Z \) of the SDE (4.1). Applying the \( C^1 \) map \( r^{-1} \), it follows from the continuous mapping theorem that \( \tilde{x}^{(e)} \to_w X \) (and hence \( x^{(e)} \to_w X \)) as required.

5. Stochastic differential equation driven by stable Lévy processes

Here, we consider the situation where the fast dynamics is not sufficiently chaotic to support the WIP. With reference to remark 2.1, this occurs when the Young tower [14] modelling the map (or the Poincaré map in the case of flows) has non-summable decay of correlations. In this case, the LDP still holds as before [24,25], but weak convergence to Brownian motion fails. However, there are instances where instead there is convergence to a stable Lévy process.

The prototypical examples are provided by Pomeau–Manneville intermittency maps [26]. For definiteness, consider the maps \( g : [0, 1] \to [0, 1] \) studied by Liverani et al. [27]:

\[
g(y) = \begin{cases} 
  y(1 + 2y^\gamma), & y \in [0, 1/2) \\
  2y - 1, & y \in [1/2, 1].
\end{cases}
\]

(5.1)

For \( y \in (0, 1) \), there exists a unique absolutely continuous invariant ergodic probability \( \mu \), and correlations decay at the rate \( n^{-(\gamma^{-1} - 1)} \). The attractor is \( A = [0, 1] \).

In particular, correlations are summable if and only if \( \gamma < 1/2 \), and in this situation all of the results in the previous sections apply. From now on, we suppose that \( \gamma \in (1/2, 1) \). Suppose that \( f_0 : A \to \mathbb{R}^d \) is Lipschitz with \( \int_A f_0 \, d\mu = 0 \), and assume further that \( f_0(0) \neq 0 \). Then, Gouëzel [28] proved that the central limit theorem fails and instead that \( n^{-\gamma} \sum_{j=0}^{n-1} f_0(y(j)) \) converges in distribution to a stable law \( Y \) of exponent \( 1/\gamma \). (More precisely, it follows from Gouëzel [28] that if \( c \in \mathbb{R}^d \) and \( c \cdot f_0(0) \neq 0 \), then \( n^{-\gamma} \sum_{j=0}^{n-1} c \cdot f_0(y(j)) \) converges in distribution to a one-dimensional
stable law of exponent 1/\gamma. Hence, there is convergence in distribution to a d-dimensional random variable Y and c·Y is stable of exponent 1/\gamma for all c. By Samorodnitsky & Taqqu [29, theorem 2.1.5(a) or (c)], Y is a d-dimensional stable distribution, and its exponent is 1/\gamma by Samorodnitsky & Taqqu [29, theorem 2.1.2]).

Let \( G = G_{1/\gamma} \) denote the corresponding stable Lévy process (independent and stationary increments with \( G(t) = \frac{1}{t^\gamma} Y \) for each t and sample paths lying in \( D([0, \infty), \mathbb{R}^d) \)). Then, it follows from Melbourne & Zweimüller [30] that \( W_n(t) = n^{-\gamma} \sum_{j=0}^{n-1} f_0(y(j)) \) converges weakly to \( G \) in \( D([0, \infty), \mathbb{R}^d) \) with the Skorokhod \( M_1 \) topology.

We proceed to consider fast–slow systems where the fast dynamics satisfies the LDP and weak convergence to a stable Lévy process of exponent 1/\gamma. First, consider the fast–slow ODE

\[
\dot{x}^{(e)} = \epsilon^{\gamma - 1} h(x^{(e)} f_0(y^{(e)})) + f(x^{(e)}, y^{(e)}), \quad x^{(e)}(0) = \xi,
\]

\[
\dot{y}^{(e)} = \epsilon^{-1} g(y^{(e)}), \quad y^{(e)}(0) = \eta.
\]

If \( h(x) \equiv 1 \), then exactly the same arguments as before yield that \( x^{(e)} \to_w X \), where X satisfies the SDE

\[
dX = dG + F(X) \, dt, \quad X(0) = \xi.
\]

If \( h = 1/r' \) is non-trivial and \( d = 1 \), then the same argument as before shows that \( x^{(e)} \to_w X \), where \( X = r(Z) \) and \( Z \) is the solution of the SDE

\[
dZ = dG + \tilde{F}(Z) \, dt, \quad Z(0) = \xi,
\]

where \( \tilde{F} = (F/h) \circ r^{-1} \). Transforming back, it is immediate that X satisfies the SDE

\[
dX = h(X) \circ dG + F(X) \, dt, \quad X(0) = \xi,
\]

provided that the stochastic integral \( h(X) \circ dG \) satisfies the usual chain rule. For Lévy processes, it turns out that neither the Itô nor Stratonovich interpretation is suitable, and the correct integral is due to Marcus [31] (see also [32]). See Applebaum [33, p. 272] for a discussion of the Marcus stochastic integral and in particular for the chain rule [33, theorem 4.4.28].

For maps, we consider

\[
\dot{x}^{(e)}(n+1) = x^{(e)}(n) + \epsilon^{\gamma} h(x^{(e)}(n)) f_0(y(n)) + \epsilon f(x^{(e)}(n), y(n)).
\]

Set \( \dot{x}^{(e)}(t) = x^{(e)}([t \epsilon^{-1}]) \). If \( h \equiv 1 \), then we obtain again that \( \dot{x}^{(e)} \) converges weakly to solutions X of the SDE

\[
dX = dG + F(X) \, dt, \quad X(0) = \xi.
\]

If \( h \) is non-trivial and \( d = 1 \), then we again obtain the Marcus SDE (5.2). (Note that the second-order expansion of \( r \) yields terms of order 2\gamma which are negligible because \( \gamma > \frac{1}{2} \). Hence, the additional correction terms that arose in the discrete case for Brownian noise are absent for Lévy noise.)

6. Numerical validation

Here, we illustrate theorem 1.3 with a numerical simulation of a suitable fast–slow map. For the fast dynamics, we could consider a Pomeau–Manneville intermittent map as in (5.1), with \( \gamma \in [0, \frac{1}{2}) \), so that the WIP and LDP are satisfied. In order to satisfy the centring condition \( \int_{\Lambda} f_0 \, d\mu = 0 \), it is more convenient to work with the following modified version of the map in (5.1):

\[
g(y) = \begin{cases} 
  y(1 + 2y^{\gamma}), & y \in [0, \frac{1}{2}), \\
  1 - 2y, & y \in [\frac{1}{2}, 1), \\
  -g(-y), & y \in [-1, 0).
\end{cases}
\]

Again, for \( \gamma \in [0, 1) \), there exists a unique absolutely continuous invariant ergodic probability measure \( \mu \). Moreover, the WIP and LDP again hold for \( \gamma \in [0, \frac{1}{2}) \); we choose \( \gamma = 0.1 \). The attractor is \( \Lambda = [-1, 1] \).
Because the map $g: A \to A$ is odd, the probability measure $\mu$ is symmetric around the origin. Hence, the condition $\int_A f_0 \, d\mu = 0$ is automatically satisfied provided $f_0 : A \to \mathbb{R}$ is odd. We choose 
\[
f_0(y) = y, \quad h(x) = x^{1/2}, \quad f(x, y, \epsilon) = \frac{1}{2}(\frac{3}{4} - x)y^2,
\]
so the slow dynamics is given by
\[
x^{(\epsilon)}(n + 1) = x^{(\epsilon)}(n) + \epsilon x^{(\epsilon)}(n)^{1/2}y(n) + \epsilon^2 \frac{1}{2}(\frac{3}{4} - x^{(\epsilon)}(n))y(n)^2.
\] 
(6.2)

According to theorem 1.3, re-scaled solutions $\hat{x}^{(\epsilon)}(t) = x^{(\epsilon)}(t\epsilon^{-2})$ of (6.2) converge weakly to solutions of the SDE
\[
dX = \sigma X^{1/2} \, dW + \frac{1}{2}(\frac{3}{4} - X)\int_A y^2 \, d\mu \, dt - \frac{1}{4} \int_A y^2 \, d\mu \, dt
\]
\[
= \sigma X^{1/2} \, dW + \frac{1}{2}(\frac{3}{4} - X)\int_A y^2 \, d\mu \, dt + \frac{1}{4} \left(\sigma^2 - \int_A y^2 \, d\mu\right) \, dt
\]
\[
= \sigma X^{1/2} \, dW + \alpha(\beta - X) \, dt
\] 
(6.3)

where $W$ is unit one-dimensional Brownian motion and
\[
\sigma^2 = \int_A y(0)^2 \, d\mu + 2 \sum_{n=1}^{\infty} \int_A y(n) y(0) \, d\mu = \lim_{n \to \infty} n^{-1} \int_A \left(\sum_{j=0}^{n-1} y(j)\right)^2 \, d\mu,
\]
\[
\alpha = \frac{1}{2} \int_A y^2 \, d\mu, \quad \beta = \frac{1}{4} \left(1 + \frac{\sigma^2}{\alpha}\right).
\]

This is the Cox–Ingersoll–Ross model [34,35] which has the closed form solution
\[
X(t) = c(t)H(t), \quad c(t) = \frac{\sigma^2}{4\alpha} (1 - e^{-\alpha t}),
\] 
(6.4)

where $H(t)$ is a non-central Chi-squared distribution with $4\alpha\beta/\sigma^2$ degrees of freedom and non-centrality parameter $c(t)^{-1}e^{-\alpha t}$. A long-time iteration of the map (6.1), taking an ensemble average, yields the approximate values $\sigma^2 = 0.085$ and $\int_A y^2 \, d\mu = 0.319$, and hence $\alpha = 0.160, \beta = 0.383$.

A consequence of weak convergence is convergence in distribution for each fixed $t > 0$, namely that for any $a \in \mathbb{R}$,
\[
\lim_{\epsilon \to 0} \mu(\hat{x}^{(\epsilon)}(t) < a) = P(X(t) < a).
\]

We proceed to verify this result numerically with $t = 10$ and the initial condition $\xi = x^{(\epsilon)}(0) = X(0) = 1$.

Computing the probability density function for $\hat{x}^{(\epsilon)}(10)$ from the numerical simulation of the full fast–slow system, and the limiting probability density function for $X(10)$ using the closed-form solution (6.4), we obtain the results shown in figure 1. We used ensembles consisting of 5,000,000 realizations (though for the fast–slow system we found that 100,000 realizations were ample).

In particular, figure 1 confirms our prediction regarding the drift term $\frac{1}{2}(\frac{3}{4} - X)\int_A y^2 \, d\mu + \frac{1}{4}(\sigma^2 - \int_A y^2 \, d\mu)$ in the limiting SDE (6.3). Figure 2 shows a comparison of the associated probability density function with those that would result from having the incorrect drift terms $\frac{1}{2}(\frac{3}{4} - X)\int_A y^2 \, d\mu$ or $\frac{1}{2}(\frac{3}{4} - X)\int_A y^2 \, d\mu + \frac{1}{4}\sigma^2$ that arise when the limiting SDE is interpreted as being Itô (as in the iid case) or Stratonovich (as in the continuous time case).

We also present numerical evidence for convergence of first moments $\mathbb{E}(\hat{x}^{(\epsilon)}(t))$. This is not a direct consequence of theorem 1.3. However, for each $t > 0$, theorem 1.3 together with boundedness (as $\epsilon$ varies) of a higher moment implies (e.g. [36, exercise 2.5, p. 86]) that $\mathbb{E}(\hat{x}^{(\epsilon)}(t))$
Figure 1. Probability density function (empirical measure) at $t = 10$ for the fast–slow map (6.1), (6.2) with $\epsilon = 0.8$ (a), $0.4$ (b), $0.2$ (c) (dashed, red) and for the SDE limit (6.3) (solid, blue). We used ensembles consisting of 5 000 000 realizations. (Online version in colour.)

Figure 2. Probability density function (empirical measure) at $t = 10$ for the limiting SDE with different drift terms corresponding to three different interpretations of the stochastic integral: the theoretically predicted drift term in (6.3) (left, blue) and those for the Itô interpretation (middle, red) and the Stratonovich interpretation (right, green). We used ensembles consisting of 5 000 000 realizations. (Online version in colour.)

Figure 3. First moment as function of time $t$ for the fast–slow map (6.1), (6.2) with $\epsilon = 0.8$ (a), $0.4$ (b), $0.2$ (c) (dashed, red) and for the SDE limit (6.5) (solid, blue). We used ensembles consisting of 100 000 realizations for the fast–slow map. (Online version in colour.)
converges, as $\epsilon \to 0$, to
\[ \mathbb{E}(X(t)) = \xi e^{-at} + \beta(1 - e^{-at}). \] (6.5)
We verified numerically that $\mathbb{E}((\dot{x}(t))²)$ is convergent and hence bounded, implying convergence of $\mathbb{E}((\dot{x}(t)))$ as demonstrated in figure 3.

7. Conclusions
The paper by Melbourne & Stuart [2] set out a programme for a rigorous investigation of homogenization for fast–slow deterministic systems under very mild assumptions on the fast dynamics. In this paper, we have extended these results in a number of ways:

1. extension from continuous time to discrete time;
2. incorporation of one-dimensional multiplicative noise;
3. inclusion of situations where the SDE is driven by a stable Lévy process; and
4. relaxation of regularity assumptions, in particular the requirement in Melbourne & Stuart [2] that certain vector fields are uniformly Lipschitz.

Items 2–4 require interpretation of certain stochastic integrals. In the case of multiplicative noise in continuous time, the stochastic integrals are of Stratonovich type as would be expected. For discrete time, it was pointed out in Givon & Kupferman [18] that the integrals are neither Itô nor Stratonovich. We recover their result in a much more general context. For multiplicative noise in the situation of item 3, where the SDE is Lévy, we obtain Marcus stochastic integrals for both discrete and continuous time.

Important directions of future research include the analysis of higher-dimensional multiplicative noise and of fully coupled systems where the fast dynamics depends on the slow variables.

Appendix A. Large deviation calculation
We describe the implications of the LDP that are required in this paper. The proofs are identical to those in the continuous time context in Melbourne & Stuart [2] and are included for completeness.

Proposition A.1. For any $a > 0$, $N \geq 1$, $n \geq 0$, $x \in \mathbb{R}^d$,
\[ \mathbb{E} \left| \frac{1}{N} \sum_{j=nN}^{(n+1)N-1} f(x, y(j)) - F(x) \right| \leq a + 2|f|_{\infty} b(a, N). \]

Proof. The proof is the same as the one in Melbourne & Stuart [2] for continuous time and is included for completeness. Let $S(N, x) = \{1/N \sum_{j=0}^{N-1} f(x, y(j)) - F(x)\}$. Then, by the LDP,
\[ \mathbb{E}S(N, x) = \int_{S(N, x) \leq a} S(N, x) \, d\mu + \int_{S(N, x) > a} S(N, x) \, d\mu \]
\[ \leq a + |S|_{\infty} b(a, N) \leq a + 2|f|_{\infty} b(a, N). \]
This proves the result for $n = 0$ and the general case follows from invariance of $\mu$.

Let $K_2^{(e)}$ be the expression appearing in the proof of theorem 1.1 in §4a.

Lemma A.2. $K_2^{(e)} \to 0$ in $L^1(D([0, T], \mathbb{R}^d); \mu)$.

Proof. Define $Q(x, y) = f(x, y, 0) - F(x)$ and note that $|Q|_{\infty} \leq 2|f|_{\infty}$ and $\text{Lip}(Q) \leq 2\text{Lip}(f)$. Then, $K_2^{(e)}(t) = \epsilon^2 \sum_{0 \leq j < [t \epsilon^{-1}]} Q(x^{(e)}(j), y(j))$. Let $N = \lfloor T/\delta \rfloor$ and write $K_2^{(e)}(t) = K_2^{(e)}(N \delta) + I_0$, where
\(I_0 = \varepsilon^2 \sum_{N \delta \varepsilon^{-2} \leq j < (N+1) \delta \varepsilon^{-2}} Q(x^e(j), y(j)).\) We have

\[|I_0| \leq (t - N \delta)|Q|_\infty \leq 2f|_\infty \delta. \tag{A 1}\]

We now estimate \(K_2^e(N \delta)\) as follows:

\begin{align*}
K_2^e(N \delta) &= \varepsilon^2 \sum_{n=0}^{N-1} \sum_{n \delta \varepsilon^{-2} \leq j < (n+1) \delta \varepsilon^{-2}} Q(x^e(j), y(j)) \\
&= \varepsilon^2 \sum_{n=0}^{N-1} \sum_{n \delta \varepsilon^{-2} \leq j < (n+1) \delta \varepsilon^{-2}} (Q(x^e(j), y(j)) - Q(x^e(n \delta \varepsilon^{-2}), y(j))) \\
&\quad + \varepsilon^2 \sum_{n=0}^{N-1} \sum_{n \delta \varepsilon^{-2} \leq j < (n+1) \delta \varepsilon^{-2}} Q(x^e(n \delta \varepsilon^{-2}), y(j)) \\
&= I_1 + I_2.
\end{align*}

For \(n \delta \varepsilon^{-2} \leq j < (n + 1) \delta \varepsilon^{-2},\) we have \(|x^e(j) - x^e(n \delta \varepsilon^{-2})| \leq (\varepsilon |f_0|_\infty + \varepsilon^2 |f|_\infty) \delta \varepsilon^{-2}.\) Hence

\[|I_1| \leq N \delta \text{Lip}(Q)(|f_0|_\infty + |f|_\infty) \delta \varepsilon^{-1} \leq 2\text{Lip}(f)(|f_0|_\infty + |f|_\infty)T \delta \varepsilon^{-1}. \tag{A 2}\]

Next,

\[I_2 = \varepsilon^2 \sum_{n=0}^{N-1} \sum_{n \delta \varepsilon^{-2} \leq j < (n+1) \delta \varepsilon^{-2}} (f(x^e(n \delta \varepsilon^{-2}), y(j), 0) - F(x^e(n \delta \varepsilon^{-2}))) = \delta \sum_{n=0}^{N-1} J_n,
\]

where

\[J_n = \varepsilon^2 \delta^{-1} \sum_{n \delta \varepsilon^{-2} \leq j < (n+1) \delta \varepsilon^{-2}} f(x^e(n \delta \varepsilon^{-2}), y(j), 0) - F(x^e(n \delta \varepsilon^{-2})).\]

Let \(a > 0.\) By proposition A.1, \(E|J_n| \leq a + 2f|_\infty b(a, \delta \varepsilon^{-2}).\) Recalling that \(N = \lceil t/\delta \rceil\) we have that

\[\sup_{t \in [0,T]} |I_2(t)| \leq \delta \sum_{n=0}^{T \delta - 1} |J_n|.
\]

It follows that

\[E \sup_{t \in [0,T]} |I_2(t)| \leq T(a + 2f|_\infty b(a, \delta \varepsilon^{-2})). \tag{A 3}\]

Finally, we set \(\delta = \varepsilon^{3/2}.\) By (A 1) and (A 2), \(\lim_{\varepsilon \to 0} E \sup_{t \in [0,T]} |I_0(t)| = \lim_{\varepsilon \to 0} E \sup_{t \in [0,T]} |I_1(t)| = 0.\) Moreover, by (A 3) and the assumption on \(b\) in the LDP, we have

\[\lim_{\varepsilon \to 0} E \sup_{t \in [0,T]} |I_2(t)| \leq Ta,
\]

and, because \(a > 0\) is arbitrary,

\[\lim_{\varepsilon \to 0} E \sup_{t \in [0,T]} |I_2(t)| = 0.
\]

Altogether, \(\lim_{\varepsilon \to 0} E \sup_{t \in [0,T]} |K_2^e(t)| = 0\) proving the claim.

The research of G.A.G. was supported in part by ARC grant (no. FT-0992214). The research of I.M. was supported in part by EPSRC grant (no. EP/F031807/1) held at the University of Surrey. I.M. is grateful to the hospitality at the University of Sydney where this research commenced in 2011. We are grateful to Ben Goldys, David Kelly and especially Andrew Stuart for helpful discussions, and also to the referees for several helpful comments and suggestions.
References

1. Pavliotis GA, Stuart AM. 2008 Multiscale methods: homogenization and averaging. Texts in
in Math. no. 470. Berlin, Germany: Springer.
(doi:10.1007/BF01389848)
Reading MA: Addison Wesley.
RM1972v027n04ABEH001383)
(doi:10.1016/0378-4371(90)90173-P)
10. Majda AM, Timofeyev I, Vanden-Eijnden E. 2006 Stochastic models for selected slow variables
11. Mackay RS. 2010 Langevin equation for slow degrees of freedom of Hamiltonian systems.
In Nonlinear dynamics and chaos: advances and perspectives. (Understanding complex systems)
12. Billingsley P. 1999 Convergence of probability measures, 2nd edn. Wiley Series in probability and
(doi:10.1007/BF02808180)
hyperbolic flows. Ergodic Theory Dyn. Syst. 32, 1091–1100. (doi:10.1017/S0143385711000174)
17. Wong E, Zakai M. 1965 On the convergence of ordinary integrals to stochastic integrals.
18. Givon D, Kupferman R. 2004 White noise limits for discrete dynamical systems driven by fast
PhysRevE.70.036120)
20. Pavliotis GA, Stuart AM. 2005 Analysis of white noise limits for stochastic systems with two
Math. Soc. 360, 6661–6676. (doi:10.1090/S0002-9947-08-04520-0)
23. Melbourne I, Nicol M. 2009 A vector-valued almost sure invariance principle for hyperbolic
(doi:10.1088/0951-7715/22/9/001)
Dyn. Syst. 19, 671–685. (doi:10.1017/S0143385799133856)


